Functional estimation of anisotropic covariance and autocovariance operators on the sphere[*]

Alessia Caponera, Julien Fageot,
Matthieu Simeoni and Victor M. Panaretos

École Polytechnique Fédérale de Lausanne
e-mail: alessia.caponera@epfl.ch*;* julien.fageot@epfl.ch*;* matthieu.simeoni@
epfl.ch*;* victor.panaretos@epfl.ch

Abstract: We propose nonparametric estimators for the second-order central moments of possibly anisotropic spherical random fields, within a functional data analysis context. We consider a measurement framework where each random field among an identically distributed collection of spherical random fields is sampled at a few random directions, possibly subject to measurement error. The collection of random fields could be i.i.d. or serially dependent. Though similar setups have already been explored for random functions defined on the unit interval, the nonparametric estimators proposed in the literature often rely on local polynomials, which do not readily extend to the (product) spherical setting. We therefore formulate our estimation procedure as a variational problem involving a generalized Tikhonov regularization term. The latter favours smooth covariance/autocovariance functions, where the smoothness is specified by means of suitable Sobolev-like pseudo-differential operators. Using the machinery of reproducing kernel Hilbert spaces, we establish representer theorems that fully characterize the form of our estimators. We determine their uniform rates of convergence as the number of random fields diverges, both for the *dense* (increasing number of spatial samples) and *sparse* (bounded number of spatial samples) regimes. We moreover demonstrate the computational feasibility and practical merits of our estimation procedure in a simulation setting, assuming a fixed number of samples per random field. Our numerical estimation procedure leverages the sparsity and second-order Kronecker structure of our setup to reduce the computational and memory requirements by approximately three orders of magnitude compared to a naive implementation would require.

AMS 2000 subject classifications: Primary 62G08; secondary 62M.
Keywords and phrases: Functional data analysis, measurement error, representer theorem, sparse sampling, spherical random field.

Received June 2022.

Contents

1. Introduction

Functional Data Analysis (FDA) comprises a wide class of statistical methods for the analysis of collections of data modelled as functions. In a classical FDA setting (see [39, 21]), we typically assume that we have access to a collection of *fully observed* realizations of continuous-domain random processes, say $X_1(\cdot), \ldots, X_n(\cdot)$, which can be modelled as random elements of some separable Hilbert space. Depending on the context, these realizations can be either independent and identically distributed (i.i.d. functional data) or exhibit some form of serial dependence (functional time series). In many practical setups however, this "fully observed" context is not suitable. For example, environmental scientists monitor temperature, salinity and currents in the Earth's oceans from spatial samples of such indicators collected by drifting floats, such as the ones of the Argo fleet [2]. While originating from a latent continuous-domain process, the data in this context are not observed in their fully functional form, but rather come in the form of *noisy* and *sparse* spatial samples *irregularly* distributed on the surface of the Earth (see Figure 1). This suggests to consider the following measurement scheme

$$W_{ij} = X_i(U_{ij}) + \epsilon_{ij}, \qquad i = 1, \ldots, n, \ j = 1, \ldots, r_i, \tag{1}$$

where U_{ij} are the sampling locations, ϵ_{ij} are noise disturbances. In this paper, we focus specifically on latent processes X_i's that constitute random fields over the *2D sphere* $\mathbb{S}^2$ and address the functional data analysis problem of estimating their first and second-order moment structure. This problem will be studied for X_i's that are independent replicates of some second-order process $X = \{X(u), \, u \in \mathbb{S}^2\}$ as well as for X_i's that constitute a finite stretch of some stationary sequence $\mathcal{X} = \{X_t(\cdot), \, t \in \mathbb{Z}\}$. In either case, we will *not* assume the process to be either strongly or weakly isotropic, hence our reference to an anisotropic setting.

Related work in FDA   The model presented in Equation (1), in the case of i.i.d. functional data $\{X_i\}$ on the domain $[0, 1]$, has been treated in several works, as for instance in [62] with the aim of building a bridge between FDA and longitudinal studies, and also in [19, 27, 63, 8, 7]. In this setting, it is not possible to recover *a priori* the entire signals by a pre-smoothing step, and
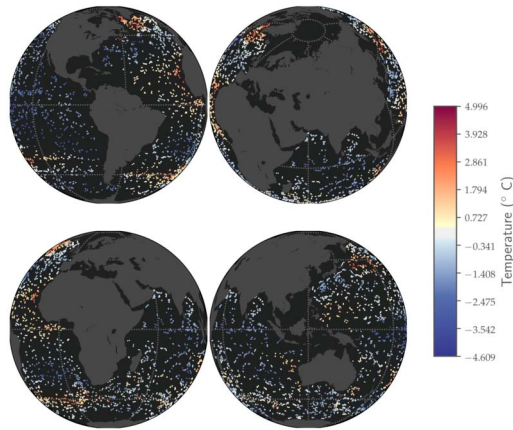
FIG 1. *Sea surface temperature anomalies recorded by Argo floats in January 2011. Float locations are marked by dots coloured according to the recorded anomaly (red=warmer temperatures, blue=colder temperatures).*

hence the statistical estimators and procedures cannot be based on the intrinsically infinite dimensional inputs and techniques as in the classical FDA setting. However, having access to n (partial) replications still makes inference possible, even if the number of points per replicate is small. The core point is that all the information coming from the measurements can be shared to estimate the main features of the underlying process, such as the mean and the covariance kernel. This idea of *borrowing information* leads naturally to the implementation of smoothing techniques. In the key paper [62], the authors propose to estimate the covariance kernel of the latent functional process by a local polynomial smoother, which is well defined in a 2-dimensional planar domain, and for such estimator they obtain rates of convergence. In the subsequent work [41], similar results for the lag-h autocovariances and spectral density kernels have been established in the serially dependent setting. However, when the domain is the product sphere, standard local polynomial smoothing techniques do not straightforwardly extend. Indeed, the covariance estimation task consists in estimating a function on a 4-dimensional curved surface (*i.e.*, $\mathbb{S}^2 \times \mathbb{S}^2$) and ideally the selected smoothing procedure would naturally incorporate the manifold geometry of the product sphere.

There exists a vast literature on (usually time-dependent) sequences of spherical random fields aimed at the characterization and parametric estimation of covariance/autocovariance functions (see [5, 15, 23, 24, 35, 36, 60, 49] and in particular [38] for a comprehensive review). The great majority of such works have focused on *isotropic* (in space) and *stationary* (in time) processes. However, in particular for the analysis of climate data, there is an attempt to move from the assumption of isotropy to that of *axial symmetry* (*i.e.*, heterogeneous spatial dependence across latitudes), but always in a purely parametric framework (see, for instance, [12, 37] and the references therein).

Very recently, the problem of investigating (isotropic and stationary) serially-dependent spherical random fields and their second-order structure has been tackled from a FDA perspective by [11, 10, 9] when fully observed functional time series are available.

Our aim is to bridge the gap between the idealised setting of these previous works and the many practical experiments which intrinsically feature irregular and/or sparse measurements of non-isotropic random fields on the sphere, such as the Argo data from Figure 1. Hence, we focus on both the i.i.d. and serially-dependent setting, specifically for functional data measured as defined in Equation (1), without assuming any form of isotropy in space. We provide a methodology which intrinsically incorporates the structure of the (product) spherical domain and also allows to estimate the mean, the covariance, and the autocovariance functions in a completely nonparametric manner. This in particular has many advantages, such as the possibility of performing spatial kriging without any parametric restrictions.

In the following, we summarize the proposed strategy and the main contributions of our work.

Variational methods for mean and covariance estimation   Following the general approach of [8, 7], our strategy for estimating the mean, covariance, and autocovariances of the spherical random fields X_i is to define the respective estimators as solutions of regularized quadratic variational problems of the form

$$\min_{f \in \mathcal{H}} \ \mathcal{L}(\boldsymbol{y}, \Phi(f)) + \lambda \|f\|_{\mathcal{H}}^2. \tag{2}$$

In this setting, the loss functional $\mathcal{L}(\boldsymbol{y}, \Phi(f))$ constraints the measurement vector $\Phi(f)$ to be close to the data $\boldsymbol{y}$ – in our case, the data are the noisy sampled values W_{ij} in (1) for the mean estimation, or pairwise products $W_{ij} W_{i'j'}$ between these quantities for the covariance estimation. The choice of the Hilbert space $\mathcal{H}$ and therefore the regularization $\|f\|_{\mathcal{H}}$ determines the smoothness of the estimators as functions over $\mathbb{S}^2$ or $\mathbb{S}^2 \times \mathbb{S}^2$. The parameter $\lambda > 0$ balances the role of the data-fidelity and the regularization.

Optimization problems of the form (2) have a rich history far beyond covariance estimation. The underlying Hilbert space $\mathcal{H}$ can be made of vectors or functions. They are known as Tikhonov regularization in inverse problems [52, 22, 54] and ridge-regression in statistics [20, 42]. Quadratic regularized optimized problems over Hilbert *function spaces* are well-known to be connected to splines since the pioneering works of Wahba in the 1970's [25]; see also [58]. The specification of the form of the solutions of problems such as (2) is known as a representer theorem [43, 44]. Quadratic representer theorems for sequences [54], continuous functions [18] and periodic functions [4] have been proposed. The case of spherical functions is covered [46, 31]. For the mean and covariance estimation of random fields over $[0, 1]$, we refer to [8, 7]. To the best of our knowledge, this setting has never been proposed for *spherical* random fields. It is worth noting that, even if we restricted our attention to the sampling regression problem (1), we can easily generalize the approach to generalized linear measurements, as

detailed for instance in [4, 46, 31]. Estimation of spherical random fields from generalized measurements is typically relevant in radio interferometric or acoustic imaging [49, 48, 50].

Contributions The specificity of our work is to apply quadratic regularized optimization problems to mean, covariance, and autocovariance estimation of spatially *spherical* – hence non Euclidean – and *serially-dependent* random fields. To the best of our knowledge, these ideas have not been applied in these contexts. Our method appears, in fact, to be the first and only thus far to tackle non-parametric and non-isotropic covariance estimation with sparsely sampled functional data on the sphere. Our main contributions are as follows.

- *Representer theorems for the estimators:* We introduce a family of regularization-based estimators for the estimation of the mean, covariance, and autocovariance functions of spatially spherical and temporally stationary Gaussian random fields. The regularizations for the first and second moments are given by $f \mapsto \|\mathscr{D}f\|^2_{L^2(\mathbb{S}^2)}$ and $g \mapsto \|(\mathscr{D} \otimes \mathscr{D})g\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$, where $\mathscr{D}$ is a smoothness-inducing pseudo-differential operator. For such regularizations, we obtain the general form of the mean and covariance estimators, expressed in terms of the regression data and the Green's functions of the regularizing pseudo-differential operators.
- *Asymptotic Theory:* Our main result is to provide an asymptotic analysis of the performance of the mean, covariance, and autocovariance estimators obtained via the representer theorems. The performance is expressed in terms of the smoothness properties of both the X_i and their second-order structure. By distinguishing between the regularity of the random fields and that of their second-order structure, we provide a refined analysis relative to existing results on $[0, 1]^2$, which for important classes of spherical random fields results in improved rates.
- *Serially Dependent Setting:* We extend our methods and theory to also cover the case of serially correlated random fields on $\mathbb{S}^2$, a.k.a. *spherical functional time series*. To the best of our knowledge, in the context of functional time series, variational methods for estimating the mean and, especially, the sequence of lag h autocovariance kernels have not been considered before.
- *Computation:* We develop a practically feasible implementation of our estimation methodology, circumventing the computational and memory challenges raised by the high dimensionality of the estimation task. More specifically, leveraging our theoretical results as well as sparse/second-order structure, we develop a *computationally tractable* and *memory-thrifty* implementation. The latter allows us to reduce the computational/memory requirements by two to three orders of magnitude approximately. We propose moreover a K-fold cross-validation procedure, both for evaluating the finite-sample performance of our estimator but also to optimally select the regularization parameter λ in (2). The simulation code is openly available and fully-reproducible, and is released in the form of a standalone

Python 3 Jupyter notebook hosted on GitHub: https://github.com/matthieumeo/sphericov/blob/main/companion_nb.ipynb. Instructions for installing the required dependencies and running the notebook are available at: https://github.com/matthieumeo/sphericov#readme.

Outline The rest of the paper is organized as follows. In Section 2, we introduce some preliminary concepts on spherical and tensorial Sobolev spaces and related pseudo-differential operators. In Section 3, we formally present the model and the proposed estimators for the mean and covariance functions in the i.i.d. setting. For such estimators we provide representer theorems in Section 4 and the asymptotic analysis in Section 5. Our procedure is then implemented in Section 6 for a simulated experiment. In Section 7 the case of serially dependent spherical random fields is examined and rates of convergence are provided for the lag-*h* autocovariance kernels. The proofs of all formal statements are collected in Section 8.

2. Mathematical background

We denote by $L^2(\mathbb{S}^2)$ the space of real-valued square-integrable functions over the 2D sphere. It is well-known that any function $f \in L^2(\mathbb{S}^2)$ can be expanded as

$$f = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_{\ell,m} Y_{\ell,m}, \tag{3}$$

where $\{Y_{\ell,m}\}_{\ell\in\mathbb{N}, -\ell\leq m\leq\ell}$ is a standard orthonormal basis of real spherical harmonics and the spherical Fourier coefficients $f_{\ell,m} = \langle f, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)} \in \mathbb{R}$ of f are such that $\|f\|^2_{L^2(\mathbb{S}^2)} := \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} f^2_{\ell,m} < \infty$ – see for instance [30, Chapter 3]. We also consider functions over the domain $\mathbb{S}^2 \times \mathbb{S}^2$, which is required to deal with the second-order moments of spherical random fields. The family $\{Y_{\ell,m} \otimes Y_{\ell',m'}\}_{\ell,\ell'\in\mathbb{N}, -\ell\leq m\leq\ell, -\ell'\leq m'\leq\ell'}$ is then an orthonormal basis of $L^2(\mathbb{S}^2 \times \mathbb{S}^2) = L^2(\mathbb{S}^2) \otimes L^2(\mathbb{S}^2)$, where $(Y_{\ell,m} \otimes Y_{\ell',m'})(u,v) = Y_{\ell,m}(u)Y_{\ell',m'}(v)$ for any $(u,v) \in \mathbb{S}^2 \times \mathbb{S}^2$.

2.1. Sobolev spaces on $\mathbb{S}^2$ and $\mathbb{S}^2 \times \mathbb{S}^2$

We characterize the smoothness properties of the random fields, means, and covariance functions in terms of Sobolev spaces, defined in Definition 1.

Definition 1. Let $p \geq 0$. The spherical Sobolev space of order p on $\mathbb{S}^2$ is defined as

$$\mathcal{H}_p = \mathcal{H}_p(\mathbb{S}^2) = \left\{ f \in L^2(\mathbb{S}^2), \|f\|^2_{\mathcal{H}_p} := \sum_{\ell=0}^{\infty}(1 + \ell(\ell+1))^p \sum_{m=-\ell}^{\ell} \langle f, Y_{\ell,m}\rangle^2 < \infty \right\}. \tag{4}$$

The tensorial Sobolev space of order p on $\mathbb{S}^2 \times \mathbb{S}^2$ is

$$\mathbb{H}_p = \mathbb{H}_p(\mathbb{S}^2 \times \mathbb{S}^2) = \mathcal{H}_p \otimes \mathcal{H}_p \subset L^2(\mathbb{S}^2 \times \mathbb{S}^2). \tag{5}$$

The spaces $\mathcal{H}_p$ and $\mathbb{H}_p$ are Hilbert spaces for their respective Hilbert norms $\| \cdot \|_{\mathcal{H}_p}$ and $\| \cdot \|_{\mathbb{H}_p}$ inherited from the inner products

$$\langle f_1, f_2 \rangle_{\mathcal{H}_p} := \sum_{\ell=0}^{\infty} (1 + \ell(\ell+1))^p \sum_{m=-\ell}^{\ell} \langle f_1, Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)} \langle f_2, Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)}, \tag{6}$$

$$\langle g_1, g_2 \rangle_{\mathbb{H}_p} := \sum_{\ell,\ell'=0}^{\infty} (1 + \ell(\ell+1))^p (1 + \ell'(\ell'+1))^p$$

$$\times \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \langle g_1, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2)} \langle g_2, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2)} \tag{7}$$

for $f_1, f_2 \in \mathcal{H}_p$ and $g_1, g_2 \in \mathbb{H}_p$. Note that the norm $\| \cdot \|_{\mathbb{H}_p}$ is such that $\|f_1 \otimes f_2\|_{\mathbb{H}_p} = \|f_1\|_{\mathcal{H}_p} \|f_2\|_{\mathcal{H}_p}$ for any $f_1, f_2 \in \mathcal{H}_p$.

The parameter p in Definition 1 quantifies the regularity: the higher p, the smoother the functions $f \in \mathcal{H}_p$ and $g \in \mathbb{H}_p$. We also define the spaces $\mathcal{S}(\mathbb{S}^2)$ and $\mathcal{S}(\mathbb{S}^2 \times \mathbb{S}^2) = \mathcal{S}(\mathbb{S}^2) \otimes \mathcal{S}(\mathbb{S}^2)$ of infinitely smooth functions on $\mathbb{S}^2$ and $\mathbb{S}^2 \times \mathbb{S}^2$, respectively. We then have that

$$\mathcal{S}(\mathbb{S}^2) = \bigcap_{p \geq 0} \mathcal{H}_p \qquad \text{and} \qquad \mathcal{S}(\mathbb{S}^2 \times \mathbb{S}^2) = \bigcap_{p \geq 0} \mathbb{H}_p. \tag{8}$$

The topological duals of $\mathcal{S}'(\mathbb{S}^2)$ and $\mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^2)$ are respectively the space of distributions $\mathcal{S}(\mathbb{S}^2)$ and $\mathcal{S}(\mathbb{S}^2 \times \mathbb{S}^2)$ and are such that

$$\mathcal{S}'(\mathbb{S}^2) = \bigcup_{p \geq 0} \mathcal{H}_p{}' \qquad \text{and} \qquad \mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^2) = \bigcup_{p \geq 0} \mathbb{H}_p{}'. \tag{9}$$

Among Sobolev spaces, we are specifically interested in the ones on which the evaluation functionals $f \mapsto f(x)$ are continuous for any x in the domain ($\mathbb{S}^2$ or $\mathbb{S}^2 \times \mathbb{S}^2$). In other terms, we shall only consider Sobolev spaces $\mathcal{H}_p$ and $\mathbb{H}_p$ that are Reproducing Kernel Hilbert Spaces (RKHS) [3]. We characterize this property in Proposition 1, whose proof is provided in Section 8.1.

Proposition 1. *Let $p \geq 0$. Then, we have the equivalences*

$$\mathcal{H}_p \text{ is a RKHS} \quad \Longleftrightarrow \quad \mathbb{H}_p \text{ is a RKHS} \quad \Longleftrightarrow \quad p > 1. \tag{10}$$

The RKHS properties allow us to provide uniform control functions on $\mathcal{H}_p$ and $\mathbb{H}_p$, as will be used for instance in Lemma 1 in Section 8.3. This implies in particular the uniform convergence of Fourier series; *e.g.*, for any $f \in \mathcal{H}_p$,

$$\sup_{u \in \mathbb{S}^2} \left| f(u) - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \langle f, Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)} Y_{\ell,m}(u) \right| \to 0, \qquad L \to \infty.$$

2.2. Spherical and tensorial admissible operators

The Tikhonov regularization used in our estimation strategy relies on linear operators that will impact the smoothness of the estimators. We now introduce the class of linear operators that are relevant for our work.

Definition 2 (Admissible Operators)**.** An *admissible operator* will be any linear operator of the form

$$\mathscr{D} : \begin{cases} \mathcal{S}(\mathbb{S}^2) \to \mathcal{S}(\mathbb{S}^2) \\ f \mapsto \mathscr{D}f = \sum_{\ell=0}^{\infty} D_\ell \sum_{m=-\ell}^{\ell} f_{\ell,m} Y_{\ell,m} \end{cases}$$

where $\{f_{\ell,m} = \langle f, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}\}_{\ell \in \mathbb{N}, -\ell \leq m \leq \ell}$ are the spherical Fourier coefficients of f and $\{D_\ell\}_{\ell \in \mathbb{N}}$ is a sequence of non-zero real numbers such that

$$C_1(1+\ell)^p \leq |D_\ell| \leq C_2(1+\ell)^p, \tag{11}$$

for some real number $p \geq 0$, some positive constants $C_1, C_2 > 0$ and every $\ell \in \mathbb{N}$. We call the D_ℓ, $\ell \in \mathbb{N}$, the *spherical Fourier coefficients* and p the *spectral growth order* of $\mathscr{D}$.

The condition (11) implies that $\mathscr{D}$ is invertible from $\mathcal{S}(\mathbb{S}^2)$ to itself. The inverse of $\mathscr{D}$ is given by $\mathscr{D}^{-1}f = \sum_{\ell=0}^{\infty} \frac{1}{D_\ell} \sum_{m=-\ell}^{\ell} f_{\ell,m} Y_{\ell,m}$. Note that Definition 2 excludes some classical pseudo-differential operators such as the Laplacian operator $\mathscr{D} = \Delta_{\mathbb{S}^2}$ itself, which is not invertible. Indeed, we have in this case that $D_\ell = -\ell(\ell+1)$, hence $D_0 = 0$ and the condition (11) is not fulfilled. It is however possible to generalize Definition 2 above to include the Laplacian and more generally spherical pseudo-differential operators with finite-dimensional null space – see for example [47, Definition 4].

Example 1. *We consider the family of Sobolev operators,* i.e.*, operators of the form $\mathscr{D} := (\mathrm{Id} - \Delta_{\mathbb{S}^2})^{p/2}$ for some $p \geq 0$. The spherical Fourier coefficients of $(\mathrm{Id} - \Delta_{\mathbb{S}^2})^{p/2}$ are given by*

$$D_\ell = (1 + \ell(\ell+1))^{p/2}, \qquad \ell \in \mathbb{N}.$$

We easily see that $(\mathrm{Id} - \Delta_{\mathbb{S}^2})^{p/2}$ satisfies the conditions of Definition 2 with spectral growth order p.

We can use the admissible operators in Definition 2 to construct operators acting on functions over $\mathbb{S}^2 \times \mathbb{S}^2$ as follows. Let $\mathscr{D}$ be an admissible operator. Then, $\mathscr{D} \otimes \mathscr{D}$ is a linear operator from $\mathcal{S}(\mathbb{S}^2 \times \mathbb{S}^2)$ to itself characterized by the relations

$$(\mathscr{D} \otimes \mathscr{D})(f_1 \otimes f_2) = (\mathscr{D}f_1) \otimes (\mathscr{D}f_2)$$

for any $f_1, f_2 \in \mathcal{S}(\mathbb{S}^2)$. The invertibility of $\mathscr{D}$ then implies the invertibility of $\mathscr{D} \otimes \mathscr{D}$ from $\mathcal{S}(\mathbb{S}^2 \times \mathbb{S}^2)$ to itself.

The Sobolev operator $\mathscr{D} = (\mathrm{Id} - \Delta_{\mathbb{S}^2})^{p/2}$ of order $p \geq 0$ is such that

$$\|f\|_{\mathcal{H}_p} = \|\mathscr{D}f\|_{L^2(\mathbb{S}^2)} \quad \text{and} \quad \|g\|_{\mathbb{H}_p} = \|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} \tag{12}$$

for any $f \in \mathcal{H}_p$ and $g \in \mathbb{H}_p$. More generally, any admissible operator $\mathscr{D}$ specifies a continuous bijection $\mathscr{D} : \mathcal{H}_p \to L^2(\mathbb{S}^2)$ (see Proposition 4 in Section 8.1 for a formal statement). This implies that $f \mapsto \|\mathscr{D}f\|_{L^2(\mathbb{S}^2)}$ specifies a norm on $\mathcal{H}_p$ which is equivalent to $\|\cdot\|_{\mathcal{H}_p}$. Similarly, $g \mapsto \|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$ specifies a norm on $\mathbb{H}_p$ which is equivalent to $\|\cdot\|_{\mathbb{H}_p}$.

Remark 1. For any admissible operator $\mathscr{D}$, the family $\{Y_{\ell,m}/D_\ell\}_{\ell \in \mathbb{N}, -\ell \leq m \leq \ell}$ is orthonormal in $(\mathcal{H}_p, \|\mathscr{D} \cdot \|_{L^2(\mathbb{S}^2)})$ This simply follows from the fact that, as easily seen from Definition 2, $\mathscr{D}Y_{\ell,m} = D_\ell Y_{\ell,m}$ and therefore,

$$\langle \mathscr{D}Y_{\ell,m}, \mathscr{D}Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2)} = D_\ell D_{\ell'}^* \langle Y_{\ell,m}, Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2)} = D_\ell^2 \delta_\ell^{\ell'} \delta_m^{m'}. \tag{13}$$

In particular, $\left\{Y_{\ell,m}/(1 + \ell(\ell+1))^{p/2}\right\}_{\ell \in \mathbb{N}, -\ell \leq m \leq \ell}$ is an orthonormal basis of $(\mathcal{H}_p, \|\cdot\|_{\mathcal{H}_p})$. The relation (13) moreover implies, for any $f \in \mathcal{H}_p$, we have the useful relation

$$\langle \mathscr{D}f, \mathscr{D}Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)} = D_\ell^2 \langle f, Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)}. \tag{14}$$

We now introduce the Green's functions and the zonal Green kernel of admissible operators in Definition 3.

Definition 3. Let $\mathscr{D}$ be an admissible operator. For any $u \in \mathbb{S}^2$, we denote by $\Psi_u^{\mathscr{D}} = \mathscr{D}^{-1}\delta_u$. Then, we call $\Psi_u^{\mathscr{D}}$ the *Green's function* of the operator $\mathscr{D}$ at position u. Moreover, there exists a function $\psi_{\mathscr{D}} : [-1, 1] \to \mathbb{R}$ such that, for any $u, v \in \mathbb{S}^2$,

$$\Psi_u^{\mathscr{D}}(v) = \psi_{\mathscr{D}}(\langle u, v \rangle), \tag{15}$$

where $\langle u, v \rangle$ is the usual inner product between points in $\mathbb{S}^2 \subset \mathbb{R}^3$. The function $\psi_{\mathscr{D}}$ is called the *zonal Green's kernel* of $\mathscr{D}$.

It is known that the Green's functions of admissible operators are continuous as soon as the spectral growth order p satisfies $p > 1$ [47, Proposition 4]. The existence of the zonal Green's kernel is for instance proved in [47, Proposition 3]; it can be expressed in terms of the spherical Fourier coefficients $\{D_\ell\}$ and the 2-dimensional ultraspherical polynomials (see [47, Eq. (16)]). The existence of the zonal Green's kernel has important practical consequences. It means in particular that the Green's functions at any positions can be easily computed from $\psi_{\mathscr{D}}$, as will be exploited in Section 6.

3. Model and estimation methodology

In this section, we present the theoretical setting in which we develop our methodology. From a purely functional data analysis perspective, we consider a second-order stochastic process $X = \{X(u), u \in \mathbb{S}^2\}$ that is a random element of $\mathcal{H}_q$, for some $q > 1$, with

$$\mathbb{E}[X(u)] = \mu(u), \qquad \mathbb{E}[X(u)X(v)] = R(u, v),$$

and covariance function

$$C(u, v) = R(u, v) - \mu(u)\mu(v),$$

for $u, v \in \mathbb{S}^2$. Recall from Proposition 1 that $\mathcal{H}_q$ with $q > 1$ is a RKHS, hence the process X inherits all the properties of a RKHS valued process, see [21, Section 7.5]. Moreover, since $q > 1$, the realizations of X are almost surely continuous on $\mathbb{S}^2$.

Now suppose we have $X_1, \ldots, X_n$ independent replicates of X and that for the i-th replicate we make measurements at r_i random locations on $\mathbb{S}^2$. Formally, we consider the following regression problem

$$W_{ij} = X_i(U_{ij}) + \epsilon_{ij}, \qquad i = 1, \ldots, n, \; j = 1, \ldots, r_i,$$

where the U_{ij}'s are independently drawn from a common distribution on $\mathbb{S}^2$, and the ϵ_{ij}'s are independent and identically distributed measurement errors of mean 0 and variance $0 < \sigma^2 < \infty$. Furthermore, the X_i's, the measurement locations, and the measurement errors are assumed to be mutually independent. Then,

$$\mathbb{E}[W_{ij}|U_{ij} = u_{ij}] = \mu(u_{ij}), \tag{16}$$

and

$$\mathbb{E}[W_{ij}W_{lk}|U_{ij} = u_{ij}, U_{lk} = u_{lk}] = \begin{cases} \mu(u_{ij})\mu(u_{lk}) & \text{if } i \neq l \\ R(u_{ij}, u_{ik}) + \sigma^2 \delta_j^k & \text{if } i = l. \end{cases} \tag{17}$$

Moreover,

$$\text{Cov}[W_{ij}, W_{lk}|U_{ij} = u_{ij}, U_{lk} = u_{lk}] = \begin{cases} 0 & \text{if } i \neq l \\ C(u_{ij}, u_{ik}) + \sigma^2 \delta_j^k & \text{if } i = l. \end{cases}$$

Below, expectations will be sometimes computed conditionally on/with respect to the whole set of U_{ij}'s, which will be denoted by $\mathcal{U}$.

It appears evident from Equations (16) and (17) that the measurements themselves and the *off-diagonal* products can be seen as unbiased estimators for the mean and second-order moment functions, respectively, computed at fixed locations. In particular, we can recover the covariance by performing smoothing on the pooled measurements (to first recover the mean) and then on the pooled product observations (to recover the second-order moment function). Note that it is important in this second step to discard the "diagonal" elements (not easy to visualize on $\mathbb{S}^2 \times \mathbb{S}^2$ as opposed to $[0, 1] \times [0, 1]$) so that potential biases arising from the noise are neglected (as in [62]). We then use the previously introduce machinery to build smoothing techniques that takes into account the geometry of $\mathbb{S}^2$.

Specifically, for a given $p \geq q$ and $\eta, \lambda > 0$, we can define the following estimators for the mean function

$$\mu_\lambda := \arg\min_{g \in \mathcal{H}_p} \frac{4\pi}{n} \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} (W_{ij} - g(U_{ij}))^2 + \lambda \|\mathscr{D}g\|_{L^2(\mathbb{S}^2)}^2, \tag{18}$$

and for the second-order moment function

$$R_\eta := \underset{g \in \mathbb{H}_p}{\arg\min} \frac{(4\pi)^2}{n} \sum_{i=1}^n \frac{1}{r_i(r_i-1)} \sum_{1 \le j \ne k \le r_i} (W_{ij}W_{ik} - g(U_{ij}, U_{ik}))^2$$
$$+ \eta \|(\mathscr{D} \otimes \mathscr{D})g\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}, \tag{19}$$

where $\mathscr{D}$ is any admissible operator with spectral growth order p (see Definition 2 in Section 2.2). An estimate of the complete covariance kernel $C(u,v) = R(u,v) - \mu(u)\mu(v)$ is then given by

$$C_{\eta,\lambda}(u,v) = R_\eta(u,v) - \mu_\lambda(u)\mu_\lambda(v). \tag{20}$$

Remark 2. In order to simplify the notation, from now on, we will not differentiate between the two penalty parameters η, λ. In particular, for the covariance estimator, we will consider $\eta = \lambda$ and write directly C_η. This will not affect in any way our asymptotic results.

We remark that our estimators are genuinely nonparametric and anisotropies in space are allowed. Moreover, the class of admissible operators is very large and, hence, this implementation gives the possibility to be flexible with respect to particular practical problems (non-asymptotic regimes).

4. Representer theorems for mean and covariance estimation

In this section, we specify the form of the solutions of the optimization problems which are the cornerstones of our estimation strategy. More precisely, representer theorems for the mean estimator (18) and the second-order moment estimator (19) are stated in Sections 4.1 and 4.2 respectively. In both cases, the form of the solution is deduced from general principles for optimization problems over Hilbert spaces, that are presented in Section 8.2.

4.1. Representer theorem for mean estimation

As we have seen in (18) and (19), our strategy for the mean and covariance estimations relies on the minimization of quadratic cost functionals with two components: (i) a data-fidelity term which constraint the solution to be consistent with the sampled observations and (ii) a regularization term which enforces some smoothness condition via the admissible operator $\mathscr{D}$. Our goal in this section and the next one is to reveal the form of the solution of the optimization problem together with their main properties. We start with the mean estimation in Theorem 1.

Theorem 1. *Let $\mathscr{D}$ be an admissible operator with spectral growth $p > 1$. Let $n \ge 1$, $r_i \ge 1$ for $i = 1, \dots n$. We consider weights $w_{ij} \in \mathbb{R}$ and pairwise distinct*

positions $u_{ij} \in \mathbb{S}^2$ for $i = 1, \ldots, n$ and $j = 1, \ldots, r_i$. We fix $\eta > 0$. Then, the optimization problem

$$\min_{f \in \mathcal{H}_p} \frac{4\pi}{n} \sum_{i=1}^{n} \frac{1}{r_i} \sum_{j=1}^{r_i} (w_{ij} - f(u_{ij}))^2 + \eta \|\mathscr{D}f\|_{L^2(\mathbb{S}^2)}^2 \tag{21}$$

has a unique solution $\mu_\eta \in \mathcal{H}_p$ which is given by

$$\mu_\eta = \sum_{i=1}^{n} \frac{1}{\sqrt{r_i}} \sum_{j=1}^{r_i} \alpha_{ij} \psi_{\mathscr{D}^* \mathscr{D}}(\langle \cdot, u_{ij} \rangle) \tag{22}$$

for some $(\alpha_{ij})_{1 \leq i \leq n, \ 1 \leq j \leq r_i}$, where $\psi_{\mathscr{D}^* \mathscr{D}}$ is the zonal Green's kernel of $\mathscr{D}^* \mathscr{D}$ (see Definition 3).

Moreover, the coefficients α_{ij} are computed as follows. Let $L = \sum_{i=1}^{n} r_i$ be the total number of measurements. We set $\boldsymbol{\alpha} \in \mathbb{R}^L$ the vectorized version of $(\alpha_{ij})_{1 \leq i \leq n, \ 1 \leq j \leq r_i}$, $\boldsymbol{y} \in \mathbb{R}^L$ the vectorized version of the normalized observations $\left(\frac{w_{ij}}{\sqrt{r_i}} \right)_{1 \leq i \leq n, \ 1 \leq j \leq r_i}$, and $\mathbf{G} \in \mathbb{R}^{L \times L}$ the matrix whose entries are given by

$$G_{\ell_1 \ell_2} = \frac{\psi_{\mathscr{D}^* \mathscr{D}}(\langle u_{i_1 j_1}, u_{i_2 j_2} \rangle)}{\sqrt{r_{i_1} r_{i_2}}}. \tag{23}$$

where the index ℓ_1 (resp. ℓ_2) corresponds to the couple (i_1, j_1) (resp. (i_2, j_2)) in the vectorization. Then, we have that

$$\boldsymbol{\alpha} = \left(\mathbf{G} + \frac{\eta n}{4\pi} \mathbf{I}_L \right)^{-1} \boldsymbol{y} \tag{24}$$

where $\mathbf{I}_L \in \mathbb{R}^{L \times L}$ is the identity matrix.

As far as the mean is concerned, the representer theorem can be deduced from known results in the literature. In particular, the general form (22) of the solution μ_η in Theorem 1 can be seen as a particular case of [46, Theorem 5.3]. The specification of the weights α_{ij} is then a finite-dimensional quadratic optimization problem (see for instance [18, Section V.A] and [4, Proposition 4] for similar results on the discretization of quadratic optimization problems over function spaces). Theorem 1 is also a special case of the general Representer Theorem over Hilbert spaces that we recall in Section 8.2 (see Theorem 7). This specification is similar – and simpler – to the one we detail in Theorem 2 for the second-order estimator, hence we do not detail it.

Remark 3. The estimator μ_η of (21) depends linearly on $\boldsymbol{\alpha}$ via (22), which depends itself linearly on the observations $(w_{ij})_{1 \leq i \leq n, \ 1 \leq j \leq r_i}$. Hence, μ_η is a linear estimator of the mean μ. We moreover observe that μ_η is a $(\mathscr{D}^* \mathscr{D})$-spline in the sense that [46, Definition 7]

$$(\mathscr{D}^* \mathscr{D})\{\mu_\eta\} = \sum_{i=1}^{n} \frac{1}{\sqrt{r_i}} \sum_{j=1}^{r_i} \alpha_{ij} \delta_{u_{ij}} \tag{25}$$

is a sum of Dirac impulses. The estimator μ_η is therefore the optimal $(\mathscr{D}^*\mathscr{D})$-spline with knots at the sampling locations. This is the adaptation in our context of a well-known fact for Tikhonov-type regularization [46, 25, 6, 4].

This reveals that the choice of the admissible operator $\mathscr{D}$ is crucial. First, the spectral growth order determines the smoothness of the estimated mean, which is in $\mathcal{H}_p$. Since $p > 1$, we deduce that μ_η is continuous over the sphere $\mathbb{S}^2$. Second, the shape of the zonal Green's kernel $\psi_{\mathscr{D}^*\mathscr{D}}$ determines the general form of the reconstruction. Distinct admissible operators with identical spectral growth order will have identical asymptotic performances (see Section 5) but can lead to distinct practical performances in the non-asymptotic regime, what will be exploited in Section 6.

4.2. Representer theorem for second-order estimation

We now present a representer theorem which gives the form of the estimator for the second-order moment function as the solution of (19). The proof of Theorem 2 is provided in Section 8.2.

Theorem 2. *Let $\mathscr{D}$ be an admissible operator with spectral growth $p > 1$. Let $n \geq 1$, $r_i \geq 1$ for $i = 1, \ldots n$. We consider weights $w_{ij} \in \mathbb{R}$ and pairwise distinct positions $u_{ij} \in \mathbb{S}^2$ for $i = 1, \ldots, n$ and $j = 1, \ldots, r_i$. We fix $\eta > 0$. Then, the optimization problem*

$$\min_{g \in \mathbb{H}_p} \frac{(4\pi)^2}{n} \sum_{i=1}^{n} \frac{1}{r_i(r_i - 1)} \sum_{1 \leq j \neq k \leq r_i} (w_{ij}w_{ik} - g(u_{ij}, u_{ik}))^2 + \eta\|(\mathscr{D} \otimes \mathscr{D})g\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$

(26)

has a unique solution $R_\eta \in \mathbb{H}_p$ which is given by

$$R_\eta = \sum_{i=1}^{n} \frac{1}{\sqrt{r_i(r_i - 1)}} \sum_{1 \leq j \neq k \leq r_i} \beta_{ijk}\psi_{\mathscr{D}^*\mathscr{D}}(\langle\cdot, u_{ij}\rangle) \otimes \psi_{\mathscr{D}^*\mathscr{D}}(\langle\cdot, u_{ik}\rangle)$$

(27)

for some $(\beta_{ijk})_{1 \leq i \leq n, \; 1 \leq j \neq k \leq r_i}$, where $\psi_{\mathscr{D}^*\mathscr{D}}$ is the zonal Green's kernel of $\mathscr{D}^*\mathscr{D}$.

The coefficients β_{ijk} are computed as follows. Let $L = \sum_{i=1}^{n} r_i(r_i - 1)$. We set $\boldsymbol{\beta} \in \mathbb{R}^L$ the vectorized version of $(\beta_{ijk})_{1 \leq i \leq n, \; 1 \leq j \neq k \leq r_i}$, $\boldsymbol{z} \in \mathbb{R}^L$ the vectorized version of the normalized observations $\left(\frac{w_{ij}w_{ik}}{\sqrt{r_i(r_i-1)}}\right)_{1 \leq i \leq n, \; 1 \leq j \neq k \leq r_i}$, and $\mathbf{H} \in \mathbb{R}^{L \times L}$ the matrix whose entries are given by

$$H_{\ell_1\ell_2} = \frac{\psi_{\mathscr{D}^*\mathscr{D}}(\langle u_{i_1j_1}, u_{i_2j_2}\rangle) \times \psi_{\mathscr{D}^*\mathscr{D}}(\langle u_{i_1k_1}, u_{i_2k_2}\rangle)}{\sqrt{r_{i_1}(r_{i_1} - 1)r_{i_2}(r_{i_2} - 1)}}$$

(28)

where the index ℓ_1 (resp. ℓ_2) corresponds to the triplet (i_1, j_1, k_1) (resp. (i_2, j_2, k_2)) in the vectorization. Then, we have that

$$\boldsymbol{\beta} = \left(\mathbf{H} + \frac{\eta n}{(4\pi)^2}\mathbf{I}_L\right)^{-1} \boldsymbol{z}.$$

(29)

Moreover, the estimator R_η is symmetric in the sense that $R_\eta(u,v) = R_\eta(v,u)$ for any $u,v \in \mathbb{S}^2$.

The practical vectorization of the optimization problem (26) for the specification of the β_{ijk}, which is only implicit in Theorem 2, is detailed in Section 6; see in particular (46).

Remark 4. The estimator R_η is symmetric. This is due to the use of a tensorial admissible operator $\mathscr{D} \otimes \mathscr{D}$ in the regularization and is consistent with the fact that the second-order moment R is symmetric by construction.

Moreover, R_η is linear with respect to the pairwise products $w_{ij}w_{ik}$, and it is a $(\mathscr{D}^*\mathscr{D}) \otimes (\mathscr{D}^*\mathscr{D})$-spline, in the sense that

$$(\mathscr{D}^*\mathscr{D}) \otimes (\mathscr{D}^*\mathscr{D})\{R_\eta\} = \sum_{i=1}^{n} \frac{1}{\sqrt{r_i(r_i-1)}} \sum_{1 \le j \ne k \le r_i} \beta_{ijk}\delta_{(u_{ij},u_{ik})}, \qquad (30)$$

with knots determined by the sampling locations u_{ij}. As an element of $\mathbb{H}_p$ for $p > 1$, the estimated second-order moment function R_η is also continuous on $\mathbb{S}^2 \times \mathbb{S}^2$.

Finally, we estimate the covariance function C given by $C(u,v) = R(u,v) - \mu(u)\mu(v) = (R - \mu \otimes \mu)(u,v)$ using the estimators R_η and μ_η via

$$C_\eta = R_\eta - \mu_\eta \otimes \mu_\eta. \qquad (31)$$

The estimator C_η is continuous and symmetric on $\mathbb{S}^2 \times \mathbb{S}^2$, is a $(\mathscr{D}^*\mathscr{D}) \otimes (\mathscr{D}^*\mathscr{D})$-spline in $\mathbb{H}_p$ (due to (25) and (30)), and is linear with respect to the pairwise products $w_{ij}w_{ik}$.

Remark 5. We remark that our nonparametric estimators R_η and C_η are *not* necessarily positive semi-definite in general. This is a standard occurrence in functional data analysis, not specific to our estimator, but common to all smoothing-based nonparametric methods (e.g. the PACE estimator [62], see also [27], the RKHS estimator [7], or their extensions to time series [41]). In practice, however, this has negligible effects: asymptotically positive semi-definiteness is recovered, as is proved in Section 5, whereas in finite-sample setups it can be recovered via a projection (which does not affect the rate) on the semi-definite cone, as is explained in Section 6. Indeed the asymptotic theory guarantees that negative eigenvalues will typically only be encountered at the tail end of the spectrum.

5. Asymptotic theory

This section contains the main results of our paper, that is, uniform rates of convergence for the mean and covariance estimators. In the following, we define the class of probability measures for which our rates are achieved.

Definition 4. Consider $p > 2$, $1 < q \leq p$. Let $\Pi_1(p, q)$ be the collection of probability measures for $\mathcal{H}_q$-valued processes such that for any X with probability law $\mathbb{P}_X \in \Pi_1(p, q)$

$$\mathbb{E}\|X\|^2_{\mathcal{H}_q} \leq M, \qquad \|\mu\|^2_{\mathcal{H}_p} \leq K,$$

for some constants $M, K > 0$. By extension, we say that $X \in \Pi_1(p, q)$ if it's probability law does.

Definition 5. Consider $p > 2$, $1 < q \leq p$. Let $\Pi_2(p, q)$ be the collection of probability measures for $\mathcal{H}_q$-valued processes, such that for any X with probability law $\mathbb{P}_X \in \Pi_2(p, q)$

$$\mathbb{E}\|X\|^4_{\mathcal{H}_q} \leq L, \qquad \|R\|^2_{\mathbb{H}_p} \leq K_1, \qquad \|\mu\|^2_{\mathcal{H}_p} \leq K_2,$$

for some constants $L, K_1, K_2 > 0$. By extension, we say that $X \in \Pi_2(p, q)$ if it's probability law does.

The realizations of the process X are presumed to lie in $\mathcal{H}_q$, for some $1 < q \leq p$. This entails that they are allowed to be "rougher" than the mean and covariance functions; indeed, q can be strictly smaller than p. As a result, the class of processes considered in [8] is a special case of that considered in Definition 4 for the mean estimation, *i.e.*, when $p = q$. Note indeed that, if $p = q$,

$$\|\mu\|^2_{\mathcal{H}_p} \leq \mathbb{E}\|X\|^2_{\mathcal{H}_p} \leq M,$$

and the covariance kernel (as well as the second-order moment) belongs to the direct product Hilbert space $\mathbb{H}_p = \mathcal{H}_p \otimes \mathcal{H}_p$, since

$$\|C\|^2_{\mathbb{H}_p} \leq \left(\mathbb{E}\|X - \mu\|^2_{\mathcal{H}_p}\right)^2 \leq \left(\mathbb{E}\|X\|^2_{\mathcal{H}_p}\right)^2 < \infty.$$

Moreover, if $p = q$ in Definition 5, we have that

$$\|C\|^2_{\mathbb{H}_p} \leq \mathbb{E}\|X\|^4_{\mathcal{H}_p} \leq L,$$

which yields to the class of processes considered in [7] for covariance estimation. These considerations will be discussed in more detail in Section 5.3.

Remark 6. The requirements in Definitions 4 and 5 are smoothness conditions, which form the basis for any nonparametric procedure. Their specific form amounts to classical Sobolev ellipsoid conditions (see, e.g., [59, Chapter 7]).

Our rates will be expressed in terms of the number of replicates n and the *average* number of measurement locations

$$r := \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{r_i}\right)^{-1},$$

defined as the harmonic mean of the r_i's. The results can thus be interpreted in both the *dense* and *sparse* sampling regimes. In a dense design, $r = r(n)$ is required to diverge with n and it is larger than some order of n; on the other hand, in a sparse design, the sampling frequency r is bounded and can be arbitrary small (as small as two).

5.1. Mean estimation

This section is devoted to the estimation of the mean $\mu(\cdot)$. In the following theorem we provide a (uniform) rate of convergence for the estimator given in Equation (18), under a suitable condition on the decay of the penalty parameter η. The proof is provided in Section 8.3.

Theorem 3. *Assume that the $U_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i$, are independent copies of $U \sim \mathrm{Unif}(\mathbb{S}^2)$. Let $p > 2$ and $1 < q \leq p$, and consider the estimation problem in Equation (18) for an admissible operator $\mathscr{D}$ of spectral growth order p. If $\eta \asymp (nr)^{-p/(p+1)}$, then*

$$\lim_{D \to \infty} \limsup_{n \to \infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}(\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 > D((nr)^{-p/(p+1)} + n^{-1})) = 0.$$

Corollary 1. *Let X be such that $\mathbb{P}_X \in \Pi_1(p,q)$. Under the same assumptions of Theorem 3,*

$$\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 = O_{\mathbb{P}}\left((nr)^{-p/(p+1)} + n^{-1}\right).$$

Theorem 3 tells us that the estimator μ_η *achieves* the rate $(nr)^{-p/(p+1)} + n^{-1}$ uniformly over the class $\Pi_1(p,q)$ and hence such rate is called *achievable* for that class (see [51]).

It is important to remark that we can observe a phase transition phenomenon with a boundary at $r = n^{1/p}$, which allows to discriminate between sparse and dense sampling regimes. Indeed, when the sampling frequency r is small, that is, $r = O(n^{1/p})$, we have

$$\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 = O_{\mathbb{P}}\left((nr)^{-p/(p+1)}\right),$$

which is equivalent to the rate of a smoothing spline estimator in nonparametric regression based on nr independent observations. In other words, the convergence rate is not affected by the spatial dependence. In the case of high sampling frequency with $r \gg n^{1/p}$, we have a parametric rate

$$\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 = O_{\mathbb{P}}\left(n^{-1}\right),$$

that does not depend on r, as in a classical FDA approach where one can observe $X_1, \ldots, X_n$ continuously on $\mathbb{S}^2$.

All the previous definitions and results apply also to hypersphere $\mathbb{S}^d$, $d > 2$. In particular, *hyperspherical harmonics* (see for instance [46, Chapter 3]) can be used to prove the rate, that is $(nr)^{-2p/(2p+d)} + n^{-1}$, for $p > d$ and $d/2 < q \leq p$.

Remark 7 (On the distribution of the measurement locations). Theorem 3 is stated and proved for independent measurament locations $U_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i$, *uniformly distributed* over the sphere. It will be clear in the proofs below that, for this specific case, one can work directly with the L^2- and $\mathcal{H}_p$- norms

introduced in Section 2, *i.e.*, $\|\cdot\|_{L^2(\mathbb{S}^2)}$ and $\|\mathscr{D}\cdot\|_{L^2(\mathbb{S}^2)}$, and the corresponding *intermediate* spaces (see [32, Chapter 4] for further details on intermediate norms and spaces). The set of spherical harmonics $\{Y_{\ell,m}\}$ also plays a crucial role, as it forms an orthonormal basis for $L^2(\mathbb{S}^2)$ and an orthogonal basis for $\mathcal{H}_p$. However, the results can be generalized to any other (common) probability density $p_U(\cdot)$ supported on $\mathbb{S}^2$ and bounded away from 0 and infinity, that is,

$$0 < \inf_{u \in \mathbb{S}^2} p_U(u) \leq \sup_{u \in \mathbb{S}^2} p_U(u) < \infty.$$

Therefore, we can define two new inner products

$$(f, g) \mapsto 4\pi \int_{\mathbb{S}^2} f(u)g(u)p_U(u)du, \tag{32}$$

$$(f, g) \mapsto 4\pi \int_{\mathbb{S}^2} f(u)g(u)p_U(u)du + \langle \mathscr{D}f, \mathscr{D}g \rangle_{L^2(\mathbb{S}^2)}, \tag{33}$$

and their corresponding norms which are respectively equivalent to $\| \cdot \|_{L^2(\mathbb{S}^2)}$ and $\|\mathscr{D} \cdot \|_{L^2(\mathbb{S}^2)}$. Such equivalence allows to establish results parallel to those in the uniform case. Indeed, it is possible to show that there exists a set of functions $\{\psi_j\}$ that forms an orthonormal basis for $L^2(\mathbb{S}^2)$ endowed with (32) and an orthogonal basis for $\mathcal{H}_p$ endowed with (33). We are also able to define intermediate norms and associated intermediate spaces. The reader is referred in particular to [28, Section 2.4] and the references therein. It is clear the the set $\{\psi_j\}$ plays the same role as the set of spherical harmonics and we want to stress that (apart from orthonormality/orthogonality) no other specific properties of spherical harmonics have been used to prove our results. Hence, the generalization can be obtained by following exactly the same steps; see also [29] for an application.

5.2. Covariance estimation

In this section, we present our main result, which concerns the estimation of the covariance function $C(\cdot, \cdot)$. The following theorem gives a (uniform) rate of convergence for the estimator given in Equation (20), under a suitable condition on the decay of the penalty parameter η. The proof is provided in Section 8.3.

Theorem 4. *Assume that $\mathbb{E}[\epsilon_{11}^4] < \infty$ and the $U_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i,$ are independent copies of $U \sim \text{Unif}(\mathbb{S}^2)$. Let $p > 2$ and $1 < q \leq p$, and consider the estimation problem in Equation (20) for an admissible operator $\mathscr{D}$ of spectral growth order p. If $\eta \asymp (nr/\log n)^{-p/(p+1)}$, then*

$$\lim_{D \to \infty} \limsup_{n \to \infty} \sup_{\mathbb{P}_X \in \Pi_2(p,q)} \mathbb{P}\left(\|C_\eta - C\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 > D\left(\left(\frac{\log n}{nr} \right)^{p/(p+1)} + \frac{1}{n} \right) \right)$$
$$= 0.$$

Corollary 2. *Let X be such that $\mathbb{P}_X \in \Pi_2(p, q)$. Under the same assumptions of Theorem 4,*

$$\|C_\eta - C\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = O_\mathbb{P}\left(\left(\frac{\log n}{nr}\right)^{p/(p+1)} + \frac{1}{n}\right).$$

Remark 8. A careful inspection of the proof reveals that any estimator $\hat{\mu}(\cdot)$ of the mean function that satisfies

$$\lim_{D\to\infty} \limsup_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_2(p,q)} \mathbb{P}\left(\|\hat{\mu} - \mu\|^2_{L^2(\mathbb{S}^2)} > D\left(\left(\frac{\log n}{nr}\right)^{p/(p+1)} + \frac{1}{n}\right)\right) = 0$$

can be used. The estimator $\mu_\eta(\cdot)$ proposed in Section 5.1 satisfies this condition both when $\eta \asymp (nr)^{-p/(p+1)}$ (see Theorem 3) or when $\eta \asymp (nr/\log n)^{-p/(p+1)}$ (see Proof of Theorem 4).

In Theorem 4 we show that the estimator C_η achieves a uniform rate $(\log n/nr)^{p/(p+1)} + n^{-1}$ on the class $\Pi_2(p, q)$. As was the case for the mean, we have a phase transition, this time at $r = n^{1/p} \log n$. Indeed, when the functions are densely sampled, *i.e.*, $r \gg n^{1/p} \log n$, the sampling frequency has no impact and the rate is

$$\|C_\eta - C\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = O_\mathbb{P}\left(n^{-1}\right),$$

which suggests that, with a sufficient number of measurement locations per surface, the covariance function can be estimated as well as if the $X_1, \ldots, X_n$ can be observed on the whole domain $\mathbb{S}^2$. On the other hand, when the functions are sparsely sampled, the rate is jointly determined by n and r, namely,

$$\|C_\eta - C\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = O_\mathbb{P}\left(\left(\frac{\log n}{nr}\right)^{p/(p+1)}\right).$$

We can observe that, the estimator is based on a total of $nr(r - 1)$ paired observations. However, the rate is expressed in terms of just nr. This can be explained by the fact that we are actually observing nr W_{ij}'s and hence there is a significant redundancy among the off-diagonal products.

As was the case for the mean, the result can be extended to the hypersphere $\mathbb{S}^d$, $d > 2$, with rate $(\log n/(nr))^{2p/(2p+d)} + n^{-1}$, for $p > d$ and $d/2 < q \le p$.

Remark 9. A similar reasoning to that expressed in Remark 7 can be applied to the covariance estimator. Indeed, if the measurement locations are independently drawn from a common distribution with probability density $p_U(\cdot)$ supported on $\mathbb{S}^2$ and bounded away from 0 and infinity, we can define the inner products

$$(f, g) \mapsto (4\pi)^2 \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} f(u, v) g(u, v) p_U(u) p_U(v) du dv, \tag{34}$$

$$(f,g) \mapsto (4\pi)^2 \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} f(u,v)g(u,v)p_U(u)p_U(v)dudv$$
$$+ \langle (\mathscr{D} \otimes \mathscr{D})f, (\mathscr{D} \otimes \mathscr{D})g \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}, \tag{35}$$

and their corresponding norms which are respectively equivalent to $\|\cdot\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$ and $\|(\mathscr{D} \otimes \mathscr{D}) \cdot \|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$. Accordingly, the set $\{\psi_j \otimes \psi_{j'}\}$ forms an orthonormal basis for $L^2(\mathbb{S}^2 \times \mathbb{S}^2)$ endowed with (34) and an orthogonal basis for $\mathbb{H}_p$ endowed with (35). All the results can be then extended.

5.3. Examples

In this section, we exemplify the previous results for two classes of spherical Gaussian random fields. This will reveal the interest of considering the classes $\Pi_2(p,q)$ with distinct values for p and q for the asymptotic analysis of the performance of our estimation strategy.

For $\beta > 1$, we say that the spherical random field X is in the class $\mathcal{A}_\beta$ if

$$X = \mathscr{D}^{-1}W \tag{36}$$

where $\mathscr{D}$ is an admissible operator with spectral growth order β and W is a Gaussian white noise such that $\mathbb{E}[\langle W, f \rangle_{L^2(\mathbb{S}^2)} \langle W, g \rangle_{L^2(\mathbb{S}^2)}] = \sigma^2 \langle f, g \rangle_{L^2(\mathbb{S}^2)}$ for any $f, g \in L^2(\mathbb{S}^2)$. In the next proposition, we quantify the rate of convergence of our estimation strategy for spherical random fields in $\mathcal{A}_\beta$. The proof of Proposition 2 is given in Section 8.3.

Proposition 2. *Let $\beta > 5/2$. Then, we have that $\mathcal{A}_\beta \subset \Pi_2(p,q)$ for any $1 \leq q \leq p$ such that $p < \beta - 1/2$ and $q < \beta - 1$. Moreover, if $X \in \mathcal{A}_\beta$ is a spherical random field with covariance C, then for any $\epsilon > 0$, there exists an estimator C_η given by (31) such that*

$$\|C_\eta - C\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = O_{\mathbb{P}}\left(\left(\frac{\log n}{nr} \right)^{\frac{\beta - 1/2}{\beta + 1/2} - \epsilon} + \frac{1}{n} \right). \tag{37}$$

The second class of spherical random fields that we consider is as follows. For $\beta > 1$, we say that $X \in \mathcal{B}_\beta$ if

$$X = \mathscr{D}^{-1}\left\{ \sum_{k=1}^{Q} \xi_k \delta_{u_k} \right\} = \sum_{k=1}^{Q} \xi_k \psi_{\mathscr{D}}(\langle \cdot, u_k \rangle) \tag{38}$$

with $\mathscr{D}$ an admissible operator of order β and with zonal Green's kernel $\psi_{\mathscr{D}}$, $Q \geq 1$, $u_1, \ldots, u_Q \in \mathbb{S}^2$ distinct, and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_Q) \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ an i.i.d. Gaussian vector. Note that X is simply a random $\mathscr{D}$-spline with Gaussian weights. In particular, X is located in the finite-dimensional space of splines with Q knots at locations $u_1, \ldots, u_Q$. The asymptotic performance for the estimation of random fields in $\mathcal{B}_\beta$ is quantified in Proposition 3, whose proof is in Section 8.3.

Proposition 3. *Let $\beta > 2$. Then, we have that $\mathcal{B}_\beta \subset \Pi_2(p,q)$ for any $1 \leq q \leq p < \beta - 1$. Moreover, if $X \in \mathcal{B}_\beta$ is a spherical random field with covariance C, then for any $\epsilon > 0$, there exists an estimator C_η given by* (31) *such that*

$$\|C_\eta - C\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 = O_{\mathbb{P}}\left(\left(\frac{\log n}{nr}\right)^{\frac{\beta-1}{\beta}-\epsilon} + \frac{1}{n}\right). \tag{39}$$

Remark 10. The spherical random fields in $\mathcal{A}_\beta$ and $\mathcal{B}_\beta$ are both in the Sobolev spaces $\mathcal{H}_q$ for any $q < \beta - 1$. However, as seen in the proofs of Propositions 2 and 3, their covariances have distinct Sobolev regularities, measured in the tensorial spaces $\mathbb{H}_p$. We see in Theorem 4 that the Sobolev regularity of the covariance is crucial for the convergence rate of our estimation strategy. For this reason, the distinction between the regularity of the random field (parameter q in $X \in \Pi_2(p,q)$) and its covariance (parameter p in $X \in \Pi_2(p,q)$) is crucial to obtain the best possible convergence rate.

In particular, Propositions 2 and 3 reveal that we can approach the critical bounds $\frac{\beta-1/2}{\beta+1/2}$ for $X \in \mathcal{A}_\beta$ and $\frac{\beta-1}{\beta}$ for $X \in \mathcal{B}_\beta$. If we achieve the state-of-the-art results of [7] for the latter, we are able to improve existing rates for the former class $\mathcal{A}_\beta$.

Finally, it is worth noting that we cannot achieve the rates $\frac{\beta-1/2}{\beta+1/2}$ (for $X \in \mathcal{A}_\beta$) and $\frac{\beta-1}{\beta}$ (for $X \in \mathcal{B}_\beta$) with our technique. Indeed, this reflects the fact that the covariances of these random fields do not reach their critical Sobolev smoothness (see the proofs in Section 8.3), hence requiring to lose $\epsilon > 0$ arbitrarily small in the rates (37) and (39). This is consistent with existing works [7].

6. Simulations and computational aspects

In this section, we investigate the practical feasibility of our estimation methodology. We use simulated sparse samples of an anisotropic Gaussian spherical field as a test case study. Since this appears to be the first and only numerical procedure for nonparametric estimation of anisotropic spherical covariance operators from sparse functional data, we do not offer a comparison. Rather, our main focus is on demonstrating the practical feasibility of our estimation methodology. Since the estimation procedure described in Section 3 decouples the estimation of the first- and second-order moments, we assume without loss of generality a *zero-mean* Gaussian random field and focus exclusively on the (computationally more challenging) second-order moment estimate (19), for which we propose, leveraging Theorem 2 and sparse/second-order structure, a *computationally tractable* and *memory-thrifty* implementation. For details regarding the implementation of the mean estimate (18) we refer to the literature on *smoothing splines*, of which (18) is a specific instance (see Remark 3). Smoothing spline estimates are standard practice in non-parametric regression [17], and their use in the spherical setting was already proposed in the literature – see for example [31, Section 6.4.2] and the references therein as well as [46, Chapter 9]

where smoothing splines are considered for the estimation of global temperature anomaly maps from sparse recordings collected by the Argo fleet [2, 26].

The simulation code is openly available and fully-reproducible, and is released in the form of a standalone Python 3 Jupyter notebook hosted on GitHub: https://github.com/matthieumeo/sphericov/blob/main/companion_nb.ipynb. Instructions for installing the required dependencies and running the notebook are available at: `https://github.com/matthieumeo/sphericov#readme`.

Simulation setup   Our simulation setup is as follows. Consider a spherical point set $\mathcal{V}_0 = \{v_1, \ldots, v_Q\} \subset \mathbb{S}^2$ and let $S_\nu^\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ denote the *Matérn function* with smoothness parameter $\nu > 0$ and scale parameter $\varepsilon > 0$ [61, Eq. 4.16]. Define moreover the *spherical Matérn function* $\psi_\nu^\varepsilon : [-1, 1] \to \mathbb{R}_+$ as [47, Section 5.3]

$$\psi_\nu^\varepsilon(t) = S_\nu^\varepsilon(\sqrt{2 - 2t}), \qquad \forall t \in [-1, 1]. \tag{40}$$

It is possible to show that ψ_ν^ε in (40) is the zonal Green's kernel of a certain admissible spherical pseudo-differential operator[1] $\mathscr{D}_\nu^\varepsilon$ with spectral growth $p = 2(\nu + 1)$ [47, Section 5.3.1]. Following the nomenclature of [47, Section 5.3.1], we refer to $\mathscr{D}_\nu^\varepsilon$ as a *Matérn operator*.

We consider an underlying spherical random field $X = \{X(u), u \in \mathbb{S}^2\}$ taking the form of a $\mathscr{D}_\nu^\varepsilon$-*sparse random field* [55]:

$$\mathscr{D}_\nu^\varepsilon X = \sum_{q=1}^Q \xi_q \delta_{v_q} \quad \Leftrightarrow \quad X(u) = \sum_{q=1}^Q \xi_q \psi_\nu^\varepsilon(\langle u, v_q \rangle) \quad \forall u \in \mathbb{S}^2, \tag{41}$$

where $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_Q] \sim \mathcal{N}_Q(\mathbf{0}, \mathbf{R})$ for some covariance matrix $\mathbf{R} \in \mathbb{R}^{Q \times Q}$ – see also Section 5.3. Roughly speaking, the random field (41) is the primitive with respect to the pseudo-differential operator $\mathscr{D}_\nu^\varepsilon$ of a random discrete measure composed of finitely many Dirac measures. It is easy to see that X is a *Gaussian random field*, with mean zero and second-order moment given by:

$$R(u, v) = \sum_{p,q=1}^Q R_{pq} \psi_\nu^\varepsilon(\langle u, v_p \rangle) \psi_\nu^\varepsilon(\langle v, v_q \rangle), \qquad (u, v) \in \mathbb{S}^2 \times \mathbb{S}^2. \tag{42}$$

Note that since X has zero mean, the bivariate function R coincides in this case with the covariance function of the field. As described in Section 3, our input data consists in simulated realizations of noisy random samples of i.i.d. replicates $X_1, \ldots, X_n$ of the random field X:

$$W_{ij} = X_i(U_{ij}) + \epsilon_{ij}, \qquad i = 1, \ldots, n, \, j = 1, \ldots, r_i,$$

where the random sampling locations are distributed uniformly over the sphere $U_{ij} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}(\mathbb{S}^2)$, and the measurement errors are distributed as $\epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ for some $\sigma > 0$. Moreover, the random fields, sampling locations, and measurement errors are all assumed mutually independent. Realizations of the random variables W_{ij} and U_{ij} are denoted in lower-case notation, *i.e.*, w_{ij} and u_{ij} respectively.

---

[1]In the sense of Definition 2.
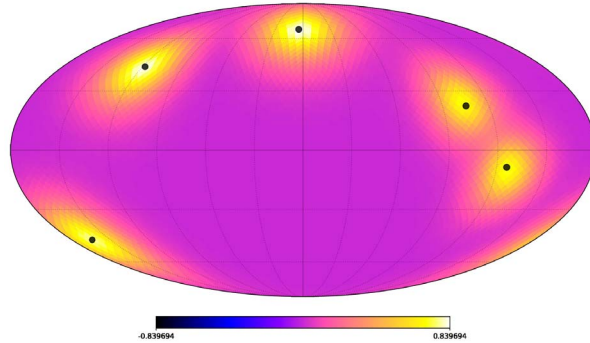
(a) Diagonal slice $u \mapsto R(u, u)$, $u \in \mathbb{S}^2$ (*i.e.*, variance function) of the second-order moment kernel. The locations of the sources $\{v_1, \ldots, v_Q\}$ in (41) are overlaid as black scatters.
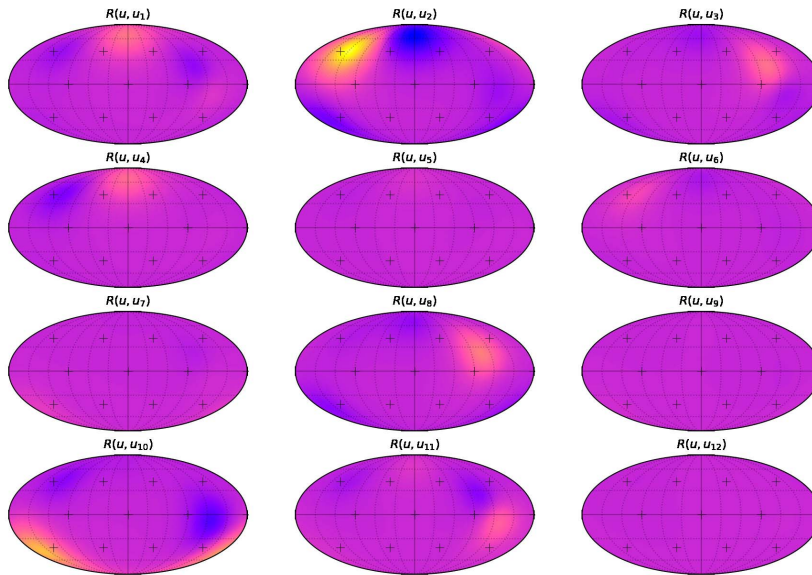
(b) Slices $u \mapsto R(u, u_p)$, $u \in \mathbb{S}^2$ of the second-order moment kernel for 12 points $\{u_1, \ldots, u_P\}$ (overlaid as black crosses).

FIG 2. *Mollweide projections of covariance kernel slices* $R : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ *of the Gaussian random field* X (41) *used in our simulations.*

Simulation parameters In our simulations, we set the various parameters listed above to the following values:

- **Matérn Function:** The smoothing parameter is set to $\nu = 5/2$. For such a value of ν, the Matérn function admits the following simple closed-form

expression:

$$S_{5/2}^{\varepsilon}(t) = \left(1 + \frac{\sqrt{5}t}{\varepsilon} + \frac{5t^2}{3\varepsilon^2}\right)\exp\left(-\frac{\sqrt{5}t}{\varepsilon}\right), \qquad \forall t \geq 0.$$

The scale parameter is set to an arbitrary value of $\varepsilon = 0.4$.

- **Random Field:** We set the number of sources to $Q = 5$, draw the spherical point set $\mathcal{V}_0$ uniformly at random, and choose the covariance matrix **R** in (41) as:

$$\mathbf{R} = \begin{bmatrix} 0.812 & -0.013 & -0.209 & -0.416 & -0.028 \\ -0.013 & 0.974 & -0.008 & -0.632 & -0.372 \\ -0.209 & -0.008 & 0.909 & -0.095 & -0.588 \\ -0.416 & -0.632 & -0.095 & 1.000 & 0.235 \\ -0.028 & -0.372 & -0.588 & 0.235 & 0.929 \end{bmatrix}.$$

We plot in Figure 2 slices of the second-order moment kernel R of the random field X obtained this way.

- **Data Simulation:** We set the number of replicates of X to $n = 64$ and consider a fixed number $r_i = r = 12$ of spatial samples per replicates. The noise level is chosen as $\sigma = 0.1$, yielding a *peak signal-to-noise ratio (PSNR)* of 10 dB.

Numerical estimation procedure We consider the following estimator for R:

$$R_\eta = \underset{g \in \mathbb{H}_p}{\arg\min} \frac{4\pi^2}{L} \sum_{i=1}^{n} \sum_{1 \leq j \neq k \leq r} (w_{ij}w_{ik} - g(u_{ij}, u_{ik}))^2 + \eta \|(\mathscr{D}_\xi^\varepsilon \otimes \mathscr{D}_\xi^\varepsilon)g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2,$$

$$(43)$$

for $L = nr(r-1)$ and some regularization parameter $\eta > 0$ (the selection of this regularization parameter is addressed in the next section). The smoothing parameter ξ of the Matérn factors $\mathscr{D}_\xi^\varepsilon$ in (43) is set to $\xi = (\nu-1)/2$, which is the minimal value so that the unknown second-order moment function R defined in (42) belongs to the search space $\mathbb{H}_p$ with $p = 2(\nu+1)$. Note that (43) coincides indeed with the second-order moment estimator (19) discussed in Section 3 when the number of random spatial samples r_i per replicate X_i is constant and equal to r (as assumed in this simulation setup). In which case indeed, the normalising constants involved in the least-square term of (19) reduce to a single factor $4\pi^2/L$, hence yielding the simpler expression (43) above. Theorem 2 applied to (43) reveals that R_η is given by:

$$R_\eta(u, v) = \sum_{i=1}^{n} \sum_{1 \leq j \neq k \leq r} \beta_{ijk} \psi_\nu^\varepsilon(\langle u, u_{ij}\rangle)\psi_\nu^\varepsilon(\langle v, u_{ik}\rangle), \qquad \forall (u, v) \in \mathbb{S}^2 \times \mathbb{S}^2. \quad (44)$$

for some coefficients $(\beta_{ijk})_{1 \leq i \leq n,\ 1 \leq j \neq k \leq r}$. The latter are moreover obtained as solutions of the following linear system of size L:

$$\left(\mathbf{H} + \frac{\eta L}{4\pi^2}\mathbf{I}_L\right)\boldsymbol{\beta} = \boldsymbol{z}, \qquad (45)$$

where $\boldsymbol{\beta} \in \mathbb{R}^L$ and $\boldsymbol{z} \in \mathbb{R}^L$ are vectorized versions of the coefficients $(\beta_{ijk})_{1 \leq i \leq n,\ 1 \leq j \neq k \leq r}$ and data $(w_{ij} w_{ik})_{1 \leq i \leq n,\ 1 \leq j \neq k \leq r}$ respectively. In practice, this vectorization is performed by mapping the multi-index triplet (i, j, k) to a single index ℓ as follows:

$$\ell = (i - 1)r(r - 1) + (k - 1)(r - 1) + j - \mathbf{1}\{j > k\}, \tag{46}$$

for $i \in [\![1, n]\!]$, $k \in [\![1, r]\!]$, $j \in [\![1, r]\!] \backslash \{k\}$, where $[\![1, n]\!] = \{1, \ldots, n\}\ \forall n \in \mathbb{N}$ and $\mathbf{1}\{j > k\} = 1$ if $j > k$ and 0 otherwise. This vectorization can easily be performed in practice by filling the off-diagonal terms of n matrices with size $r \times r$ with the coefficients/data (*e.g.*, β_{ijk} is put in the i-th $r \times r$ matrix, at row j and column k), and then considering the vector formed by stacking vertically flattened versions of the matrices, obtained by stacking for each matrix the r columns on top of one another and skipping the diagonal terms. This process is of course reversible, and the map to obtain the multi-index triplets (i, j, k) from the index ℓ can be shown to be given by:

$$\begin{cases} i = 1 + (\ell \mathbin{/\!/} (r(r - 1))) \\ k = 1 + (\ell \mathbin{\%} (r(r - 1))) \mathbin{/\!/} (r - 1) \\ h = (\ell \mathbin{\%} (r(r - 1))) \mathbin{\%} (r - 1) \\ j = h + \mathbf{1}\{h \geq k\} \end{cases}, \qquad \ell \in [\![1, L]\!], \tag{47}$$

where $\mathbin{/\!/}$ and $\mathbin{\%}$ are the *floor-divide* and *remainder* operators defined as:

$$n \mathbin{/\!/} m = \lfloor n/m \rfloor, \qquad n \mathbin{\%} m = n - \lfloor n/m \rfloor, \qquad \forall (n, m) \in \mathbb{N} \times \mathbb{N}.$$

With the vectorization scheme (46) and (47), it is possible to show that the matrix $\mathbf{H} \in \mathbb{R}^{L \times L}$ is given by

$$\mathbf{H} = \mathbf{S} \begin{bmatrix} \mathbf{J}_{11} \otimes \mathbf{J}_{11} & \cdots & \mathbf{J}_{1n} \otimes \mathbf{J}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{J}_{n1} \otimes \mathbf{J}_{n1} & \cdots & \mathbf{J}_{nn} \otimes \mathbf{J}_{nn} \end{bmatrix} \mathbf{S}^T = \mathbf{S} \left(\mathbf{J} * \mathbf{J} \right) \mathbf{S}^T, \tag{48}$$

where $\otimes$ and $*$ denote the *Kronecker* and *block Kronecker* or *Khatri-Rao* products respectively, and:

- $\mathbf{S} \in \mathbb{R}^{L \times nr^2}$ is a subsampled nr^2 identity matrix, where the rows with indices given by $m = (i - 1)r^2 + (k - 1)r + k$, $i \in [\![1, n]\!]$, $k \in [\![1, r]\!]$ have been removed.
- $\{\mathbf{J}_{pq} \in \mathbb{R}^{r \times r}, (p, q) \in [\![1, n]\!]^2\}$ are Gram matrices, with entries given by:

$$(\mathbf{J}_{pq})_{gh} = \psi_\nu^\varepsilon(\langle u_{pg}, u_{qh} \rangle), \qquad \forall (g, h) \in [\![1, r]\!]^2. \tag{49}$$

- $\mathbf{J} \in \mathbb{R}^{nr \times nr}$ is a block matrix given by:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{11} & \cdots & \mathbf{J}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{J}_{n1} & \cdots & \mathbf{J}_{nn} \end{bmatrix}.$$

Notice moreover that the submatrices $\mathbf{J}_{pq}$ defined in (49) are *sparse*[2]. Indeed, the Matérn zonal function ψ_ν^ε is such that $\psi_\nu^\varepsilon(\cos\theta) \simeq 0$ when $|\theta| \geq 3\arccos\varepsilon$ [47, Section 5.3.1], and hence the entries of the submatrices $\mathbf{J}_{pq}$ are zero whenever the angle $\theta_{pg,qh} = \arccos\langle u_{pg}, u_{qh}\rangle$ is sufficiently large, *i.e.*, the sampling locations u_{pg}, u_{qh} are sufficiently far apart on the sphere. In practice, we can leverage this sparsity and the block-Kronecker structure to compute (48) efficiently and represent it as a sparse matrix with a relatively low memory footprint. Our procedure is as follows:

1. Compute the submatrices defined in (49) and sparsify them by setting to zeros all the entries smaller than $0.01 \times \psi_\nu^\varepsilon(1)$.
2. Build $\mathbf{J}$ and store it as a sparse matrix [45].
3. Compute in a multi-threaded fashion the sparse block Kronecker product $\mathbf{J} * \mathbf{J}$.
4. Form $\mathbf{H}$ by multiplying the sparse output of the block Kronecker product by $\mathbf{S}$ and $\mathbf{S}^T$ from the left and right respectively.

These steps are performed in practice by leveraging the `sparse` [1] and `dask` [40] Python libraries, used respectively for sparse and distributed computations. Note that without this memory- and compute-efficient procedure, computing $\mathbf{H}$ in practice could quickly become intractable, due to its high dimensionality. For $n = 128$ and $r = 64$ already, the matrix $\mathbf{H}$ with size $524'288 \times 524'288$ would require TFlops/TBytes of computation/memory to be computed/stored as a dense matrix. Leveraging the sparse and block Kronecker structure allows us to bring down this computation/memory footprint by two to three orders of magnitude approximately, *i.e.*, GFlops/GBytes of computation/memory required.

Upon calculation of the matrix $\mathbf{H}$, we solve (45) by means of conjugate gradient descent. Our implementation leverages the routine `cg()` from Scipy's module `scipy.sparse.linalg` [57], which is compatible with sparse matrices. For the simulation setup under consideration, solving (45) numerically took on the order of a few dozen seconds.

Accuracy and selection of the penalty parameter   We assess the accuracy of our estimator by means of *4-fold cross-validation*. More specifically, we shuffle the data vector $\mathbf{z}$ by applying a random permutation to the latter, split it in 4 groups of equal size, train the model on three of these groups and compute the *out-of-sample mean squared error (MSE)* on the last group. We moreover loop (in parallel) through every combination of training/test configurations and consider the average MSE score over all configurations, yielding the final cross-validation score. We investigate different values of penalty parameters in the range $[1, 6]$ and select the one yielding an estimator with minimal cross-validation score, that is $\tilde{\eta} \simeq 2.363$. The cross-validation scores of the various estimators are plotted in Figure 3.

---

[2]In practice we observe that fewer than 9 % of the entries of $\mathbf{J}$ are significantly different from zero.
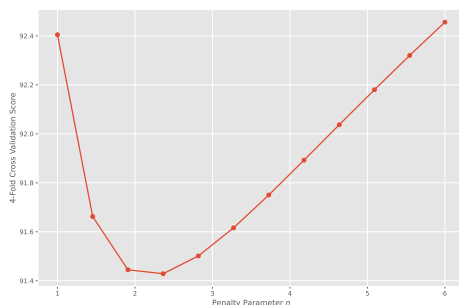
FIG 3. *4-fold Cross-validation score of various estimates R_η for $\eta \in [1, 6]$.*

Results and discussion   We plot in Figure 4 slices of the second-order moment estimator $R_{\tilde\eta}$ corresponding to the optimal penalty parameter value $\tilde\eta \simeq$ 2.363 determined by cross-validation. A comparison of Figures 2 and 4 reveals that the estimate $R_{\tilde\eta}$ approximates fairly well the actual second-order function R. Indeed, most of the main features of the variance/covariance maps in Figure 2 are still clearly visible in Figure 4, although slightly blurred and/or less contrasted. One issue with the estimator however is that it is *not* positive definite, which can be problematic if the latter is used in a kriging context. As a matter of fact, the diagonal $R_{\tilde\eta}(u, u)$ is not even always positive, as can be seen in Figure 4a where negative values in the field are marked by a semi-transparent overlay. This phenomenon is likely to arise in practical setups with finite sample sizes, since the positive definiteness of our estimator is only guaranteed asymptotically (see Theorems 2 and 4). One potential remedy consists in projecting the estimator $R_{\tilde\eta}$ on the positive semi-definite cone. This can be achieved approximately in practice by discretizing $R_{\tilde\eta}$ on a fine (quasi-)uniform spherical grid [46] (*e.g.*, the HEALPix grid [16]), and setting the negative eigenvalues of the discretized operator to zero. This however, requires computing the eigenvalue decomposition of a potentially very high dimensional operator, which can reveal very computationally and memory intensive. Figure 5 shows the projection $R_{\tilde\eta}^+$ of the estimator $R_{\tilde\eta}$ onto the positive semi-definite cone. The effect of the projection onto the estimate is mostly visible onto the variance map, which no longer exhibits negative regions and aligns slightly better with the underlying sources.

7. Extension to the serially dependent case

This section provides an extension of the previous results in the presence of serial dependence. We thus relax the assumption of independence among the $X_1, \ldots, X_n$ and we introduce serial dependence, assuming that the indices represent *time*. In doing so, we assume that the resulting temporal sequence is time-stationary (*i.e.*, a stationary functional time series).

Consider the collection of second order processes $X_t = \{X_t(u), u \in \mathbb{S}^2\}$,
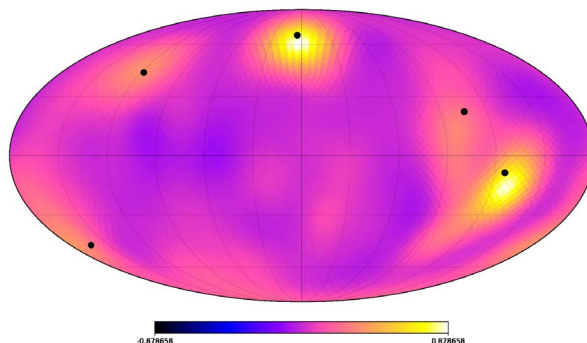
(a) Diagonal slice $u \mapsto R_{\tilde{\eta}}(u, u)$, $u \in \mathbb{S}^2$ of the second-order moment estimator. Negative values in the field are marked by a semi-transparent overlay. The locations of the sources in (41) are overlaid as black scatters.
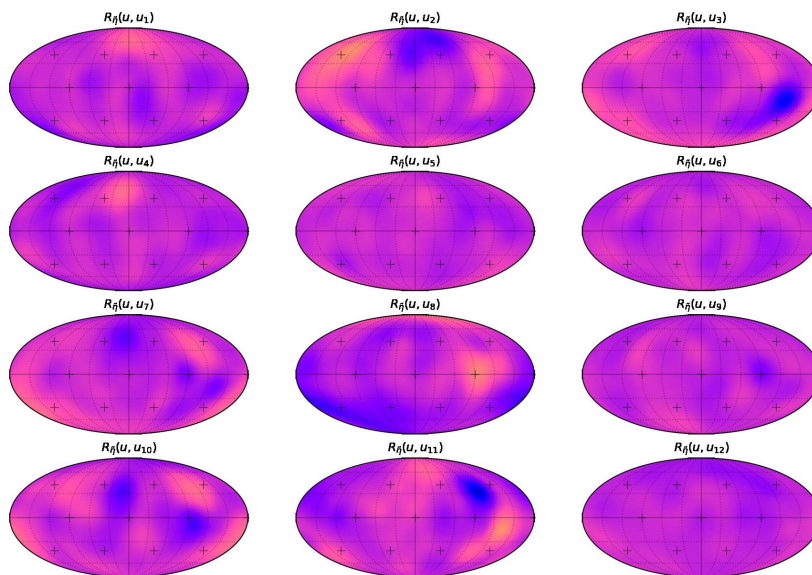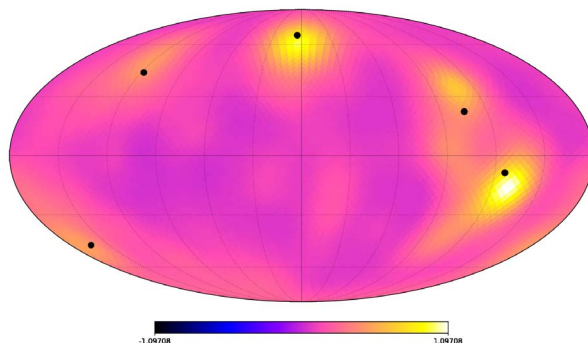

(b) Slices $u \mapsto R_{\tilde{\eta}}(u, u_p)$, $u \in \mathbb{S}^2$ of the second-order moment estimator for 12 points $\{u_1, \ldots, u_P\}$ (overlaid as black crosses).
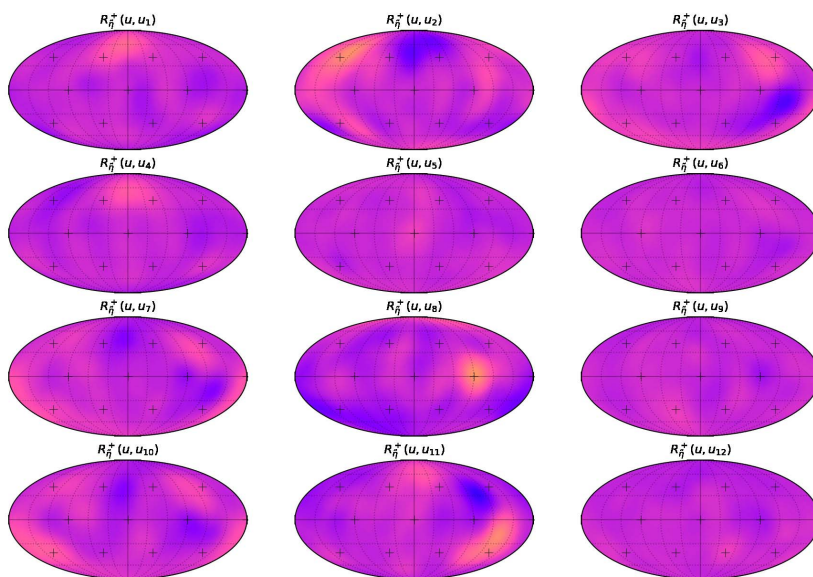
Fig 4. *Mollweide projections of second-order moment estimator slices $R_{\tilde{\eta}} : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$.*

$t \in \mathbb{Z}$, which are also random elements of $\mathcal{H}_q$, $q > 1$. Throughout, we assume that the sequence $\mathcal{X} = \{X_t, t \in \mathbb{Z}\}$ is *strictly stationary*: for any finite set of indices $I \subset \mathbb{Z}$ and any $h \in \mathbb{Z}$, the joint law of $\{X_t, t \in I\}$ coincides with that of $\{X_{t+h}, t \in I\}$. Then, the first and second order properties of $\mathcal{X}$ are summarized by

$$\mathbb{E}[X_t(u)] = \mu(u), \qquad \mathbb{E}[X_{t+h}(u)X_t(v)] = R_h(u, v),$$

(a) Diagonal slice $u \mapsto R_{\tilde{\eta}}^+(u,u)$, $u \in \mathbb{S}^2$ of the projected second-order moment estimator. The locations of the sources in (41) are overlaid as black scatters.

(b) Slices $u \mapsto R_{\tilde{\eta}}^+(u, u_p)$, $u \in \mathbb{S}^2$ of the projected second-order moment estimator for 12 points $\{u_1, \dots, u_P\}$ (overlaid as black crosses).

FIG 5. *Mollweide projections of the projected second-order moment estimator slices $R_{\tilde{\eta}}^+$: $\mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$.*

$$C_h(u,v) = R_h(u,v) - \mu(u)\mu(v),$$

for $u, v \in \mathbb{S}^2$, respectively the mean function, the second-order moment function at *lag* $h \in \mathbb{Z}$ and autocovariance function at *lag* $h \in \mathbb{Z}$.

If $\mathbb{E}\|X_0\|_{L^2(\mathbb{S}^2)}^k < \infty$ (or $\mathbb{E}\|X_0\|_{\mathcal{H}_q}^k < \infty$) for $k \geq 1$, the k-th order cumulant

kernel is well defined in a L^2 sense (or pointwise)

$$\mathrm{Cum}[X_{h_1}(u_1), \ldots, X_{h_k}(u_k)] = \sum_\pi (-1)^{|\pi|-1}(|\pi|-1)! \prod_{B \in \pi} \mathbb{E}\left[\prod_{j \in B} X_{h_j}(u_j)\right],$$

where the first sum is taken over the list of all unordered partitions of $\{1, \ldots, k\}$, and $|\pi|$ is the cardinality of the partition π. For $k \geq 1$, we can also define the $2k$-th order cumulant operator $\mathcal{C}_{h_1, \ldots, h_{2k-1}} : L^2((\mathbb{S}^2)^k) \to L^2((\mathbb{S}^2)^k)$ as

$$(\mathcal{C}_{h_1, \ldots, h_{2k-1}} f)(u_1, \ldots, u_k) = \int_{(\mathbb{S}^2)^k} \mathrm{Cum}[X_{h_1}(u_1), \ldots, X_{h_{2k-1}}(u_{2k-1}), X_0(u_{2k})]$$

$$\times f(u_{k+1}, \ldots, u_{2k}) du_{k+1} \cdots du_{2k},$$

and the $(2k+1)$-th order cumulant operator $\mathcal{C}_{h_1, \ldots, h_{2k}} : L^2((\mathbb{S}^2)^{k+1}) \to L^2((\mathbb{S}^2)^k)$ as

$$(\mathcal{C}_{h_1, \ldots, h_{2k}} f)(u_1, \ldots, u_k) = \int_{(\mathbb{S}^2)^{k+1}} \mathrm{Cum}[X_{h_1}(u_1), \ldots, X_{h_{2k}}(u_{2k}), X_0(u_{2k+1})]$$

$$\times f(u_{k+1}, \ldots, u_{2k+1}) du_{k+1} \cdots du_{2k+1}.$$

Cumulant kernels and operators in L^2 spaces were first introduced in [33] for even orders. We extended the definition to odd-order cumulants and we also give an alternative definition in the Sobolev spaces described in Section 2.

Denote with $\otimes^k \mathcal{H}_q$ the tensor product space of k copies of $\mathcal{H}_q$. Clearly, $\otimes^1 \mathcal{H}_q = \mathcal{H}_q$ and $\otimes^2 \mathcal{H}_q = \mathbb{H}_q$. The inner product $\langle \cdot, \cdot \rangle_{\otimes^k \mathcal{H}_q}$ and the norm $\|\cdot\|_{\otimes^k \mathcal{H}_q}$ are defined as natural extension of (6) and (7).

If K is the reproducing kernel of $\mathcal{H}_q$, we can define $\mathscr{C}_{h_1, \ldots, h_{2k-1}} : \otimes^k \mathcal{H}_q \to \otimes^k \mathcal{H}_q$ as the operator such that

$$\mathrm{Cum}[X_{h_1}(u_1), \ldots, X_{h_{2k-1}}(u_{2k-1}), X_0(u_{2k})]$$
$$= \langle \mathscr{C}_{h_1, \ldots, h_{2k-1}} K(\cdot, u_{k+1}) \otimes \cdots \otimes K(\cdot, u_{2k}), K(\cdot, u_1) \otimes \cdots \otimes K(\cdot, u_k) \rangle_{\otimes^k \mathcal{H}_q},$$

and $\mathscr{C}_{h_1, \ldots, h_{2k}} : \otimes^{k+1} \mathcal{H}_q \to \otimes^k \mathcal{H}_q$ such that

$$\mathrm{Cum}[X_{h_1}(u_1), \ldots, X_{h_{2k}}(u_{2k}), X_0(u_{2k+1})]$$
$$= \langle \mathscr{C}_{h_1, \ldots, h_{2k}} K(\cdot, u_{k+1}) \otimes \cdots \otimes K(\cdot, u_{2k+1}), K(\cdot, u_1) \otimes \cdots \otimes K(\cdot, u_k) \rangle_{\otimes^k \mathcal{H}_q}.$$

This mapping can be extended to every $f_1, \ldots, f_{2k+1} \in \mathcal{H}_q$, so that

$$\mathrm{Cum}[\langle X_{h_1}, f_1 \rangle_{\mathcal{H}_q}, \ldots, \langle X_{h_{2k-1}}, f_{2k-1} \rangle_{\mathcal{H}_q}, \langle X_0, f_{2k} \rangle_{\mathcal{H}_q}]$$
$$= \langle \mathscr{C}_{h_1, \ldots, h_{2k-1}} f_{k+1} \otimes \cdots \otimes f_{2k}, f_1 \otimes \cdots \otimes f_k \rangle_{\otimes^k \mathcal{H}_q},$$

and

$$\mathrm{Cum}[\langle X_{h_1}, f_1 \rangle_{\mathcal{H}_q}, \ldots, \langle X_{h_{2k}}, f_{2k} \rangle_{\mathcal{H}_q}, \langle X_0, f_{2k+1} \rangle_{\mathcal{H}_q}]$$
$$= \langle \mathscr{C}_{h_1, \ldots, h_{2k}} f_{k+1} \otimes \cdots \otimes f_{2k+1}, f_1 \otimes \cdots \otimes f_k \rangle_{\otimes^k \mathcal{H}_q}.$$

We can define the Hilbert-Schmidt and trace norms of $\mathscr{C}_{h_1,\ldots,h_{2k-1}}$ and $\mathscr{C}_{h_1,\ldots,h_{2k}}$ as usual for operators defined in Hilbert spaces, considering the fact that the set

$$\left\{ \frac{Y_{\ell_1,m_1}}{(1+\ell_1(\ell_1+1))^{q/2}} \otimes \cdots \otimes \frac{Y_{\ell_k,m_k}}{(1+\ell_k(\ell_k+1))^{q/2}} \right\}_{\ell_1,\ldots,\ell_k,m_1,\ldots,m_k}$$

forms an orthonormal basis for $(\otimes^k \mathcal{H}_q, \|\cdot\|_{\otimes^k \mathcal{H}_q})$. In particular, the trace norms will be denoted by $\|\mathscr{C}_{h_1,\ldots,h_{2k-1}}\|_{\mathrm{TR},\otimes^k \mathcal{H}_q}$ and $\|\mathscr{C}_{h_1,\ldots,h_{2k}}\|_{\mathrm{TR},\otimes^{k+1} \mathcal{H}_q}$.

There is a link between the operators defined in L^2 and the operators defined in $\mathcal{H}_q$. For instance, for the $2k$-th order cumulants we have

$$\langle \mathcal{C}_{h_1,\ldots,h_{2k-1}} Y_{\ell_{k+1},m_{k+1}} \otimes \cdots \otimes Y_{\ell_{2k},m_{2k}}, Y_{\ell_1,m_1} \otimes \cdots \otimes Y_{\ell_k,m_k} \rangle_{L^2((\mathbb{S}^2)^k)}$$

$$= \frac{1}{\prod_{i=1}^k (1+\ell_i(\ell_i+1))^{2q}}$$

$$\times \langle \mathscr{C}_{h_1,\ldots,h_{2k-1}} Y_{\ell_{k+1},m_{k+1}} \otimes \cdots \otimes Y_{\ell_{2k},m_{2k}}, Y_{\ell_1,m_1} \otimes \cdots \otimes Y_{\ell_k,m_k} \rangle_{\otimes^k \mathcal{H}_q}. \quad (50)$$

Analogous property holds if we equip $\otimes^k \mathcal{H}_p$ with $\|\otimes^k \mathscr{D}\cdot\|_{L^2((\mathbb{S}^2)^k)}$.

Consider now the regression problem

$$W_{tj} = X_t(U_{tj}) + \epsilon_{tj}, \qquad j=1,\ldots,r_t, \, t=1,\ldots,n,$$

where the the the U_{tj}'s are independently drawn from a common distribution on $\mathbb{S}^2$, and the ϵ_{tj}'s are independent and identically distributed measurement errors of mean 0 and variance $0 < \sigma^2 < \infty$. As before, the process $\mathcal{X} = \{X_t, \, t \in \mathbb{Z}\}$, the measurement locations and the measurement errors are assumed to be mutually independent. Then,

$$\mathbb{E}[W_{tj}|U_{tj} = u_{tj}] = \mu(u_{tj}),$$

and

$$\mathbb{E}[W_{t+h,j}W_{tk}|U_{t+h,j} = u_{t+h,j}, U_{tk} = u_{tk}] = \begin{cases} R_h(u_{t+h,j}, u_{tk}) & \text{if } h \neq 0 \\ R_0(u_{tj}, u_{tk}) + \sigma^2 \delta_j^k & \text{if } h = 0. \end{cases}$$

The estimators for μ and R_0 are defined as in Equations (18) and (19), respectively. In addition here, we define the estimators of the lag-h autocovariance kernels for $h > 0$, as follows. For $h = 1,\ldots,n-1$, we first compute

$$R_{h;\eta} := \arg\min_{g \in \mathbb{H}_p} \frac{(4\pi)^2}{n-h} \sum_{t=1}^{n-h} \frac{1}{r_t^2} \sum_{j=1}^{r_t} \sum_{k=1}^{r_t} (W_{t+h,j}W_{tk} - g(U_{t+h,j}, U_{tk}))^2$$
$$+ \eta\|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2, \quad (51)$$

and then we obtain

$$C_{h;\eta}(u,v) = R_{h;\eta}(u,v) - \mu_\eta(u)\mu_\eta(v). \quad (52)$$

For $h < 0$, we set $C_{h;\eta} : (u,v) \mapsto C_{-h;\eta}(v,u)$. Observe that the diagonal terms are not removed when $h \neq 0$. Similarly to Theorem 2, we can obtain a representer

theorem for the unique solution of (51). In particular, $R_{h;\eta}$ is a continuous function in $\mathbb{S}^2 \times \mathbb{S}^2$ and is a $(\mathscr{D} \otimes \mathscr{D})$-spline.

Differently from the i.i.d. case, consistency and rates of convergence for the mean and autocovariance estimators can be proved with additional conditions in the time-domain. For the mean, a form of functional weak dependence is assumed, *i.e.*, summability of the autocovariance operators (in trace norm). For the autocovariance kernels this concept is extended up to the fourth-order cumulant operators. Note that cumulants mixing conditions arise naturally in this context, since we are essentially dealing with moment-based estimators. See for instance [34, 41].

Definition 6. Consider $p > 2$, $1 < q \leq p$. Let $\Pi_1^s(p, q)$ be the collection of probability measures for stationary sequences of $\mathcal{H}_q$-valued processes such that for any $\mathcal{X} = \{X_t, \, t \in \mathbb{Z}\}$ with probability law $\mathbb{P}_{\mathcal{X}} \in \Pi_1^s(p, q)$

$$\sum_{h \in \mathbb{Z}} \|\mathscr{C}_h\|_{\mathrm{TR}, \mathcal{H}_q} \leq M, \qquad \|\mu\|_{\mathcal{H}_p}^2 \leq K,$$

for some constants $M, K > 0$.

Note that, for any $t \in \mathbb{Z}$,

$$\mathbb{E}\|X_t - \mu\|_{\mathcal{H}_q}^2 = \|\mathscr{C}_0\|_{\mathrm{TR}, \mathcal{H}_q}.$$

Definition 7. Consider $p > 2$, $1 < q \leq p$. Let $\Pi_2^s(p, q)$ be the collection of probability measures for stationary sequences of $\mathcal{H}_q$-valued processes such that for any $\mathcal{X} = \{X_t, \, t \in \mathbb{Z}\}$ with probability law $\mathbb{P}_{\mathcal{X}} \in \Pi_2^s(p, q)$

$$\sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \sum_{h_3 \in \mathbb{Z}} \|\mathscr{C}_{h_1, h_2, h_3}\|_{\mathrm{TR}, \mathbb{H}_q} \leq L_1, \qquad \sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \|\mathscr{C}_{h_1, h_2}\|_{\mathrm{TR}, \mathbb{H}_q} \leq L_2,$$

$$\sum_{h \in \mathbb{Z}} \|\mathscr{C}_h\|_{\mathrm{TR}, \mathcal{H}_q} \leq M, \qquad \sup_{h \in \mathbb{Z}} \|R_h\|_{\mathbb{H}_p}^2 \leq K_1, \qquad \|\mu\|_{\mathcal{H}_p}^2 \leq K_2,$$

for some constants $L_1, L_2, M, K_1, K_2 > 0$.

Remark 11. A weaker but less interpretable condition for the autocovariances is to replace the summability of $\mathscr{C}_h, \mathscr{C}_{h_1, h_2}, \mathscr{C}_{h_1, h_2, h_3}$ with the summability of the operator associated with the kernel $\mathrm{Cov}[X_{h_1}(u)X_{h_2}(v), X_{h_3}(w)X_0(z)]$. Note also that, when the mean is zero, conditions on odd-order cumulants can be discarded.

Now we are able to state the analogue of Theorem 3 and Theorem 4 in the time-dependent setting. Proofs are provided in Section 8.4.

Theorem 5. *Assume that the $U_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i$, are independent copies of $U \sim \mathrm{Unif}(\mathbb{S}^2)$. Let $p > 2$ and $1 < q \leq p$, and consider the estimation problem in Equation (18) for an admissible operator $\mathscr{D}$ of spectral growth order p. If $\eta \asymp (nr)^{-p/(p+1)}$, then*

$$\lim_{D \to \infty} \limsup_{n \to \infty} \sup_{\mathbb{P}_{\mathcal{X}} \in \Pi_1^s(p, q)} \mathbb{P}(\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 > D((nr)^{-p/(p+1)} + n^{-1})) = 0.$$

Theorem 6. *Assume that $\mathbb{E}[\epsilon_{11}^4] < \infty$ and the $U_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i$, are independent copies of $U \sim \mathrm{Unif}(\mathbb{S}^2)$. Let $p > 2$ and $1 < q \leq p$, and consider the estimation problem in Equation (52) for an admissible operator $\mathscr{D}$ of spectral growth order p. If $\eta \asymp (nr/\log n)^{-p/(p+1)}$, then*

$$\lim_{D \to \infty} \limsup_{n \to \infty} \sup_{\mathbb{P}_{\mathcal{X}} \in \Pi_2^s(p,q)} \mathbb{P}\left(\|C_{h;\eta} - C_h\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 > D\left(\left(\frac{\log n}{nr}\right)^{p/(p+1)} + \frac{1}{n} \right) \right)$$

$$= 0,$$

for any $h \in \mathbb{Z}$.

Remark 12. The previous result gives a rate of convergence which holds for any fixed lag $h \in \mathbb{Z}$. However, its dependence on the lag order h can be made explicit. Write $n_h = n - h$. If $\eta \asymp (n_h r/\log n_h)^{-p/(p+1)}$, then

$$\|C_{h;\eta} - C_h\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 = O_\mathbb{P}\left(\left(\frac{\log n_h}{n_h r}\right)^{p/(p+1)} + \frac{1}{n_h} \right),$$

provided that $n_h \to \infty$. This informs us on how many lags we can estimate uniformly with high level of precision.

8. Proofs of formal statements

8.1. Proofs of Section 2

Proof of Proposition 1. The fact that $\mathcal{H}_p$ is a RKHS if and only if $p > 1$ is classical; it is for instance a particular case of [46, Lemma 5.5]. We therefore simply have to prove that $\mathcal{H}_p$ is a RKHS if and only if $\mathbb{H}_p$ is.

Let $\mathbb{H}_p'$ be the topological dual of $\mathbb{H}_p$. It is itself a Hilbert space for the norm

$$\|g\|_{\mathbb{H}_p'} := \sum_{\ell,\ell'=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \frac{\langle g, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle^2}{(1 + \ell(\ell+1))^p(1 + \ell'(\ell'+1))^p}, \quad g \in \mathcal{S}'(\mathbb{S}^2 \times \mathbb{S}^2).$$

Then, $\mathbb{H}_p$ is a RKHS if and only if the evaluation functionals $g \mapsto g(u,v)$ are continuous from $\mathbb{H}_p$ to $\mathbb{R}$ for any $(u,v) \in \mathbb{S}^2 \times \mathbb{S}^2$. This is therefore equivalent to $\delta_{(u,v)} \in \mathbb{H}_p'$, *i.e.*, $\|\delta_{(u,v)}\|_{\mathbb{H}_p'} < \infty$ for any $(u,v) \in \mathbb{S}^2 \times \mathbb{S}^2$. Since $\langle \delta_{(u,v)}, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2)} = Y_{\ell,m}(u)Y_{\ell',m'}(v)$, we have that

$$\|\delta_{(u,v)}\|_{\mathbb{H}_p'}^2 = \left(\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{Y_{\ell,m}(u)^2}{(1+\ell(\ell+1))^p} \right) \left(\sum_{\ell'=0}^{\infty} \sum_{m'=-\ell'}^{\ell'} \frac{Y_{\ell',m'}(v)^2}{(1+\ell'(\ell'+1))^p} \right)$$

$$= \|\delta_x\|_{\mathcal{H}_p'}\|\delta_y\|_{\mathcal{H}_p'} \tag{53}$$

where $\|\cdot\|_{\mathcal{H}_p'}$ is the dual norm on the topological dual $\mathcal{H}_p'$ of $\mathcal{H}_p$.

The relation (53) then shows that $\|\delta_{(u,v)}\|_{\mathbb{H}_p'} < \infty$ for any $(u,v) \in \mathbb{S}^2 \times \mathbb{S}^2$ if and only if $\|\delta_x\|_{\mathcal{H}_p'} < \infty$ for any $x \in \mathbb{S}^2$, *i.e.*, $\mathbb{H}_p$ is a RKHS if and only if $\mathcal{H}_p$ is. $\square$

In Proposition [4], we formalize the invertibility properties of admissible operators. This allows us to deduce that $f \mapsto \|\mathscr{D}f\|_{L^2(\mathbb{S}^2)}$ is a norm equivalent to $f \mapsto \|f\|_{\mathcal{H}_p}$ on $\mathcal{H}_p$, as was used in Sections [2.2] and [4].

Proposition 4. *Let $\mathscr{D}$ be an admissible operator with spectral growth order $p \geq 0$. Then, $\mathscr{D}$ can be uniquely extended as a diffeomorphism $\mathscr{D} : \mathcal{H}_p \to L^2(\mathbb{S}^2)$ and there exist constants $0 < A_1 \leq A_2$ such that, for any $f \in \mathcal{H}_p$,*

$$A_1\|f\|_{\mathcal{H}_p} \leq \|\mathscr{D}f\|_{L^2(\mathbb{S}^2)} \leq A_2\|f\|_{\mathcal{H}_p}. \tag{54}$$

Similarly, $\mathscr{D} \otimes \mathscr{D}$ can be uniquely extended as a diffeomorphism $\mathscr{D} \otimes \mathscr{D} : \mathbb{H}_p \to L^2(\mathbb{S}^2 \times \mathbb{S}^2)$ and there exists $0 < B_1 \leq B_2$ such that, for any $g \in \mathbb{H}_p$,

$$B_1\|g\|_{\mathbb{H}_p} \leq \|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2)} \leq B_2\|g\|_{\mathbb{H}_p}. \tag{55}$$

Proof. Let $f \in \mathcal{S}(\mathbb{S}^2)$. We have that

$$\|\mathscr{D}f\|_{L^2(\mathbb{S}^2)}^2 = \sum_{\ell=0}^{\infty} |D_\ell|^2 \sum_{m=-\ell}^{\ell} |f_{\ell,m}|^2.$$

Then, the existence of A_1, A_2 in (54) for any $f \in \mathcal{S}(\mathbb{S}^2)$ easily follows from the condition (11) and the definition of $\mathcal{H}_p$. The space $\mathcal{S}(\mathbb{S}^2)$ being dense in the Hilbert space $\mathcal{H}_p$, this means in particular that $\mathscr{D}$ can be extended in $\mathcal{H}_p$ and that $\mathscr{D} : \mathcal{H}_p \to L^2(\mathbb{S}^2)$ linearly and continuously. Finally, (54) also implies that $\mathscr{D}$ is a continuous bijection with continuous inverse $\mathscr{D}^{-1} : L^2(\mathbb{S}^2) \to \mathcal{H}_p$. The arguments for tensorial operators $\mathscr{D} \otimes \mathscr{D}$ are similar. □

8.2. *Proofs of Section [4]*

Both Theorem [1] and Theorem [2] can be deduced from general principles on regularized cost functionals over Hilbert spaces, as recalled in the next theorem.

Theorem 7 (Representer Theorem on Hilbert Spaces)**.** *Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\|\cdot\|_{\mathcal{H}}$, $\lambda > 0$, $L \geq 1$, $\boldsymbol{y} = (y_1, \ldots, y_L) \in \mathbb{R}^L$, and $\nu_1, \ldots, \nu_L$ be linearly independent elements in $\mathcal{H}$. Then, the optimization problem*

$$\min_{f \in \mathcal{H}} \quad \sum_{\ell=1}^{L} (y_\ell - \langle \nu_\ell, f \rangle_{\mathcal{H}})^2 + \lambda\|f\|_{\mathcal{H}}^2 \tag{56}$$

has a unique solution $\widehat{f}$ such that

$$\widehat{f} = \sum_{\ell=1}^{L} \alpha_\ell \nu_\ell \quad where \quad \boldsymbol{\alpha} = (\mathbf{G} + \lambda\mathbf{I}_L)^{-1}\boldsymbol{y} \tag{57}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L) \in \mathbb{R}^L$ and $\mathbf{G} \in \mathbb{R}^{L \times L}$ is such that $G_{\ell_1, \ell_2} = \langle \nu_{\ell_1}, \nu_{\ell_2} \rangle_{\mathcal{H}}$ for $1 \leq \ell_1, \ell_2 \leq L$.

Remark 13 (On Theorem 7). The existence and uniqueness of the solution of (56) can be deduced from the Hilbert projection theorem. The exact form (56) of the solution is classical. It is for instance proved in [53, Section 3.2] in two steps. First, the optimization problem (56) is shown to admit a unique solution of the form $\widehat{f} = \sum_{\ell=1}^{L} \alpha_\ell \nu_\ell$. Second, injecting the form of this solution in (56), we observe that $\|\widehat{f}\|_{\mathcal{H}} = \langle \boldsymbol{\alpha}, \mathbf{G}\boldsymbol{\alpha} \rangle$ and $\langle \nu_\ell, \widehat{f} \rangle = (\mathbf{G}\boldsymbol{\alpha})_\ell$. Hence, $\boldsymbol{\alpha}$ is solution of the optimization problem

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^L} \|\boldsymbol{y} - \mathbf{G}\boldsymbol{\alpha}\|_2^2 + \lambda \langle \boldsymbol{\alpha}, \mathbf{G}\boldsymbol{\alpha} \rangle. \tag{58}$$

Then, we easily show that the optimizer of the finite-dimensional quadratic optimization problem (58) is $\boldsymbol{\alpha} = (\mathbf{G} + \lambda \mathbf{I}_L)^{-1} \boldsymbol{y}$. Note that this requires the invertibility of $\mathbf{G}$, which is true because the ν_ℓ are assumed to be linearly independent.

Proof of Theorem 2. The proof is divided in two steps. We first prove that the unique solution is given by (27) and then show that R_η is symmetric.

   Form of the solution. We show that Theorem 2 is a particular case of Theorem 7. According to Proposition 4, the norms $\|(\mathscr{D} \otimes \mathscr{D}) \cdot \|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$ and $\| \cdot \|_{\mathbb{H}_p}$ being equivalent, $(\mathbb{H}_p, \|(\mathscr{D} \otimes \mathscr{D}) \cdot \|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)})$ is a Hilbert space.

   Let $\psi_{\mathscr{D}^* \mathscr{D}}$ the zonal Green's kernel of $\mathscr{D}^* \mathscr{D}$ (see (15)). For any $1 \leq i \leq n$, $1 \leq j \neq k \leq r_i$, we set $\nu_{ijk} = \psi_{\mathscr{D}^* \mathscr{D}}(\langle \cdot, u_{ij} \rangle) \otimes \psi_{\mathscr{D}^* \mathscr{D}}(\langle \cdot, u_{ik} \rangle)$ and we observe that

$$(\mathscr{D} \otimes \mathscr{D})^* (\mathscr{D} \otimes \mathscr{D})\{\nu_{ijk}\} = (\mathscr{D}^* \mathscr{D}\{\psi_{\mathscr{D}^* \mathscr{D}}(\langle \cdot, u_{ij} \rangle)\}) \otimes (\mathscr{D}^* \mathscr{D}\{\psi_{\mathscr{D}^* \mathscr{D}}(\langle \cdot, u_{ik} \rangle)\})$$
$$= \delta_{u_{ij}} \otimes \delta_{u_{ik}}.$$

This implies that, for any $g \in \mathcal{H} = \mathbb{H}_p$,

$$\langle g, \nu_{ijk} \rangle_{\mathcal{H}} = \langle (\mathscr{D} \otimes \mathscr{D})\{g\}, (\mathscr{D} \otimes \mathscr{D})\{\nu_{ijk}\} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = g(u_{ij}, u_{ik}).$$

We deduce that (26) can be recast as

$$\min_{g \in \mathcal{H}} \sum_{i=1}^{n} \sum_{1 \leq j \neq k \leq r_i} \left(\frac{w_{ij} w_{ik}}{\sqrt{r_i(r_i - 1)}} - \left\langle g, \frac{\nu_{ijk}}{\sqrt{r_i(r_i - 1)}} \right\rangle_{\mathcal{H}} \right)^2 + \lambda \|g\|_{\mathcal{H}} \tag{59}$$

with $\lambda = \frac{n\eta}{(4\pi)^2}$.

   The optimization problem (59) corresponds to (56) where, for any $1 \leq \ell \leq L = \sum_{i=1}^{n} r_i(r_i - 1)$ and its corresponding (i, j, k) in the vectorization, we set $\nu_\ell = \nu_{ijk}/\sqrt{r_i(r_i - 1)}$. Note that the ν_ℓ are linearly independent since the Dirac impulses $\delta_{u_{ij}} \otimes \delta_{u_{ik}} = (\mathscr{D}^* \mathscr{D} \otimes \mathscr{D}^* \mathscr{D})\nu_{ijk}$ are. We moreover observe that

$$\langle \nu_{i_1 j_1 k_1}, \nu_{i_2 j_2 k_2} \rangle_{\mathcal{H}} = \langle \nu_{i_1 j_1 k_1}, (\mathscr{D} \otimes \mathscr{D})^* (\mathscr{D} \otimes \mathscr{D}) \nu_{i_2 j_2 k_2} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$
$$= \langle \nu_{i_1 j_1 k_1}, \delta_{u_{i_2 j_2}} \otimes \delta_{u_{i_2 k_2}} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$
$$= \nu_{i_1 j_1 k_1}(u_{i_2 j_2}, u_{i_2 k_2})$$
$$= \psi_{\mathscr{D}^* \mathscr{D}}(\langle u_{i_1 j_1}, u_{i_2 j_2} \rangle) \times \psi_{\mathscr{D}^* \mathscr{D}}(\langle u_{i_1 k_1}, u_{i_2 k_2} \rangle),$$

which implies that

$$
\begin{aligned}
H_{\ell_1,\ell_2} = \langle \nu_{\ell_1}, \nu_{\ell_2} \rangle_{\mathcal{H}} &= \frac{\langle \nu_{i_1 j_1 k_1}, \nu_{i_2 j_2 k_2} \rangle_{\mathcal{H}}}{\sqrt{r_{i_1}(r_{i_1}-1) r_{i_2}(r_{i_2}-1)}} \\
&= \frac{\psi_{\mathscr{D}^*\mathscr{D}}(\langle u_{i_1 j_1}, u_{i_1 k_1} \rangle) \times \psi_{\mathscr{D}^*\mathscr{D}}(\langle u_{i_2 j_2}, u_{i_2 k_2} \rangle)}{\sqrt{r_{i_1}(r_{i_1}-1) r_{i_2}(r_{i_2}-1)}}
\end{aligned}
$$

where ℓ_1 (resp. ℓ_2) corresponds to (i_1, j_1, k_1) (resp. (i_2, j_2, k_2)) in the vectorization. With these identifications, Theorem 7 gives (27) and (29).

Symmetry of R_η. Any $g \in \mathbb{H}_p$ can be uniquely decomposed as

$$
g(u,v) = g_s(u,v) + g_a(u,v) = \left(\frac{g(u,v) + g(v,u)}{2} \right) + \left(\frac{g(u,v) - g(v,u)}{2} \right) \quad (60)
$$

where g_s is symmetric ($g_s(u,v) = g_s(v,u)$) and g_a is antisymmetric ($g_a(u,v) = -g_a(v,u)$). Then, we have that $(\mathscr{D} \otimes \mathscr{D})g_s$ (resp. $(\mathscr{D} \otimes \mathscr{D})g_a$) is symmetric (resp. antisymmetric). Indeed, the operator $(\mathscr{D} \otimes \mathscr{D})$ maps symmetric (resp. antisymmetric) functions to symmetric (resp. antisymmetric) ones. This fact is obvious for functions $g = f_1 \otimes f_2$ since $(\mathscr{D} \otimes \mathscr{D})g = \mathscr{D}f_1 \otimes \mathscr{D}f_2$ and extended to any function by density of the span of separable functions in $\mathbb{H}_p$. Therefore

$$
\|(\mathscr{D} \otimes \mathscr{D})g\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = \|(\mathscr{D} \otimes \mathscr{D})g_a\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} + \|(\mathscr{D} \otimes \mathscr{D})g_s\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} \quad (61)
$$

due to the fact that the inner product between symmetric and antisymmetric bivariate functions is 0, hence $\langle (\mathscr{D} \otimes \mathscr{D})g_s, (\mathscr{D} \otimes \mathscr{D})g_a \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = 0$.

For $i = 1, \ldots, n$, let $\Sigma_i \in \mathbb{R}^{r_i \times r_i}$ be the matrix such that $\Sigma_i[j,k] = w_{ij} w_{ik} \delta_j^k$. We also define the operator $\Phi_i : \mathbb{H}_p \to \mathbb{R}^{r_i \times r_i}$ such that $\Phi_i(g)[j,k] = g(u_{ij}, u_{ik}) \delta_j^k$. Then, the data fidelity term in (26) can be rewritten as

$$
\sum_{i=1}^{n} \frac{1}{r_i(r_i-1)} \sum_{1 \le j \ne k \le r_i} (W_{ij} W_{ik} - g(U_{ij}, U_{ik}))^2 = \sum_{i=1}^{n} \frac{1}{r_i(r_i-1)} \|\Sigma_i - \Phi_i(g)\|_2^2. \tag{62}
$$

Then, the matrix $\Sigma_i - \Phi_i(g_s)$ (resp. $\Phi_i(g_a)$) is symmetric (resp. antisymmetric), which implies that, for any $i = 1, \ldots, n$,

$$
\|\Sigma_i - \Phi_i(g)\|_2^2 = \|(\Sigma_i - \Phi_i(g_s)) - \Phi_i(g_a)\|_2^2 = \|(\Sigma_i - \Phi_i(g_s))\|_2^2 + \|\Phi_i(g_a)\|_2^2, \quad (63)
$$

the inner product between one symmetric and one antisymmetric matrix being 0.

If we denote by $J(g)$ the cost functional to be optimized in (26), we deduce from (61) and (63) that

$$
J(g) = J(g_s) + \frac{(4\pi)^2}{n} \sum_{i=1}^{n} \frac{\|\Phi_i(g_a)\|_2^2}{r_i(r_i-1)} + \eta \|(\mathscr{D} \otimes \mathscr{D})g_a\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} \ge J(g_s). \quad (64)
$$

In other term, for any $g \in \mathbb{H}_p$, there exists $g_s \in \mathbb{H}_p$ symmetric such that $J(g_s) \le J(g)$. This ensures that the unique minimizer of (26) is symmetric, as expected. $\square$

8.3. Proofs of Section 5

The following lemma will be used extensively for proving the rates of the mean and covariance estimators.

Lemma 1. *Let $p > 1$. There exists $B_1 > 0$ such that, for any $f \in \mathcal{H}_p$,*

$$\sup_{u \in \mathbb{S}^2} |f(u)| \leq B_1 \|f\|_{\mathcal{H}_p}. \tag{65}$$

Similarly, there exists $B_2 > 0$ such that, for any $g \in \mathbb{H}_p$,

$$\sup_{(u,v) \in \mathbb{S}^2 \times \mathbb{S}^2} |g(u,v)| \leq B_2 \|g\|_{\mathbb{H}_p}. \tag{66}$$

Proof. According to Proposition 1, the condition $p > 1$ ensures that $\mathcal{H}_p$ is a RKHS. For any $u \in \mathbb{S}^2$, we denote by $K : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ the reproducing kernel such that $f(u) = \langle K(\cdot, u), f \rangle_{\mathcal{H}_p}$. Then, the Cauchy-Schwarz inequality implies that

$$|f(u)| = |\langle K(\cdot, u), f \rangle_{\mathcal{H}_p}| \leq \|K(\cdot, u)\|_{\mathcal{H}_p} \|f\|_{\mathcal{H}_p}. \tag{67}$$

The norm of $\mathcal{H}_p$ is isotropic in the sense that $\|\mathcal{R}f\|_{\mathcal{H}_p} = \|f\|_{\mathcal{H}_p}$ for any rotation $\mathcal{R}$. This implies that $\|K(\cdot, u)\|_{\mathcal{H}_p}$ does not depend on $u \in \mathbb{S}^2$. Hence, (67) implies (65). The proof for Equation (66) is similar. $\square$

Proof of Theorem 3. Without loss of generality, we will prove the theorem for $D_\ell = (1 + \ell(\ell+1))^{p/2}$, which leads to a penalization term in the $\mathcal{H}_p$ norm. However, all the following steps can be generalized to every admissible operator with spectral growth order p in Definition 2.

Define

$$F_\eta(g) := \frac{4\pi}{n} \sum_{i=1}^{n} \frac{1}{r_i} \sum_{j=1}^{r_i} (W_{ij} - g(U_{ij}))^2 + \eta \|g\|_{\mathcal{H}_p}^2,$$

$$\bar{F}_\eta(g) := \mathbb{E}[F_\eta(g)] = 4\pi \mathbb{E}|W_{11} - \mu(U_{11})|^2 + \|\mu - g\|_{L^2(\mathbb{S}^2)}^2 + \eta \|g\|_{\mathcal{H}_p}^2,$$

and let

$$\bar{\mu}_\eta := \arg\min_{g \in \mathcal{H}_p} \bar{F}_\eta(g).$$

Also, define

$$\tilde{\mu}_\eta = \bar{\mu}_\eta - (\bar{F}_\eta'')^{-1} F_\eta'(\bar{\mu}_\eta). \tag{68}$$

The definitions of F_η' and $\bar{F}_\eta''$ will be given in Lemma 2, while the existence of $(\bar{F}_\eta'')^{-1}$ will be discussed in Lemma 3.

Now, write $\mu_\eta - \mu = \mu_\eta - \tilde{\mu}_\eta + \tilde{\mu}_\eta - \bar{\mu}_\eta + \bar{\mu}_\eta - \mu$. We shall prove that

1. $\|\bar{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 \leq M_1 \left((nr)^{-p/(p+1)} + n^{-1}\right)$,
2. $\mathbb{E}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_{L^2(\mathbb{S}^2)}^2 \leq M_2 \left((nr)^{-p/(p+1)} + n^{-1}\right)$,
3. $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}\left(\|\mu_\eta - \tilde{\mu}_\eta\|_{L^2(\mathbb{S}^2)}^2 > \varepsilon \left((nr)^{-p/(p+1)} + n^{-1}\right)\right) = 0$$

whenever $\eta \asymp (nr)^{-p/(p+1)}$, for any choice of the sampling distribution in $\Pi_1(p,q)$.

Note that, for $t > 0$,

$$
\begin{aligned}
\mathbb{P}(\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 > t) &= \mathbb{P}(\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)} > \sqrt{t}) \\
&\leq \mathbb{P}(\|\mu_\eta - \tilde{\mu}_\eta\|_{L^2(\mathbb{S}^2)} > \sqrt{t}/2) \\
&\quad + \mathbb{P}(\|\tilde{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)} > \sqrt{t}/2).
\end{aligned}
$$

The second term satisfies

$$
\mathbb{P}(\|\tilde{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)} > t') \leq \frac{\mathbb{E}\|\tilde{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)}}{t'} \leq \frac{\mathbb{E}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_{L^2(\mathbb{S}^2)}}{t'} + \frac{\|\bar{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)}}{t'}
$$

where $t' = \sqrt{t}/2$. By choosing $t = D\left((nr)^{-p/(p+1)} + n^{-1}\right)$, then

$$
\mathbb{P}(\|\tilde{\mu}_\eta - \mu\|_{L^2(\mathbb{S}^2)} > t') \leq \frac{c_0}{\sqrt{D}},
$$

where c_0 is a positive constant not depending on the choice of $\mathbb{P}_X \in \Pi_1(p,q)$. Moreover, from 3, $\forall \varepsilon > 0$,

$$
\lim_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}(\|\mu_\eta - \tilde{\mu}_\eta\|_{L^2(\mathbb{S}^2)}^2 > \varepsilon\left((nr)^{-p/(p+1)} + n^{-1}\right)) = 0.
$$

Hence, by taking $\varepsilon = D/4$ and $t = D\left((nr)^{-p/(p+1)} + n^{-1}\right)$,

$$
\begin{aligned}
&\lim_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}(\|\mu_\eta - \tilde{\mu}_\eta\|_{L^2(\mathbb{S}^2)} > \sqrt{t}/2) \\
&= \lim_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}(\|\mu_\eta - \tilde{\mu}_\eta\|_{L^2(\mathbb{S}^2)}^2 > t/4) = 0.
\end{aligned}
$$

For the rest of the proof, it is useful to define an *intermediate* norm $\|\cdot\|_\alpha$, $\alpha \in [0,1]$, between $\|\cdot\|_{L^2(\mathbb{S}^2)}$ and $\|\cdot\|_{\mathcal{H}_p}$. Let $g \in L^2(\mathbb{S}^2)$,

$$
\|g\|_\alpha^2 := \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell D_\ell^{2\alpha} \langle g, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2.
$$

Since $|D_\ell| \geq 1$, for all $\ell \in \mathbb{N}$,

$$
\|g\|_{L^2(\mathbb{S}^2)} = \|g\|_0 \leq \|g\|_\alpha \leq \|g\|_1 = \|g\|_{\mathcal{H}_p};
$$

moreover, $g \mapsto \|g\|_\alpha$ specifies a norm on $\mathcal{H}_{p\alpha}$ which is equivalent to $\|\cdot\|_{\mathcal{H}_{p\alpha}}$. Note that, since we are considering $D_\ell = (1 + \ell(\ell+1))^{p/2}$, the two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_{\mathcal{H}_{p\alpha}}$ are actually identical; however, for generality purposes, we will maintain such distinction in our notation.

Proof of 1 follows immediately from the fact that, for $\alpha \in [0,1]$,

$$
\|\mu - \bar{\mu}_\eta\|_\alpha^2 \leq \eta^{1-\alpha}\|\mu\|_{\mathcal{H}_p}^2. \tag{69}
$$

Indeed, for $g \in \mathcal{H}_p$,

$$\|\mu - g\|_{L^2(\mathbb{S}^2)}^2 + \eta\|g\|_{\mathcal{H}_p}^2 = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} |\mu_{\ell,m} - g_{\ell,m}|^2 + \eta\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} D_\ell^2 |g_{\ell,m}|^2,$$

where $\mu_{\ell,m} = \langle \mu, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}$ and $g_{\ell,m} = \langle g, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}$. The minimizer is then given by

$$\bar{\mu}_\eta = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \frac{\mu_{\ell,m}}{1 + \eta D_\ell^2} Y_{\ell,m}.$$

Hence,

$$\|\mu - \bar{\mu}_\eta\|_\alpha^2 = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} D_\ell^{2\alpha} \left|\mu_{\ell,m} - \mu_{\ell,m}(1 + \eta D_\ell^2)^{-1}\right|^2$$

$$\leq \eta^{1-\alpha}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} D_\ell^2 |\mu_{\ell,m}|^2 (\eta D_\ell^2)^{1+\alpha}(1 + \eta D_\ell^2)^{-2}$$

$$\leq \eta^{1-\alpha}\|\mu\|_{\mathcal{H}_p}^2.$$

and, setting $\alpha = 0$,

$$\|\mu - \bar{\mu}_\eta\|_{L^2(\mathbb{S}^2)}^2 \leq \eta\|\mu\|_{\mathcal{H}_p}^2 \leq K\eta,$$

which gives the claimed result under the assumptions on η.

In order to prove 2, we first show that

$$\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\alpha^2 = \frac{1}{4}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}(F_\eta'(\bar{\mu}_\eta)Y_{\ell,m})^2. \qquad (70)$$

By the definitions of $\tilde{\mu}_\eta$ and $\|\cdot\|_\alpha$,

$$\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\alpha^2 = \|(\bar{F}_\eta'')^{-1}F_\eta'(\bar{\mu}_\eta)\|_\alpha^2$$

$$= \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} D_\ell^{2\alpha}\langle(\bar{F}_\eta'')^{-1}F_\eta'(\bar{\mu}_\eta), Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2.$$

However,

$$\langle(\bar{F}_\eta'')^{-1}F_\eta'(\bar{\mu}_\eta), Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)} = \frac{1}{D_\ell^2}\langle(\bar{F}_\eta'')^{-1}F_\eta'(\bar{\mu}_\eta), Y_{\ell,m}\rangle_{\mathcal{H}_p}$$

$$= \frac{1}{D_\ell^2}\langle \mathcal{Q}F_\eta'(\bar{\mu}_\eta), (\tilde{F}_\eta'')^{-1}Y_{\ell,m}\rangle_{\mathcal{H}_p},$$

since $\mathcal{Q}F_\eta'(\bar{\mu}_\eta)$ is the representer of $F_\eta'(\bar{\mu}_\eta)$ and $(\tilde{F}_\eta'')^{-1}$ is self-adjoint. From Lemma 3,

$$(\tilde{F}_\eta'')^{-1}Y_{\ell,m} = \frac{1}{2}\frac{D_\ell^2}{1 + \eta D_\ell^2}Y_{\ell,m};$$

thus,

$$\langle (\bar{F}_\eta'')^{-1} F_\eta'(\bar{\mu}_\eta), Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)} = \frac{1}{2(1 + \eta D_\ell^2)} \langle \mathcal{Q} F_\eta'(\bar{\mu}_\eta), Y_{\ell,m} \rangle_{\mathcal{H}_p}$$

$$= \frac{1}{2(1 + \eta D_\ell^2)} F_\eta'(\bar{\mu}_\eta) Y_{\ell,m}.$$

Now observe that $\bar{F}_\eta'(\bar{\mu}_\eta) = 0$ (see [21, Theorem 3.6.3]). Then, an application of Lemma 2 reveals that, for any $g \in \mathcal{H}_p$,

$$F_\eta'(\bar{\mu}_\eta)g = F_\eta'(\bar{\mu}_\eta)g - \bar{F}_\eta'(\bar{\mu}_\eta)g$$

$$= -\frac{8\pi}{n} \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} (W_{ij} - \bar{\mu}_\eta(U_{ij}))g(U_{ij}) + 2\langle \mu - \bar{\mu}_\eta, g \rangle_{L^2(\mathbb{S}^2)}. \quad (71)$$

Consequently,

$$\mathbb{E}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}] = \mathbb{E}_{\mathcal{U}}\mathbb{E}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|\mathcal{U}] = 0$$

and

$$\mathbb{E}|F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|^2 = \mathrm{Var}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}]$$

$$= \frac{(8\pi)^2}{n^2} \sum_{i=1}^n \frac{1}{r_i^2} \mathrm{Var} \left[\sum_{j=1}^{r_i} (W_{ij} - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij}) \right].$$

Using the law of total variance, for a generic i we can write

$$\mathrm{Var} \left[\sum_{j=1}^{r_i} (W_{ij} - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij}) \right]$$

$$= \mathrm{Var} \left[\sum_{j=1}^{r_i} (\mu(U_{ij}) - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij}) \right]$$

$$+ \mathbb{E}_{\mathcal{U}} \left[\mathrm{Var} \left[\sum_{j=1}^{r_i} (W_{ij} - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij}) \Big| \mathcal{U} \right] \right].$$

For the first term on the right hand side of this expression, we have

$$\mathrm{Var} \left[\sum_{j=1}^{r_i} (\mu(U_{ij}) - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij}) \right]$$

$$= r_i \mathrm{Var} \left[(\mu(U_{ii}) - \bar{\mu}_\eta(U_{ii}))Y_{\ell,m}(U_{ii}) \right]$$

$$\leq \frac{r_i}{4\pi} \int_{\mathbb{S}^2} |\mu(u) - \bar{\mu}_\eta(u)|^2 |Y_{\ell,m}(u)|^2 du$$

$$\leq \frac{Br_i}{4\pi} \|\mu - \bar{\mu}_\eta\|_{\mathcal{H}_p}^2$$

$$\leq \frac{BKr_i}{4\pi},$$

where the last two inequalities are justified by Lemma 1, for some $B > 0$, and Equation (69) with $\alpha = 1$. Then, for the second term,

$$\mathbb{E}_{\mathcal{U}}\left[\operatorname{Var}\left[\sum_{j=1}^{r_i}(W_{ij} - \bar{\mu}_\eta(U_{ij}))Y_{\ell,m}(U_{ij})\Big|\mathcal{U}\right]\right]$$

$$\leq \sum_{j=1}^{r_i}\sum_{j'=1}^{r_i}\mathbb{E}_{\mathcal{U}}[Y_{\ell,m}(U_{ij})Y_{\ell,m}(U_{ij'})\mathbb{E}[W_{ij}W_{ij'}|\mathcal{U}]]$$

$$= \frac{r_i(r_i - 1)}{(4\pi)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$+ \frac{r_i}{4\pi}\int_{\mathbb{S}^2}R(u, u)|Y_{\ell,m}(u)|^2 du + \frac{r_i}{4\pi}\sigma^2$$

$$\leq \frac{r_i(r_i - 1)}{(4\pi)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$+ \frac{B'r_i}{4\pi}\mathbb{E}\|X\|_{\mathcal{H}_q}^2 + \frac{r_i}{4\pi}\sigma^2,$$

again by applying Lemma 1, for some $B' > 0$. Hence, combining all the bounds,

$$\mathbb{E}|F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|^2 \leq \frac{4}{n}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2 + O\left(\frac{1}{nr}\right),$$

and

$$\mathbb{E}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\alpha^2 \leq \frac{1}{n}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$+ O\left(\frac{1}{nr}\right)\sum_{\ell=0}^{\infty}\frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}(2\ell + 1),$$

where r is the harmonic mean of $r_1, \ldots, r_n$. Now, for $\alpha \leq q/p$,

$$\sum_{\ell,m}\frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2 \leq \mathbb{E}\left[\sum_{\ell,m}D_\ell^{2\alpha}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2\right] \leq \mathbb{E}\|X\|_{q/p}^2,$$

which is bounded (possibly up to an arbitrary constant) by $\mathbb{E}\|X\|_{\mathcal{H}_q}^2$. Moreover, from [29],

$$\sum_{\ell=0}^{\infty}\frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}(2\ell + 1) = O\left(1 + \eta^{-(\alpha+1/p)}\right).$$

Thus, we have that

$$\mathbb{E}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\alpha^2 \leq M_2\left((nr)^{-1}\eta^{-(\alpha+1/p)} + n^{-1}\right),$$

where M_2 is a positive constant not depending on the choice of $\mathbb{P}_X \in \Pi_1(p, q)$. By choosing $\alpha = 0$, we obtain the claimed rate.

Now, let us prove 3. The first step is to obtain a useful analytic form for $\mu_\eta - \tilde{\mu}_\eta$, by observing that

$$
\begin{aligned}
\mu_\eta - \tilde{\mu}_\eta &= \mu_\eta - \bar{\mu}_\eta + (\bar{F}''_\eta)^{-1} F'_\eta(\bar{\mu}_\eta) \\
&= (\bar{F}''_\eta)^{-1} \left[\bar{F}''_\eta(\mu_\eta - \bar{\mu}_\eta) + F'_\eta(\bar{\mu}_\eta) \right].
\end{aligned}
$$

Since μ_η minimizes F'_η, it holds $F'_\eta(\mu_\eta) = 0$ (see [21, Theorem 3.6.3]). Then, for any $g \in \mathcal{H}_p$,

$$
\begin{aligned}
&\left[\bar{F}''_\eta(\mu_\eta - \bar{\mu}_\eta) + F'_\eta(\bar{\mu}_\eta) \right] g \\
&= \left[\bar{F}''_\eta(\mu_\eta - \bar{\mu}_\eta) + F'_\eta(\bar{\mu}_\eta) - F'_\eta(\mu_\eta) \right] g \\
&= \left[\bar{F}''_0(\mu_\eta - \bar{\mu}_\eta) - F''_0(\mu_\eta - \bar{\mu}_\eta) \right] g.
\end{aligned}
$$

where we used Lemma 2; in other words,

$$
\mu_\eta - \tilde{\mu}_\eta = (\bar{F}''_\eta)^{-1} \left[\bar{F}''_0(\mu_\eta - \bar{\mu}_\eta) - F''_0(\mu_\eta - \bar{\mu}_\eta) \right].
$$

Now, the same argument that leads to (70) gives us

$$
\| \mu_\eta - \tilde{\mu}_\eta \|_\alpha^2 = \frac{1}{4} \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2} \left(\left[\bar{F}''_0(\mu_\eta - \bar{\mu}_\eta) - F''_0(\mu_\eta - \bar{\mu}_\eta) \right] Y_{\ell,m} \right)^2,
$$

with

$$
\begin{aligned}
&\left[\bar{F}''_0(\mu_\eta - \bar{\mu}_\eta) - F''_0(\mu_\eta - \bar{\mu}_\eta) \right] Y_{\ell,m} \\
&= 2 \langle \mu_\eta - \bar{\mu}_\eta, Y_{\ell,m} \rangle_{L^2(\mathbb{S}^2)} - \frac{8\pi}{n} \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} (\mu_\eta(U_{ij}) - \bar{\mu}_\eta(U_{ij})) Y_{\ell,m}(U_{ij}).
\end{aligned}
$$

Now since $\mu_\eta - \bar{\mu}_\eta \in \mathcal{H}_p$, we can write $\mu_\eta - \bar{\mu}_\eta = \sum_{\ell',m'} h_{\ell',m'} Y_{\ell',m'}$, where the convergence is both in $L^2(\mathbb{S}^2)$ and pointwise. Then,

$$
\| \mu_\eta - \tilde{\mu}_\eta \|_\alpha^2 = \sum_{\ell,m} \frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2} \left(\sum_{\ell',m'} h_{\ell',m'} V_{\ell,\ell',m,m'} \right)^2,
$$

where

$$
V_{\ell,\ell',m,m'} = \delta_\ell^{\ell'} \delta_m^{m'} - \frac{4\pi}{n} \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} Y_{\ell,m}(U_{ij}) Y_{\ell',m'}(U_{ij}).
$$

By applying the Cauchy–Schwarz inequality for arbitrary $\theta \in (1/p, 1]$, we obtain

$$
\| \mu_\eta - \tilde{\mu}_\eta \|_\alpha^2 \le \| \mu_\eta - \bar{\mu}_\eta \|_\theta^2 \sum_{\ell,m} \frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2} \sum_{\ell',m'} D_{\ell'}^{-2\theta} V_{\ell,\ell',m,m'}^2.
$$

It is readily seen that $\mathbb{E}[V_{\ell,\ell',m,m'}] = 0$ and

$$
\begin{aligned}
\mathbb{E}[V_{\ell,\ell',m,m'}^2] = \mathrm{Var}[V_{\ell,\ell',m,m'}] &= \frac{(4\pi)^2}{nr} \mathrm{Var}[Y_{\ell,m}(U_{ii})Y_{\ell',m'}(U_{ii})] \\
&\leq \frac{4\pi}{nr} \int_{\mathbb{S}^2} |Y_{\ell,m}(u)|^2 |Y_{\ell',m'}(u)|^2 du \\
&\leq D_\ell^{2\theta} O\left(\frac{1}{nr}\right),
\end{aligned}
$$

by Lemma 1, since $Y_{\ell,m} \in \mathcal{H}_{p\theta}$, and $p\theta > 1$. Hence, we obtain

$$
\sum_{\ell,m} \frac{D_\ell^{2\alpha}}{(1+\eta D_\ell^2)^2} \sum_{\ell',m'} D_{\ell'}^{-2\theta} \mathbb{E}[V_{\ell,\ell',m,m'}^2] \leq \frac{M_3}{nr\eta^{\alpha+\theta+1/p}}.
$$

Let us define
$$
A_n := \sum_{\ell,m} \frac{D_\ell^{2\alpha}}{(1+\eta D_\ell^2)^2} \sum_{\ell',m'} D_{\ell'}^{-2\theta} V_{\ell,\ell',m,m'}^2,
$$

$a_n := \frac{1}{nr\eta^{\alpha+\theta+1/p}}$ and $\gamma_n := \frac{1}{nr\eta^{\alpha+1/p}} + \frac{1}{n}$. Note that $A_n = \sup_{\mathbb{P}_X \in \Pi_1(p,q)} A_n$ and $A_n = o_\mathbb{P}(1)$. Then, $\|\mu_\eta - \tilde{\mu}_\eta\|_\alpha^2 \leq A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2$ and therefore

$$
\begin{aligned}
\mathbb{P}\left(\|\mu_\eta - \tilde{\mu}_\eta\|_\alpha^2 > \varepsilon\,\gamma_n\right) &\leq \mathbb{P}\left(A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n\right) \\
&= \mathbb{P}\left(A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n, A_n < 1\right) \\
&\quad + \mathbb{P}\left(A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n, A_n \geq 1\right) \\
&\leq \mathbb{P}\left(A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n, A_n < 1\right) + \mathbb{P}\left(A_n \geq 1\right).
\end{aligned}
$$

If $A_n < 1$,

$$
\begin{aligned}
\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\theta \geq \|\mu_\eta - \bar{\mu}_\eta\|_\theta - \|\mu_\eta - \tilde{\mu}_\eta\|_\theta \\
\geq (1 - \sqrt{A_n})\|\mu_\eta - \bar{\mu}_\eta\|_\theta,
\end{aligned}
$$

which allows to write

$$
\begin{aligned}
&\mathbb{P}\left(A_n \|\mu_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n, A_n < 1\right) \\
\leq &\mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n, A_n < 1\right) \\
\leq &\mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\theta^2 > \varepsilon\,\gamma_n\right).
\end{aligned}
$$

Let us now define $B_n := \|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\theta^2$ and $b_n := \frac{1}{nr\eta^{\theta+1/p}} + \frac{1}{n}$. Recall that $\mathbb{E}[B_n] \leq M_2 b_n$, for $\theta \leq q/p$. Moreover, $|1 - \sqrt{A_n}|^{-2} = O_\mathbb{P}(1)$. We can observe that

$$
a_n b_n = \frac{1}{nr\eta^{2\theta+1/p}}\left(\frac{1}{nr\eta^{\alpha+1/p}} + \frac{\eta^{\theta-\alpha}}{n}\right),
$$

so that $c_n := a_n b_n / \gamma_n \to 0$, assuming $\theta < 1/2$ (recall that $p > 2$) and $\alpha \in [0, \theta]$. Then,

$$\mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}B_n > \varepsilon\,\gamma_n\right) = \mathbb{P}\left(\frac{A_n|1 - \sqrt{A_n}|^{-2}B_n}{a_n b_n} > \frac{\varepsilon}{c_n}\right)$$

$$\leq \mathbb{P}\left(\frac{B_n}{b_n} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right) + \mathbb{P}\left(\frac{A_n}{a_n} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right)$$

$$+ \mathbb{P}\left(|1 - \sqrt{A_n}|^{-2} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right)$$

$$\leq M_2\frac{c_n^{1/3}}{\varepsilon^{1/3}} + M_3\frac{c_n^{1/3}}{\varepsilon^{1/3}}$$

$$+ \mathbb{P}\left(|1 - \sqrt{A_n}|^{-2} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right).$$

Clearly $c_n^{1/3}|1 - \sqrt{A_n}|^{-2} = o_{\mathbb{P}}(1)$, hence

$$\lim_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_1(p,q)} \mathbb{P}\left(\|\mu_\eta - \tilde{\mu}_\eta\|_\alpha^2 > \varepsilon\,\gamma_n\right) = 0.$$

By taking $\alpha = 0$ we obtain the claimed result. $\qquad\qquad\qquad\square$

The next two lemmas refer to

$$F_\eta(g) := \frac{4\pi}{n}\sum_{i=1}^n \frac{1}{r_i}\sum_{j=1}^{r_i}(W_{ij} - g(U_{ij}))^2 + \eta\|\mathscr{D}g\|_{L^2(\mathbb{S}^2)}^2,$$

$$\bar{F}_\eta(g) := \mathbb{E}[F_\eta(g)] = 4\pi\mathbb{E}|W_{11} - \mu(U_{11})|^2 + \|\mu - g\|_{L^2(\mathbb{S}^2)}^2 + \eta\|\mathscr{D}g\|_{L^2(\mathbb{S}^2)}^2.$$

Let $\mathbb{X}_1$ and $\mathbb{X}_2$ be normed spaces. We will use $\mathcal{B}(\mathbb{X}_1, \mathbb{X}_2)$ to denote the set of all linear and bounded operators from $\mathbb{X}_1$ to $\mathbb{X}_2$. Here, we consider $\mathcal{H}_p$ endowed with $\|\mathscr{D}\cdot\|_{L^2(\mathbb{S}^2)}$.

Lemma 2. *Let* f, g, g_1, g_2 *be arbitrary elements of* $\mathcal{H}_p$.

1. *The Fréchet derivative of* F_η *at* f *is the element* $F'_\eta(f)$ *of* $\mathcal{B}(\mathcal{H}_p, \mathbb{R})$ *characterized by*

$$F'_\eta(f)g = -\frac{8\pi}{nr}\sum_{i=1}^n\sum_{j=1}^r(W_{ij} - f(U_{ij}))g(U_{ij}) + 2\eta\langle\mathscr{D}f, \mathscr{D}g\rangle_{L^2(\mathbb{S}^2)}$$

The second Fréchet derivative $F''_\eta \in \mathcal{B}(\mathcal{H}_p, \mathcal{B}(\mathcal{H}_p, \mathbb{R}))$ *is characterized by*

$$F''_\eta g_1 g_2 = \frac{8\pi}{nr}\sum_{i=1}^n\sum_{j=1}^r g_1(U_{ij})g_2(U_{ij}) + 2\eta\langle\mathscr{D}g_1, \mathscr{D}g_2\rangle_{L^2(\mathbb{S}^2)}.$$

2. *The Fréchet derivative of $\bar{F}_\eta$ at f is the element $\bar{F}'_\eta(f)$ of $\mathcal{B}(\mathcal{H}_p, \mathbb{R})$ characterized by*

$$\bar{F}'_\eta(f)g = -2\langle \mu - f, g \rangle_{L^2(\mathbb{S}^2)} + 2\eta\langle \mathscr{D}f, \mathscr{D}g \rangle_{L^2(\mathbb{S}^2)}$$

The second Fréchet derivative $\bar{F}''_\eta \in \mathcal{B}(\mathcal{H}_p, \mathcal{B}(\mathcal{H}_p, \mathbb{R}))$ is characterized by

$$\bar{F}''_\eta g_1 g_2 = 2\langle g_1, g_2 \rangle_{L^2(\mathbb{S}^2)} + 2\eta\langle \mathscr{D}g_1, \mathscr{D}g_2 \rangle_{L^2(\mathbb{S}^2)}. \tag{72}$$

Proof. The proof is a direct application of Theorem 3.6.4 in [21]. See also Lemma 8.3.3. □

The evaluation of $\tilde{\mu}_\eta$, given in Equation (68), involves the inverse of the operator $\bar{F}''_\eta$. To this purpose, it is convenient to invoke the Riesz representation theorem (see [21, Theorem 3.2.1]), which tells us that there is an invertible norm-preserving mapping $\mathscr{Q}$ such that $\mathscr{Q}\mathcal{B}(\mathcal{H}_p, \mathbb{R}) = \mathcal{H}_p$. Thus,

$$\tilde{F}''_\eta := \mathscr{Q}\bar{F}''_\eta \tag{73}$$

is an element of $\mathcal{B}(\mathcal{H}_p, \mathcal{H}_p)$ and it is invertible if and only if $\bar{F}''_\eta$ is invertible.

Lemma 3. *The operator $\tilde{F}''_\eta$ in (73) is an invertible element of $\mathcal{B}(\mathcal{H}_p, \mathcal{H}_p)$ and, for any $g \in \mathcal{H}_p$,*

$$(\tilde{F}''_\eta)^{-1}g = \frac{1}{2}\sum_{\ell=0}^{\ell}\sum_{m=-\ell}^{\ell} \frac{D_\ell^2}{1 + \eta D_\ell^2}\langle g, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}Y_{\ell,m}.$$

Notice that $(\tilde{F}''_\eta)^{-1}$ is a self-adjoint operator.

Proof. Take $g_1 \in \mathcal{H}_p$, so that $\bar{F}''_\eta g_1$ belongs to $\mathcal{B}(\mathcal{H}_p, \mathbb{R})$, with representer $\tilde{F}''_\eta g_1 \in \mathcal{H}_p$. Then, for any $g_2 \in \mathcal{H}_p$,

$$\bar{F}''_\eta g_1 g_2 = \langle \mathscr{D}\tilde{F}''_\eta g_1, \mathscr{D}g_2 \rangle_{L^2(\mathbb{S}^2)} = 2\langle g_1, g_2 \rangle_{L^2(\mathbb{S}^2)} + 2\eta\langle \mathscr{D}g_1, \mathscr{D}g_2 \rangle_{L^2(\mathbb{S}^2)},$$

where the last equality comes from (72). We can hence write the expansion in $\mathcal{H}_p$

$$\tilde{F}''_\eta g_1 = 2\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \frac{1 + \eta D_\ell^2}{D_\ell^2}\langle g_1, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}Y_{\ell,m},$$

which suggests that $\tilde{F}''_\eta$ is invertible and

$$(\tilde{F}''_\eta)^{-1}g_1 = \frac{1}{2}\sum_{\ell=0}^{\ell}\sum_{m=-\ell}^{\ell} \frac{D_\ell^2}{1 + \eta D_\ell^2}\langle g_1, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}Y_{\ell,m}. \qquad \square$$

Proof of Theorem 4. As for Theorem 3, without loss of generality, we will consider $D_\ell = (1 + \ell(\ell + 1))^{p/2}$, which leads to a penalization term in the $\mathbb{H}_p$ norm. However, all the following steps can be generalized to every spherical pseudo-differential operator in Definition 2.

In keeping with the notation that was used for proving Theorem 3, we first define

$$F_\eta(g) := \frac{(4\pi)^2}{n} \sum_{i=1}^n \frac{1}{r_i(r_i-1)} \sum_{1 \le j \ne k \le r_i} (W_{ij}W_{ik} - g(U_{ij}, U_{ik}))^2 + \eta\|g\|_{\mathbb{H}_p}^2,$$

$$\bar{F}_\eta(g) := \mathbb{E}[F_\eta(g)] = (4\pi)^2 \operatorname{Var}[W_{11}W_{12}] + \|R - g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 + \eta\|g\|_{\mathbb{H}_p}^2,$$

and we let

$$\bar{R}_\eta := \argmin_{g \in \mathbb{H}_p} \bar{F}_\eta(g).$$

We also define

$$\tilde{R}_\eta = \bar{R}_\eta - (\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta). \tag{74}$$

The definitions of F_η' and $\bar{F}_\eta''$ will be given in Lemma 4, while the existence of $(\bar{F}_\eta'')^{-1}$ will be discussed in Lemma 5.

Now, we can write $R_\eta - R = R_\eta - \tilde{R}_\eta + \tilde{R}_\eta - \bar{R}_\eta + \bar{R}_\eta - R$. In parallel to Proof of Theorem 3, we must show the following

1. $\|\bar{R}_\eta - R\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 \le M_1 \left((nr/\log n)^{-p/(p+1)} + n^{-1}\right)$,
2. $\mathbb{E}\|\tilde{R}_\eta - \bar{R}_\eta\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 \le M_2 \left((nr/\log n)^{-p/(p+1)} + n^{-1}\right)$,
3. $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \sup_{\mathbb{P}_X \in \Pi_2(p,q)} \mathbb{P}\left(\|R_\eta - \tilde{R}_\eta\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 > \varepsilon\left(\left(\frac{\log n}{nr}\right)^{p/(p+1)} + \frac{1}{n}\right)\right) = 0,$$

whenever $\eta \asymp (nr/\log n)^{-p/(p+1)}$, for any choice of the sampling distribution in $\Pi_2(p,q)$.

At this point, we define the *intermediate* norm $\|\cdot\|_\alpha$, $\alpha \in [0,1]$, between $\|\cdot\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$ and $\|\cdot\|_{\mathbb{H}_p}$. Let $g \in L^2(\mathbb{S}^2 \times \mathbb{S}^2)$, then

$$\|g\|_\alpha^2 := \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \sum_{\ell'=0}^\infty \sum_{m'=-\ell'}^{\ell'} D_\ell^{2\alpha} D_{\ell'}^{2\alpha} \langle g, Y_{\ell,m} \otimes Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2,$$

which satisfies

$$\|g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} = \|g\|_0 \le \|g\|_\alpha \le \|g\|_1 = \|g\|_{\mathbb{H}_p}.$$

Similarly as in Proof of Theorem 3, $g \mapsto \|g\|_\alpha$ specifies a norm on $\mathbb{H}_{p\alpha}$ which is equivalent to $\|\cdot\|_{\mathbb{H}_{p\alpha}}$.

An argument analogous to that used for proving Equation (69) shows that, for $\alpha \in [0,1]$,

$$\|R - \bar{R}_\eta\|_\alpha^2 \le \eta^{1-\alpha}\|R\|_{\mathbb{H}_p}^2. \tag{75}$$

Hence, setting $\alpha = 0$,

$$\|R - \bar{R}_\eta\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2 \le \eta\|R\|_{\mathbb{H}_p}^2 \le K_1\eta,$$

which gives the claimed result under the assumptions on η.

We now prove 2, by first showing that

$$\|\tilde{R}_\eta - \bar{R}_\eta\|_\alpha^2 = \frac{1}{4} \sum_{\ell,m} \sum_{\ell',m'} \frac{D_\ell^{2\alpha} D_{\ell'}^{2\alpha}}{(1 + \eta D_\ell^2 D_{\ell'}^2)^2} (F_\eta'(\bar{R}_\eta) Y_{\ell,m} \otimes Y_{\ell',m'})^2. \tag{76}$$

By the definitions of $\tilde{R}_\eta$ and $\|\cdot\|_\alpha$,

$$\|\tilde{R}_\eta - \bar{R}_\eta\|_\alpha^2 = \|(\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta)\|_\alpha^2$$
$$= \sum_{\ell,m} \sum_{\ell',m'} D_\ell^{2\alpha} D_{\ell'}^{2\alpha} \langle (\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta), Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2.$$

However,

$$\langle (\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta), Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$
$$= \frac{1}{D_\ell^2 D_{\ell'}^2} \langle (\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta), Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{\mathbb{H}_p}$$
$$= \frac{1}{D_\ell^2 D_{\ell'}^2} \langle \mathcal{Q} F_\eta'(\bar{R}_\eta), (\tilde{F}_\eta'')^{-1} Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{\mathbb{H}_p},$$

since $\mathcal{Q} F_\eta'(\bar{R}_\eta)$ is the representer of $F_\eta'(\bar{R}_\eta)$ and $(\tilde{F}_\eta'')^{-1}$ is self-adjoint. From Lemma 5,

$$(\tilde{F}_\eta'')^{-1} Y_{\ell,m} \otimes Y_{\ell',m'} = \frac{1}{2} \frac{D_\ell^2 D_{\ell'}^2}{1 + \eta D_\ell^2 D_{\ell'}^2} Y_{\ell,m} \otimes Y_{\ell',m'};$$

thus,

$$\langle (\bar{F}_\eta'')^{-1} F_\eta'(\bar{R}_\eta), Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$
$$= \frac{1}{2(1 + \eta D_\ell^2 D_{\ell'}^2)} \langle \mathcal{Q} F_\eta'(\bar{R}_\eta), Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{\mathbb{H}_p}$$
$$= \frac{1}{2(1 + \eta D_\ell^2 D_{\ell'}^2)} F_\eta'(\bar{R}_\eta) Y_{\ell,m} \otimes Y_{\ell',m'}.$$

Now observe that $\bar{F}_\eta'(\bar{R}_\eta) = 0$ (see [21, Theorem 3.6.3]). Then, an application of Lemma 4 reveals that, for any $g \in \mathbb{H}_p$,

$$F_\eta'(\bar{R}_\eta)g = F_\eta'(\bar{R}_\eta)g - \bar{F}_\eta'(\bar{R}_\eta)g$$
$$= -\frac{2(4\pi)^2}{n} \sum_{i=1}^n \frac{1}{r_i(r_i-1)} \sum_{1 \le j \ne k \le r_i} (W_{ij} W_{ik} - \bar{R}_\eta(U_{ij}, U_{ik})) g(U_{ij}, U_{ik})$$
$$+ 2\langle R - \bar{R}_\eta, g \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}.$$

Consequently,

$$\mathbb{E}[F_\eta'(\bar{R}_\eta) Y_{\ell,m} \otimes Y_{\ell',m'}] = \mathbb{E}_{\mathcal{U}} \mathbb{E}[F_\eta'(\bar{R}_\eta) Y_{\ell,m} \otimes Y_{\ell',m'} | \mathcal{U}] = 0$$

and

$$\mathbb{E}|F'_\eta(\bar{R}_\eta)Y_{\ell,m} \otimes Y_{\ell',m'}|^2 = \mathrm{Var}[F'_\eta(\bar{R}_\eta)Y_{\ell,m} \otimes Y_{\ell',m'}] = \frac{4(4\pi)^4}{n^2}\sum_{i=1}^{n}\frac{1}{r_i^2(r_i-1)^2}$$

$$\times \mathrm{Var}\left[\sum_{1\leq j\neq k\leq r_i}(W_{ij}W_{ik} - \bar{R}_\eta(U_{ij},U_{ik}))Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right].$$

Using the law of total variance, for a generic i we can write

$$\mathrm{Var}\left[\sum_{1\leq j\neq k\leq r_i}(W_{ij}W_{ik} - \bar{R}_\eta(U_{ij},U_{ik}))Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right]$$

$$= \mathrm{Var}\left[\sum_{1\leq j\neq k\leq r_i}(R(U_{ij},U_{ik}) - \bar{R}_\eta(U_{ij},U_{ik}))Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right]$$

$$+ \mathbb{E}_{\mathcal{U}}\left[\mathrm{Var}\left[\sum_{1\leq j\neq k\leq r_i}W_{ij}W_{ik}Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\Big|\mathcal{U}\right]\right]. \tag{77}$$

In what follows, we will handle sums over four indices j, k, j', k'. It is then useful to identify the distinct cases which lead to terms of different orders. Recall that $j \neq k$, $j' \neq k'$, then we have

1. terms of order $r_i(r_i - 1)$:
    (a) $j = j'$, $k = k'$
    (b) $j = k'$, $j' = k$
2. terms of order $r_i(r_i - 1)(r_i - 2)$:
    (a) $j = j'$, $k \neq k'$
    (b) $j \neq j'$, $k = k'$
    (c) $j \neq j'$, $k \neq k'$, $j = k'$, $j' \neq k$
    (d) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' = k$
3. terms of order $r_i(r_i - 1)(r_i - 2)(r_i - 3)$:
    (a) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' \neq k$

Now, for the first term on the right hand side of Equation (77), we obtain

$$\mathrm{Var}\left[\sum_{1\leq j\neq k\leq r_i}(R(U_{ij},U_{ik}) - \bar{R}_\eta(U_{ij},U_{ik}))Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right]$$

$$= \mathbb{E}\left(\sum_{1\leq j\neq k\leq r_i}(R(U_{ij},U_{ik}) - \bar{R}_\eta(U_{ij},U_{ik}))Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right)^2$$

$$-\frac{r_i^2(r_i-1)^2}{(4\pi)^4}\langle R-\bar{R}_\eta, Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}$$

$$\leq B\frac{2r_i(r_i-1)+4r_i(r_i-1)(r_i-2)}{(4\pi)^2}\|R-\bar{R}_\eta\|^2_{\mathbb{H}_p}$$

$$+\frac{r_i(r_i-1)(r_i-2)(r_i-3)-r_i^2(r_i-1)^2}{(4\pi)^4}\langle R-\bar{R}_\eta, Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}$$

$$\leq BK_1\frac{2r_i(r_i-1)+4r_i(r_i-1)(r_i-2)}{(4\pi)^2},$$

where the last two inequalities are justified by Lemma 1, for some $B>0$, and Equation (75) with $\alpha=1$. For the second term, write

$$\sum_{1\leq j\neq k\leq r_i} W_{ij}W_{ik}Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})$$

$$=\sum_{1\leq j\neq k\leq r_i} X_i(U_{ij})X_i(U_{ik})Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})$$

$$+\sum_{1\leq j\neq k\leq r_i} \epsilon_{ij}X_i(U_{ik})Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})$$

$$+\sum_{1\leq j\neq k\leq r_i} X_i(U_{ij})\epsilon_{ik}Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})$$

$$+\sum_{1\leq j\neq k\leq r_i} \epsilon_{ij}\epsilon_{ik}Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik}).$$

Denote the four terms in this last expression by S_1, S_2, S_3, and S_4 with indices corresponding to their location in the sum. Then, by an application of the Cauchy-Schwartz inequality,

$$\mathbb{E}_{\mathcal{U}}[\mathrm{Var}\,[S_1+S_2+S_3+S_4|\mathcal{U}]]\leq 4\left(\mathbb{E}|S_1|^2+\mathbb{E}|S_2|^2+\mathbb{E}|S_3|^2+\mathbb{E}|S_4|^2\right).$$

We will illustrate how to derive the bound for S_1, since the other three are somewhat simpler to handle and of order at most r_i^3. Thus, following the scheme previously described, we obtain

$$\mathbb{E}\left(\sum_{1\leq j\neq k\leq r_i} X_i(U_{ij})X_i(U_{ik})Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})\right)^2$$

$$=\sum_{1\leq j\neq k\leq r_i}\sum_{1\leq j'\neq k'\leq r_i} \mathbb{E}_{\mathcal{U}}[Y_{\ell,m}(U_{ij})Y_{\ell',m'}(U_{ik})Y_{\ell,m}(U_{ij'})Y_{\ell',m'}(U_{ik'})$$

$$\times\mathbb{E}[X_i(U_{ij})X_i(U_{ik})X_i(U_{ij'})X_i(U_{ik'})|\mathcal{U}]]$$

$$\leq\frac{r_i(r_i-1)(r_i-2)(r_i-3)}{(4\pi)^4}\mathbb{E}\langle X,Y_{\ell,m}\rangle^2_{L^2(\mathbb{S}^2)}\langle X,Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2)}$$

$$+B'\frac{2r_i(r_i-1)+4r_i(r_i-1)(r_i-2)}{(4\pi)^2}\mathbb{E}\|X\|^4_{\mathcal{H}_q},$$

for some $B' > 0$. Indeed, for instance, when we consider cases 2.c-2.d,

$$\left| \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} \mathbb{E}\left[X(u)X(v)|X(w)|^2 \right] Y_{\ell,m}(w)Y_{\ell',m'}(u)Y_{\ell,m}(v)Y_{\ell',m'}(w)dudvdw \right|$$

$$\leq B' \, \mathbb{E}\|X\|_{\mathcal{H}_q}^2 \int_{\mathbb{S}^2} |Y_{\ell',m'}(u)|du \int_{\mathbb{S}^2} |Y_{\ell,m}(v)|dv \int_{\mathbb{S}^2} |Y_{\ell,m}(w)||Y_{\ell',m'}(w)|dw$$

$$\leq B'(4\pi)\mathbb{E}\|X\|_{\mathcal{H}_q}^4,$$

where again we have used Lemma 1. Hence, combining all the bounds,

$$\mathbb{E}|F_\eta'(\bar{R}_\eta)Y_{\ell,m} \otimes Y_{\ell',m'}|^2 \leq \frac{4}{n}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2 \langle X, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$+ O\left(\frac{1}{nr}\right),$$

and

$$\mathbb{E}\|\tilde{R}_\eta - \bar{R}_\eta\|_\alpha^2 \leq \frac{1}{n}\sum_{\ell,m}\sum_{\ell',m'} \frac{D_\ell^{2\alpha}D_{\ell'}^{2\alpha}}{(1+\eta D_\ell^2 D_{\ell'}^2)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2 \langle X, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$+ O\left(\frac{1}{nr}\right)\sum_{\ell,\ell'} \frac{D_\ell^{2\alpha}D_{\ell'}^{2\alpha}}{(1+\eta D_\ell^2 D_{\ell'}^2)^2}(2\ell+1)(2\ell'+1),$$

where again r is the harmonic mean of $r_1, \ldots, r_n$. Now, for $\alpha \leq q/p$,

$$\sum_{\ell,m}\sum_{\ell',m'} \frac{D_\ell^{2\alpha}D_{\ell'}^{2\alpha}}{(1+\eta D_\ell^2 D_{\ell'}^2)^2}\mathbb{E}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2 \langle X, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}^2$$

$$\leq \mathbb{E}\left(\sum_{\ell,m} D_\ell^{2\alpha}\langle X, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}^2\right)^2$$

$$\leq \mathbb{E}\|X\|_{p/q}^4,$$

which is bounded (possibly up to an arbitrary constant) by $\mathbb{E}\|X\|_{\mathcal{H}_q}^4$. Moreover, from [29],

$$\sum_{\ell=0}^{\infty}\sum_{\ell'=0}^{\infty} \frac{D_\ell^{2\alpha}D_{\ell'}^{2\alpha}}{(1+\eta D_\ell^2 D_{\ell'}^2)^2}(2\ell+1)(2\ell'+1) = O\left(\eta^{-(\alpha+1/p)}\log(1/\eta) + 1\right).$$

Thus, we have that

$$\mathbb{E}\|\tilde{R}_\eta - \bar{R}_\eta\|_\alpha^2 \leq M_2\left((nr)^{-1}\eta^{-(\alpha+1/p)}\log(1/\eta) + n^{-1}\right),$$

where M_2 is a positive constant not depending on the choice of $\mathbb{P}_X \in \Pi_2(p,q)$. By choosing $\alpha = 0$, we obtain the claimed rate.

Now, let us prove 3. The first step is to obtain a useful analytic form for $R_\eta - \tilde{R}_\eta$, by observing that

$$R_\eta - \tilde{R}_\eta = R_\eta - \bar{R}_\eta + (\bar{F}_\eta'')^{-1}F_\eta'(\bar{R}_\eta)$$

$$= (\bar{F}''_\eta)^{-1} \left[\bar{F}''_\eta (R_\eta - \bar{R}_\eta) + F'_\eta(\bar{R}_\eta) \right].$$

Since R_η minimizes F'_η, it holds $F'_\eta(R_\eta) = 0$ (see [21, Theorem 3.6.3]). Then, for any $g \in \mathbb{H}_p$,

$$
\begin{aligned}
&\left[\bar{F}''_\eta (R_\eta - \bar{R}_\eta) + F'_\eta(\bar{R}_\eta) \right] g \\
&= \left[\bar{F}''_\eta (R_\eta - \bar{R}_\eta) + F'_\eta(\bar{R}_\eta) - F'_\eta(R_\eta) \right] g \\
&= \left[\bar{F}''_0 (R_\eta - \bar{R}_\eta) - F''_0(R_\eta - \bar{R}_\eta) \right] g.
\end{aligned}
$$

where we used Lemma 4; in other words,

$$R_\eta - \tilde{R}_\eta = (\bar{F}''_\eta)^{-1} \left[\bar{F}''_0 (R_\eta - \bar{R}_\eta) - F''_0(R_\eta - \bar{R}_\eta) \right].$$

Now, the same argument that leads to (76) gives us

$$
\begin{aligned}
\|R_\eta - \tilde{R}_\eta\|_\alpha^2 = \frac{1}{4} \sum_{\ell,m} \sum_{\ell',m'} &\frac{D_\ell^{2\alpha} D_{\ell'}^{2\alpha}}{(1 + \eta D_\ell^2 D_{\ell'}^2)^2} \\
&\times \left(\left[\bar{F}''_0 (R_\eta - \bar{R}_\eta) - F''_0(R_\eta - \bar{R}_\eta) \right] Y_{\ell,m} \otimes Y_{\ell',m'} \right)^2,
\end{aligned}
$$

with

$$
\begin{aligned}
&\left[\bar{F}''_0 (R_\eta - \bar{R}_\eta) - F''_0(R_\eta - \bar{R}_\eta) \right] Y_{\ell,m} \otimes Y_{\ell',m'} \\
&= 2\langle R_\eta - \bar{R}_\eta, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} - \frac{2(4\pi)^2}{n} \sum_{i=1}^n \frac{1}{r_i(r_i - 1)} \\
&\times \sum_{1 \le j \ne k \le r_i} (R_\eta(U_{ij}, U_{ik}) - \bar{R}_\eta(U_{ij}, U_{ik})) Y_{\ell,m}(U_{ij}) Y_{\ell',m'}(U_{ik}).
\end{aligned}
$$

Now, $R_\eta - \bar{R}_\eta = \sum_{\ell,m} \sum_{\ell',m'} h_{\ell,\ell',m,m'} Y_{\ell,m} Y_{\ell',m'}$. Then,

$$
\|R_\eta - \tilde{R}_\eta\|_\alpha^2 = \sum_{\ell_1,m_1} \sum_{\ell_2,m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2} \left(\sum_{\ell_3,m_3} \sum_{\ell_4,m_4} h_{\ell_3,\ell_4,m_3,m_4} V_{\ell_{1:4},m_{1:4}} \right)^2,
$$

where

$$
\begin{aligned}
V_{\ell_{1:4},m_{1:4}} = \delta_{\ell_1}^{\ell_3} \delta_{m_1}^{m_3} \delta_{\ell_2}^{\ell_4} \delta_{m_2}^{m_4} &- \frac{(4\pi)^2}{n} \sum_{i=1}^n \frac{1}{r_i(r_i - 1)} \\
&\times \sum_{1 \le j \ne k \le r_i} Y_{\ell_1,m_1}(U_{ij}) Y_{\ell_2,m_2}(U_{ik}) Y_{\ell_3,m_3}(U_{ij}) Y_{\ell_4,m_4}(U_{ik}).
\end{aligned}
$$

By applying the Cauchy–Schwarz inequality for arbitrary $\theta \in (1/p, 1]$, we obtain

$$
\|R_\eta - \tilde{R}_\eta\|_\alpha^2 \le \|R_\eta - \bar{R}_\eta\|_\theta^2 \sum_{\ell_1,m_1} \sum_{\ell_2,m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2}
$$

$$\times \sum_{\ell_3, m_3} \sum_{\ell_4, m_4} D_{\ell_3}^{-2\theta} D_{\ell_4}^{-2\theta} V_{\ell_{1:4}, m_{1:4}}^2.$$

It is readily seen that $\mathbb{E}[V_{\ell_{1:4}, m_{1:4}}] = 0$ and

$$\mathbb{E}[V_{\ell_{1:4}, m_{1:4}}^2] = -\delta_{\ell_1}^{\ell_3} \delta_{m_1}^{m_3} \delta_{\ell_2}^{\ell_4} \delta_{m_2}^{m_4} + \frac{(4\pi)^4}{n^2} \sum_{i=1}^n \frac{1}{r_i^2 (r_i - 1)^2}$$

$$\times \mathbb{E}\left(\sum_{1 \leq j \neq k \leq r_i} Y_{\ell_1, m_1}(U_{ij}) Y_{\ell_2, m_2}(U_{ik}) Y_{\ell_3, m_3}(U_{ij}) Y_{\ell_4, m_4}(U_{ik}) \right)^2.$$

Thus, it is possible to show that

$$\mathbb{E}[V_{\ell_{1:4}, m_{1:4}}^2] \leq D_{\ell_1}^{2\theta} D_{\ell_2}^{2\theta} O\left(\frac{1}{nr} \right),$$

and hence

$$\sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2} \sum_{\ell_3, m_3} \sum_{\ell_4, m_4} D_{\ell_3}^{-2\theta} D_{\ell_4}^{-2\theta} \mathbb{E}[V_{\ell_{1:4}, m_{1:4}}^2]$$

$$= O\left(\frac{\log(1/\eta)}{nr\eta^{\alpha + \theta + 1/p}} \right).$$

Let us define

$$A_n := \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2} \sum_{\ell_3, m_3} \sum_{\ell_4, m_4} D_{\ell_3}^{-2\theta} D_{\ell_4}^{-2\theta} V_{\ell_{1:4}, m_{1:4}}^2,$$

$a_n := \frac{\log(1/\eta)}{nr\eta^{\alpha + \theta + 1/p}}$ and $\gamma_n := \frac{\log(1/\eta)}{nr\eta^{\alpha + 1/p}} + \frac{1}{n}$. Note that $A_n = \sup_{\mathbb{P}_X \in \Pi_2(p,q)} A_n$ and $A_n = o_{\mathbb{P}}(1)$.

Then, $\|R_\eta - \tilde{R}_\eta\|_\alpha^2 \leq A_n \|R_\eta - \bar{R}_\eta\|_\theta^2$ and therefore

$$\mathbb{P}\left(\|R_\eta - \tilde{R}_\eta\|_\alpha^2 > \varepsilon \gamma_n \right) \leq \mathbb{P}\left(A_n \|R_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon \gamma_n \right)$$

$$= \mathbb{P}\left(A_n \|R_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon \gamma_n, A_n < 1 \right)$$

$$+ \mathbb{P}\left(A_n \|R_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon \gamma_n, A_n \geq 1 \right)$$

$$\leq \mathbb{P}\left(A_n \|R_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon \gamma_n, A_n < 1 \right) + \mathbb{P}\left(A_n \geq 1 \right).$$

If $A_n < 1$,

$$\|\tilde{R}_\eta - \bar{R}_\eta\|_\theta \geq \|R_\eta - \bar{R}_\eta\|_\theta - \|R_\eta - \tilde{R}_\eta\|_\theta$$

$$\geq (1 - \sqrt{A_n}) \|R_\eta - \bar{R}_\eta\|_\theta,$$

which allows to write

$$\mathbb{P}\left(A_n \|R_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon \gamma_n, A_n < 1 \right)$$

$$\leq \mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}\|\tilde{R}_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon\gamma_n, A_n < 1\right)$$

$$\leq \mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}\|\tilde{R}_\eta - \bar{R}_\eta\|_\theta^2 > \varepsilon\gamma_n\right).$$

Let us now define $B_n := \|\tilde{R}_\eta - \bar{R}_\eta\|_\theta^2$ and $b_n := \frac{\log(1/\eta)}{nr\eta^{\theta+1/p}} + \frac{1}{n}$. Recall that $\mathbb{E}[B_n] \leq M_2 b_n$, for $\theta \leq q/p$. Moreover, $|1 - \sqrt{A_n}|^{-2} = O_\mathbb{P}(1)$. We can observe that

$$a_n b_n = \frac{\log(1/\eta)}{nr\eta^{2\theta+1/p}}\left(\frac{\log(1/\eta)}{nr\eta^{\alpha+1/p}} + \frac{\eta^{\theta-\alpha}}{n}\right),$$

so that $c_n := a_n b_n/\gamma_n \to 0$, assuming $\theta < 1/2$ (recall that $p > 2$) and $\alpha \in [0, \theta]$. Then,

$$\mathbb{P}\left(A_n|1 - \sqrt{A_n}|^{-2}B_n > \varepsilon\gamma_n\right) = \mathbb{P}\left(\frac{A_n|1 - \sqrt{A_n}|^{-2}B_n}{a_n b_n} > \frac{\varepsilon}{c_n}\right)$$

$$\leq \mathbb{P}\left(\frac{B_n}{b_n} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right) + \mathbb{P}\left(\frac{A_n}{a_n} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right)$$

$$+ \mathbb{P}\left(|1 - \sqrt{A_n}|^{-2} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right)$$

$$\leq M_2 \frac{c_n^{1/3}}{\varepsilon^{1/3}} + M_3 \frac{c_n^{1/3}}{\varepsilon^{1/3}}$$

$$+ \mathbb{P}\left(|1 - \sqrt{A_n}|^{-2} > \frac{\varepsilon^{1/3}}{c_n^{1/3}}\right).$$

Clearly $c_n^{1/3}|1 - \sqrt{A_n}|^{-2} = o_\mathbb{P}(1)$, hence

$$\lim_{n\to\infty} \sup_{\mathbb{P}_X \in \Pi_2(p,q)} \mathbb{P}\left(\|R_\eta - \tilde{R}_\eta\|_\alpha^2 > \varepsilon\gamma_n\right) = 0.$$

By taking $\alpha = 0$ we obtain the claimed result.

Recall now that, when the mean is $\mu \neq 0$, an estimate of the complete covariance kernel $C(u, v) = R(u, v) - \mu(u)\mu(v)$ is given by

$$C_\eta(u, v) = R_\eta(u, v) - \mu_\eta(u)\mu_\eta(v).$$

Moreover, observe that

$$\|C_\eta - C\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)} \leq \|R_\eta - R\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)} + \|\mu_\eta \otimes \mu_\eta - \mu \otimes \mu\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)},$$

by the triangle inequality; hence,

$$\mathbb{P}\left(\|C_\eta - C\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}^2 > t\right) \leq \mathbb{P}\left(\|R_\eta - R\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}^2 > t\right)$$

$$+ \mathbb{P}\left(\|\mu_\eta \otimes \mu_\eta - \mu \otimes \mu\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}^2 > t\right).$$

For the second term on the right-hand side, we have

$$\|\mu_\eta \otimes \mu_\eta - \mu \otimes \mu\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)} = \|\mu_\eta \otimes \mu_\eta \pm \mu_\eta \otimes \mu - \mu \otimes \mu\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}$$

$$\leq \|\mu_\eta\|_{L^2(\mathbb{S}^2)}\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}$$
$$+ \|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}\|\mu\|_{L^2(\mathbb{S}^2)}$$
$$\leq \|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}^2 + 2\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}\|\mu\|_{L^2(\mathbb{S}^2)}$$
$$= O_{\mathbb{P}}\left(\|\mu_\eta - \mu\|_{L^2(\mathbb{S}^2)}\right).$$

If we repeat the same steps in Proof of Theorem 3, but this time with $\eta \asymp (nr/\log n)^{-p/(p+1)}$, we see that the (uniform) rate for μ_η is exactly $(nr/\log n)^{-p/(p+1)} + n^{-1}$, which concludes the proof. $\qquad\square$

The next two lemmas are referred to

$$F_\eta(g) := \frac{(4\pi)^2}{n} \sum_{i=1}^{n} \frac{1}{r_i(r_i - 1)} \sum_{1 \leq j \neq k \leq r_i} (W_{ij}W_{ik} - g(U_{ij}, U_{ik}))^2$$
$$+ \eta\|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2,$$

$$\bar{F}_\eta(g) := \mathbb{E}[F_\eta(g)] = (4\pi)^2 \operatorname{Var}[W_{11}W_{12}] + \|R - g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2$$
$$+ \eta\|(\mathscr{D} \otimes \mathscr{D})g\|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}^2.$$

Recall that $\mathcal{B}(\mathbb{X}_1, \mathbb{X}_2)$ denote the set of all linear and bounded operators from $\mathbb{X}_1$ to $\mathbb{X}_2$, being normed spaces. Here, we consider $\mathbb{H}_p$ endowed with $\|(\mathscr{D} \otimes \mathscr{D}) \cdot \|_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$.

Lemma 4. *Let f, g, g_1, g_2 be arbitrary elements of $\mathbb{H}_p$.*

1. *The Fréchet derivative of F_η at f is the element $F'_\eta(f)$ of $\mathcal{B}(\mathbb{H}_p, \mathbb{R})$ characterized by*

$$F'_\eta(f)g = -\frac{2(4\pi)^2}{n} \sum_{i=1}^{n} \frac{1}{r_i(r_i - 1)} \sum_{1 \leq j \neq k \leq r_i} (W_{ij}W_{ik} - f(U_{ij}, U_{ik}))g(U_{ij}, U_{ik})$$
$$+ 2\eta\langle(\mathscr{D} \otimes \mathscr{D})f, (\mathscr{D} \otimes \mathscr{D})g\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$

The second Fréchet derivative $F''_\eta \in \mathcal{B}(\mathbb{H}_p, \mathcal{B}(\mathbb{H}_p, \mathbb{R}))$ is characterized by

$$F''_\eta g_1 g_2 = \frac{2(4\pi)^2}{n} \sum_{i=1}^{n} \frac{1}{r_i(r_i - 1)} \sum_{1 \leq j \neq k \leq r_i} g_1(U_{ij}, U_{ik})g_2(U_{ij}, U_{ik})$$
$$+ 2\eta\langle(\mathscr{D} \otimes \mathscr{D})g_1, (\mathscr{D} \otimes \mathscr{D})g_2\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}.$$

2. *The Fréchet derivative of $\bar{F}_\eta$ at f is the element $\bar{F}'_\eta(f)$ of $\mathcal{B}(\mathbb{H}_p, \mathbb{R})$ characterized by*

$$\bar{F}'_\eta(f)g = -2\langle R - f, g\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} + 2\eta\langle(\mathscr{D} \otimes \mathscr{D})f, (\mathscr{D} \otimes \mathscr{D})g\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$

The second Fréchet derivative $\bar{F}''_\eta \in \mathcal{B}(\mathbb{H}_p, \mathcal{B}(\mathbb{H}_p, \mathbb{R}))$ is characterized by

$$\bar{F}''_\eta g_1 g_2 = 2\langle g_1, g_2\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} + 2\eta\langle(\mathscr{D} \otimes \mathscr{D})g_1, (\mathscr{D} \otimes \mathscr{D})g_2\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}. \quad (78)$$

Proof. The proof is a direct application of Theorem 3.6.4 in [21]. See also Lemma 8.3.3. □

The evaluation of $\tilde{R}_\eta$ involves the inverse of the operator $\bar{F}''_\eta$. To this purpose, it is convenient to invoke the Riesz representation theorem (see [21, Theorem 3.2.1]), which tells us that there is an invertible norm-preserving mapping $\mathscr{Q}$ such that $\mathscr{Q}\mathcal{B}(\mathbb{H}_p, \mathbb{R}) = \mathbb{H}_p$. Thus,

$$\tilde{F}''_\eta := \mathscr{Q}\bar{F}''_\eta \tag{79}$$

is an element of $\mathcal{B}(\mathbb{H}_p, \mathbb{H}_p)$ and it is invertible if and only if $\bar{F}''_\eta$ is invertible.

Lemma 5. *The operator $\tilde{F}''_\eta$ in* (79) *is an invertible element of $\mathcal{B}(\mathbb{H}_p, \mathbb{H}_p)$ and, for any $g \in \mathbb{H}_p$,*

$$(\tilde{F}''_\eta)^{-1} g = \frac{1}{2} \sum_{\ell,\ell'=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \frac{D_\ell^2 D_{\ell'}^2}{1 + \eta D_\ell^2 D_{\ell'}^2} \langle g, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} Y_{\ell,m} Y_{\ell'm'}.$$

Notice that $(\tilde{F}''_\eta)^{-1}$ is a self-adjoint operator.

Proof. Take $g_1 \in \mathbb{H}_p$, so that $\bar{F}''_\eta g_1$ belongs to $\mathcal{B}(\mathbb{H}_p, \mathbb{R})$, with representer $\tilde{F}''_\eta g_1 \in \mathbb{H}_p$. Then, for any $g_2 \in \mathbb{H}_p$,

$$\bar{F}''_\eta g_1 g_2 = \langle (\mathscr{Q} \otimes \mathscr{Q}) \tilde{F}''_\eta g_1, (\mathscr{Q} \otimes \mathscr{Q}) g_2 \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)}$$
$$= 2 \langle g_1, g_2 \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} + 2\eta \langle (\mathscr{Q} \otimes \mathscr{Q}) g_1, (\mathscr{Q} \otimes \mathscr{Q}) g_2 \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)},$$

where the last equality comes from (78). We can hence write the expansion in $\mathbb{H}_p$

$$\tilde{F}''_\eta g_1 = 2 \sum_{\ell,\ell'=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \frac{1 + \eta D_\ell^2 D_{\ell'}^2}{D_\ell^2 D_{\ell'}^2} \langle g_1, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} Y_{\ell,m} Y_{\ell'm'},$$

which suggests that $\tilde{F}''_\eta$ is invertible and

$$(\tilde{F}''_\eta)^{-1} g_1 = \frac{1}{2} \sum_{\ell,\ell'=0}^{\infty} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \frac{D_\ell^2 D_{\ell'}^2}{1 + \eta D_\ell^2 D_{\ell'}^2} \langle g_1, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} Y_{\ell,m} Y_{\ell'm'}.$$
□

We now obtain the asymptotic performance for the two classes of spherical random fields introduced in Section 5.3.

Proof of Proposition 2. We fix $q \leq p$ such that $2 < p < \beta - 1/2$ and $1 < q < \beta - 1$. Note that p and q exist due to the assumption that $\beta > 5/2$. The proof is divided in two parts.

1. $\mathcal{A}_\beta \subset \Pi_2(p, q)$. Let $X \in \mathcal{A}_\beta$. The Gaussian white noise over the d-dimensional hypersphere is almost surely in the Sobolev space with negative smoothness $W \in \mathcal{H}_{-d/2-\epsilon}$ for any $\epsilon > 0$. This result is well-known for Gaussian white noises over $\mathbb{S}^1$ or more generally the d-dimensional torus $\mathbb{T}^d = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$;

see for instance [56, Theorem 3.4] or [14, Theorem 5]. More generally, it is true for the Gaussian white noise over a compact Riemannian manifold with no boundary, as is demonstrated for instance in the proof of Proposition 3.8 in [13]. In our setting, we deduce that $W \in \mathcal{H}_{-1-\epsilon}$ almost surely for any $\epsilon > 0$. Moreover, any admissible operator $\mathscr{D}$ with spectral growth order β is such that $\mathscr{D}^{-1}\mathcal{H}_{-1-\epsilon} = \mathcal{H}_{\beta-1-\epsilon}$, hence $X \in \mathcal{H}_{\beta-1-\epsilon}$ almost surely. For $\epsilon > 0$ small enough, we have that $\beta - 1 - \epsilon \geq q$, hence $\mathcal{H}_{\beta-1-\epsilon} \subset \mathcal{H}_q$. In particular, $X \in \mathcal{H}_q$ almost surely.

Our goal is now to prove that the covariance C of X is in $\mathbb{H}_p$. We first prove that $C(u,v) = K(u,v)$, for any $u, v \in \mathbb{S}^2$, where $K : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ is the reproducing kernel of $(\mathcal{H}_\beta, \|\mathscr{D} \cdot \|_{L^2(\mathbb{S}^2)})$. Using that $X = \mathscr{D}^{-1}W$ and $X(u) = \langle X, \delta_u \rangle_{L^2(\mathbb{S}^2)}$, we have that, for $u, v \in \mathbb{S}^2$,

$$
\begin{aligned}
\mathbb{E}[X(u)X(v)] &= \mathbb{E}[\langle \mathscr{D}^{-1}W, \delta_u \rangle_{L^2(\mathbb{S}^2)} \langle \mathscr{D}^{-1}W, \delta_v \rangle_{L^2(\mathbb{S}^2)}] \\
&= \mathbb{E}[\langle W, (\mathscr{D}^{-1})^* \delta_u \rangle_{L^2(\mathbb{S}^2)} \langle W, (\mathscr{D}^{-1})^* \delta_v \rangle_{L^2(\mathbb{S}^2)}] \\
&= \langle (\mathscr{D}^{-1})^* \delta_u, (\mathscr{D}^{-1})^* \delta_v \rangle_{L^2(\mathbb{S}^2)} \\
&= \langle \mathscr{D}\mathscr{D}^{-1}(\mathscr{D}^{-1})^* \delta_u, \mathscr{D}\mathscr{D}^{-1}(\mathscr{D}^{-1})^* \delta_v \rangle_{L^2(\mathbb{S}^2)} \\
&= \langle (\mathscr{D}^*\mathscr{D})^{-1} \delta_u, (\mathscr{D}^*\mathscr{D})^{-1} \delta_v \rangle_{\mathcal{H}_\beta}.
\end{aligned}
$$

We then observe that $(\mathscr{D}^*\mathscr{D})^{-1}\delta_u = K(\cdot, u)$, $u \in \mathbb{S}^2$ since it satisfy the reproducing property

$$
\langle (\mathscr{D}^*\mathscr{D})^{-1}\delta_u, f \rangle_{\mathcal{H}_\beta} = \langle \mathscr{D}(\mathscr{D}^*\mathscr{D})^{-1}\delta_u, \mathscr{D}f \rangle_{L^2(\mathbb{S}^2)} = \langle \delta_u, f \rangle_{L^2(\mathbb{S}^2)} = f(u),
$$

for every $f \in \mathcal{H}_\beta$. Thus we have that

$$
C(u,v) = \mathbb{E}[X(u)X(v)] = \langle K(\cdot, u), K(\cdot, v) \rangle_{\mathcal{H}_\beta} = K(u,v).
$$

We therefore deduce that, since the Fourier coefficients of K are given by $K_{\ell,m} = 1/|D_\ell|^2$, we have that

$$
\|C\|^2_{\mathbb{H}_p} = \|K\|^2_{\mathbb{H}_p} = \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell (1+\ell(\ell+1))^{2p}|K_{\ell,m}|^2 = \sum_{\ell=0}^\infty (2\ell+1)(1+\ell(\ell+1))^{2p}D_\ell^{-4},
$$

Finally, using that the D_ℓ satisfies (11) (with $p = \beta$), the previous sum is finite if and only if $4\beta - 4p - 1 > 1$, which is true due to the assumption that $p < \beta - 1/2$. Finally, we have shown that $X \in \Pi_2(p, q)$.

2. Existence of C_η. Let $\mathscr{D}$ be an admissible operator of order p such that $2 < p < \beta - 1/2$. Then, we have seen that, for such p and for $q \leq p$, $1 < q < \beta - 1$, $X \in \Pi_2(p, q)$. Then, the estimator $C_\eta = R_\eta$ associated to the operator $\mathscr{D}$ verifies the assumptions of Theorem 4 and the bound of Corollary 2 is achieved. This is true for $p < \beta - 1/2$ arbitrarily close to $\beta - 1/2$. For such p, using that $x \mapsto \frac{x}{x+1}$ is increasing, we have that $\frac{\beta-1/2}{\beta+1/2} - \frac{p}{p+1} > 0$. By setting

$$
\epsilon = \frac{\beta - 1/2}{\beta + 1/2} - \frac{p}{p+1} > 0 \tag{80}
$$

we obtain (37) and concludes the proof. $\square$

Proof of Proposition 3. Let $p < \beta - 1$ Then, for any $u \in \mathbb{S}^2$, the Green's function $\psi_u^{\mathscr{D}}$ of $\mathscr{D}$ is in $\mathcal{H}_{\beta-1-\epsilon}$. This follows from that, $\delta_u \in \mathcal{H}_{-1-\epsilon}$ and the fact that, as we have seen in the proof of Proposition 2, $\psi_u^{\mathscr{D}} = \mathscr{D}^{-1}\delta_u \in \mathcal{H}_{\beta-1-\epsilon} \subset \mathcal{H}_p$ for $\epsilon > 0$ small enough. In particular, $X \in \mathcal{H}_p$ as a (random) linear combination of Green's functions of $\mathscr{D}$. We moreover easily see that the covariance C of X is a linear combinations of tensorial functions $\psi_u^{\mathscr{D}} \otimes \psi_v^{\mathscr{D}}$ for $u, v \in \mathbb{S}^2$ and is therefore in $\mathbb{H}_p$. This shows that $X \in \Pi_2(p, p) \subset \Pi_2(p, q)$ for any $1 \leq q \leq p$ and concludes the first part of Proposition 3. The existence of C_η is then obtained in same way as for the proof of Proposition 2, except that $p < \beta - 1$ and therefore (39) follows with $\epsilon = \frac{\beta-1}{\beta} - \frac{p}{p+1}$. $\qquad\square$

8.4. Proofs of Section 7

Proof of Theorem 5. For simplicity, we consider the case $r_1 = r_2 = \cdots = r_n$.

This proof follows the same lines of Proof of Theorem 3 in Section 8.3. The only different part is the one referred to Point 2. Starting from Equation (71), we have that

$$\mathbb{E}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}] = \mathbb{E}_{\mathcal{U}}\mathbb{E}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|\mathcal{U}] = 0$$

and

$$\mathbb{E}|F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|^2 = \mathrm{Var}[F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}]$$

$$= \frac{(8\pi)^2}{(nr)^2}\mathrm{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(W_{tj} - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\right].$$

Using the law of total variance, we can write

$$\mathrm{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(W_{tj} - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\right]$$

$$= \mathrm{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(\mu(U_{tj}) - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\right]$$

$$+ \mathbb{E}_{\mathcal{U}}\left[\mathrm{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(W_{tj} - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\Big|\mathcal{U}\right]\right].$$

For the first term on the right hand side of this expression, we have

$$\mathrm{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(\mu(U_{tj}) - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\right]$$

$$= nr\,\mathrm{Var}\left[(\mu(U_{11}) - \bar{\mu}_\eta(U_{11}))Y_{\ell,m}(U_{11})\right]$$

$$\leq \frac{nr}{4\pi}\int_{\mathbb{S}^2}|\mu(u) - \bar{\mu}_\eta(u)|^2|Y_{\ell,m}(u)|^2 du$$

$$\leq \frac{Bnr}{4\pi} \|\mu - \bar{\mu}_\eta\|_{\mathcal{H}_p}^2$$

$$\leq \frac{BKnr}{4\pi},$$

where the last two inequalities are justified by Lemma 1, for some $B > 0$, and Equation (69) with $\alpha = 1$. Then, for the second term,

$$\mathbb{E}_{\mathcal{U}}\left[\text{Var}\left[\sum_{t=1}^{n}\sum_{j=1}^{r}(W_{tj} - \bar{\mu}_\eta(U_{tj}))Y_{\ell,m}(U_{tj})\Big|\mathcal{U}\right]\right]$$

$$= \sum_{t=1}^{n}\sum_{t'=1}^{n}\sum_{j=1}^{r}\sum_{j'=1}^{r}\mathbb{E}_{\mathcal{U}}[Y_{\ell,m}(U_{tj})Y_{\ell,m}(U_{t'j'})\,\text{Cov}[W_{tj}, W_{t'j'}|\mathcal{U}]]$$

$$= \frac{r(r-1)}{(4\pi)^2}\sum_{t=1}^{n}\text{Var}[\langle X_t, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}]$$

$$+ \frac{r^2}{(4\pi)^2}\sum_{t\neq t'}\text{Cov}[\langle X_t, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}]$$

$$+ \frac{nr}{4\pi}\int_{\mathbb{S}^2}C_0(u,u)|Y_{\ell,m}(u)|^2 du + \frac{nr}{4\pi}\sigma^2$$

$$\leq \frac{r^2}{(4\pi)^2}\sum_{t=1}^{n}\sum_{t'=1}^{n}\text{Cov}[\langle X_t, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}]$$

$$+ \frac{B'nr}{4\pi}\mathbb{E}\|X_0\|_{\mathcal{H}_q}^2 + \frac{nr}{4\pi}\sigma^2,$$

again by applying Lemma 1, for some $B' > 0$. Now, observe that, by stationarity,

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{t'=1}^{n}|\text{Cov}[\langle X_t, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}]|$$

$$= \frac{1}{n}\sum_{t=1}^{n}\sum_{t'=1}^{n}|\langle \mathcal{C}_{t-t'}Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$

$$= \sum_{|\xi|<n}\left(1 - \frac{|\xi|}{n}\right)|\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$

$$\leq \sum_{\xi\in\mathbb{Z}}|\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|.$$

Hence, combining all the bounds,

$$\mathbb{E}|F_\eta'(\bar{\mu}_\eta)Y_{\ell,m}|^2 \leq \frac{4}{n}\sum_{\xi\in\mathbb{Z}}|\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}| + O\left(\frac{1}{nr}\right),$$

and

$$\mathbb{E}\|\tilde{\mu}_\eta - \bar{\mu}_\eta\|_\alpha^2 \leq \frac{1}{n}\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\frac{D_\ell^{2\alpha}}{(1 + \eta D_\ell^2)^2}\sum_{\xi\in\mathbb{Z}}|\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$

$$+ O\left(\frac{1}{nr}\right) \sum_{\ell=0}^{\infty} \frac{D_\ell^{2\alpha}}{(1+\eta D_\ell^2)^2}(2\ell+1)$$

$$\leq \frac{1}{n} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} D_\ell^{2\alpha} \sum_{\xi\in\mathbb{Z}} |\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$

$$+ O\left(\frac{1}{nr}\right) \sum_{\ell=0}^{\infty} \frac{D_\ell^{2\alpha}}{(1+\eta D_\ell^2)^2}(2\ell+1).$$

Then, for $\alpha \leq q/p$ and using the property in Equation (50), we obtain

$$\sum_{\xi\in\mathbb{Z}} \sum_{\ell=0}^{\infty} D_\ell^{2\alpha} \sum_{m=-\ell}^{\ell} |\langle \mathcal{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}| = \sum_{\xi\in\mathbb{Z}} \sum_{\ell=0}^{\infty} \frac{1}{D_\ell^{2\alpha}} \sum_{m=-\ell}^{\ell} |\langle \mathscr{C}_\xi Y_{\ell,m}, Y_{\ell,m}\rangle_{\mathcal{H}_q}|$$

$$\leq \sum_{\xi\in\mathbb{Z}} \|\mathscr{C}_\xi\|_{\mathrm{TR},\mathcal{H}_q}$$

(possibly up to an arbitrary constant). The rest of the proof follows exactly as in Proof of Theorem 3 in Section 8.3. □

Proof of Theorem 6. For simplicity, we consider the case $r_1 = r_2 = \cdots = r_n$ and we define

$$N_h = \begin{cases} nr(r-1) & h=0 \\ (n-h)r^2 & h\neq 0 \end{cases}.$$

Recall that the estimator for R_h at fixed lag $h \in \{0,1,\ldots,n-1\}$ is given by

$$R_{h;\eta} := \operatorname*{arg\,min}_{g\in\mathbb{H}_p} F_{h;\eta}(g),$$

$$F_{h;\eta}(g) := \frac{(4\pi)^2}{N_h} \sum_{t=1}^{n-h} \sum_{j=1}^{r} \sum_{k=1}^{r} (W_{t+h,j}W_{tk} - g(U_{t+h,j}, U_{tk}))^2 + \eta\|g\|_{\mathbb{H}_p}^2.$$

In keeping with the notation that was used in the previous proof, we first define

$$\bar{F}_{h;\eta}(g) := \mathbb{E}[F_{h;\eta}(g)] = (4\pi)^2 \operatorname{Var}[W_{h+1,1}W_{12}] + \|R_h - g\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}^2 + \eta\|g\|_{\mathbb{H}_p}^2$$

and we let

$$\bar{R}_{h;\eta} := \operatorname*{arg\,min}_{g\in\mathbb{H}_p} \bar{F}_{h;\eta}(g).$$

Also, define $\tilde{R}_{h;\eta} = \bar{R}_{h;\eta} - (\bar{F}_{h;\eta}'')^{-1} F_{h;\eta}'(\bar{R}_{h;\eta})$. Then, write

$$R_{h;\eta} - R_h = R_{h;\eta} - \tilde{R}_{h;\eta} + \tilde{R}_{h;\eta} - \bar{R}_{h;\eta} + \bar{R}_{h;\eta} - R_h.$$

The first part of the proof follows the same lines of Proof of Theorem 4 in Section 8.3. We first define the intermediate norm $\|\cdot\|_\alpha$, $\alpha \in [0,1]$. We then show that

$$\|R_h - \bar{R}_{h;\eta}\|_\alpha^2 \leq \eta^{1-\alpha}\|R_h\|_{\mathbb{H}_p}^2, \tag{81}$$

and hence that

$$\|R_h - \bar{R}_{h\eta}\|^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} \le \eta \|R_h\|^2_{\mathbb{H}_p} \le K_1 \eta,$$

which proves the analogous of Point 1. t Afterwards, we show that

$$\|\tilde{R}_{h;\eta} - \bar{R}_{h;\eta}\|^2_\alpha = \frac{1}{4} \sum_{\ell,m} \sum_{\ell',m'} \frac{D_\ell^{2\alpha} D_{\ell'}^{2\alpha}}{(1 + \eta D_\ell^2 D_{\ell'}^2)^2} (F'_{h;\eta}(\bar{R}_{h;\eta}) Y_{\ell,m} \otimes Y_{\ell',m'})^2. \quad (82)$$

Now, we consider the following quantity

$$F'_{h;\eta}(\bar{R}_{h;\eta})g = F'_{h;\eta}(\bar{R}_{h;\eta})g - \bar{F}'_{h;\eta}(\bar{R}_{h;\eta})g$$

$$= -\frac{2(4\pi)^2}{N_h} \sum_{t=1}^{n-h} \sum_{\substack{j=1 \\ j \ne k \text{ if } h=0}}^{r} \sum_{k=1}^{r} (W_{t+h,j} W_{tk} - \bar{R}_{h;\eta}(U_{t+h,j}, U_{tk})) g(U_{t+h,j}, U_{tk})$$

$$+ 2\langle R_h - \bar{R}_{h;\eta}, g\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)},$$

$g \in \mathbb{H}_p$. When $g = Y_{\ell,m} \otimes Y_{\ell',m'}$, we have

$$\mathbb{E}[F'_{h;\eta}(\bar{R}_{h;\eta}) Y_{\ell,m} \otimes Y_{\ell',m'}] = \mathbb{E}_{\mathcal{U}} \mathbb{E}[F'_{h;\eta}(\bar{R}_{h;\eta}) Y_{\ell,m} \otimes Y_{\ell',m'}|\mathcal{U}] = 0$$

and

$$\mathbb{E}|F'_{h;\eta}(\bar{R}_{h;\eta}) Y_{\ell,m} \otimes Y_{\ell',m'}|^2 = \mathrm{Var}[F'_{h;\eta}(\bar{R}_{h;\eta}) Y_{\ell,m} \otimes Y_{\ell',m'}] = \frac{4(4\pi)^4}{N_h^2}$$

$$\times \mathrm{Var}\left[\sum_{t=1}^{n-h} \sum_{\substack{j=1 \\ j \ne k \text{ if } h=0}}^{r} \sum_{k=1}^{r} (W_{t+h,j} W_{tk} - \bar{R}_{h;\eta}(U_{t+h,j}, U_{tk})) Y_{\ell,m}(U_{t+h,j}) Y_{\ell',m'}(U_{tk})\right].$$

Similarly to Equation (77), we use the law of total variance by conditioning on $\mathcal{U}$.

In what follows, we will handle sums over four indices j, k, j', k'. It is then useful to identify the distinct cases which lead to terms of different orders, that is,

1. terms of order r:

    (a) $j = j' = k = k'$

2. terms of order $r(r-1)$:

    (a) $j = j' \ne k = k'$

    (b) $j = k' \ne j' = k$

    (c) $j = k \ne j' = k'$

    (d) $j = k = j' \ne k'$

    (e) $j = k = k' \ne j'$

(f) $j' = k' = j \neq k$

(g) $j' = k' = k \neq j$

3. terms of order $r(r-1)(r-2)$:

(a) $j = j'$, $k \neq k'$, $j \neq k$, $j' \neq k'$

(b) $j \neq j'$, $k = k'$, $j \neq k$, $j' \neq k'$

(c) $j \neq j'$, $k \neq k'$, $j = k'$, $j' \neq k$, $j \neq k$, $j' \neq k'$

(d) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' = k$, $j \neq k$, $j' \neq k'$

(e) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' \neq k$, $j = k$, $j' \neq k'$

(f) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' \neq k$, $j \neq k$, $j' = k'$

4. terms of order $r(r-1)(r-2)(r-3)$:

(a) $j \neq j'$, $k \neq k'$, $j \neq k'$, $j' \neq k$, $j \neq k$, $j' \neq k'$

Recall that, when $h = 0$, we have $j \neq k$, $j' \neq k'$. Thus, cases 1.a, 2.c–2.g, 3.e and 3.f have not to be considered.

Variance of the conditional expectation. When $h = 0$, the sum over i is composed of independent terms, hence it follows exactly as in Theorem 4 that

$$
\mathrm{Var}\left[\sum_{t=1}^{n} \sum_{1 \leq j \neq k \leq r} (R_0(U_{tj}, U_{tk}) - \bar{R}_{0;\eta}(U_{tj}, U_{tk})) Y_{\ell,m}(U_{tj}) Y_{\ell',m'}(U_{tk}) \right]
$$

$$
= n \, \mathrm{Var}\left[\sum_{1 \leq j \neq k \leq r} (R_0(U_{ij}, U_{ik}) - \bar{R}_{0;\eta}(U_{ij}, U_{ik})) Y_{\ell,m}(U_{ij}) Y_{\ell',m'}(U_{ik}) \right]
$$

$$
\leq O(nr^3).
$$

When $h \neq 0$, for the last case 4.a, we have

$$
\langle R_h - \bar{R}_{h;\eta}, Y_{\ell,m} \otimes Y_{\ell',m'} \rangle^2_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)},
$$

regardless of t and t' (cardinality $(n-h)^2$). The same for the other cases when jointly $t \neq t', t \neq t' + h, t' \neq t + h$ (cardinality $(n-h)^2 - (n-h) - 2(n-2h) \leq (n-h)^2$). Moreover, we have

- terms of order $(n-h)r$, $(n-h)r(r-1)$, $(n-h)r(r-1)(r-2)$, when $t = t'$
- terms of order $(n-2h)r$, $(n-2h)r(r-1)$, $(n-2h)r(r-1)(r-2)$, when $t = t' + h$
- terms of order $(n-2h)r$, $(n-2h)r(r-1)$, $(n-2h)r(r-1)(r-2)$, when $t' = t + h$

Hence,

$$
\mathrm{Var}\left[\sum_{t=1}^{n-h} \sum_{j=1}^{r} \sum_{k=1}^{r} (R_h(U_{t+h,j}, U_{tk}) - \bar{R}_{h;\eta}(U_{t+h,j}, U_{tk})) Y_{\ell,m}(U_{t+h,j}) Y_{\ell',m'}(U_{tk}) \right]
$$

$$
=\mathbb{E}\left(\sum_{t=1}^{n-h}\sum_{j=1}^{r}\sum_{k=1}^{r}(R_h(U_{t+h,j},U_{tk})-\bar{R}_{h;\eta}(U_{t+h,j},U_{tk}))Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})\right)^2
$$

$$
-\frac{(n-h)^2r^4}{(4\pi)^4}\langle R_h-\bar{R}_{h;\eta},Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}
$$

$$
\leq B\frac{3(n-h)(r+7r(r-1)+6r(r-1)(r-2))}{(4\pi)^2}\|R_h-\bar{R}_{h;\eta}\|^2_{\mathbb{H}_p}
$$

$$
+\frac{(n-h)^2r(r-1)(r-2)(r-3)}{(4\pi)^4}\langle R_h-\bar{R}_{h;\eta},Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}
$$

$$
+\frac{(n-h)^2(r+7r(r-1)+6r(r-1)(r-2))}{(4\pi)^4}\langle R_h-\bar{R}_{h;\eta},Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}
$$

$$
-\frac{(n-h)^2r^4}{(4\pi)^4}\langle R_h-\bar{R}_{h;\eta},Y_{\ell,m}\otimes Y_{\ell',m'}\rangle^2_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}
$$

$$
\leq O(nr^3),
$$

where the last two inequalities are justified by Lemma 1, for some $B>0$, and Equation (81) with $\alpha=1$.

Expectation of the conditional variance. Write

$$
\sum_{\substack{t=1\\ }}^{n-h}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{r}\sum_{k=1}^{r}W_{t+h,j}W_{tk}Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})
$$

$$
=\sum_{\substack{t=1\\ }}^{n-h}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{r}\sum_{k=1}^{r}X_{t+h}(U_{t+h,j})X_t(U_{tk})Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})
$$

$$
+\sum_{\substack{t=1\\ }}^{n-h}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{r}\sum_{k=1}^{r}\epsilon_{t+h,j}X_t(U_{tk})Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})
$$

$$
+\sum_{\substack{t=1\\ }}^{n-h}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{r}\sum_{k=1}^{r}X_{t+h}(U_{t+h,j})\epsilon_{tk}Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})
$$

$$
+\sum_{\substack{t=1\\ }}^{n-h}\sum_{\substack{j=1\\ j\neq k \text{ if } h=0}}^{r}\sum_{k=1}^{r}\epsilon_{t+h,j}\epsilon_{tk}Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk}).
$$

Denote the four terms in this last expression by S_1, S_2, S_3, and S_4 with indices corresponding to their location in the sum. Then,

$$
\mathbb{E}_{\mathcal{U}}[\text{Var}\,[S_1+S_2+S_3+S_4|\mathcal{U}]]
$$

$$
\leq 4\left(\mathbb{E}_{\mathcal{U}}[\text{Var}[S_1|\mathcal{U}]]+\mathbb{E}_{\mathcal{U}}[\text{Var}[S_2|\mathcal{U}]]+\mathbb{E}_{\mathcal{U}}[\text{Var}[S_3|\mathcal{U}]]+\mathbb{E}_{\mathcal{U}}[\text{Var}[S_4|\mathcal{U}]]\right).
$$

We will illustrate how to derive the bound for S_1, since the other three are somewhat simpler to handle and of order at most nr^3. We start by looking at

$$\mathbb{E}_{\mathcal{U}}[Y_{\ell,m}(U_{i+h,j})Y_{\ell',m'}(U_{ik})Y_{\ell,m}(U_{i'+h,j'})Y_{\ell',m'}(U_{i'k'})$$
$$\times \text{Cov}[X_{i+h}(U_{i+h,j})X_i(U_{ik}), X_{i'+h}(U_{i'+h,j'})X_{i'}(U_{i'k'})|U]],$$

for different configurations of the indices t, t', j, j', k, k'

For 4.a, regardless of t and t', and for 1-2-3 when jointly $t \neq t'$, $t \neq t'+h$, $t' \neq t+h$, we have

$$\mathbb{E}_{\mathcal{U}}[Y_{\ell,m}(U_{t+h,j})Y_{\ell',m'}(U_{tk})Y_{\ell,m}(U_{t'+h,j'})Y_{\ell',m'}(U_{t'k'})$$
$$\times \text{Cov}\left[X_{t+h}(U_{t+h,j})X_t(U_{tk}), X_{t'+h}(U_{t'+h,j'})X_{t'}(U_{t'k'})|\mathcal{U}\right]]$$
$$= \text{Cov}[\langle X_{t+h}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}\langle X_t, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'+h}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}\langle X_{t'}, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}].$$

It is possible to show that for non-centered (real-valued) random variables X_1, X_2, X_3, X_4, with means $\mu_1, \mu_2, \mu_3, \mu_4$, the following holds

$$\begin{aligned}
\text{Cov}[X_1 X_2, X_3 X_4] = {} & \text{Cum}[X_1, X_2, X_3, X_4] \\
& + \mu_4 \, \text{Cum}[X_3, X_1, X_2] + \mu_3 \, \text{Cum}[X_4, X_1, X_2] \\
& + \mu_2 \, \text{Cum}[X_1, X_3, X_4] + \mu_1 \, \text{Cum}[X_2, X_3, X_4] \\
& + \text{Cov}[X_1, X_3] \, \text{Cov}[X_2, X_4] \\
& + \text{Cov}[X_1, X_4] \, \text{Cov}[X_2, X_3] \\
& + \mu_2 \mu_4 \, \text{Cov}[X_1, X_3] + \mu_1 \mu_3 \, \text{Cov}[X_2, X_4] \\
& + \mu_2 \mu_3 \, \text{Cov}[X_1, X_4] + \mu_1 \mu_4 \, \text{Cov}[X_2, X_3].
\end{aligned}$$

In the following, we show how to bound one of the terms where third-order cumulant appears. We can observe that, by stationarity,

$$\frac{1}{n-h} \sum_{t=1}^{n-h} \sum_{t'=1}^{n-h} |\langle \mu, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}$$
$$\times \text{Cum}[\langle X_{t+h}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'+h}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}, \langle X_{t'}, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}]|$$
$$= |\langle \mu, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}| \sum_{|\xi|<n-h} \left(1 - \frac{|\xi|}{n-h}\right) |\langle \mathcal{C}_{\xi+h,h}Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$
$$\leq |\langle \mu, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}| \sum_{\xi \in \mathbb{Z}} |\langle \mathcal{C}_{\xi+h,h}Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|.$$

Then, for $\alpha \leq p/q$ and using the property in Equation (50), we obtain

$$\sum_{\xi \in \mathbb{Z}} \sum_{\ell,m} \sum_{\ell',m'} \frac{D_\ell^{2\alpha} D_{\ell'}^{2\alpha}}{(1 + \eta D_\ell^2 D_{\ell'}^2)^2} |\langle \mu, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}| |\langle \mathcal{C}_{\xi+h,h}Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$
$$\leq \sum_{\xi \in \mathbb{Z}} \sum_{\ell,m} \sum_{\ell',m'} D_\ell^{2\alpha} D_{\ell'}^{2\alpha} |\langle \mu, Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2)}| |\langle \mathcal{C}_{\xi+h,h}Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{L^2(\mathbb{S}^2)}|$$

$$\leq \|\mu\|_{\mathcal{H}_q} \sum_{\xi \in \mathbb{Z}} \|\mathscr{C}_{\xi+h,h}\|_{\mathrm{TR},\mathbb{H}_q}$$

(possibly up to an arbitrary constant). The last inequality is justified by the fact that, if $\mathcal{C}_{h_1,h_2}$ is trace class (and hence compact), it has singular value decomposition

$$\mathcal{C}_{h_1,h_2} g = \sum_j \lambda_j \langle g, f_{1j}\rangle_{\mathbb{H}_q} f_{2j}, \qquad g \in \mathbb{H}_q,$$

see [21, Theorem 4.3.1]. Then,

$$\sum_{\ell,m} \sum_{\ell',m'} |\langle \mu, Y_{\ell'm'}\rangle_{\mathcal{H}_q}||\langle \mathcal{C}_{h_1,h_2} Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{\mathcal{H}_q}|$$

$$= \sum_j \lambda_j \sum_{\ell,m} \sum_{\ell',m'} |\langle \mu, Y_{\ell'm'}\rangle_{\mathcal{H}_q}||\langle Y_{\ell,m} \otimes Y_{\ell',m'}, f_{1j}\rangle_{\mathbb{H}_q}||\langle f_{2j}, Y_{\ell,m}\rangle_{\mathcal{H}_q}|$$

$$\leq \sum_j \lambda_j \left(\sum_{\ell,m} \sum_{\ell',m'} |\langle \mu, Y_{\ell'm'}\rangle_{\mathcal{H}_q}|^2 |\langle f_{2j}, Y_{\ell,m}\rangle_{\mathcal{H}_q}|^2 \right)^{1/2} \|f_{1j}\|_{\mathbb{H}_q}$$

$$= \|\mu\|_{\mathcal{H}_q} \sum_j \lambda_j \|f_{2j}\|_{\mathcal{H}_q} \|f_{1j}\|_{\mathbb{H}_q}.$$

Since $\{f_{1j}\}$ and $\{f_{2j}\}$ are orthonormal, we have

$$\sum_{\ell,m} \sum_{\ell',m'} |\langle \mu, Y_{\ell'm'}\rangle_{\mathcal{H}_q}||\langle \mathcal{C}_{h_1,h_2} Y_{\ell,m} \otimes Y_{\ell',m'}, Y_{\ell,m}\rangle_{\mathcal{H}_q}| \leq \|\mu\|_{\mathcal{H}_q} \|\mathcal{C}_{h_1,h_2}\|_{\mathrm{TR},\mathbb{H}_q}.$$

Similarly it holds for the other terms in the sum.

Now, following Proof of Theorem 4 in Section 8.3, we can bound the terms corresponding to cases 1-2-3 when $t = t'$, $t = t' + h$, $t' = t + h$. Indeed, for instance, when $t = t' + h$ for $h \neq 0$ and $j' = k$, we have

$$\int_{\mathbb{S}^2} \int_{\mathbb{S}^2} \int_{\mathbb{S}^2} Y_{\ell,m}(u) Y_{\ell',m'}(w) Y_{\ell,m}(w) Y_{\ell',m'}(v)$$
$$\times \mathrm{Cov}\left[X_{t+h}(u)X_t(w), X_t(w)X_{t-h}(v)\right] du dv dw, \tag{83}$$

and

$$|\mathrm{Cov}[X_{t+h}(u)X_t(w), X_t(w)X_{t-h}(v)]|$$
$$\leq (\mathrm{Var}\left[X_{t+h}(u)X_t(w)\right])^{1/2} (\mathrm{Var}\left[X_t(w)X_{t-h}(v)\right])^{1/2}$$
$$\leq \left(\mathbb{E}[|X_{t+h}(u)|^2|X_t(w)|^2]\right)^{1/2} \left(\mathbb{E}[|X_t(w)|^2|X_{t-h}(u)|^2]\right)^{1/2}$$
$$\leq \left(\mathbb{E}|X_{t+h}(u)|^4\right)^{1/4} \left(\mathbb{E}|X_t(w)|^4\right)^{1/2} \left(\mathbb{E}|X_{t-h}(v)|^4\right)^{1/4}.$$

Thus, the expression in Equation (83) can be bounded by $B'(4\pi)\mathbb{E}\|X_0\|_{\mathcal{H}_q}^4$, for all $t \in \mathbb{Z}$, again by applying Lemma 1, for some $B' > 0$.

In conclusion, we have

$$
\mathbb{E}\|\tilde{R}_{h;\eta} - \bar{R}_{h;\eta}\|_\alpha^2 \leq \frac{1}{n-h}\Bigg\{ \sum_{\xi_1,\xi_2,\xi_3 \in \mathbb{Z}} \|\mathscr{C}_{\xi_1,\xi_2,\xi_3}\|_{\mathrm{TR},\mathbb{H}_q}
$$
$$
+ 4\|\mu\|_{\mathcal{H}_q} \sum_{\xi_1,\xi_2 \in \mathbb{Z}} \|\mathscr{C}_{\xi_1,\xi_2}\|_{\mathrm{TR},\mathbb{H}_q}
$$
$$
+ 2\left(\sum_{\xi \in \mathbb{Z}} \|\mathscr{C}_\xi\|_{\mathrm{TR},\mathcal{H}_q}\right)^2 + 4\|\mu\|_{\mathcal{H}_q}^2 \sum_{\xi \in \mathbb{Z}} \|\mathscr{C}_\xi\|_{\mathrm{TR},\mathcal{H}_q}\Bigg\}
$$
$$
+ O\left(\frac{1}{nr}\right) O\left(\eta^{-(\alpha+1/p)}\log(1/\eta) + 1\right),
$$

which gives the result.

Now, let us prove 3. The first step is to obtain a useful analytic form for $R_{h;\eta} - \tilde{R}_{h;\eta}$, by observing that

$$
R_{h;\eta} - \tilde{R}_{h;\eta} = R_{h;\eta} - \bar{R}_{h;\eta} + (\bar{F}_{h;\eta}'')^{-1} F_{h;\eta}'(\bar{R}_{h;\eta})
$$
$$
= (\bar{F}_{h;\eta}'')^{-1} \left[\bar{F}_{h;\eta}''(R_{h;\eta} - \bar{R}_{h;\eta}) + F_{h;\eta}'(\bar{R}_{h;\eta})\right].
$$

Since $R_{h;\eta}$ minimizes $F_{h;\eta}'$, it holds $F_{h;\eta}'(R_{h;\eta}) = 0$ (see [21, Theorem 3.6.3]). Then, for any $g \in \mathbb{H}_p$,

$$
\left[\bar{F}_{h;\eta}''(R_{h;\eta} - \bar{R}_{h;\eta}) + F_{h;\eta}'(\bar{R}_{h;\eta})\right] g
$$
$$
= \left[\bar{F}_{h;\eta}''(R_{h;\eta} - \bar{R}_{h;\eta}) + F_{h;\eta}'(\bar{R}_{h;\eta}) - F_{h;\eta}'(R_{h;\eta})\right] g
$$
$$
= \left[\bar{F}_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta}) - F_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta})\right] g.
$$

where we used Lemma 4; in other words,

$$
R_{h;\eta} - \tilde{R}_{h;\eta} = (\bar{F}_{h;\eta}'')^{-1} \left[\bar{F}_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta}) - F_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta})\right].
$$

Now, the same argument that leads to (82) gives us

$$
\|R_{h;\eta} - \tilde{R}_{h;\eta}\|_\alpha^2 = \frac{1}{4} \sum_{\ell,m} \sum_{\ell',m'} \frac{D_\ell^{2\alpha} D_{\ell'}^{2\alpha}}{(1 + \eta D_\ell^2 D_{\ell'}^2)^2}
$$
$$
\times \left(\left[\bar{F}_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta}) - F_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta})\right] Y_{\ell,m} \otimes Y_{\ell',m'}\right)^2,
$$

with

$$
\left[\bar{F}_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta}) - F_{h;0}''(R_{h;\eta} - \bar{R}_{h;\eta})\right] Y_{\ell,m} \otimes Y_{\ell',m'}
$$
$$
= 2\langle R_{h;\eta} - \bar{R}_{h;\eta}, Y_{\ell,m} \otimes Y_{\ell',m'}\rangle_{L^2(\mathbb{S}^2 \times \mathbb{S}^2)} - \frac{2(4\pi)^2}{N_h}
$$
$$
\times \sum_{\substack{t=1 \\ j \neq k \text{ if } h=0}}^{n-h} \sum_{j=1}^{r} \sum_{k=1}^{r} (R_{h;\eta}(U_{t+h,j}, U_{tk}) - \bar{R}_{h;\eta}(U_{t+h,j}, U_{tk})) Y_{\ell,m}(U_{t+h,j}) Y_{\ell',m'}(U_{tk}).
$$

Now, $R_{h;\eta} - \bar{R}_{h;\eta} = \sum_{\ell,m} \sum_{\ell',m'} h_{\ell,\ell',m,m'} Y_{\ell,m} Y_{\ell',m'}$. Then,

$$\|R_{h;\eta} - \tilde{R}_{h;\eta}\|_\alpha^2 = \sum_{\ell_1,m_1} \sum_{\ell_2,m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2}$$

$$\times \left(\sum_{\ell_3,m_3} \sum_{\ell_4,m_4} h_{\ell_3,\ell_4,m_3,m_4} V_{\ell_{1:4},m_{1:4}} \right)^2,$$

where

$$V_{\ell_{1:4},m_{1:4}} = \delta_{\ell_1}^{\ell_3} \delta_{m_1}^{m_3} \delta_{\ell_2}^{\ell_4} \delta_{m_2}^{m_4} - \frac{(4\pi)^2}{N_h}$$

$$\times \sum_{t=1}^{n-h} \sum_{\substack{j=1 \\ j \neq k \text{ if } h=0}}^{r} \sum_{k=1}^{r} Y_{\ell_1,m_1}(U_{t+h,j}) Y_{\ell_2,m_2}(U_{tk}) Y_{\ell_3,m_3}(U_{t+h,j}) Y_{\ell_4,m_4}(U_{tk}).$$

By applying the Cauchy–Schwarz inequality for arbitrary $\theta \in (1/p, 1]$, we obtain

$$\|R_{h;\eta} - \tilde{R}_{h;\eta}\|_\alpha^2 \leq \|R_\eta - \bar{R}_\eta\|_\theta^2 \sum_{\ell_1,m_1} \sum_{\ell_2,m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2}$$

$$\times \sum_{\ell_3,m_3} \sum_{\ell_4,m_4} D_{\ell_3}^{-2\theta} D_{\ell_4}^{-2\theta} V_{\ell_{1:4},m_{1:4}}^2.$$

It is readily seen that $\mathbb{E}[V_{\ell_{1:4},m_{1:4}}] = 0$ and

$$\mathbb{E}[V_{\ell_{1:4},m_{1:4}}^2] = -\delta_{\ell_1}^{\ell_3} \delta_{m_1}^{m_3} \delta_{\ell_2}^{\ell_4} \delta_{m_2}^{m_4} + \frac{(4\pi)^4}{N_h^2}$$

$$\times \mathbb{E} \left(\sum_{t=1}^{n-h} \sum_{\substack{j=1 \\ j \neq k \text{ if } h=0}}^{r} \sum_{k=1}^{r} Y_{\ell_1,m_1}(U_{t+h,j}) Y_{\ell_2,m_2}(U_{tk}) Y_{\ell_3,m_3}(U_{t+h,j}) Y_{\ell_4,m_4}(U_{tk}) \right)^2.$$

Thus, it is possible to show that

$$\mathbb{E}[V_{\ell_{1:4},m_{1:4}}^2] \leq D_{\ell_1}^{2\theta} D_{\ell_2}^{2\theta} O\left(\frac{1}{nr} \right),$$

and hence

$$\sum_{\ell_1,m_1} \sum_{\ell_2,m_2} \frac{D_{\ell_1}^{2\alpha} D_{\ell_2}^{2\alpha}}{(1 + \eta D_{\ell_1}^2 D_{\ell_2}^2)^2} \sum_{\ell_3,m_3} \sum_{\ell_4,m_4} D_{\ell_3}^{-2\theta} D_{\ell_4}^{-2\theta} \mathbb{E}[V_{\ell_{1:4},m_{1:4}}^2]$$

$$= O\left(\frac{\log(1/\eta)}{nr\eta^{\alpha+\theta+1/p}} \right).$$

The rest of the proof follows exactly as in Proof of Theorem 4 in Section 8.3. $\square$

Acknowledgments

References

[1] ABBASI, H. (2018). Sparse: A more modern sparse array library. In *Proceedings of the 17th Python in Science Conference* 27–30.

[2] ARGO (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). *SEANOE*.

[3] ARONSZAJN, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society* **68** 337–404. MR0051437

[4] BADOUAL, A., FAGEOT, J. and UNSER, M. (2018). Periodic splines and Gaussian processes for the resolution of linear inverse problems. *IEEE Transactions on Signal Processing* **66** 6047–6061. MR3887349

[5] BERG, C. and PORCU, E. (2017). From Schoenberg coefficients to Schoenberg functions. *Constructive Approximation* **45** 217–241. MR3619442

[6] BERLINET, A. and THOMAS-AGNAN, C. (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media. MR2239907

[7] CAI, T. T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data Technical Report.

[8] CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics* **39** 2330–2355. MR2906870

[9] CAPONERA, A. (2021). SPHARMA approximations for stationary functional time series on the sphere. *Statistical Inference for Stochastic Processes* **24** 609–634. MR4321852

[10] CAPONERA, A., DURASTANTI, C. and VIDOTTO, A. (2021). LASSO estimation for spherical autoregressive processes. *Stochastic Processes and their Applications* **137** 167–199. MR4244190

[11] CAPONERA, A. and MARINUCCI, D. (2021). Asymptotics for spherical functional autoregressions. *The Annals of Statistics* **49** 346–369. MR4206681

[12] CASTRUCCIO, S. and STEIN, M. L. (2013). Global space-time models for climate ensembles. *The Annals of Applied Statistics* **7** 1593–1611. MR3127960

[13] DELLO SCHIAVO, L., KOPFER, E. and STURM, K.-T. (2020). A discovery tour in random Riemannian geometry. *arXiv: 2012.06796*.

[14] FAGEOT, J., UNSER, M. and WARD, J. P. (2017). On the Besov regularity of periodic Lévy noises. *Applied and Computational Harmonic Analysis* **42** 21–36. MR3574559

[15] GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97** 590–600. MR1941475

[16] Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M. and Bartelmann, M. (2005). HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *The Astrophysical Journal* **622** 759.

[17] Green, P. J. and Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* Chapman and Hall/CRC. MR1270012

[18] Gupta, H., Fageot, J. and Unser, M. (2018). Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization. *IEEE Transactions on Signal Processing* **66** 4670–4684. MR3861607

[19] Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34** 1493–1517. MR2278365

[20] Hoerl, A. E. (1962). Application of the ridge analysis to regression problems. *Chemical Engineering Progress* **58** 54–59.

[21] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* John Wiley & Sons. MR3379106

[22] Ito, K. and Jin, B. (2014). *Inverse Problems: Tikhonov Theory and Algorithms* **22**. World Scientific. MR3244283

[23] Jeong, J. and Jun, M. (2015). A class of Matérn-like covariance functions for smooth processes on a sphere. *Spatial Statistics* **11** 1–18. MR3311853

[24] Jun, M. (2014). Matérn-based nonstationary cross-covariance models for global processes. *Journal of Multivariate Analysis* **128** 134–146. MR3199833

[25] Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications* **33** 82–95. MR0290013

[26] Kuusela, M. and Stein, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proceedings of the Royal Society A* **474** 20180400.

[27] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38** 3321–3351. MR2766854

[28] Lin, Y. (1998). Tensor product space ANOVA models in multivariate function estimation, PhD thesis, University of Pennsylvania. MR2697355

[29] Lin, Y. (2000). Tensor product space ANOVA models. *The Annals of Statistics* **28** 734–755. MR1792785

[30] Marinucci, D. and Peccati, G. (2011). *Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications. London Mathematical Society Lecture Note Series.* Cambridge University Press. MR2840154

[31] Michel, V. (2012). *Lectures on Constructive Approximation: Fourier, Spline, and Wavelet methods on the Real Line, the Sphere, and the Ball.* Springer Science & Business Media. MR2987774

[32] ODEN, J. T. and REDDY, J. N. (1976). *An Introduction to the Mathematical Theory of Finite Elements*. Wiley, New York. MR0461950

[33] PANARETOS, V. M. and TAVAKOLI, S. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41** 568–603. MR3099114

[34] PANARETOS, V. M. and TAVAKOLI, S. (2013). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications* **123** 2779–2807. MR3054545

[35] PORCU, E., ALEGRÍA, A. and FURRER, R. (2018). Modeling temporally evolving and spatially globally dependent data. *International Statistical Review* **86** 344–377. MR3852415

[36] PORCU, E., BEVILACQUA, M. and GENTON, M. G. (2016). Spatiotemporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association* **111** 888–898. MR3538713

[37] PORCU, E., CASTRUCCIO, S., ALEGRIA, A. and CRIPPA, P. (2019). Axially symmetric models for global data: A journey between geostatistics and stochastic generators: Axially symmetric models. *Environmetrics* **30** e2555. MR3908104

[38] PORCU, E., FURRER, R. and NYCHKA, D. (2020). 30 years of space-time covariance functions. *WIREs Computational Statistics* e1512. MR4218945

[39] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer, New-York. MR2168993

[40] ROCKLIN, M. (2015). Dask: Parallel Computation with Blocked Algorithms and Task Scheduling. In *Proceedings of the 14th Python in Science Conference* **130** 136.

[41] RUBÍN, T. and PANARETOS, V. M. (2020). Sparsely observed functional time series: estimation and prediction. *Electronic Journal of Statistics* **14** 1137–1210. MR4069991

[42] SALEH, A. M. E., ARASHI, M. and KIBRIA, B. G. (2019). *Theory of Ridge Regression Estimation with Applications* **285**. John Wiley & Sons.

[43] SCHÖLKOPF, B., HERBRICH, R. and SMOLA, A. J. (2001). A generalized representer theorem. In *International Conference on Computational Learning Theory* 416–426. MR2042050

[44] SCHOLKOPF, B. and SMOLA, A. J. (2018). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Adaptive Computation and Machine Learning Series.

[45] SHAIKH, M. A. H. and HASAN, K. A. (2015). Efficient storage scheme for n-dimensional sparse array: GCRS/GCCS. In *2015 International Conference on High Performance Computing & Simulation (HPCS)* 137–142. MR3447298

[46] SIMEONI, M. (2020). Functional inverse problems on spheres: Theory, algorithms and applications Technical Report, EPFL.

[47] SIMEONI, M. (2021). Functional penalised basis pursuit on spheres. *Applied and Computational Harmonic Analysis* **53** 1–53. MR4214856

[48] SIMEONI, M. and HURLEY, P. (2019). Graph spectral clustering of convolution artefacts in radio interferometric images. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4260–4264.

[49] SIMEONI, M. and HURLEY, P. (2021). SIML: Sieved Maximum Likelihood for array signal processing. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4535–4539.

[50] SIMEONI, M., KASHANI, S., HURLEY, P. and VETTERLI, M. (2019). DeepWave: A recurrent neural-network for real-time acoustic imaging. *Advances In Neural Information Processing Systems 32 (Nips 2019)* **32**.

[51] STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8** 1348–1360. MR0594650

[52] TIKHONOV, A. (1963). On the solution of ill-posed problems and the regularization method. *Soviet Mathematics Doklady* **4** 1035–1038. MR0162378

[53] UNSER, M. (2021). A unifying representer theorem for inverse problems and machine learning. *Foundations of Computational Mathematics* **21** 941–960. MR4298236

[54] UNSER, M., FAGEOT, J. and GUPTA, H. (2016). Representer theorems for sparsity-promoting ℓ_1 regularization. *IEEE Transactions on Information Theory* **62** 5167–5180. MR3549472

[55] UNSER, M. and TAFTI, P. D. (2014). *An Introduction to Sparse Stochastic Processes.* Cambridge University Press. MR3495257

[56] VERAAR, M. C. (2011). Regularity of Gaussian white noise on the *d*-dimensional torus. *Marcinkiewicz Centenary Volume* **95** 385–398. MR2918350

[57] VIRTANEN, P., GOMMERS, R., OLIPHANT, T. E., HABERLAND, M., REDDY, T., COURNAPEAU, D., BUROVSKI, E., PETERSON, P., WECKESSER, W., BRIGHT, J. et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods* **17** 261–272.

[58] WAHBA, G. (1990). *Spline Models for Observational Data.* SIAM. MR1045442

[59] WASSERMAN, L. (2006). *All of Nonparametric Statistics.* Springer Science & Business Media. MR2172729

[60] WHITE, P. and PORCU, E. (2019). Towards a complete picture of stationary covariance functions on spheres cross time. *Electronic Journal of Statistics* **13** 2566–2594. MR3988087

[61] WILLIAMS, C. K. and RASMUSSEN, C. E. (2006). *Gaussian Processes for Machine Learning* **2**. MIT press Cambridge, MA. MR2514435

[62] YAO, F., MÜLLER, H. G. and WANG, J. L. (2005). Functional data analysis of sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590. MR2160561

[63] ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44** 2281–2321. MR3546451