LASSO estimation for spherical autoregressive processes

Alessia Caponera[a,b], Claudio Durastanti[c,*], Anna Vidotto[b,d]

[a] *Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Italy*
[b] *Dipartimento di Matematica, Università di Roma Tor Vergata, Italy*
[c] *Dipartimento SBAI, Sapienza Università di Roma, Italy*
[d] *Dipartimento di Economia, Università "G. D'Annunzio", Chieti–Pescara, Italy*

Abstract

The purpose of the present paper is to investigate a class of spherical functional autoregressive processes in order to introduce and study LASSO (Least Absolute Shrinkage and Selection Operator) type estimators for the corresponding autoregressive kernels, defined in the harmonic domain by means of their spectral decompositions. Some crucial properties for these estimators are proved, in particular, consistency and oracle inequalities.
ⓒ 2021 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Background and motivations

In recent years, growing attention has been paid to space–time processes built over various domains as, for example, regular grids, Euclidean spaces, and Riemannian manifolds (see, among others, [3,7,16,24,28,37]). In particular, space–time random fields defined over the two-dimensional unit sphere $\mathbb{S}^2$ find a wide set of applications in Cosmology, Geophysics, and Medical Imaging, providing a tool to perform analysis of data evolving with time and

---

* Corresponding author.
 E-mail addresses: alessia.caponera@uniroma1.it (A. Caponera), claudio.durastanti@uniroma1.it
(C. Durastanti), anna.vidotto@unich.it (A. Vidotto).

distributed over the sphere representing, for instance, either the Universe (see [35]), the planet Earth (see [17]) or the human brain (see [43]). Moreover, statistical techniques dealing with the analysis of – both purely spatial and depending on time – spherical random fields have received considerable attention, for example, in [1,22,31,34] and in [15,21], respectively.

In the last two decades, classes of space–time covariance functions has been defined and examined in the perspective to build the so-called sphere-cross-time random fields. The reader is referred to [16,24,29,37,38,41] and the references therein for some neat examples. Another construction involving sphere-cross-time random fields has been presented in [21] (see also [9]), where quantitative central limit theorems for linear and non-linear statistics based on spherical time-dependent Poisson random fields have been established.

A sphere-cross-time random field is denoted by a collection of random variables

$$\{T(x,t) : (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\},$$

and it can be described in the so-called frequency domain by its harmonic expansion

$$T(x,t) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} a_{\ell,m}(t) Y_{\ell,m}(x), \quad (x,t) \in \mathbb{S}^2 \times \mathbb{Z}, \tag{1.1}$$

where $\{Y_{\ell,m} : \ell \geq 0; m = -\ell, \ldots, \ell\}$ is the standard orthonormal basis of real spherical harmonics for $L^2(\mathbb{S}^2) = L^2(\mathbb{S}^2, \mathrm{d}x)$, the space of square-integrable functions on the sphere with respect to the spherical Lebesgue measure $\mathrm{d}x$ (see Section 2 and the Refs. [42,48]). Harmonic analysis has already been proved to be a valid tool to perform statistical analysis on the sphere (see, for instance, [20,30]) and allows one to describe a random field as the linear combination of spherical harmonics, weighted by the corresponding time-varying harmonic coefficients (see [7]). The stochastic information of the random field T is then contained in the set of harmonic coefficients $\{a_{\ell,m}(t) : \ell \geq 0; m = -\ell, \ldots, \ell\}$, given by

$$a_{\ell,m}(t) = \langle T(\cdot, t), Y_{\ell,m} \rangle_{\mathbb{S}^2},$$

where $\langle \cdot, \cdot \rangle_{\mathbb{S}^2}$ is the standard inner product over the sphere.

In this paper, the object of study is a class of sphere-cross-time random fields, introduced in the literature by [15], which are functional autoregressive processes defined over $L^2(\mathbb{S}^2)$ (see also [8] and the references therein). As the name suggests, the spherical autoregressive model of order p, from now on shortened to SPHAR(p), specifies the output field $T(\cdot, t)$ as an infinite-dimensional linear transformation of its p previous realizations (p lags) added to an independent *spherical white noise* $Z(\cdot, t)$. More formally, the SPHAR(p) equation is given by

$$T(x,t) = \sum_{j=1}^{p} \left(\Phi_j T(\cdot, t-j) \right)(x) + Z(x,t), \quad (x,t) \in \mathbb{S}^2 \times \mathbb{Z}, \tag{1.2}$$

where, for $j = 1, \ldots, p$, the autoregressive kernel operator $\Phi_j : L^2(\mathbb{S}^2) \to L^2(\mathbb{S}^2)$ is defined as

$$\left(\Phi_j f \right)(x) = \int_{\mathbb{S}^2} k_j(\langle x, y \rangle) f(y) \, \mathrm{d}y, \quad f \in L^2(\mathbb{S}^2), \tag{1.3}$$

with $k_j : [-1, 1] \to \mathbb{R}$ the corresponding autoregressive kernel, assumed to be continuous. Then, for $j = 1, \ldots, p$, the estimation of Φ_j can be reduced to the one of k_j. Note that, for any $j = 1, \ldots, p$, the kernel k_j is isotropic, that is, it depends only on $z = \langle x, y \rangle$, the

standard inner product on $\mathbb{R}^3$. As a consequence, the following spectral representation holds in the L^2-sense

$$k_j(z) = \sum_{\ell \geq 0} \phi_{\ell;j} \frac{2\ell + 1}{4\pi} P_\ell(z), \tag{1.4}$$

where $P_\ell : [-1, 1] \to \mathbb{R}$ is the Legendre polynomial of order ℓ, and, for $j = 1, \ldots, p$, the coefficients $\{\phi_{\ell;j} : \ell \geq 0\}$ are the eigenvalues of the operator Φ_j. The reader is referred to Section 2 for a detailed discussion. Moreover, the random field Z in (1.2) is defined such that, for any $x, y \in \mathbb{S}^2$, the covariance function $\Gamma_Z(x, y) = \mathbb{E}[Z(x, t)Z(y, t)]$ is isotropic and independent on the choice of $t \in \mathbb{Z}$ (see [48]).

As a standard consequence of the so-called *duplication property* for spherical harmonics, we can characterize functional spherical autoregressions in the frequency domain as

$$a_{\ell,m}(t) = \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t - j) + a_{\ell,m;Z}(t). \tag{1.5}$$

In [15], estimators for the kernels $\{k_j : j = 1, \ldots, p\}$ have been defined according to a functional L^2-minimization criterion, exploiting their spectral decomposition (1.4); some asymptotic properties, such as consistency, quantitative central limit theorem, and weak convergence, have then been established under additional regularity conditions.

In this work, we approach the estimation of the autoregressive kernels under a *sparsity* assumption, that is, for any $j = 1, \ldots, p$, we assume that only a few of the components of k_j in (1.4) are non-zero (see Definition 1.1). Assuming sparsity conditions can lead to considerable advantages in estimation problems (see, for example, [26]). In this case, the proper identification of the null components allows to perform regularization on the functional estimates, preserving accuracy; moreover, sparsity enhances computational efficiency. The latter is a crucial aspect in spatial and space–time statistics, often related to the estimation of covariance functions; see for instance [5,6] for recent results.

LASSO – or ℓ_1-regularized – regression, introduced in the statistical literature by the celebrated paper [44], is among the most popular penalization techniques to estimate sparse models. In particular, it corrects the L^2-loss for sparse models by adding a convex penalty term and, then, constraining the estimation process and selecting the most significant variables. In the framework of independent and identically distributed (i.i.d.) observations, LASSO has been proved to be extremely efficient both from the point of view of theoretical properties and in terms of applications (see [25,47] and references therein). The connections between LASSO, ridge regression, best subset selection and other ℓ_q-based penalization methods, as well as further links between LASSO and other nonparametric statistical techniques, such as soft and hard thresholding, have been widely investigated, for instance, in [11,26].

Applications of LASSO in the framework of time series and stochastic processes represent a much more recent development. A pioneering contribution in this area has been given in [2], where the authors explore the properties of ℓ_1-regularized estimators in the settings of stochastic regression with serially correlated errors and vector autoregressive (VAR) models (see also [18,40] for related ideas). Their results can be seen as a successful extension of the standard LASSO technique to the framework of non-i.i.d. observations. More specifically, in [2], under sparsity constraints, ℓ_1-regularized estimators have been investigated by introducing a measure of stability for stationary processes, a very powerful tool to study the correlation structure of

multivariate processes, and crucial to settle some useful deviation bounds for dependent data. Further details on the stability of autoregressions can be found, among others, in [10,33] as well as in [8] for the functional case (see also Section 3). In turn, these deviation bounds are instrumental to establish concentration properties of the estimators, and so-called *oracle inequalities*.

The aim of this work is to define and study LASSO-type estimators for spherical autoregressive kernels under sparsity assumptions. Our approach does not require any specific functional form for the kernel k_j; in this sense, the estimation procedure can be viewed as fully nonparametric, see also [39,45]. It is important to stress that, given the nonparametric nature of the model (1.2), we are dealing with a *functional* penalized regression problem, hence differing from the framework of VAR(p) processes, where estimators assume a vectorial form (see also Remark 3.2). More specifically, our oracle inequalities will involve functions rather than scalar or vectorial parameters (see Section 1.2). Exploiting the harmonic expansion for the spherical autoregressions (1.2) and the isotropy assumption on $\{k_j : j = 1, \ldots, p\}$ in (1.4), together with an extension of the concept of stability measure introduced in [2], we will be able to establish concentration properties in functional norms for the autoregressive kernels (see Section 4). Moreover, the sparsity enforcement properties of LASSO procedures will avoid overfitting and will select only the most relevant components of each kernel function spectral decomposition (1.4).

1.2. Background and main results

First of all, let us set some standard notation, necessary to state our main findings. For two positive sequences $\{a_n\}_{n \in \mathbb{N}}, \{b_n\}_{n \in \mathbb{N}}$, we write $a_n \succeq b_n$ if there exists an absolute constant $c > 0$, which does not depend on the model parameters, such that $a_n \geq c\, b_n$, for all $n \in \mathbb{N}$; while we use $a_n \simeq b_n$ if there exist $M_1, M_2 > 0$, which may depend on the model, such that $M_1 b_n \leq a_n \leq M_2 b_n$, for all $n \in \mathbb{N}$. For a vector $v \in \mathbb{R}^d$, $\|v\|_q$ indicates the ℓ_q-norm of v,

$$\|v\|_q = \left(\sum_{i=1}^d |v_i|^q \right)^{\frac{1}{q}}; \quad \|v\|_0 = \sum_{i=1}^d \mathbb{1}\{v_i \neq 0\}; \quad \|v\|_\infty = \max_{i=1,\ldots,d} |v_i|,$$

for $0 < q < \infty$, $q = 0$ and $q = \infty$ respectively. We say that v is a r-sparse vector, $1 \leq r \leq d$, if $\|v\|_0 = r$. Unless stated otherwise, for the sake of simplicity, $\|\cdot\|$ denotes the ℓ_2-norm of v. Let $f : [-1, 1] \to \mathbb{R}^p$, $f \in L^q([-1, 1], \rho(\mathrm{d}z))$, where $\rho(\mathrm{d}z)$ is the Lebesgue measure over $[-1, 1]$. Then, for $1 \leq q < \infty$, the L^q-norm of f is given by

$$\|f\|_{L^q} = \left(\int_{-1}^1 \|f(z)\|^q \rho(\mathrm{d}z) \right)^{\frac{1}{q}}.$$

Analogously, the L^∞-norm of f is given by $\|f\|_{L^\infty} = \sup_{z \in [-1,1]} \|f(z)\|$. The spherical L^q-norms are defined by

$$\|f\|_{L^q(\mathbb{S}^2)} = \left(\int_{\mathbb{S}^2} |f(x)|^q \, \mathrm{d}x \right)^{\frac{1}{q}}, \quad f \in L^q(\mathbb{S}^2),$$

where $\mathrm{d}x$ is the Lebesgue measure over $\mathbb{S}^2$. Finally, the Hilbert–Schmidt and the trace class norms of a compact self-adjoint operator $\mathscr{T} : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ is a separable Hilbert space,

are given respectively by

$$\|\mathcal{T}\|_{\mathrm{HS}} = \sum_{i \geq 0} |\lambda_i|^2 \, ; \quad \|\mathcal{T}\|_{\mathrm{TR}} = \sum_{i \geq 0} |\lambda_i| \, ,$$

where $\{\lambda_i\}_{i \in \mathbb{N}}$ are the eigenvalues of $\mathcal{T}$ (see, for example, [27]).

We provide now the definition of sparsity set, which can be understood as an instrumental characterization of sparsity.

Definition 1.1 (*Sparsity Set*). For any $\ell \geq 0$ and $\boldsymbol{\phi}_\ell = \big(\phi_{\ell;1}, \dots, \phi_{\ell;p} \big)$, we define $q_\ell = \big\| \boldsymbol{\phi}_\ell \big\|_0$ the ℓth sparsity index, which satisfies $0 \leq q_\ell \leq p$. We call $\{q_\ell : \ell \geq 0\}$ the sparsity set.

Remark 1.2.   Following [8], to ensure identifiability we assume that there exists at least one $\ell \geq 0$ such that $\phi_{\ell;p} \neq 0$, so that $\mathbb{P} \big(\Phi_p T \, (\cdot, t) \neq 0 \big) > 0$, for all $t \in \mathbb{Z}$. As a consequence, for some $\ell \geq 0$, we can have $\boldsymbol{\phi}_\ell = 0$ and hence $\big\| \boldsymbol{\phi}_\ell \big\|_0 = q_\ell = 0$; however, $q = \max_{\ell \geq 0} q_\ell \geq 1$. Note that, in the simplest case $p = 1$, the sparsity set gives information about the null components of the single kernel.

Set the sequence of polynomials $\phi_\ell : \mathbb{C} \to \mathbb{C}$, $\ell \geq 0$, associated with the subprocesses defined in (1.5), as follows

$$\phi_\ell(z) = 1 - \phi_{\ell;1} z - \cdots - \phi_{\ell;p} z^p. \tag{1.6}$$

We must introduce some conditions on the model, essential to achieve our results. Condition 1.3 will require that the eigenvalues of the covariance operator associated with $\Gamma_Z(\cdot, \cdot)$ are strictly positive the operator is injective (see [27, Section 7.3]). Condition 1.5 will guarantee the existence of a unique isotropic and stationary solution for (1.2). For rigorous proofs, the reader is referred to [8,13,14].

Condition 1.3 (*Identifiability*). Let $Z(x, t)$ be the spherical white noise used in (1.2) and let its covariance function be given by

$$\Gamma_Z(x, y) = \mathrm{Cov}\,(Z(x, t), Z(y, t)), \qquad (x, t), (y, t) \in \mathbb{S}^2 \times \mathbb{Z}.$$

It holds that

$$\int_{\mathbb{S}^2 \times \mathbb{S}^2} \Gamma_Z(x, y)\, f(x)\, f(y)\, \mathrm{d}x \mathrm{d}y > 0,$$

for any $f \in L^2 \big(\mathbb{S}^2 \big)$ such that $f(\cdot) \neq 0$.

Remark 1.4.   Note that, as a consequence of [4, Lemma 4.1], Condition 1.3 is equivalent to state that the covariance function Γ_Z is strictly positive definite.

Condition 1.5 (*Causality/Stationarity*).   The sequence of polynomials (1.6) is such that

$$|z| \leq 1 \implies \phi_\ell(z) \neq 0, \qquad \text{for all } \ell \geq 0.$$

More explicitly, there are no roots in the unit disk, uniformly over ℓ.

Remark 1.6.   Condition 1.5, together with the spectral decomposition (1.4) for the kernels $\{k_j, j = 1, \dots, p\}$, implies that, if $\xi_{\ell;1}, \dots, \xi_{\ell;d_\ell}$ are the roots of the d_ℓ-degree polynomial (1.6), $1 \leq d_\ell \leq p$, then

$$|\xi_{\ell;j}| \geq \xi_* > 1,$$

uniformly over ℓ. In other words, there exists $\delta > 0$ such that

$$|z| < 1 + \delta \implies \phi_\ell(z) \neq 0, \qquad \text{for all } \ell \geq 0.$$

Condition 1.7 (*Smoothness*). For all $j = 1, \ldots, p$, Φ_j is a nuclear operator, that is, its trace norm is finite:

$$\|\Phi_j\|_{\mathrm{TR}} = \sum_{\ell \geq 0} (2\ell + 1)|\phi_{\ell;j}| < \infty, \tag{1.7}$$

see again [27].

As experimental setting, for any $\ell \geq 0$, we assume that the harmonic coefficients $\{a_{\ell,m}(t) : m = -\ell, \ldots, \ell\}$ can be observed over a finite set of times $\{1, \ldots, n\} \subset \mathbb{Z}$. The vector of functions

$$\mathbf{k} = (k_1, \ldots, k_p)$$

contains all the autoregressive kernels described above. We will focus on the following penalized minimization problem:

$$\widehat{\mathbf{k}}_N^{\mathrm{lasso}} = \underset{\mathbf{k} \in \mathcal{P}_N^p}{\mathrm{argmin}} \frac{1}{N} \sum_{t=p+1}^{n} \left\| T(\cdot, t) - \sum_{j=1}^{p} \Phi_j T(\cdot, t - j) \right\|_{L^2(\mathbb{S}^2)}^2 + \lambda \sum_{j=1}^{p} \|\Phi_j\|_{\mathrm{TR}}, \tag{1.8}$$

where $N = n - p$ can be read as the effective number of observations, and $\lambda \in \mathbb{R}^+$ is the penalty parameter. The space $\mathcal{P}_N^p$ is the Cartesian product of p copies of

$$\mathrm{span}\{\frac{2\ell + 1}{4\pi} P_\ell(\cdot) : \ell = 0, \ldots, L_N - 1\}, \tag{1.9}$$

where the integer $L_N > 0$ is the truncation level, which corresponds to the frequency of the highest component in (1.4) estimated by (1.8), see Section 3.1 for a detailed discussion. Let us also define

$$k_{j,N}(z) = \sum_{\ell=0}^{L_N - 1} \phi_{\ell;j} \frac{2\ell + 1}{4\pi} P_\ell(z)$$

and, accordingly, $\mathbf{k}_N = (k_{1,N}, \ldots, k_{p,N})$.

Our main result is extensively stated in Theorem 4.12 and can be compactly formulated as follows.

Theorem 1.8. *Consider the estimation problem* (1.8)*, assume that Conditions* 1.3 *and* 1.5 *hold, and suppose that*

$$N \succeq b_1 \, q \log (p L_N) \,,$$

where $q = \max_{\ell \geq 0} q_\ell$. Then, for any penalty parameter $\lambda = \lambda_N \geq b_2 \sqrt{\frac{\log(p L_N)}{N}}$, any solution $\widehat{\mathbf{k}}_N^{\mathrm{lasso}}$ satisfies

$$\mathbb{P}\left(\left\| \widehat{\mathbf{k}}_N^{\mathrm{lasso}} - \mathbf{k} \right\|_{L^2}^2 \leq \frac{18}{\pi^2} \lambda_N^2 \sum_{\ell=0}^{L_N - 1} \frac{q_\ell}{\alpha_\ell^2} (2\ell + 1) + \sum_{\ell \geq L_N} \|\phi_\ell\|_2^2 \frac{2\ell + 1}{8\pi^2} \right)$$

$$\geq 1 - c_1 e^{-c_2 \log(p L_N)}, \tag{1.10}$$

where the sequence $\{\alpha_\ell : \ell = 0, \ldots, L_N - 1\}$ is defined in (4.11), and $c_1, c_2 > 0$ are absolute constants. Moreover, under the additional Condition *1.7, it holds that*

$$\mathbb{P}\left(\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^\infty} \leq \frac{3}{\pi} \lambda_N \sum_{\ell=0}^{L_N-1} \frac{\sqrt{q_\ell}}{\alpha_\ell} (2\ell + 1) + \sum_{\ell \geq L_N} \|\boldsymbol{\phi}_\ell\|_2 \frac{2\ell + 1}{4\pi} \right)$$
$$\geq 1 - c_1 e^{-c_2 \log(pL_N)}.$$
(1.11)

Remark 1.9. The constants b_1 and b_2 depend on the model. In particular, $b_1 = \max\{\omega^2, 1\}$ and $b_2 = 4\mathcal{F}$, where $\mathcal{F}, \omega > 0$ are tightly connected to the stability measure introduced in Section 4.1. The reader is referred to Section 4.2 for further details and comments.

Our findings provide upper bounds for the L^2- and the L^∞-distances between the LASSO-type estimator $\widehat{\mathbf{k}}_N^{\text{lasso}}$ and the "true" $\mathbf{k}$. These upper bounds consist of the sum of two terms. The first summand represents the error due to the approximation of the first L_N components of $\mathbf{k}$ with $\widehat{\mathbf{k}}_N^{\text{lasso}}$. The second one arises because $\widehat{\mathbf{k}}_N^{\text{lasso}}$ provides an estimation of $\mathbf{k}$ truncated at the multipole L_N. In this sense, we can draw an analogy with standard nonparametric statistics and refer to them as the stochastic and the bias error, respectively (see [39]).

The upper bounds are non-asymptotic and they hold with high-probability, in the sense that, for a fixed N sufficiently large, the probability on the left side of (1.10) and (1.11) is arbitrarily close to 1. An appropriate choice of L_N leads both the upper bounds to converge to 0 and their probabilities to converge to 1, as $N \to \infty$; as a consequence, our result can be also read in terms of asymptotic consistency.

1.3. Plan of the paper

This paper is organized as follows. In Section 3, we present the LASSO estimators for spherical autoregressive kernels under sparsity assumptions. Section 4 contains the main results of this work, that is, how the classical LASSO-scheme fits in our setting, using the concept of stability measure, as well as our oracle inequalities. In Section 5 we briefly show the performance of the LASSO estimators under sparsity assumptions. Finally, Section 6 collects the proofs.

2. General setting

2.1. Harmonic analysis on the sphere and space–time spherical random fields

This section includes some well-established results concerning harmonic analysis on the sphere and provides the reader with an overview of the construction of space–time spherical random fields.

The reader is referred to [31,35,42,46] and the references therein for further details concerning harmonic analysis on the sphere and spherical random fields. Sphere-cross-time random fields have been discussed, among others, in [24,29,37,41].

Let us start by setting some useful notation and some important concepts related to the harmonic analysis on the sphere. Each point x on the sphere is identified by two angular coordinates, that is, $x = (\vartheta, \varphi)$, where $\vartheta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ are the colatitude and the longitude, respectively. The spherical Lebesgue measure is labeled by $\mathrm{d}x = \sin \vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\varphi$,

while $L^2\left(\mathbb{S}^2\right) = L^2\left(\mathbb{S}^2, \mathrm{d}x\right)$ describes the space of square-integrable functions over the sphere with respect to the measure $\mathrm{d}x$.

Following for example [35,42,46], we denote by $\{\mathcal{H}_\ell : \ell \geq 0\}$ the set of spaces of homogeneous harmonic polynomials of degree ℓ restricted to $\mathbb{S}^2$. For any $\ell \geq 0$, $\mathcal{H}_\ell$ is spanned by the set of spherical harmonics $\{Y_{\ell,m} : m = -\ell, \ldots, \ell\}$. The index $\ell \in \mathbb{N}$ is the so-called multipole, while $m = -\ell, \ldots, \ell$ is the "azimuth" number. The set $\{\mathcal{H}_\ell : \ell \geq 0\}$ is dense in $L^2\left(\mathbb{S}^2\right)$ (see again [35, Proposition 3.33, p.73]), thus the following decomposition holds

$$L^2\left(\mathbb{S}^2\right) = \bigoplus_{\ell \geq 0} \mathcal{H}_\ell$$

and spherical harmonics provide an orthonormal basis for $L^2\left(\mathbb{S}^2\right)$. For the sake of simplicity, here we will make use of the so-called real spherical harmonics. For any $\ell \in \mathbb{N}$ and $m = -\ell, \ldots, \ell$, the spherical harmonic $Y_{\ell,m}$ can be written as the normalized product of the Legendre associated function $P_{\ell,m} : [-1, 1] \to \mathbb{R}$ of degree ℓ and order m, which depends only on the colatitude ϑ, and a trigonometric function depending only on the longitude φ, namely,

$$Y_{\ell,m}(\vartheta, \varphi) = \begin{cases} \sqrt{\frac{(2\ell+1)}{2\pi} \frac{(\ell-m)!}{(\ell+m)!}} \, P_{\ell,m}(\cos \vartheta) \cos(m\varphi) & \text{for } m \in \{1, \ldots, \ell\} \\ \sqrt{\frac{(2\ell+1)}{4\pi}} \, P_\ell(\cos \vartheta) & \text{for } m = 0 \\ \sqrt{\frac{(2\ell+1)}{2\pi} \frac{(\ell+m)!}{(\ell-m)!}} \, P_{\ell,-m}(\cos \vartheta) \sin(-m\varphi) & \text{for } m \in \{-\ell, \ldots, -1\} \end{cases},$$

where

$$P_{\ell,m}(u) = \frac{1}{2^\ell \ell!} \left(1 - u^2\right)^{\frac{m}{2}} \frac{\mathrm{d}^{\ell+m}}{\mathrm{d}u^{\ell+m}} \left(u^2 - 1\right)^\ell, \quad u \in [-1, 1].$$

It is a well-known fact that the following addition formula holds

$$\sum_{m=-\ell}^{\ell} Y_{\ell,m}(x) Y_{\ell,m}(y) = \frac{2\ell+1}{4\pi} P_\ell(\langle x, y \rangle), \quad x, y \in \mathbb{S}^2,$$

where P_ℓ is the Legendre polynomial of order ℓ, given by

$$P_\ell(u) = \frac{1}{2^\ell \ell!} \frac{\mathrm{d}^\ell}{\mathrm{d}u^\ell} \left(u^2 - 1\right)^\ell, \quad u \in [-1, 1].$$

Legendre polynomials are orthogonal on $L^2([-1, 1])$, that is,

$$\int_{-1}^{1} P_\ell(u) P_{\ell'}(u) \, \mathrm{d}u = \frac{2}{2\ell+1} \delta_\ell^{\ell'}. \tag{2.1}$$

Recall that we are considering a sphere-cross-time random fields, which is a real-valued collection of random variables

$$\{T(x, t) : (x, t) \in \mathbb{S}^2 \times \mathbb{Z}\}.$$

Remark 2.1. All the results presented in this paper can be easily extended to the case of complex-valued spherical random fields, after a proper definition of complex spherical harmonics (see, for example, [35, Theorem 5.13, p.123]).

From now on, we will consider real-valued, centered, mean-square continuous, Gaussian random fields. Moreover, T is assumed to be isotropic in the spatial domain and stationary in the time domain. A spherical random field is said to be isotropic when it is invariant in

distribution with respect to rotations, while stationarity guarantees that the stochastic properties of the process do not change over time. In other words

$$T\left(R\cdot,\cdot+\tau\right)\overset{d}{=}T\left(\cdot,\cdot\right),\tag{2.2}$$

where $\tau \in \mathbb{Z}$, R belongs to the special group of rotations $SO\left(3\right)$, and $\overset{d}{=}$ denotes equality in distribution.

As mentioned in Section 1, under isotropy, for any fixed $t \in \mathbb{Z}$, the random field T can be rewritten in terms of its harmonic expansion, given by Eq. (1.1) and here recalled

$$T\left(x,t\right)=\sum_{\ell\geq 0}\sum_{m=-\ell}^{\ell}a_{\ell,m}\left(t\right)Y_{\ell,m}\left(x\right),\quad \left(x,t\right)\in\mathbb{S}^2\times\mathbb{Z}.$$

The set of the harmonic coefficients $\{a_{\ell,m} : \ell \geq 0, m = -\ell, \ldots, \ell\}$ contains all the stochastic information related to T and can be computed explicitly by

$$a_{\ell,m}\left(t\right)=\int_{\mathbb{S}^2}T\left(x,t\right)Y_{\ell,m}\left(x\right)\mathrm{d}x.$$

Since T is centered, that is, $\mathbb{E}\left[T\left(x,t\right)\right]=0$ for all $\left(x,t\right)\in\mathbb{S}^2\times\mathbb{Z}$, it follows that

$$\mathbb{E}\left[a_{\ell,m}\left(t\right)\right]=0\quad\text{for }\ell\in\mathbb{N},\quad m=-\ell,\ldots,\ell,\quad t\in\mathbb{Z}.$$

The covariance function of T will be denoted by $\Gamma : \left(\mathbb{S}^2\times\mathbb{Z}\right)\times\left(\mathbb{S}^2\times\mathbb{Z}\right)\to\mathbb{R}$. If the space–time spherical random field is isotropic and stationary, then there exists a function $\Gamma_0 : [-1,1]\times\mathbb{Z}\to\mathbb{R}$, so that (2.2) yields

$$\Gamma\left(x,t,y,s\right)=\Gamma_0\left(\langle x,y\rangle,t-s\right),\quad \left(x,t\right),\left(y,s\right)\in\mathbb{S}^2\times\mathbb{Z}.$$

Furthermore, for any $\ell,\ell'\in\mathbb{N}$, $m=-\ell,\ldots,\ell$, $m'=-\ell',\ldots,\ell'$, the elements of the covariance matrix constructed over the harmonic coefficients of T are given by

$$\mathbb{E}\left[a_{\ell,m}\left(t\right)a_{\ell',m'}\left(s\right)\right]=C_\ell\left(t-s\right)\delta_\ell^{\ell'}\delta_m^{m'},\quad t,s\in\mathbb{Z}.\tag{2.3}$$

Observe that, for $t = s$, $C_\ell\left(0\right)$ in (2.3) corresponds to the so-called angular power spectrum, the spectral decomposition of the covariance function of a purely spatial spherical random field (see, for example, [35, Remark 5.15, p.124; Remark 6.16, p.147]). Hence, from now on, we will use the notation $C_\ell\left(0\right)=C_\ell$.

Covariance functions of isotropic-stationary sphere-cross-time random fields have a spectral decomposition in terms of Legendre polynomials, namely,

$$\Gamma\left(x,t,y,s\right)=\sum_{\ell\geq 0}C_\ell\left(t-s\right)\frac{2\ell+1}{4\pi}P_\ell\left(\langle x,y\rangle\right),\quad \left(x,t\right),\left(y,s\right)\in\mathbb{S}^2\times\mathbb{Z},$$

see [4].

2.2. Spherical autoregressions

This section lists those conditions which, from now on, will characterize the class of sphere-cross-time random fields studied along this paper. Hence, they will be implicitly assumed in all our statements.

Following [15], we consider isotropic-stationary random fields $\{T(x, t)\,(x, t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ that satisfy the functional autoregressive equation (see also [8]) given by Eq. (1.2) and here recalled

$$T(x, t) = \sum_{j=1}^{p} \left(\Phi_j T(\cdot, t-j) \right)(x) + Z(x, t),$$

where:

- for $j = 0, \ldots, p$, $\Phi_j : L^2\left(\mathbb{S}^2\right) \to L^2\left(\mathbb{S}^2\right)$ is a linear and bounded operator defined by Eq. (1.3), i.e.,

$$\left(\Phi_j f \right)(x) = \int_{\mathbb{S}^2} K_j(x, y) f(y) \, \mathrm{d}y, \quad f \in L^2\left(\mathbb{S}^2\right).$$

- The kernel $K_j : \mathbb{S}^2 \times \mathbb{S}^2 \to \mathbb{R}$ is continuous and isotropic, that is, there exists a function $k_j : [-1, 1] \to \mathbb{R}$ so that $K_j(x, y) = k_j(\langle x, y \rangle)$, for all $x, y \in \mathbb{S}^2$. Moreover, the following decomposition holds (in the L^2-sense and pointwise under Condition 1.7)

$$K_j(x, y) = k_j(\langle x, y \rangle) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \phi_{\ell;j} Y_{\ell,m}(x) Y_{\ell,m}(y)$$
$$= \sum_{\ell \geq 0} \phi_{\ell;j} \frac{2\ell+1}{4\pi} P_\ell(\langle x, y \rangle), \tag{2.4}$$

  where the coefficients $\{\phi_{\ell;j} : \ell \geq 0\}$ are the eigenvalues of the operator Φ_j.

- The field Z is a Gaussian spherical white noise, that is, $\{Z(\cdot, t), t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed Gaussian isotropic spherical random fields, defined so that

  (i) for every fixed $t \in \mathbb{Z}$, $Z(\cdot, t)$ is a Gaussian, zero-mean isotropic random field, with covariance function given by

$$\Gamma_Z(x, y) = \sum_{\ell \geq 0} \frac{2\ell+1}{4\pi} C_{\ell;Z} P_\ell(\langle x, y \rangle), \quad \text{such that } \sum_{\ell \geq 0} \frac{2\ell+1}{4\pi} C_{\ell;Z} < \infty,$$

  where $\{C_{\ell;Z} : \ell \geq 0\}$ is the angular power spectrum of $Z(\cdot, t)$ such that $C_{\ell;Z} > 0$ for any $\ell \geq 0$ (cf. Condition 1.3);

  (ii) for every $t \neq s$, the random fields $Z(\cdot, t)$ and $Z(\cdot, s)$ are independent, so that, for any $x, y \in \mathbb{S}^2$,

$$\mathbb{E}\left[Z(x, t) Z(y, s)\right] = 0.$$

For any $t \in \mathbb{Z}$, (2.4) yields

$$\left(\Phi_j T(\cdot, t-j) \right)(x) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \phi_{\ell;j} a_{\ell,m}(t-j) Y_{\ell,m}(x), \quad x \in \mathbb{S}^2,$$

namely, the spectral representation of $\left(\Phi_j T(\cdot, t-j) \right)(\cdot)$ is characterized by the following set of harmonic coefficients $\{\phi_{\ell;j} a_{\ell,m}(t-j) : \ell \geq 0; m = -\ell, \ldots, \ell\}$. Consequently, from (1.1), it follows that

$$a_{\ell,m}(t) = \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) + a_{\ell,m;Z}(t), \tag{2.5}$$

that is, that is, the coefficients $\{a_{\ell,m}(t) : t \in \mathbb{Z}\}$ follows an autoregressive model of order p, for any $\ell \geq 0$, $m = -\ell, \ldots, \ell$.

Remark 2.2. An alternative formulation can be used to described our setting. It is well known that real-valued autoregressive models of order p can be represented as order 1 vector autoregressive models, thus defined in $\mathbb{R}^p$. In a functional framework, our SPHAR(p) can be rewritten as a process defined in the Cartesian product of p copies of $L^2(\mathbb{S}^2)$ (see [8, Section 5.1]).

Before concluding this section, let us fix the so-called truncation frequency $L \in \mathbb{N}$. Then, the truncated random field is defined by

$$
T_L(x, t) = \sum_{\ell=0}^{L-1} \sum_{m=-\ell}^{\ell} a_{\ell,m}(t) Y_{\ell,m}(x).
$$

Remark 2.3. The truncated random field $T_L(x, t)$ describes exactly a band-limited random field, with band-width $L_0 < L$. Moreover, as mentioned for example in [36], it provides a very good approximation of a smooth random field, in the sense that its covariance decays fast as ℓ grows to infinity.

3. LASSO estimation on the sphere

Here we present LASSO-type estimators for spherical autoregressive kernels under sparsity assumptions. More specifically, merging the techniques based on the stability measure presented in [2] with the properties of SPHAR(p) processes enables the construction of functional estimators for the kernels $\{k_j : j = 1, \ldots, p\}$.

3.1. The estimator construction

As introduced in Section 1, the spectral decomposition on the sphere allows to reduce the functional penalized minimization problem to the equivalent ℓ_1-penalized problems in the space of the harmonic coefficients, see (3.1). Let us recall the definition of our LASSO estimator.

Definition 3.1 (*LASSO Estimator*). The functional LASSO estimator for the vector of kernels $\mathbf{k} = (k_1, \ldots, k_p)$ is defined by Eq. (1.8), that is,

$$
\widehat{\mathbf{k}}_N^{\text{lasso}} = \underset{\mathbf{k} \in \mathcal{P}_N^p}{\text{argmin}} \ \frac{1}{N} \sum_{t=p+1}^{n} \left\| T(\cdot, t) - \sum_{j=1}^{p} \Phi_j T(\cdot, t-j) \right\|_{L^2(\mathbb{S}^2)}^2 + \lambda \sum_{j=1}^{p} \| \Phi_j \|_{\text{TR}},
$$

where $\mathcal{P}_N^p$ is given by (1.9).

More in details, the first part of (1.8) can be interpreted as a *functional residual sum of squares*, where each term has spectral decomposition

$$
T(\cdot, t) - \sum_{j=1}^{p} \left(\Phi_j T(\cdot, t-j) \right) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \left(a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right) Y_{\ell,m}.
$$

As a consequence,

$$
\sum_{t=p+1}^{n} \left\| T\left(\cdot, t\right) - \sum_{j=1}^{p} \Phi_j \left(T\left(\cdot, t-j\right)\right) \right\|_{L^2(\mathbb{S}^2)}^2
$$

$$
= \sum_{t=p+1}^{n} \sum_{\ell \geq 0} \sum_{m=-\ell}^{\ell} \left| a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right|^2 .
$$

Recalling that the truncation level induced by $\mathcal{P}_N^p$ is $L = L_N$, which depends on the number of observations N, the functional minimization problem can be reformulated as

$$
\widehat{\mathbf{k}}_N^{\text{lasso}} = \sum_{\ell=0}^{L_N-1} \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \frac{2\ell+1}{4\pi} P_\ell,
$$

with

$$
\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} = \operatorname*{argmin}_{\boldsymbol{\phi}_\ell \in \mathbb{R}^p} \frac{1}{N} \sum_{t=p+1}^{n} \left| a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right|^2 + \lambda(2\ell+1) \left\| \boldsymbol{\phi}_\ell \right\|_1
$$

$$
= \operatorname*{argmin}_{\boldsymbol{\phi}_\ell \in \mathbb{R}^p} \frac{1}{N(2\ell+1)} \sum_{t=p+1}^{n} \left| a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right|^2 + \lambda \left\| \boldsymbol{\phi}_\ell \right\|_1 . \tag{3.1}
$$

Remark 3.2 (*Important Remark on the High Dimensional Nature of the Procedure*). The formula (3.1) could lead to a misleading interpretation of the minimization problem. Although from a computational point of view the procedure is *separable*, i.e. it can be solved separately for each ℓ, the problem still has a high-dimensional nature. First of all, the penalty parameter λ does not depend on ℓ. Moreover, our penalization problem applies also to SPHAR(1) processes, which means that it can be used as a penalization procedure only on the first L_N multipoles and not on the lags. This fact is made very clear by Figs. 1–4 in Section 5 where it shows up that, also with very small lag dimension ($p = 2$), the LASSO has the effect to regularize the kernel estimates. To better understand our point, take $p = 1$ and

$$
\widehat{\phi}_\ell^{\text{lasso}} = \operatorname*{argmin}_{\phi_\ell \in \mathbb{R}} \frac{1}{N(2\ell+1)} \sum_{t=2}^{n} \sum_{m=-\ell}^{\ell} \left| a_{\ell,m}(t) - \phi_\ell a_{\ell,m}(t-1) \right|^2 + \lambda \left| \phi_\ell \right| ;
$$

then the solution can be computed as follows

$$
\widehat{\phi}_\ell^{\text{lasso}} =
\begin{cases}
\widehat{\phi}_\ell^{\text{ols}} + \dfrac{\lambda}{2\widehat{C}_\ell} & \text{for } \widehat{\phi}_\ell^{\text{ols}} \widehat{C}_\ell < -\dfrac{\lambda}{2} \\[2mm]
0 & \text{for } \left| \widehat{\phi}_\ell^{\text{ols}} \widehat{C}_\ell \right| \leq \dfrac{\lambda}{2} \\[2mm]
\widehat{\phi}_\ell^{\text{ols}} - \dfrac{\lambda}{2\widehat{C}_\ell} & \text{for } \widehat{\phi}_\ell^{\text{ols}} \widehat{C}_\ell > \dfrac{\lambda}{2}
\end{cases} .
$$

Note that the procedure shrinks the estimate towards zero not only if the *true* value ϕ_ℓ is close to zero, but also if the variance C_ℓ is very small, which clearly happens at high frequencies, because of the summability of the C_ℓ's. On the contrary, for classical LASSO problems, one assumes that the regressors are standardized, which surely cannot happen in our case at fixed

ℓ. Indeed, since here we start our penalization procedure from Eq. (1.8), we can at most rescale by the total variance of the spherical process.

Remark 3.3. In our functional minimization problem, the penalization term implicitly assigns a uniform weight to all the p linear operators. Indeed, all the kernels k_j, $j \in \{1, \ldots, p\}$, are assumed to be sparsely decomposable. To take into account different structures for the k_j's and capture the potential complexity of the model, several strategies can be hinted, such as adaptive LASSO, elastic-net or group-Lasso variants; see also [2, Section 5.3] for a similar discussion related to VAR processes.

Remark 3.4. Note that the penalization procedure given by Definition 3.1 preserves isotropy. This is a consequence of the fact that we are considering a *block-sparsity* model; indeed, given the structure of the predictor $\Phi_j T(\cdot, t - j)$, where all the a_ℓ share the same $\phi_{\ell;j}$, the procedure will automatically select only the relevant multipoles ℓ. As a result, a multipole is either removed entirely or not removed at all from the jth component of $\widehat{\mathbf{k}}_N^{\text{lasso}}$. The reader is referred for further discussions to [12], where it is shown that not all the ℓ_1-penalized problems have solutions which are isotropic, and to [23] for sparsity enforcing procedures for isotropic spherical random fields.

3.2. Matrix notation

An alternative form for the minimization problem given by Definition 3.1 can be introduced as follows. First, we define the following $N(2\ell + 1)$-dimensional vectors,

$$\mathbf{Y}_{\ell,N} = \left(a_{\ell,-\ell}(n), \ldots, a_{\ell,-\ell}(p+1), \ldots, a_{\ell,\ell}(p+1)\right)',$$

$$\mathbf{Y}_{\ell,N}(h) = \left(a_{\ell,-\ell}(n-h), \ldots, a_{\ell,-\ell}(p+1-h), \ldots, a_{\ell,\ell}(p+1-h)\right)',$$
$$h = 1, \ldots, p,$$

$$\mathbf{E}_{\ell,N} = \left(a_{\ell,-\ell;Z}(n), \ldots, a_{\ell,-\ell;Z}(p+1), \ldots, a_{\ell,\ell;Z}(p+1)\right),$$

where we recall that $N = n - p$. We can thus define the $(N(2\ell+1) \times p)$ matrix

$$X_{\ell,N} = \{\mathbf{Y}_{\ell,N}(1) : \cdots : \mathbf{Y}_{\ell,N}(p)\}, \tag{3.2}$$

so that the LASSO problem (1.8) reduces to

$$\widehat{\mathbf{k}}_N^{\text{lasso}} = \sum_{\ell=0}^{L_N-1} \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \frac{2\ell+1}{4\pi} P_\ell,$$

with

$$\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} = \underset{\boldsymbol{\phi}_\ell \in \mathbb{R}^p}{\text{argmin}} \frac{1}{N(2\ell+1)} \left\| \mathbf{Y}_{\ell,N} - X_{\ell,N}\boldsymbol{\phi}_\ell \right\|_2^2 + \lambda \left\| \boldsymbol{\phi}_\ell \right\|_1. \tag{3.3}$$

Fixed $\ell = 0, \ldots, L_N - 1$, we define the covariance matrix Γ_ℓ, that is, the $p \times p$ matrix with generic ijth element $C_\ell(i-j)$. We can use (3.2) to define its unbiased estimator

$$\widehat{\Gamma}_{\ell,N} = \frac{X'_{\ell,N} X_{\ell,N}}{N(2\ell+1)}.$$

Let us now consider the product $X'_{\ell,N}\mathbf{E}_{\ell,N}/N(2\ell+1)$. Observe that $\mathbf{E}_{\ell,N}$ is related to the error random field Z, so that we can read this random object as the process obtained from

the product of the stochastic data matrix $X_{\ell,N}$ and the noise vector. Indeed, it represents the so-called *empirical process* (see [11] and the references therein), associated with the multipole ℓ. Furthermore, observe that the ℓth empirical process corresponds to the following sum

$$\frac{X'_{\ell,N}\mathbf{E}_{\ell,N}}{N(2\ell+1)} = \frac{X'_{\ell,N}\mathbf{Y}_{\ell,N}}{N(2\ell+1)} - \frac{X'_{\ell,N}X_{\ell,N}}{N(2\ell+1)}\boldsymbol{\phi}_\ell = \widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell, \tag{3.4}$$

where

$$\widehat{\gamma}_{\ell,N} = \frac{X'_{\ell,N}\mathbf{Y}_{\ell,N}}{N(2\ell+1)}.$$

Establishing an upper bound for the empirical processes will be crucial for the proof of the consistency property for the LASSO estimators.

4. Main results

In this section, we present the main results of this paper, which consist of properties for the LASSO-type estimator of **k** given in (1.8). First of all, in Section 4.1, we introduce the concept of stability measure, a powerful tool firstly proposed in the LASSO framework in [2], to obtain some bounds on the concentration of the sample covariances and the empirical processes around their expected values. Then, in Section 4.2, we will follow the standard scheme of LASSO-techniques, see [11,26], to establish a *basic inequality*, a *deviation condition* and a *compatibility condition*. Finally, in Section 4.3, we present our main theorem, that is, the so-called *oracle inequalities* for $\widehat{\mathbf{k}}_N^{\text{lasso}}$.

4.1. Stability measure on the sphere and deviation bounds

Here, we discuss the stability measure for SPHAR(p) random fields. Intuitively, a stability measure quantifies the dependencies between the variables of the process and, hence, how stable the autocovariance matrix is. The more intricate the dependencies between the variables are, the less stable should the process result. Several proposals aiming to represent and measure the stability of a given process have been suggested in the literature over the years, mostly involving set of mixing conditions, in order to establish for how long in time the dependence between the components is effective (see Appendix E in the supplementary file of [2]). The stability measure considered here is in line with the one defined by [2].

Recall that $\{a_{\ell,m}(t), t \in \mathbb{Z}\}$ can be read as a real-valued autoregressive process of order p (see (2.5)). Under standard stationarity assumptions (see [10, page 123]), we can define its spectral density as

$$f_\ell(\nu) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} C_\ell(\tau) e^{-i\nu\tau} = \frac{1}{2\pi} \frac{C_{\ell;Z}}{\left|\phi_\ell(e^{-i\nu})\right|^2}, \qquad \nu \in [-\pi, \pi],$$

which is bounded and continuous. Upper and lower extrema of the spectral density over the unit circle are hence given by

$$\mathcal{M}(f_\ell) = \max_{\nu \in [-\pi,\pi]} f_\ell(\nu),$$
$$\mathfrak{m}(f_\ell) = \min_{\nu \in [-\pi,\pi]} f_\ell(\nu).$$

In what follows, we adopt $\mathcal{M}(f_\ell)$ as a measure of the stability of the process $\{a_{\ell,m}(t), t \in \mathbb{Z}\}$. Generalizing [2], we can consider this as a *band limited* stability measure, in the sense that it

refers only to the subprocesses belonging to the multipole ℓ. A *global* stability measure can be obtained by considering jointly all the multipoles $\ell \geq 0$ via the following definition

$$\mathcal{M} = \mathcal{M}(T) = \max_{\ell \geq 0} \mathcal{M}(f_\ell);$$

whereas we can refer to $\mathcal{M}_N = \max_{\ell < L_N} \mathcal{M}(f_\ell)$ as the *observed* stability measure. Let us now define the following p-dimensional process

$$\tilde{a}_{\ell,m}(t) = \left(a_{\ell,m}(t), \ldots, a_{\ell,m}(t-p+1)\right)',$$

with spectral density and corresponding stability measure given by

$$\tilde{f}_\ell(v) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \Gamma_\ell(\tau) e^{-i\tau v} \quad \text{and} \quad M\left(\tilde{f}_\ell\right) = \max_{v \in [-\pi,\pi]} \Lambda_{\max}(\tilde{f}_\ell(v)),$$

where

$$\Gamma_\ell(\tau) = \mathbb{E}\left[\tilde{a}_{\ell,m}(t+\tau)\tilde{a}'_{\ell,m}(t)\right]$$

and $\Gamma_\ell = \Gamma_\ell(0)$. We can therefore construct r-dimensional subprocesses of $\{\tilde{a}_{\ell,m}(t) : t \in \mathbb{Z}\}$ as follows. We fix a r-dimensional index $J = (j_1, \ldots, j_r)$, so that $J \in \{1, \ldots, p\}^r$, and $j_1 < \cdots < j_r$. Then, we define

$$\tilde{a}^J_{\ell,m}(t) = \left(\left(\tilde{a}_{\ell,m}(t)\right)_{j_1}, \ldots, \left(\tilde{a}_{\ell,m}(t)\right)_{j_r}\right)',$$

where $\left(\tilde{a}_{\ell,m}(t)\right)_i$ is the ith component of $\{\tilde{a}_{\ell,m}(t) : t \in \mathbb{Z}\}$. This subprocess has spectral density $\tilde{f}^J_\ell(v)$. We can finally introduce

$$\begin{aligned}
M\left(\tilde{f}_\ell, r\right) &= \max_{J \subset \{1, \ldots, p\}, |J| \leq r} M\left(\tilde{f}^J_\ell\right), \\
\widetilde{\mathcal{M}}(r) &= \max_{\ell \geq 0} M\left(\tilde{f}_\ell, r\right), \\
\widetilde{\mathcal{M}}_N(r) &= \max_{\ell < L_N} M\left(\tilde{f}_\ell, r\right),
\end{aligned}$$

respectively, the band-limited, the global and the observed stability measures of the subprocess $\{\tilde{a}^r_{\ell,m}(t) : t \in \mathbb{Z}\}$. Notice that $M\left(\tilde{f}_\ell\right) = M\left(\tilde{f}_\ell, p\right)$, while, for the sake of completeness, we define $M\left(\tilde{f}_\ell, r\right) = M(\tilde{f}_\ell)$, for all $r > p$. Moreover, it can be shown that

$$M\left(\tilde{f}_\ell, 1\right) \leq M\left(\tilde{f}_\ell, 2\right) \leq \cdots \leq M\left(\tilde{f}_\ell, p\right) = M\left(\tilde{f}_\ell\right).$$

The following quantities are also well-defined

$$\mu_{\min;\ell} = \min_{z \in \mathbb{C}:|z|=1} |\phi_\ell(z)|^2, \quad \mu_{\max;\ell} = \max_{z \in \mathbb{C}:|z|=1} |\phi_\ell(z)|^2, \quad \text{for} \quad \ell \geq 0. \tag{4.1}$$

Remark 4.1. Note that, in line with Remark 1.6, there exists a positive constant c such that

$$\max_{\ell \geq 0} \mu_{\max;\ell} = \max_{\ell \geq 0} \max_{z \in \mathbb{C}:|z|=1} \left|1 - \sum_{j=1}^{p} \phi_{\ell;j} z^j\right|^2 \leq \left(1 + \sum_{j=1}^{p} \max_{\ell \geq 0} |\phi_{\ell;j}|\right)^2 \leq c,$$

and

$$\min_{\ell \geq 0} \mu_{\min;\ell} = \min_{\ell \geq 0} \min_{z \in \mathbb{C}:|z|=1} \left| 1 - \sum_{j=1}^{p} \phi_{\ell;j} z^j \right|^2 = \min_{\ell \geq 0} \min_{z \in \mathbb{C}:|z|=1} \prod_{j=1}^{d_\ell} |1 - \xi_{\ell;j}^{-1} z|^2 \geq (1 - \xi_*^{-1})^{2p}$$

$$> 0,$$

recalling that $\xi_* = \min_{\ell \geq 0} \min_{j=1,\dots,d_\ell} |\xi_{\ell;j}|$. These bounds represent the moral counterpart of Proposition 2.2 in [2] and show how the global stability $\mathcal{M}$ behaves for SPHAR models. Indeed, for a SPHAR(1) process, the operator Φ_1 plays the same role of the matrix A_1 in [2] and $\xi_*^{-1} = \max_{\ell \geq 0} |\phi_\ell| = \|\Phi_1\|_{\mathrm{op}}$. As a consequence $\mathcal{M}$ is bounded as long as the operator norm of Φ_1 (the counterpart of the spectral radius of A_1) is bounded away from 1.

We apply now the idea of stability measure to establish some relevant deviation bounds on the covariance estimators and the empirical processes, which will be pivotal to analyze our regression problem. Note that, $\{a_{\ell,m}(t) : t \in \mathbb{Z}\}$, $m = -\ell, \dots, \ell$, can be seen as a tool to provide an alternative notation for the empirical process (3.4). Indeed the hth component of $X_{\ell,N}' \mathbf{E}_{\ell,N}/N(2\ell+1)$ is given by

$$\frac{\mathbf{Y}_{\ell,N}(h)' \mathbf{E}_{\ell,N}}{N(2\ell+1)} = \frac{1}{N(2\ell+1)} \sum_{m=-\ell}^{\ell} \sum_{t=p+1}^{n} a_{\ell,m}(t-h) \, a_{\ell,m;Z}(t). \tag{4.2}$$

Proposition 4.2 (*Deviation Bounds*)**.** *Assume that Conditions 1.3 and 1.5 hold. Then, there exists a constant $c > 0$ such that for any $\ell \in \mathbb{N}_0$, any r-sparse vectors $u, v \in \mathbb{R}^p$ with $\|u\|, \|v\| \leq 1$, $r \geq 1$ and any $\eta \geq 0$, it holds that*

$$\mathbb{P}\left(\left| v'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right) v \right| > 2\pi \mathcal{M}\left(\widetilde{f}_\ell, r\right) \eta \right) \leq 2e^{-cN(2\ell+1)\min\{\eta^2,\eta\}}, \tag{4.3}$$

$$\mathbb{P}\left(\left| u'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right) v \right| > 6\pi \mathcal{M}\left(\widetilde{f}_\ell, 2r\right) \eta \right) \leq 6e^{-cN(2\ell+1)\min\{\eta^2,\eta\}}. \tag{4.4}$$

In particular, for any $i, j \in \{1, \dots, p\}$, it holds that

$$\mathbb{P}\left(\left| \left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)_{ij} \right| > 6\pi \mathcal{M}(\widetilde{f}_\ell, 2)\eta \right) \leq 6e^{-cN(2\ell+1)\min\{\eta^2,\eta\}}. \tag{4.5}$$

Moreover, for all $1 \leq h \leq p$, it holds that

$$\mathbb{P}\left(\left| \frac{\mathbf{Y}_{\ell,N}'(h) \mathbf{E}_{\ell,N}}{N(2\ell+1)} \right| > 2\pi \, C_{\ell;Z} \left(1 + \frac{1 + \mu_{\max;\ell}}{\mu_{\min;\ell}} \right) \eta \right) \leq 6e^{-cN(2\ell+1)\min\{\eta^2,\eta\}}, \tag{4.6}$$

where $\mu_{\max;\ell}$ and $\mu_{\min;\ell}$ are defined by (4.1).

Remark 4.3. Note that the deviation bounds can be also stated in terms of the global or observed stability measure $\widetilde{\mathcal{M}}(r)$ or $\widetilde{\mathcal{M}}_N(r)$.

The analogous result presented in [2] is very general, since it deals with stationary Gaussian random processes on $\mathbb{R}^d$, while here we focus on deviation bounds for our specific empirical covariance matrices and empirical processes. The main technical difference between our results and the ones in [2] is that, in our framework, we use observations from a group of $2\ell+1$ stationary processes, namely, $\{a_{\ell,m}(t) : t \in \mathbb{Z}\}$, $m = -\ell, \dots, \ell$, to estimate the same covariance matrix Γ_ℓ, exploiting the isotropy of the random field.

Similarly to the work [2], (4.3) and (4.6) quantify how the underlying estimators concentrate around their expected values. In particular, (4.3) will be used to verify the compatibility

condition (see Proposition 4.6), while (4.6) will be used to prove the deviation condition (see Proposition 4.8). In the i.i.d. case, bounds on the empirical process can be easily established, since the data matrix is deterministic and the randomness comes only from the noise vector. In our case, similarly to [2], the ℓth empirical process is the product of a dependent noise vector and a stochastic data matrix. Therefore, proving consistency requires a bound on both these two random objects.

4.2. Bounds for LASSO techniques

We are now in the position to present the classical path of LASSO in our setting.

The very first result concerns the so-called *basic inequality*, an elementary yet essential result, which does not require any condition or assumption, except the existence of a linear underlying model, and it is simply based on the definition of the LASSO estimator.

Proposition 4.4 (*Basic Inequality*)**.** *Consider the estimation problem in Definition* 3.1*. For any* $\ell = 0, \ldots, L_N - 1$*, set* $v_\ell = \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell$*. Then, the following basic inequality holds*

$$v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell \leq \frac{2 v_\ell' X_{\ell,N}' \mathbf{E}_{\ell,N}}{N(2\ell + 1)} + \lambda \left[\left\| \boldsymbol{\phi}_\ell \right\|_1 - \left\| \boldsymbol{\phi}_\ell + v_\ell \right\|_1 \right] \quad a.s. \tag{4.7}$$

This simple result implies that the prediction error $v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell$ is bounded by the sum of two factors. The first one is random and it depends on the empirical process $X_{\ell,N}' \mathbf{E}_{\ell,N} / N(2\ell + 1)$. The second one is deterministic and its value depends on the penalty parameter λ, on N and on the chosen linear model itself.

The second step consists in defining an event $\mathscr{S}_N$ such that the fluctuations of the random factors

$$\frac{2 v_\ell' X_{\ell,N}' \mathbf{E}_{\ell,N}}{N(2\ell + 1)}, \quad \ell = 0, \ldots, L_N - 1,$$

when conditioned to $\mathscr{S}_N$, are all controlled by the same deterministic quantity. Moreover, we need to prove that this event has a high probability, implying that a bound on the prediction errors can be obtained in most cases. The event $\mathscr{S}_N$ is defined as follows.

Definition 4.5. In the setting previously described, let

$$\mathscr{S}_N = \bigcap_{\ell=0}^{L_N-1} \{ \left\| \widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N} \boldsymbol{\phi}_\ell \right\|_\infty \leq \mathcal{F}_N \sqrt{\frac{\log p L_N}{N}} \},$$

where $\mathcal{F}_N$ is a deterministic function depending only on the parameters $\left(\boldsymbol{\phi}_0, \ldots, \boldsymbol{\phi}_{L_N-1} \right)$ and noise variances $\left(C_{0;Z}, \ldots, C_{L_N-1;Z} \right)$. The deviation condition is said to hold if the event $\mathscr{S}_N$ happens.

The following theorem shows that, for an appropriate choice of $\mathcal{F}_N$ and L_N, the event $\mathscr{S}_N$ has high probability to occur.

Proposition 4.6 (*Deviation Condition*)**.** *Consider the estimation problem in Definition* 3.1 *and assume that Conditions* 1.3 *and* 1.5 *hold. There exist some constants* $c_0, c_1, c_2 > 0$ *such that, if we define*

$$\mathcal{F} \left(\boldsymbol{\phi}_\ell, C_{\ell;Z} \right) = c_0 \left[C_{\ell;Z} \left(1 + \frac{1 + \mu_{\max;\ell}}{\mu_{\min;\ell}} \right) \right], \qquad \mathcal{F}_N = \max_{\ell < L_N} \mathcal{F} \left(\boldsymbol{\phi}_\ell, C_{\ell;Z} \right),$$

and if $N \succeq \log p L_N$, then

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1} \left\|\widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell\right\|_\infty \leq \mathcal{F}_N\sqrt{\frac{\log p L_N}{N}}\right) \geq 1 - c_1 e^{-c_2 \log(p L_N)}.$$

Remark 4.7. Observe that:

(i) the results presented in Proposition 4.6 also hold for a different choice of $\mathcal{F}_N$, that is,

$$\mathcal{F}_N = c_0 \left[\left(\max_{\ell < L_N} C_{\ell;Z}\right)\left(1 + \frac{1 + \max_{\ell < L_N} \mu_{\max;\ell}}{\min_{\ell < L_N} \mu_{\min;\ell}}\right)\right],$$

which corresponds to the one used in [2];

(ii) in order for this bound to make sense, we need that

$$\log(p L_N) = o(N).$$

The third and final step is to establish a *compatibility condition* that, whenever verified on the event $\mathscr{S}_N$, will allow us to bound both the prediction errors $\{\left\|X_{\ell,N}\left(\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell\right)\right\|_2^2\}$ and the estimation errors $\{\left\|\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell\right\|_2^2\}$ by the same quantity. In this sense, it makes the errors compatible.

A symmetric $d \times d$ matrix A satisfies the compatibility condition, also called restricted eigenvalue (RE) condition, with curvature $\alpha > 0$ and tolerance $\tau > 0$ ($A \sim RE(\alpha, \tau)$), if, for any $\vartheta \in \mathbb{R}^d$,

$$\vartheta' A \vartheta \geq \alpha \|\vartheta\|_2^2 - \tau \|\vartheta\|_1^2. \tag{4.8}$$

The next result gives some sufficient conditions in order to have

$$\bigcap_{\ell=0}^{L_N-1} \{\widehat{\Gamma}_{\ell,N} \sim RE(\alpha_\ell, \tau_\ell)\},$$

for some α_ℓ's and τ_ℓ's, with high probability.

Proposition 4.8 (*Compatibility Condition*). *Consider the estimation problem in Definition 3.1 and assume that Conditions 1.3 and 1.5 hold. Define $q_N = \max_{\ell < L_N} q_\ell$. There exist some constants $c_1, c_2, c_3 > 0$ such that, if*

$$N \succeq \max\{\omega_N^2, 1\} q_N \log(p L_N), \quad \text{with} \quad \omega_N = c_3 \max_{\ell < L_N} \frac{\mu_{\max;\ell}}{\mu_{\min;\ell}}, \tag{4.9}$$

then

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1} \{\widehat{\Gamma}_{\ell,N} \sim RE(\alpha_\ell, \tau_\ell)\}\right) \geq 1 - c_1 e^{-c_2 N \min\{\omega_N^{-2}, 1\}}, \tag{4.10}$$

with

$$\alpha_\ell = \frac{C_{\ell;Z}}{2\,\mu max}, \quad \text{and} \quad \tau_\ell = \alpha_\ell \max\{\omega_N^2, 1\}\frac{\log(p L_N)}{N}. \tag{4.11}$$

Remark 4.9. The results presented in Proposition 4.8 also hold for

$$\omega_N = c_3 \frac{\max_{\ell < L_N} \mu_{\max;\ell}}{\min_{\ell < L_N} \mu_{\min;\ell}},$$

analogously to the findings in [2].

Remark 4.10. The compatibility condition on $\widehat{\Gamma}_{\ell,N}$ is a requirement on its smallest eigenvalue, which can be seen as a measure of the dependence of the random matrix columns. The sufficient condition (4.9) ensures that, with high probability, the (sample) minimum eigenvalue of the matrix $\widehat{\Gamma}_{\ell,N}$ is bounded away from zero.

Remark 4.11. Note that $0 \leq \mathcal{F}_N \leq \mathcal{F}$, where $\mathcal{F} = \max_{\ell \geq 0} \mathcal{F}\left(\boldsymbol{\phi}_\ell, C_{\ell;Z}\right)$ exists finite. Indeed,

$$\mathcal{F}\left(\boldsymbol{\phi}_\ell, C_{\ell;Z}\right) = c_0 \left[C_{\ell;Z} \left(1 + \frac{1 + \mu_{\max;\ell}}{\mu_{\min;\ell}} \right) \right] \longrightarrow 0, \qquad \text{as } \ell \longrightarrow \infty,$$

since $C_{\ell;Z}$ converges to zero as ℓ goes to infinity and $a \leq \mu_{\min;\ell} \leq \mu_{\max;\ell} \leq b$, with a, b positive constants (independent of ℓ), see Remark 4.1. Similarly, it holds that $0 \leq \omega_N \leq \omega$, where

$$\omega = c_3 \max_{\ell \geq 0} \frac{\mu_{\max;\ell}}{\mu_{\min;\ell}},$$

and $q_N \leq q = \max_{\ell \geq 0} q_\ell$. In particular, all the results presented in this paper can be stated using $\mathcal{F}, \omega, q$ instead of $\mathcal{F}_N, \omega_N, q_N$. Without loss of generality, we can assume $q_N \geq 1$.

4.3. Oracle inequalities

Oracle inequalities are used to estimate the accuracy of the $\widehat{\mathbf{k}}_N^{\text{lasso}}$. In general, $\widehat{\mathbf{k}}_N^{\text{lasso}}$ depends on the penalty parameter λ, according to (1.8). As a consequence, given a proper choice of λ, our oracle inequalities produce upper bounds for the estimation error in the L^2- and the L^∞-norms, resulting in the sum of two terms. The first summand describes the stochastic error due to the approximation of the first L_N components of $\mathbf{k}$ with $\widehat{\mathbf{k}}_N^{\text{lasso}}$; the second term is a deterministic error, related to the not estimated components of $\mathbf{k}$, with $\ell \geq L_N$. It is worth noting that, just like in the multivariate (i.e. non-functional) case, such upper bounds are characterized by a multiplicative factor $\log(pL_N)$; roughly speaking, this factor is the cost of not knowing explicitly the set of non-zero coefficients.

Theorem 4.12. *Consider the estimation problem in* Definition 3.1 *and assume that* Conditions 1.3 *and* 1.5 *hold. Moreover, suppose that,*

$$\bigcap_{\ell=0}^{L_N-1} \{\widehat{\Gamma}_\ell \sim RE(\alpha_\ell, \tau_\ell)\} \quad a.s., \quad with \quad q_\ell \tau_\ell \leq \alpha_\ell/32,$$

and the deviation condition is satisfied almost surely, that is,

$$\bigcap_{\ell=0}^{L_N-1} \{\left\| \widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell \right\|_\infty \leq \mathcal{F}_N \sqrt{\frac{\log\left(pL_N\right)}{N}}\} \quad a.s. \tag{4.12}$$

Then, for any penalty parameter $\lambda = \lambda_N \geq 4\mathcal{F}_N\sqrt{\log(pL_N)/N}$, any solution $\widehat{\mathbf{k}}_N^{\text{lasso}}$ satisfies

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^2}^2 \leq \frac{18}{\pi^2} \lambda_N^2 \sum_{\ell=0}^{L_N-1} \frac{q_\ell}{\alpha_\ell^2}(2\ell+1) + \sum_{\ell \geq L_N} \left\| \boldsymbol{\phi}_\ell \right\|_2^2 \frac{2\ell+1}{8\pi^2} \quad a.s.; \tag{4.13}$$

moreover, under the additional Condition 1.7, *it holds that*

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^\infty} \leq \frac{3}{\pi} \lambda_N \sum_{\ell=0}^{L_N-1} \frac{\sqrt{q_\ell}}{\alpha_\ell} (2\ell+1) + \sum_{\ell \geq L_N} \left\| \boldsymbol{\phi}_\ell \right\|_2 \frac{2\ell+1}{4\pi} \quad a.s. \tag{4.14}$$

To discuss the possible rates of convergence in (4.13)–(4.14), let us choose $\lambda_N = 4\mathcal{F}_N \sqrt{\log(pL_N)/N}$ and $L_N \simeq N^d$. Moreover, we impose some semiparametric structure to the set $\{\phi_{\ell;j} : \ell \geq 0\}$, that is,

$$\left| \phi_{\ell;j} \right| \leq G_j \ell^{-\beta_j},$$

where $\beta_j > 1$ and $G_j > 0$; see also [31] for an interpretation in terms of regularity of the kernels. Note that, since we are looking at the asymptotic behavior as $N \to \infty$, the sufficient condition (4.9) automatically holds and the curvature $\alpha_\ell \simeq C_{\ell;Z}$. In particular, a standard assumption for the behavior of the power spectrum of a spherical white noise is $C_{\ell;Z} \simeq \ell^{-\alpha}$, with $\alpha > 2$, see [35]. As a consequence, in this framework, we have

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^2}^2 \leq \frac{18}{\pi^2} \lambda_N^2 \sum_{\ell=0}^{L_N-1} \frac{q_\ell}{\alpha_\ell^2} (2\ell+1) + \sum_{\ell \geq L_N} \left\| \boldsymbol{\phi}_\ell \right\|_2^2 \frac{2\ell+1}{8\pi^2}$$

$$= O\left(\frac{\log pL_N}{N} \sum_{\ell=0}^{L_N-1} \ell^{2\alpha} (2\ell+1) + \sum_{\ell \geq L_N} \ell^{-2\beta}(2\ell+1) \right)$$

$$= O\left(\log N \, N^{2d(\alpha+1)-1} + N^{2d(1-\beta)} \right),$$

where $\beta = \min_{j=1,\dots,p} \beta_j$, and, in order to ensure consistency, we can choose

$$0 < d < \frac{1}{2(\alpha+1)}.$$

Analogously, imposing this time $\beta_j > 2$, one has that

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^\infty} \leq \frac{3}{\pi} \lambda_N \sum_{\ell=0}^{L_N-1} \frac{\sqrt{q_\ell}}{\alpha_\ell} (2\ell+1) + \sum_{\ell \geq L_N} \left\| \boldsymbol{\phi}_\ell \right\|_2 \frac{2\ell+1}{4\pi}$$

$$= O\left((\log N)^{1/2} \, N^{d(\alpha+2)-\frac{1}{2}} + N^{d(2-\beta)} \right)$$

and in this case the consistency in the supremum norm is reached for any

$$0 < d < \frac{1}{2(\alpha+2)}.$$

Moreover, we stress that the parameter α can be estimated via a Whittle-like procedure, see [19,20], which could be useful to choose a suitable rate for L_N.

5. Some numerical results

In this section we illustrate the finite sample performance of the LASSO estimator $\widehat{\mathbf{k}}_N^{\text{lasso}}$ under various degrees of sparsity. In particular, fixed $N = 300$ and $L_N = L = 50$, we are concerned with the empirical evaluation of the L^2-risk and L^∞-risk of $\widehat{\mathbf{k}}_N^{\text{lasso}}$ for the penalty parameters

$$\lambda_{2i-1} = 10^{i-5} \quad \text{and} \quad \lambda_{2i} = 5 \times 10^{i-5}, \quad \text{for} \quad i = 1, \dots, 4,$$

in comparison with the one of the non-penalized estimator described in [15], corresponding to the case $\lambda_0 = 0$. We remark that our simulations can be considered as a hint, in view of future applications on real data.

In what follows, we consider seven different case studies, all belonging to the class of SPHAR(2) processes, so that (1.2) becomes

$$T(x, t) = \Phi_1 (T(\cdot, t-1))(x) + \Phi_2 (T(\cdot, t-2))(x) + Z(x, t).$$

The noise structure is kept fixed, with $C_{\ell;Z} \propto \ell^{-2}$, and the operators Φ_1 and Φ_2 are selected to ensure that the process satisfy Condition 1.5 with a signal-to-noise ratio $\max_{\ell \geq 0} C_\ell / C_{\ell;Z}$ not exceeded 2.

In the first case, the random field T_1 is strongly sparse, in the sense that the only non-null eigenvalues are

$$\phi_{2;1} = -0.7, \quad \phi_{3;2} = 0.5.$$

In the fourth case, the random field T_4 is characterized by less sparsity and homogeneity between the two kernels; in particular, the non-null coefficients are

$$\phi_{18;1} = -0.5, \quad \phi_{19;1} = 0.3, \quad \phi_{20;1} = 0.65, \quad \phi_{2;2} = 0.25,$$
$$\phi_{3;2} = -0.67, \quad \phi_{5;2} = -0.7.$$

The remaining fields are defined as follows:

- T_2: $\phi_{\ell;j} \neq 0$ for $\ell < 10$ and 0 otherwise, $j = 1, 2$;
- T_3: non-sparse version of T_2 with $\phi_{\ell,j} \propto \ell^{-3}$ for $\ell \geq 10$, $j = 1, 2$;
- T_5: $\phi_{\ell;j} \neq 0$ for $\ell < 20$ and 0 otherwise, $j = 1, 2$;
- T_6: non-sparse version of T_5 with $\phi_{\ell,j} \propto \ell^{-3}$ for $\ell \geq 20$, $j = 1, 2$;
- T_7: $\phi_{\ell;j} \neq 0$ for $\ell < 50$ and $\phi_{\ell,j} \propto \ell^{-3}$ for $\ell \geq 50$, $j = 1, 2$.

Note that, although T_3 and T_6 are non-sparse, they are very similar to T_2 and T_5, respectively, since their ϕ_ℓ's at higher multipoles are non-null, but small.

For each scenario, we run $B = 1000$ replications and we compute the following quantities

$$\text{MSE}\left(\widehat{\mathbf{k}}_N^{\text{lasso}}, \mathbf{k}\right) = \frac{1}{B} \sum_{b=1}^{B} \left\{ \frac{1}{G} \sum_{g=1}^{G} \sum_{j=1}^{2} \left(\widehat{k}_{j,b}^{\text{lasso}}(z_g) - k_j(z_g)\right)^2 \right\},$$

$$\text{SUP}\left(\widehat{\mathbf{k}}_N^{\text{lasso}}, \mathbf{k}\right) = \frac{1}{B} \sum_{b=1}^{B} \left\{ \sup_{g \in \{1, \ldots, G\}} \sqrt{\sum_{j=1}^{2} \left(\widehat{k}_{j,b}^{\text{lasso}}(z_g) - k_j(z_g)\right)^2} \right\},$$

where $\{z_1, \ldots, z_G\}$, $G = 2000$, is an equally spaced grid over $[-1, 1]$. As usually, the MSE can be decomposed into the sum of

$$\text{VAR}\left(\widehat{\mathbf{k}}_N^{\text{lasso}}, \mathbf{k}\right) = \frac{1}{B} \sum_{b=1}^{B} \left\{ \frac{1}{G} \sum_{g=1}^{G} \sum_{j=1}^{2} \left(\widehat{k}_{j,b}^{\text{lasso}}(z_g) - \bar{k}_j^{\text{lasso}}(z_g)\right)^2 \right\},$$

$$\text{BIAS}^2\left(\widehat{\mathbf{k}}_N^{\text{lasso}}, \mathbf{k}\right) = \frac{1}{G} \sum_{g=1}^{G} \sum_{j=1}^{2} \left(\bar{k}_j^{\text{lasso}}(z_g) - k_j(z_g)\right)^2,$$

where $\bar{k}_j^{\text{lasso}}(z_g)$ is the average over the B replicates of $\widehat{k}_{j,b}^{\text{lasso}}(z_g)$.

Table 1 collects the values of the empirical L^2-error (MSE) and L^∞-error (SUP) associated with the seven models of interest; the bold entries correspond to their minima over λ, which

Table 1
Values of the L^2-error (MSE) and L^∞-error (SUP) for the case studies $T_1, \ldots, T_7$, by varying the penalty parameter λ.

| MSE	T_1	T_2	T_3	T_4	T_5	T_6	T_7
λ_0	0.00209	0.00200	0.00202	0.00205	0.00195	0.00195	**0.00179**
λ_1	0.00173	0.00166	0.00167	0.00171	0.00161	0.00162	0.00185
λ_2	0.00107	0.00101	0.00103	0.00106	0.00102	0.00102	0.00334
λ_3	0.00081	0.00076	0.00077	0.00082	**0.00083**	**0.00084**	0.00791
λ_4	0.00043	0.00042	0.00044	**0.00061**	0.00119	0.00119	0.14259
λ_5	0.00033	**0.00037**	**0.00039**	0.00095	0.00273	0.00272	0.40857
λ_6	0.00018	0.00090	0.00091	0.01518	0.04605	0.04597	2.93066
λ_7	**0.00015**	0.00223	0.00224	0.05287	0.11930	0.11931	3.63458
λ_8	0.00021	0.00964	0.00961	0.19060	0.42671	0.42670	4.32159
SUP	T_1	T_2	T_3	T_4	T_5	T_6	T_7
λ_0	0.22810	0.22510	0.22724	0.22756	0.22454	0.22316	**0.21326**
λ_1	0.20965	0.20698	0.20896	0.21022	0.20693	0.20551	0.21841
λ_2	0.16302	0.16163	0.16294	0.16629	0.16529	0.16314	0.34847
λ_3	0.13592	0.13364	0.13559	0.13906	**0.14416**	**0.14526**	0.60061
λ_4	0.07845	0.08183	**0.09875**	**0.10212**	0.18168	0.19659	2.78143
λ_5	0.06127	**0.07211**	0.10213	0.12481	0.29815	0.31749	4.81152
λ_6	0.03384	0.11782	0.14852	0.46755	1.56045	1.58153	11.5330
λ_7	**0.02824**	0.18457	0.20981	0.79587	2.54149	2.56116	12.7524
λ_8	0.03475	0.39983	0.40388	1.23177	4.06218	4.06166	13.9386
λ_9	0.05762	0.67875	0.66424	1.23839	4.02005	4.02358	13.8951
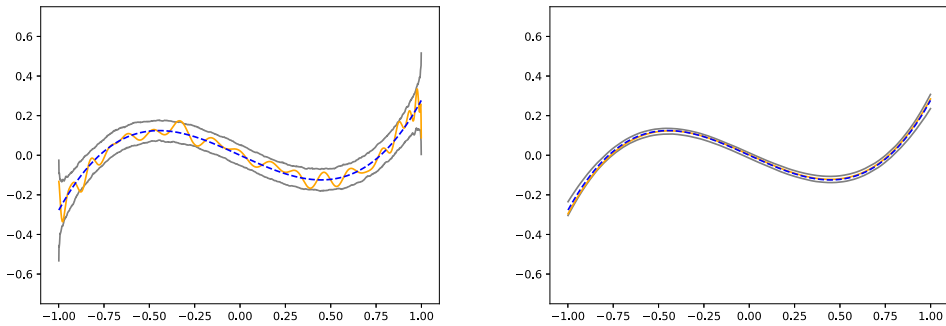
Fig. 1. *Model T_1*. 95% confidence bands (gray lines) for least squares (left panel) and LASSO with $\lambda = \lambda_7$ (right panel), based on $B = 1000$ replicates. The blue dashed line is the true k_2, while the orange straight one corresponds to one estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

decrease as the level of sparsity grows. As expected, LASSO-type estimators outperform least squares uniformly in all the cases, except in the last one (highly non-sparse). Even for the non-sparse models T_3 and T_6, LASSO achieves good performance, as for T_2 and T_5, due to its smoothing effect on the higher multipoles.
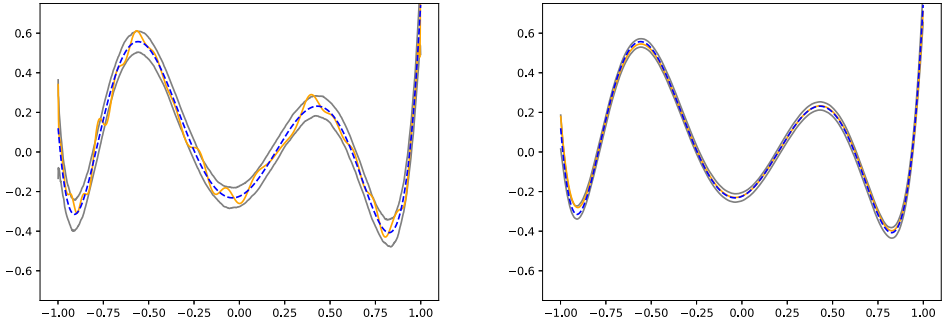
Fig. 2. *Model T_2*. 95% confidence bands (gray lines) for least squares (left panel) and LASSO with $\lambda = \lambda_5$ (right panel), based on $B = 1000$ replicates. The blue dashed line is the true k_2, while the orange straight one corresponds to one estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
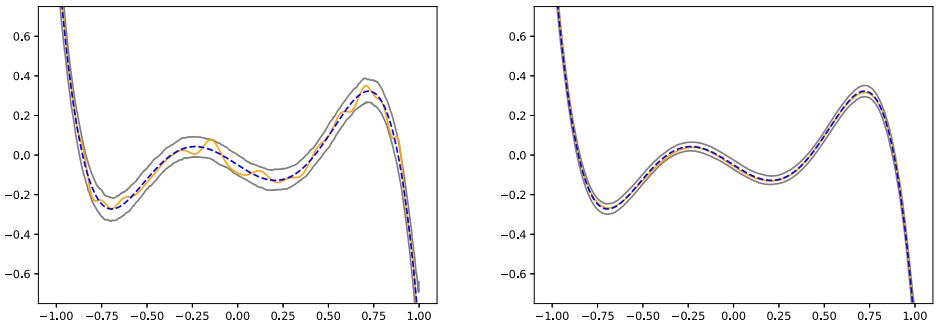

Fig. 3. *Model T_4*. 95% confidence bands (gray lines) for least squares (left panel) and LASSO with $\lambda = \lambda_4$ (right panel), based on $B = 1000$ replicates. The blue dashed line is the true k_2, while the orange straight one corresponds to one estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. *Model T_5*. 95% confidence bands (gray lines) for least squares (left panel) and LASSO with $\lambda = \lambda_3$ (right panel), based on $B = 1000$ replicates. The blue dashed line is the true k_2, while the orange straight one corresponds to one estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
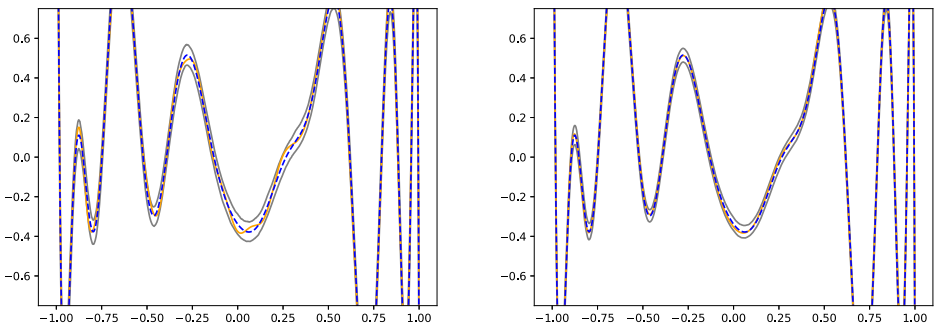
Table 2

Values of the variance (VAR) and squared bias (BIAS^2) for the case studies $T_1, \ldots, T_7$, by varying the penalty parameter λ.

VAR	T_1	T_2	T_3	T_4	T_5	T_6	T_7
λ_0	0.00208	0.00200	0.00201	0.00205	0.00194	0.00195	0.00178
λ_1	0.00173	0.00166	0.00167	0.00170	0.00161	0.00162	0.00178
λ_2	0.00107	0.00101	0.00102	0.00106	0.00101	0.00102	0.00177
λ_3	0.00081	0.00075	0.00077	0.00081	0.00081	0.00081	0.00176
λ_4	0.00043	0.00041	0.00042	0.00045	0.00067	0.00068	0.00165
λ_5	0.00033	0.00034	0.00035	0.00035	0.00067	0.00067	0.00160
λ_6	0.00018	0.00030	0.00029	0.00025	0.00067	0.00068	0.00106
λ_7	0.00015	0.00027	0.00027	0.00021	0.00060	0.00060	0.00095
λ_8	0.00010	0.00025	0.00025	0.00012	0.00052	0.00052	0.00052

BIAS^2	T_1	T_2	T_3	T_4	T_5	T_6	T_7
λ_0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
λ_1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00006
λ_2	0.00000	0.00000	0.00000	0.00000	0.00001	0.00001	0.00157
λ_3	0.00000	0.00000	0.00000	0.00001	0.00002	0.00002	0.00615
λ_4	0.00000	0.00001	0.00001	0.00016	0.00052	0.00051	0.14094
λ_5	0.00000	0.00003	0.00004	0.00060	0.00206	0.00205	0.40697
λ_6	0.00000	0.00060	0.00062	0.01494	0.04538	0.04529	2.92960
λ_7	0.00001	0.00196	0.00197	0.05265	0.11870	0.11870	3.63364
λ_8	0.00011	0.00938	0.00936	0.19048	0.42618	0.42618	4.32107

Table 3

Values of the area under the ROC curve (AUROC) for the sparse models T_1, T_2, T_4, T_5, computed in correspondence of the best λ.

	AUROC
T_1	0.96874
T_2	0.97728
T_4	0.88277
T_5	0.90257

Figs. 1–4 illustrate the functional forms of the estimated kernel, compared to the real one, and the 95% confidence bands based on the $B = 1000$ replicates. The left panels present the non-penalized estimates of k_2 while the right panels contain the corresponding penalized ones. Heuristically, the (best) penalized estimates reconstruct the true kernel function better than the non-penalized ones, which show an oscillatory behavior (undersmoothing) due to the lack of selection of the relevant multipoles. Indeed, as usual for LASSO estimators, the variance is reduced by paying the price of a slightly larger bias, see also Table 2, leading anyway to an improvement in terms of MSE. This also suggests that, in the non-penalized case, the value of L_N must be tuned more carefully.

Moreover, also in terms of model selection, the values of the area under the ROC curve (AUROC), averaged over the $B = 1000$ replicates and reported in Table 3, reflect that the (best) penalized estimates perform pretty well.

Remark 5.1 (*Penalty Parameter Selection*). Looking at Table 1 it is clear that the role of the penalty parameter is crucial. Our theoretical results suggest possible rates of the penalty parameter that imply consistency of the kernel estimator. However, we cannot directly compute the optimal value of λ due to the presence of unknown constants. As a consequence, a careful selection of the penalty parameter using data driven methods needs to be conducted in order to make our procedure applicable in practice. The idea could be to implement a *functional cross-validation*, which takes into account all the frequencies simultaneously; this is not the focus of our work but it can be addressed in future more applied works.

6. Proofs

In the present section we prove the bounds showed in Section 4.2 as well as our main theorem. Many of the proofs have arguments which are broadly similar to those given for related results in [2].

Proof of Proposition 4.2. Let us define $J = \text{supp}(v) = \{j_1, \ldots, j_r\} \subset \{1, \ldots, p\}$, $r \geq 1$, and

$$W_{\ell,J} = X_{\ell,N} v = \sum_{j \in J} v_j \mathbf{Y}_{\ell,N}(j).$$

Then, $Q_{\ell,J} = \mathbb{E}\left[W_{\ell,J} W'_{\ell,J}\right] = B_{\ell,J} \otimes I_{2\ell+1}$, where $B_{\ell,J}$ is the covariance matrix of the random vector

$$\sum_{j \in J} v_j \begin{bmatrix} a_{\ell,m}(n-j) \\ \vdots \\ a_{\ell,m}(p+1-j) \end{bmatrix} \qquad \text{for any } m = -\ell, \ldots, \ell.$$

As a consequence, $\left\|Q_{\ell,J}\right\|_{op} = \left\|B_{\ell,J}\right\|_{op} \leq 2\pi \mathcal{M}(\widetilde{f}_\ell, r)$ (see [2, Proposition 2.4]) and (4.3) is proved. To prove (4.4), note that

$$2\left|u'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)v\right| \leq \left|u'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)u\right|$$
$$+ \left|v'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)v\right| + \left|(u+v)'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)(u+v)\right|,$$

and $u + v$ is $2r$-sparse with $|u + v| \leq 2$. The result follows by applying (4.3) separately on each of the three terms on the right. The element-wise deviation bound (4.5) is obtained by choosing $u = e_i$, $v = e_j$.

Let us now prove (4.6). Recall (4.2); the following decomposition holds

$$\frac{2}{N(2\ell+1)} \sum_{m=-\ell}^{\ell} \sum_{t=p+1}^{n} a_{\ell,m}(t-h) a_{\ell,m;Z}(t)$$

$$= \left[\frac{1}{N(2\ell+1)} \sum_{m=-\ell}^{\ell} \sum_{t=p+1}^{n} \left(a_{\ell,m}(t-h) + a_{\ell,m;Z}(t)\right)^2 - (C_\ell + C_{\ell;Z})\right]$$

$$- \left[\frac{1}{N(2\ell+1)} \sum_{m=-\ell}^{\ell} \sum_{t=p+1}^{n} a_{\ell,m}(t-h)^2 - C_\ell\right]$$

$$- \left[\frac{1}{N(2\ell+1)} \sum_{m=-\ell}^{\ell} \sum_{t=p}^{T} a_{\ell,m;Z}(t)^2 - C_{\ell;Z}\right].$$

Implementing (4.3) for $v = e_h$, we have

$$\mathbb{P}\left(\left|\frac{1}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p+1}^{n}a_{\ell,m}(t-h)^2 - C_\ell\right| > 2\pi\mathcal{M}(\widetilde{f_\ell},1)\eta\right) \le 2e^{-c\,N(2\ell+1)\min\{\eta^2,\eta\}},$$

which implies

$$\mathbb{P}\left(\left|\frac{1}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p+1}^{n}a_{\ell,m}(t-h)^2 - C_\ell\right| > 2\pi\mathcal{M}(f_\ell)\eta\right) \le 2e^{-c\,N(2\ell+1)\min\{\eta^2,\eta\}},$$

where we used the fact that $\mathcal{M}(\widetilde{f_\ell}, 1) = \mathcal{M}(f_\ell)$. Following steps that are analogous to the ones that led to (4.3) (setting $v \in \mathbb{R}$, $v = 1$, and obviously $r = 1$), one can show that

$$\mathbb{P}\left(\left|\frac{1}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p}^{T}a_{\ell,m;Z}(t)^2 - C_{\ell;Z}\right| > 2\pi\mathcal{M}(f_{\ell;Z})\eta\right) \le 2e^{-c\,N(2\ell+1)\min\{\eta^2,\eta\}},$$

and that, for any fixed $h = 1, \ldots, p$,

$$\mathbb{P}\left(\left|\frac{1}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p+1}^{n}\left(a_{\ell,m}(t-h)+a_{\ell,m;Z}(t)\right)^2 - (C_\ell + C_{\ell;Z})\right| > 2\pi\mathcal{M}(f_{\ell;T+Z})\eta\right)$$
$$\le 2e^{-c\,N(2\ell+1)\min\{\eta^2,\eta\}},$$

Moreover,

$$f_{\ell;T+Z}(\lambda) = f_\ell(\lambda) + f_{\ell;Z}(\lambda) + 2f_{\ell;(T,Z)}(\lambda),$$

which implies, $\mathcal{M}(f_{\ell;T+Z}) \le \mathcal{M}(f_\ell) + \mathcal{M}(f_{\ell;Z}) + \mathcal{M}(f_{\ell;(T,Z)})$, where $f_{\ell;T+Z}(\lambda)$ is the spectral density of the process $\{a_{\ell,m}(t-h) + a_{\ell,m;Z}(t),\, t \in \mathbb{Z}\}$ and $f_{\ell;(T,Z)}$ is the spectral density of the joint process $\{\left(a_{\ell,m}(t-h), a_{\ell,m;Z}(t)\right),\, t \in \mathbb{Z}\}$.

Now, using the obvious implications

$$\{|X_1 + X_2 + X_3| > a\} \subset \{|X_1| + |X_2| + |X_3| > a\} \subset \bigcup_{i=1}^{3}\{|X_i| > \frac{a}{3}\},$$

we have

$$\mathbb{P}\left(\left|\frac{2}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p+1}^{n}a_{\ell,m}(t-h)a_{\ell,m;Z}(t)\right| > 2\pi\left(\mathcal{M}(f_\ell) + \mathcal{M}(f_{\ell;Z}) + \mathcal{M}(f_{\ell;(T,Z)})\right)\eta\right)$$
$$\le 6e^{-c\,N(2\ell+1)\min\{\eta^2,\eta\}},$$

Following the last steps of the proof of Proposition 2.4 in [2], we obtain

$$2\pi\mathcal{M}(f_\ell) \le \frac{C_{\ell;Z}}{\mu_{\min;\ell}}, \quad 2\pi\mathcal{M}(f_{\ell;Z}) = C_{\ell;Z} \quad \text{and} \quad 2\pi\mathcal{M}\left(f_{\ell;(T,Z)}\right) \le \frac{C_{\ell;Z}\,\mu_{\max;\ell}}{\mu_{\min;\ell}},$$

which finally implies (4.6). □

Proof of Proposition 4.4. Since $\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}}$ is the solution of the minimization problem (3.3), it follows that

$$\frac{\left\|\mathbf{Y}_{\ell,N} - X_{\ell,N}\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}}\right\|_2^2}{N(2\ell+1)} + \lambda\left\|\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}}\right\|_1 \le \frac{\left\|\mathbf{Y}_{\ell,N} - X_{\ell,N}\boldsymbol{\phi}_\ell\right\|_2^2}{N(2\ell+1)} + \lambda\left\|\boldsymbol{\phi}_\ell\right\|_1,$$

which, using the definition of $\mathbf{Y}_{\ell,N}$, becomes

$$
\frac{\left\| X_{\ell,N}\boldsymbol{\phi}_\ell + \mathbf{E}_{\ell,N} - X_{\ell,N}\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \right\|_2^2}{N(2\ell+1)} + \lambda \left\| \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \right\|_1 \leq \frac{\left\| \mathbf{E}_{\ell,N} \right\|_2^2}{N(2\ell+1)} + \lambda \left\| \boldsymbol{\phi}_\ell \right\|_1 .
$$

Now, we have

$$
\frac{\left\| X_{\ell,N}\boldsymbol{\phi}_\ell - X_{\ell,N}\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \right\|_2^2}{N(2\ell+1)} + \frac{\left\| \mathbf{E}_{\ell,N} \right\|_2^2}{N(2\ell+1)} - 2\frac{\left(\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell\right)' X_{\ell,N}'\mathbf{E}_{\ell,N}}{N(2\ell+1)} + \lambda \left\| \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \right\|_1
$$

$$
\leq \frac{\left\| \mathbf{E}_{\ell,N} \right\|_2^2}{N(2\ell+1)} + \lambda \left\| \boldsymbol{\phi}_\ell \right\|_1 ,
$$

and finally, using the notation v_ℓ, it holds that

$$
\frac{v_\ell' X_{\ell,N}' X_{\ell,N} v_\ell}{N(2\ell+1)} - 2\frac{v_\ell' X_{\ell,N}'\mathbf{E}_{\ell,N}}{N(2\ell+1)} + \lambda \left\| \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} \right\|_1 \leq \lambda \left\| \boldsymbol{\phi}_\ell \right\|_1 . \quad \square
$$

Proof of Proposition 4.6. First of all, we have

$$
\left\| \widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell \right\|_\infty = \frac{1}{N(2\ell+1)} \left\| X_{\ell,N}'\mathbf{E}_{\ell,N} \right\|_\infty = \max_{1\leq h\leq p} \left| \frac{\mathbf{Y}_{\ell,N}'(h)\mathbf{E}_{\ell,N}}{N(2\ell+1)} \right| .
$$

Now, using (4.6), we obtain that, for any $\eta \geq 0$ and $c > 0$,

$$
\mathbb{P}\left(\left| \frac{\mathbf{Y}_{\ell,N}'(h)\mathbf{E}_{\ell,N}}{N(2\ell+1)} \right| > C_{\ell;Z}\left(1 + \frac{1+\mu_{max;\ell}}{\mu_{\min;\ell}} \right)\eta \right) \leq 6\exp\left(-c\, N(2\ell+1)\, \min\{\eta, \eta^2\} \right) .
$$

Thus it follows that

$$
\mathbb{P}\left(\max_{1\leq h\leq p} \left| \frac{\mathbf{Y}_{\ell,N}'(h)\mathbf{E}_{\ell,N}}{N(2\ell+1)} \right| > C_{\ell;Z}\left(1 + \frac{1+\mu_{max;\ell}}{\mu_{\min;\ell}} \right)\eta \right)
$$
$$
\leq 6\, p\, \exp\left(-c\, N(2\ell+1)\, \min\{\eta, \eta^2\} \right) .
$$

Since, for every $\ell < L_N$,

$$
c_0 C_{\ell;Z}\left(1 + \frac{1+\mu_{max;\ell}}{\mu_{\min;\ell}} \right) \leq \mathcal{F}_N ,
$$

we have

$$
\mathbb{P}\left(\max_{\ell<L_N}\max_{1\leq h\leq p} \left| \frac{\mathbf{Y}_{\ell,N}'(h)\mathbf{E}_{\ell,N}}{N(2\ell+1)} \right| > \frac{1}{c_0}\mathcal{F}_N\eta \right) \leq 6\, p\, L_N\, \exp\left(-c\, N\, \min\{\eta, \eta^2\} \right) .
$$

Hence, for $\eta = c_0 \sqrt{\frac{\log pL_N}{N}}$

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1} \left\|\widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell\right\|_\infty \le \mathcal{F}_N \sqrt{\frac{\log pL_N}{N}}\right)$$

$$= \mathbb{P}\left(\max_{\ell < L_N} \left\|\widehat{\gamma}_{\ell,N} - \widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell\right\|_\infty \le \mathcal{F}_N \sqrt{\frac{\log pL_N}{N}}\right)$$

$$= \mathbb{P}\left(\max_{\ell < L_N} \max_{1 \le h \le p} \left|\frac{\mathbf{Y}_{\ell,N}(h)'\mathbf{E}_{\ell,N}}{N(2\ell+1)}\right| \le \mathcal{F}_N \sqrt{\frac{\log pL_N}{N}}\right)$$

$$\ge 1 - 6\,p\,L_N \exp\left(-c\,N\,\min\{c_0, c_0^2\}\frac{\log pL_N}{N}\right)$$

$$= 1 - 6\,p\,L_N\,e^{-c\,\min\{c_0, c_0^2\}\log pL_N}$$

$$= 1 - 6\,e^{-\left(c\,\min\{c_0, c_0^2\}-1\right)\log pL_N},$$

and the statement is proved with $c_1 = 6$, $c_2 = c\,\min\{c_0, c_0^2\} - 1$, where c_0 is any positive constant that satisfies $c_2 > 0$. $\square$

Proof of Proposition 4.8. Our goal is to prove (4.10), which can be rewritten as follows

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1} A_\ell\right) \ge 1 - c_1\,e^{-c_2\,N\,\min\{\omega^{-2}, 1\}},$$

where $A_\ell = \{v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell \ge \alpha_\ell \|v_\ell\|_2^2 - \tau_\ell \|v_\ell\|_1^2, \ \forall\, v_\ell \in \mathbb{R}^p\}$.

We start from Eq. (4.3), and considering that

$$2\pi \mathcal{M}(\widetilde{f}_\ell, r) \le 2\pi \mathcal{M}(\widetilde{f}_\ell) \le p\,2\pi \mathcal{M}(f_\ell) \le p\,\frac{C_{\ell;Z}}{\mu_{\min;\ell}}, \quad r \ge 1,$$

we have

$$\mathbb{P}\left(\left|v_\ell'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)v_\ell\right| > \frac{p\,C_{\ell;Z}}{\mu_{\min;\ell}}\eta\right) \le 2e^{-c\,N(2\ell+1)\min\{\eta^2, \eta\}}.$$

Using Lemma F.2 in the supplementary material of [2] yields

$$\mathbb{P}\left(\sup_{v_\ell \in \mathcal{K}(2s)} \left|v_\ell'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)v_\ell\right| > \eta\,\frac{p\,C_{\ell;Z}}{\mu_{\min;\ell}}\right) \le 2e^{-cN(2\ell+1)\min\{\eta^2, \eta\}+2s\,\min\left\{\log p, \log\left(\frac{21\,e\,p}{2s}\right)\right\}},$$

where $\mathcal{K}(2s) = \{v \in \mathbb{R}^p : \|v\|_2 \le 1, \|v\|_0 \le 2s\}$, for an integer $s \ge 1$. Now we set

$$c_3 = 54 \quad \eta = \omega_\ell^{-1} = \frac{1}{54\,p}\frac{\mu_{\min;\ell}}{\mu_{\max;\ell}},$$

to obtain

$$\mathbb{P}\left(\sup_{v_\ell \in \mathcal{K}(2s)} \left|v_\ell'\left(\widehat{\Gamma}_{\ell,N} - \Gamma_\ell\right)v_\ell\right| \le \frac{1}{54}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\right)$$

$$\ge 1 - 2e^{-cN(2\ell+1)\min\{1, \omega_\ell^{-2}\}+2s\,\min\left\{\log p, \log\left(\frac{21\,e\,p}{2s}\right)\right\}},$$

and apply Lemma 12 of the supplementary material of [32] to get

$$\mathbb{P}\left(\left|v_\ell'\left(\widehat{\Gamma}_{\ell,N}-\Gamma_\ell\right)v_\ell\right| \le \frac{1}{2}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\{\|v_\ell\|_2^2+\frac{1}{s}\|v_\ell\|_1^2\},\ \forall v_\ell\in\mathbb{R}^p\right)$$
$$\ge 1-2e^{-cN(2\ell+1)\min\{1,\omega_\ell^{-2}\}+2s\min\{\log p,\log\left(\frac{21\,e\,p}{2s}\right)\}}.$$

Moreover, it holds that

$$\left|v_\ell'\left(\widehat{\Gamma}_{\ell,N}-\Gamma_\ell\right)v_\ell\right|=\left|v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell-v_\ell'\Gamma_\ell v_\ell\right|=\left|v_\ell'\Gamma_\ell v_\ell-v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\right|\ge\left|v_\ell'\Gamma_\ell v_\ell\right|$$
$$-\left|v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\right|$$
$$\ge\Lambda_{\min}\left(\Gamma_\ell\right)\|v_\ell\|_2^2-\left|v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\right|\ge\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_2^2-\left|v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\right|$$
$$=\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_2^2-v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell,$$

which implies that

$$\mathbb{P}\left(\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_2^2-v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\le\frac{1}{2}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\{\|v_\ell\|_2^2+\frac{1}{s}\|v_\ell\|_1^2\},\ \forall v_\ell\in\mathbb{R}^p\right)$$
$$\ge 1-2e^{-cN(2\ell+1)\min\{1,\omega_\ell^{-2}\}+2s\min\{\log p,\log\left(\frac{21\,e\,p}{2s}\right)\}}.$$

Hence, we can rearrange the terms in the previous relation to have

$$\mathbb{P}\left(v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\ge\frac{1}{2}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_2^2-\frac{1}{2s}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_1^2,\ \forall v_\ell\in\mathbb{R}^p\right)$$
$$\ge 1-2e^{-cN(2\ell+1)\min\{1,\omega_\ell^{-2}\}+2s\min\{\log p,\log\left(\frac{21\,e\,p}{2s}\right)\}}.$$

Now, taking $\omega_N=\max_{\ell<L_N}\omega_\ell$, we obtain

$$\mathbb{P}\left(\bigcup_{\ell=0}^{L_N-1}\overline{A}_\ell\right)\le\sum_{\ell=0}^{L_N-1}[1-\mathbb{P}(A_\ell)]$$
$$\le 2e^{-cN\min\{1,\omega_N^{-2}\}+2s\min\{\log pL_N,\log\left(\frac{21\,e\,pL_N}{2s}\right)\}},$$

which is the desired conclusion, once we set

$$s=\frac{c\,N\,\min\{\omega_N^{-2},1\}}{4\log pL_N}.\quad\square$$

Proof of Theorem 4.12. Set $v_\ell=\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}}-\boldsymbol{\phi}_\ell$. Using the basic inequality (4.7) and the deviation condition (4.12), we obtain that, almost surely,

$$v_\ell'\widehat{\Gamma}_{\ell,N}v_\ell\le 2v_\ell'\left(\widehat{\gamma}_{\ell,N}-\widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell\right)+\lambda_N\left(\|\boldsymbol{\phi}_\ell\|_1-\|\boldsymbol{\phi}_\ell+v_\ell\|_1\right)$$
$$\le 2\|v_\ell\|_1\|\widehat{\gamma}_{\ell,N}-\widehat{\Gamma}_{\ell,N}\boldsymbol{\phi}_\ell\|_\infty+\lambda_N\left(\|\boldsymbol{\phi}_\ell\|_1-\|\boldsymbol{\phi}_\ell+v_\ell\|_1\right)$$
$$\le 2\|v_\ell\|_1\mathcal{F}_N\sqrt{\frac{\log(pL_N)}{N}}+\lambda_N\left(\|\boldsymbol{\phi}_\ell\|_1-\|\boldsymbol{\phi}_\ell+v_\ell\|_1\right). \tag{6.1}$$

Now, let $J=\text{supp}(\boldsymbol{\phi}_\ell)=\{j_1,\ldots,j_{q_\ell}\}$ be such that $|J|=q_\ell$, then $J^c=\{1,\ldots,p\}\setminus J$, $\|\boldsymbol{\phi}_{\ell,J}\|_1=\|\boldsymbol{\phi}_\ell\|_1$ and $\|\boldsymbol{\phi}_{\ell,J^c}\|_1=0$. Consequently, it holds that

$$\|\boldsymbol{\phi}_\ell+v_\ell\|_1=\|\boldsymbol{\phi}_{\ell,J}+v_{\ell,J}\|_1+\|v_{\ell,J^c}\|_1$$
$$\ge\|\boldsymbol{\phi}_{\ell,J}\|_1-\|v_{\ell,J}\|_1+\|v_{\ell,J^c}\|_1,$$

which implies

$$\lambda_N \left(\left\| \boldsymbol{\phi}_\ell \right\|_1 - \left\| \boldsymbol{\phi}_\ell + v_\ell \right\|_1 \right) \leq \lambda_N \left(\left\| \boldsymbol{\phi}_{\ell,J} \right\|_1 - \left\| \boldsymbol{\phi}_{\ell,J} \right\|_1 + \left\| v_{\ell,J} \right\|_1 - \left\| v_{\ell,J^c} \right\|_1 \right)$$

$$\leq \lambda_N \left(\left\| v_{\ell,J} \right\|_1 - \left\| v_{\ell,J^c} \right\|_1 \right).$$

Having explicitly required that $\lambda_N \geq 4\mathcal{F}_N \sqrt{\log(pL_N)/N}$, Eq. (6.1) becomes

$$0 \leq v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell \leq \frac{\lambda_N}{2} \left\| v_\ell \right\|_1 + \lambda_N \left(\left\| v_{\ell,J} \right\|_1 - \left\| v_{\ell,J^c} \right\|_1 \right)$$

$$= \frac{\lambda_N}{2} \left(\left\| v_{\ell,J} \right\|_1 + \left\| v_{\ell,J^c} \right\|_1 \right) + \lambda_N \left(\left\| v_{\ell,J} \right\|_1 - \left\| v_{\ell,J^c} \right\|_1 \right)$$

$$= \frac{3\lambda_N}{2} \left\| v_{\ell,J} \right\|_1 - \frac{\lambda_N}{2} \left\| v_{\ell,J^c} \right\|_1 \leq \frac{3}{2} \lambda_N \left\| v_\ell \right\|_1. \tag{6.2}$$

This ensures that $\left\| v_{\ell,J^c} \right\|_1 \leq 3 \left\| v_{\ell,J} \right\|_1$ and hence, adding $\left\| v_{\ell,J} \right\|_1$ on both sides, that $\| v_\ell \|_1 \leq 4 \left\| v_{\ell,J} \right\|_1$, which implies that

$$\| v_\ell \|_1 \leq 4 \sqrt{q_\ell} \, \| v_\ell \|_2,$$

from Cauchy–Schwarz inequality.

Now we use this property into the (RE) inequality (4.8), keeping in mind that we specifically required that $q_\ell \tau_\ell \leq \alpha_\ell/32$, and we obtain

$$v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell \geq \alpha_\ell \| v_\ell \|_2^2 - \tau_\ell \| v_\ell \|_1^2 \geq \alpha_\ell \| v_\ell \|_2^2 - 16 q_\ell \tau_\ell \| v_\ell \|_2^2$$

$$\geq \alpha_\ell \| v_\ell \|_2^2 - \frac{\alpha_\ell}{2} \| v_\ell \|_2^2 \geq \frac{\alpha_\ell}{2} \| v_\ell \|_2^2. \tag{6.3}$$

Hence, combining Eqs. (6.2) and (6.3), we get

$$\frac{\alpha_\ell}{2} \| v_\ell \|_2^2 \leq v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell \leq \frac{3}{2} \lambda_N \| v_\ell \|_1 \leq 6 \sqrt{q_\ell} \, \lambda_N \| v_\ell \|_2,$$

which results in the following estimate for the norm of the error

$$\frac{\alpha_\ell}{3} \| v_\ell \|_2^2 \leq \lambda_N \| v_\ell \|_1 \leq \lambda_N 4 \sqrt{q_\ell} \, \| v_\ell \|_2.$$

As a consequence,

$$\| v_\ell \|_2 \leq 12 \sqrt{q_\ell} \, \frac{\lambda_N}{\alpha_\ell}, \tag{6.4}$$

$$\| v_\ell \|_1 \leq 4 \sqrt{q_\ell} \, \| v_\ell \|_2 \leq 48 \, q_\ell \, \frac{\lambda_N}{\alpha_\ell},$$

$$v_\ell' \widehat{\Gamma}_{\ell,N} v_\ell \leq \frac{3}{2} \lambda_N \| v_\ell \|_1 \leq 72 \, \frac{\lambda_N^2}{\alpha_\ell}.$$

It is readily seen that

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^2}^2 = \left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k}_N \right\|_{L^2}^2 + \left\| \mathbf{k} - \mathbf{k}_N \right\|_{L^2}^2$$

$$= \sum_{\ell=0}^{L_N-1} \sum_{\ell'=0}^{L_N-1} \left\langle \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell, \widehat{\boldsymbol{\phi}}_{\ell',N}^{\text{lasso}} - \boldsymbol{\phi}_{\ell'} \right\rangle \int_{-1}^1 \frac{2\ell+1}{4\pi} P_\ell(z) \frac{2\ell'+1}{4\pi} P_{\ell'}(z) \mathrm{d}z$$

$$+ \sum_{\ell \geq L_N} \sum_{\ell' \geq L_N} \langle \boldsymbol{\phi}_\ell, \boldsymbol{\phi}_{\ell'} \rangle \int_{-1}^{1} \frac{2\ell + 1}{4\pi} P_\ell(z) \frac{2\ell' + 1}{4\pi} P_{\ell'}(z) \mathrm{d}z$$

$$= \sum_{\ell=0}^{L_N - 1} \left\| \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell \right\|_2^2 \frac{2\ell + 1}{8\pi^2} + \sum_{\ell \geq L_N} \|\boldsymbol{\phi}_\ell\|_2^2 \frac{2\ell + 1}{8\pi^2},$$

by the orthonormality of Legendre polynomials (see (2.1)). Moreover, under the additional Condition 1.7,

$$\left\| \mathbf{k} - \mathbf{k}_N \right\|_{L^\infty} \leq \sup_{u \in [-1,1]} \sum_{\ell \geq L_N} \|\phi_\ell\|_2 \frac{2\ell + 1}{4\pi} |P_\ell(u)| \leq \sum_{\ell \geq L_N} \|\phi_\ell\|_2 \frac{2\ell + 1}{4\pi} < \infty,$$

since $|P_\ell(u)| \leq 1$, for all $u \in [-1, 1]$. Then, using triangle inequality,

$$\left\| \widehat{\mathbf{k}}_N^{\text{lasso}} - \mathbf{k} \right\|_{L^\infty} \leq \sum_{\ell=0}^{L_N - 1} \left\| \widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell \right\|_2 \frac{2\ell + 1}{4\pi} + \sum_{\ell \geq L_N} \|\phi_\ell\|_2 \frac{2\ell + 1}{4\pi}.$$

Recalling that $\widehat{\boldsymbol{\phi}}_{\ell,N}^{\text{lasso}} - \boldsymbol{\phi}_\ell = v_\ell$, by (6.4), the proof is concluded. $\quad\square$

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] P. Baldi, G. Kerkyacharian, D. Marinucci, D. Picard, Adaptive density estimation for directional data using needlets, Ann. Statist. 37 (6A) (2009) 3362–3395.

[2] S. Basu, G. Michailidis, Regularized estimation in sparse high-dimensional time series models, Ann. Statist. 43 (4) (2015) 1535–1567.

[3] J. Beran, Y. Feng, S. Ghosh, R. Kulik, Spatial and Space–Time Processes, Springer, 2013, pp. 753–769.

[4] C. Berg, E. Porcu, From Schoenberg coefficients to Schoenberg functions, Constr. Approx. 45 (2017) 217–241.

[5] M. Bevilacqua, C. Caamaño-Carrillo, E. Porcu, Unifying compactly supported and matern covariance functions in spatial statistics, 2020, arXiv:2008.02904.

[6] M. Bevilacqua, T. Faouzi, R. Furrer, E. Porcu, Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics, Ann. Statist. 47 (2) (2019) 828–856.

[7] N.H. Bingham, T.L. Symons, Gaussian random fields on the sphere and sphere cross line, Stochastic Process. Appl. (2019+) (in press).

[8] D. Bosq, Linear Processes in Function Spaces. Theory and Applications, Springer-Verlag, 2000.

[9] S. Bourguin, C. Durastanti, D. Marinucci, G. Peccati, Gaussian approximations of nonlinear statistics on the sphere, J. Math. Anal. Appl. 436 (2016) 1121–1148.

[10] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer, 1991.

[11] P. Bühlmann, S. van de Geer, Statistics for High Dimensional Data, Springer, 2011.

[12] V. Cammarota, D. Marinucci, The stochastic properties of ℓ^1-regularized spherical gaussian fields, Appl. Comput. Harmon. Anal. 38 (2) (2015) 262–283.

[13] A. Caponera, Statistical Inference for Spherical Functional Autoregressions (Ph.D. thesis), Sapienza University of Rome, 2020.

[14] A. Caponera, SPHARMA approximations for stationary functional time series on the sphere, 2020, arXiv:20 09.13189.

[15] A. Caponera, D. Marinucci, Asymptotics for spherical functional autoregressions, Ann. Statist. 49 (2021) 346–369.

[16] J. Clarke De la Cerda, A. Alegría, E. Porcu, Regularity properties and simulations of gaussian random fields on the sphere cross time, Electron. J. Stat. 12 (2018) 399–426.

[17] G. Christakos, Random Field Models in Earth Sciences, Elsevier, 2005.

[18] R.A. Davis, P. Zang, T. Zheng, Sparse vector autoregressive modeling, J. Comput. Graph. Statist. 25 (4) (2016) 1077–1096.

[19] C. Durastanti, X. Lan, D. Marinucci, Needlet-whittle estimates on the unit sphere, Electron. J. Stat. 7 (2013) 597–646.

[20] C. Durastanti, X. Lan, D. Marinucci, Gaussian semiparametric estimates on the unit sphere, Bernoulli 20 (2014) 28–77.

[21] C. Durastanti, D. Marinucci, G. Peccati, Normal approximations for wavelet coefficients on spherical Poisson fields, J. Math. Anal. Appl. 409 (2014) 212–227.

[22] M. Fan, D. Paul, T.C.M. Lee, T. Matsuo, A multi-resolution model for non-gaussian random fields on a sphere with application to ionospheric electrostatic potentials, Ann. Appl. Stat. 1 (2018).

[23] Q.T.L. Gia, I.H. Sloan, R.S. Womersley, Y.G. Wang, Sparse isotropic regularization for spherical harmonic representations of random fields on the sphere, Appl. Comput. Harmon. Anal. 49 (2020) 257–278.

[24] T. Gneiting, Nonseparable, stationary covariance functions for space–time data, J. Amer. Statist. Assoc. 97 (2002) 590–600.

[25] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009.

[26] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, 2015.

[27] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, Wiley & Sons, 2015.

[28] R.H. Jones, Y. Zhang, Models for continuous stationary space–time processes, in: Modelling Longitudinal and Spatially Correlated Data, Springer, 1997, pp. 289–298.

[29] M. Jun, Matérn-based nonstationary cross-covariance models for global processes, J. Multivariate Anal. 128 (2014) 134–146.

[30] P.T. Kim, J.-Y. Koo, Optimal spherical deconvolution, J. Multivariate Anal. 80 (2002) 21–42.

[31] A. Lang, C. Schwab, Isotropic gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations, Ann. Appl. Probab. 25 (2015) 3047–3094.

[32] P.-L. Loh, M.J. Wainwright, High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity, Ann. Statist. 40 (3) (2012) 1637–1664.

[33] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer, 2005.

[34] A. Malyarenko, Invariant random fields in vector bundles and applications to cosmology, Ann. Inst. H. Poincaré Probab. Stat. 47 (2011) 1068–1095.

[35] D. Marinucci, G. Peccati, Random Fields on the Sphere: Representation, Limit Theorems and Cosmological Applications, in: London Mathematical Society Lecture Note Series, Cambridge University Press, 2011.

[36] J.D. McEwen, Y. Wiaux, A novel sampling theorem on the sphere, IEEE Trans. Signal Process. 59 (2011) 5876–5887.

[37] E. Porcu, M. Bevilacqua, M.G. Genton, Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere, J. Amer. Statist. Assoc. 111 (2016) 888–898.

[38] E. Porcu, R. Furrer, D. Nychka, 30 years of space time covariance functions, in: WIREs Computational Statistics, 2020, e1512.

[39] B.L.S. Prakasa Rao, Nonparametric Functional Estimation, Academic Press, 1983.

[40] S. Song, P.J. Bickel, Large vector auto regressions, 2011, preprint.

[41] M.L. Stein, On a class of space–time intrinsic random functions, Bernoulli 19 (2013).

[42] E.M. Stein, G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces., Princeton University Press, 1971.

[43] J.E. Taylor, K.J. Worsley, Detecting sparse signals in random fields, with an application to brain mapping, J. Amer. Statist. Assoc. 102 (2007) 913–928.

[44] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.

[45] A. Tsybakov, Introduction to Nonparametric Estimation, Springer, 2009.

[46] N.J. Vilenkin, A.U. Klimyk, Representation of Lie Groups and Special Functions, Kluwer, 1991.

[47] M.J. Wainwright, High-Dimensional Statistics: A Non-Asymptotic Viewpoint, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2019.

[48] M.I. Yadrenko, Spectral Theory of Random Fields, Optimization Software Inc., 1983.