SAPIENZA
UNIVERSITÀ DI ROMA

Statistical Inference for Spherical Functional Autoregressions

Scuola di Scienze Statistiche

Dottorato di Ricerca in Statistica Metodologica – XXXII Ciclo

Candidate

Alessia Caponera
ID number 1478089

Thesis Advisor

Prof. Pierpaolo Brutti

Co-Advisor

Prof. Domenico Marinucci

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics

October 2019

Thesis defended on 17 February 2020
in front of a Board of Examiners composed by:

Prof. Roberto Rocci (chairman)

Prof. Bruno Toaldo

Prof. Simone Vantini

Statistical Inference for Spherical Functional Autoregressions

Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: February 17, 2020

Author's email: alessia.caponera@uniroma1.it

A mia madre,
che ha preso in mano la sua vita
e l'ha trascinata in salvo

Ringraziamenti

Sono felice di essere con te, Samvise Gamgee.
Qui, alla fine di ogni cosa.
— Il Signore degli Anelli

Il ringraziamento più importante va certamente al prof. Domenico Marinucci. Grazie per aver dato fiducia a una dottoranda alle prime armi, che neanche un cumulante sapeva calcolare! Grazie per avermi reso consapevole di amare profondamente questo lavoro. Grazie per avermi trasmesso cosa vuol dire perdersi nella bellezza della matematica. Grazie per tutto ciò che in questi 2 anni e 12 giorni mi hai insegnato.

Un altro ringraziamento speciale va al prof. Pierpaolo Brutti. Per i suggerimenti, le idee, il confronto; per aver creduto in me, anche quando le serie di Fourier mi "spaventavano". Per le bustine di camomilla, il cioccolato di SAID and, last but not least, la crostata ricotta e visciole. Ma soprattutto per la tua amicizia.

Un immenso grazie a Claudio Durastanti ed Anna Vidotto, che hanno contribuito in modo significativo alla stesura di questa tesi. Il vostro aiuto è stato prezioso. Aforementioned *a parte, siamo una bella squadra.*

Sempre in ambito accademico, vorrei ringraziare Max Werner, che mi ha fatto sperimentare per la prima volta il vero significato di "fare ricerca".

Una dedica particolare va al quarto piano *del Dipartimento di Scienze Statistiche, che mi ha accolto per otto lunghi anni. In particolare, a Fulvio, Luca, Stefania e Valeria per il loro affetto e agli studenti di Inferenza che porterò sempre nel cuore. Grazie a tutti i miei colleghi dottorandi e ex, che ho visto passare per le stanze 40 e 41, per aver condiviso con me questi anni; specialmente a Dotto per esser stato il mio primo punto di riferimento e a Nina, vera amica e complice.*

Ringrazio anche il "Gruppo vacanze": Anna, Anna Paola, Claudio, Marco, Maurizia, Nicola e Valentina. Per avermi accolto sin da subito con tanto affetto, senza avermi fatto mai sentire l'ultima arrivata.

E infine il grazie più grande alla mia famiglia. A mio padre che mi ha consigliato di iscrivermi a Statistica. Sono contenta di aver portato a termine ciò che tu *avevi cominciato. A mia madre che, nonostante le preoccupazioni e le difficoltà di questi anni, mi ha sempre sostenuto. Alle mie Gemme, per le quali vorrei poter esser sempre un esempio da seguire. Vi voglio bene.*

Contents

Introduction and Overview

In recent years, a lot of interest has been drawn by the statistical analysis of spherical isotropic random fields. These investigations have been motivated by a wide array of applications arising in many different areas, including, in particular, Cosmology, Astrophysics, Geophysics, Climate and Atmospheric Sciences, and many others, see, e.g., [4, 8, 19, 20, 24, 25, 26, 39, 61, 62]. Most papers in Cosmology and Astrophysics have focused so far on spherical random fields with no temporal dependence; the next generation of cosmological experiments is however going to make this aspect much more relevant. On the other hand, applications in Climate and Atmospheric Sciences, Geophysics, and several other areas have always been naturally modeled in terms of a double-dependence in the spatial and temporal domains. In many works of these fields, the attention has been focused on the definition of wide classes of space-time covariance functions, and then on the derivation of likelihood functions; the literature on these themes is vast and we make no attempt to a complete list of references, see, for instance, [8, 20, 26, 33, 62] and the references therein.
The main purpose of this thesis is to address some foundational questions regarding time-dependent spherical random fields and, then, to investigate some new estimation procedures for the class of *spherical functional autoregressions*.
More precisely, we start in Chapter 1 with a review of a number of recent developments in probability and mathematical statistics, which will be instrumental for the derivation of our results in the chapters to follow. Indeed, after recalling some basic facts on stationary time series and their spectral representation (see [12]), we review some less standard materials on harmonic expansion of spherical random fields (see [43]). We introduce here the orthonormal system of *spherical harmonics* and we discuss some properties of the corresponding family of random coefficients. In this same chapter we review also some recent, very powerful techniques to establish *quantitative central limit theorems* (the so-called Stein-Malliavin approach, [49]) and the basic tools of operator theory for functional data analysis (see [32]). All these instruments will be deeply exploited in the chapters to follow.
In Chapter 2, we investigate harmonic properties of time-varying spherical random fields. This chapter builds on earlier results [56, 55], see also [8], and it is principally divided into two main parts. The first section provides some results on a double spectral representation, with respect to both the temporal and spatial components of the field, while the other section focuses on properties of a specific class of sphere-cross-time random fields which can be also interpreted as functional autoregressive processes (see [10]) taking values in $L^2(\mathbb{S}^2)$.
In Chapter 3, we present our first estimation procedure for functional spherical autoregressions, see also [18]. In particular, we exploit isotropy assumptions and

the so-called *duplication property* of the spherical harmonics basis to derive a more transparent representation of these autoregressions; hence we transform a nonparametric kernel estimation problem into the investigation of a growing sequence of spectral parameters. We are then able to investigate three kinds of asymptotics: we first establish a general consistency result (with rates) in L^2 and L^∞ norms; we then prove a quantitative central limit theorem for our nonparametric estimator, and finally, under stronger smoothness conditions, we prove a weak convergence result (finite dimensional distributions and tightness) for the kernel estimators. These results are also validated by a small numerical experiment.

Chapter 4 is developed under the same framework as Chapter 3, but imposing some further *sparsity* constraints. In particular, we are assuming that only a limited range of *multipoles* (i.e., spatial frequencies) are actually relevant in the functional autoregressions, and then we implement a convex regularization procedure of LASSO (Least Absolute Shrinkage and Selection Operator) type, similarly to a recent important contribution by [7]. We are then able to establish concentration results, holding with probability arbitrarily close to one, in L^2 and L^∞ norms.

The numerical results given in this thesis are obtained by means of the python package `healpy`, based on the so-called HEALPix software (see [27]). While this software is now very popular and standard in Cosmology and Astrophysics, its use in the mathematical and statistical community has so far been rather limited; because of this, in Appendix A, we provide a quick guide explaining its basic operating principles.

Chapter 1

Background

1.1 Spectral Analysis of Stationary Time Series

In this section we recall some basic notions of time series analysis, such as the concepts of stationarity and the popular Spectral Representation Theorem, together with some examples. Even if already widely known, they will be constantly used throughout the thesis; so we report them for completeness and better readability. For a in-depth discussion on these topics, see, for example, [12].

The fundamental idea of time series analysis is to model the data obtained from observations collected sequentially over time as a realization (or part of a realization) of a stochastic process $\{X_t,\ t \in \mathcal{T}\}$ defined on a complete probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, where $\mathcal{T}$ is generally a subset of $\mathbb{R}$. In particular, we will focus on the specific case $\mathcal{T} = \mathbb{Z}$. We shall frequently use the term time series to mean both the data and the process of which it is a realization.

1.1.1 Stationary Time Series

We start with the two standard definitions of stationarity and second-order properties of stationary processes.

Definition 1.1.1 (see, e.g., [12]). *The process $\{X_t,\ t \in \mathbb{Z}\}$ is said to be strongly stationary if the joint distributions of $(X_{t_1}, \dots, X_{t_k})'$ and $(X_{t_1+h}, \dots, X_{t_k+h})'$ are the same for all positive integers k and for all $t_1, \dots, t_k, h \in \mathbb{Z}$.*

We consider only finite variance processes, for which the autocovariance function $\gamma(t,s) = \mathrm{Cov}[X_t, X_s]$ is well-defined.

Definition 1.1.2 (see, e.g., [12]). *The process $\{X_t,\ t \in \mathbb{Z}\}$, is said to be stationary if*

(i) $\mathbb{E}|X_t|^2 < \infty$, *for all* $t \in \mathbb{Z}$,

(ii) $\mathbb{E}[X_t] = m$, *for all* $t \in \mathbb{Z}$,

(iii) $\gamma(t,s) = \gamma(t+h, s+h)$, *for all* $t, s, h \in \mathbb{Z}$.

If $\{X_t\}$ is strongly stationary it immediately follows, on taking $k = 1$ in Definition 1.1.1, that X_t has the same distribution for each $t \in \mathbb{Z}$. If $\mathbb{E}|X_t|^2 < \infty$, this implies in particular that $\mathbb{E}[X_t]$ and $\mathbb{V}[X_t]$ are both constant. Moreover, taking $k = 2$ in Definition 1.1.1, we find that X_{t+h} and X_t have the same joint distribution and hence the same covariance for all $h \in \mathbb{Z}$. Thus a strongly stationary process with finite second moments is stationary. The converse of the previous statement is not true; however, in the Gaussian case weak stationarity implies strong stationarity (see, for instance, [12].
It is worth to notice that in both cases $\gamma(t, s) = \gamma(t - s, 0),\ t, s \in \mathbb{Z}$. It is therefore convenient to redefine the autocovariance function of a stationary process as the function of just one variable,

$$\gamma(h) \equiv \gamma(h, 0) = \mathrm{Cov}[X_{t+h}, X_t], \qquad h, t \in \mathbb{Z},$$

that is, the autocovariance function of $\{X_t\}$ at *lag* h. Clearly, $\gamma(0) \geq 0$; moreover, $|\gamma(h)| \leq \gamma(0)$, as a consequence of Cauchy-Schwartz inequality, $\gamma(\cdot)$ is an even function, i.e $\gamma(h) = \gamma(-h),\ h \in \mathbb{Z}$. As every autocovariance function, it is also positive definite, namely, $\sum_{i,j=1}^{n} a_i \gamma(i - j) a_j \geq 0$ for all positive integers n and for all vectors $(a_1, \ldots, a_n)' \in \mathbb{R}^n$.
The last two properties relate to a standard characterization of autocovariance functions for stationary time series (see [12, Theorem 4.1.1]). Moreover, Herglotz's theorem, which we recall here, characterizes them as those functions that have a frequency domain representation with respect to a finite Radon measure.

Theorem 1.1.3 (Herglotz's Theorem, [12])**.** *A complex-valued function $\gamma(\cdot)$ defined on the integers is positive definite if and only if*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda), \qquad \textit{for all } h \in \mathbb{Z},$$

where $F(\cdot)$ is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ and $F(-\lambda) = 0$.

Here, the so-called spectral distribution function $F(\cdot)$ (with $F(-\pi) = 0$) is uniquely determined by $\gamma(\cdot)$. Furthermore, it is well-known that if $\{X_t\}$ is a stationary time series whose autocovariance function $\gamma(\cdot)$ satisfies $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the related spectral density exists and it is defined as

$$f(\lambda) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma(h), \qquad \pi \leq \lambda \leq \pi.$$

A very wide class of stationary processes is provided by the so-called autoregressive-moving average processes, which will play a key role throughout the next chapters. We recall that an autoregressive-moving average process of order (p, q) (or ARMA(p, q)) can be generated by using a white noise $\{Z_t, t \in \mathbb{Z}\}$ (for a formal definition, see [12, Chapter 3]) as the forcing term in a set of linear difference equations of the form

$$\phi(B) X_t = \theta(B) Z_t, \tag{1.1.1}$$

where $\phi(\cdot)$ and $\theta(\cdot)$ are the p-th and q-th degree polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

and B is the backward shift operator defined by $B^j X_t = X_{t-j},\ t \in \mathbb{Z}$.
Particular cases of such a process can be obtained by letting one of the two polynomials constant and equal to 1. For instance, if $\phi(z) \equiv 1$, we obtain a moving-average process of order q (or MA(q)) described by

$$X_t = \theta(B) Z_t; \tag{1.1.2}$$

whereas, If $\theta(z) \equiv 1$, then we have the so-called autoregressive process of order p (or AR(p)) with

$$\phi(B) X_t = Z_t, \tag{1.1.3}$$

For a MA(q) process, it is quite clear that the difference equations (1.1.2) has a unique stationary solution; instead, in the last case (as in the general case) the existence and uniqueness of a stationary solution of (1.1.3) needs closer investigation. In the next few lines, we summarily discuss the conditions to ensure the existence of a unique stationary solution of the ARMA equations; this will be useful in Chapter 2, specifically Section 2.2, to have an insight into the proof of the same results generalized to a particular class of functional autoregressive processes.
First recall the definition of a causal ARMA process; see also in [12, Chapter 3].

Definition 1.1.4. *An* ARMA(p, q) *process defined by the equations* $\phi(B)X_t = \theta(B)Z_t,\ t \in \mathbb{Z}$, *is said to be causal if there exists a sequence of constants* $\{\psi_j\}$ *such that* $\sum_{j=0}^{\infty} |\psi_j| < \infty$ *and*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \qquad t \in \mathbb{Z}, \tag{1.1.4}$$

where the series converges in mean square and absolutely with probability one.

It should be noted that causality is a property not of the process $\{X_t\}$ alone but rather of the relationship between the two processes $\{X_t\}$ and $\{Z_t\}$ appearing in the defining ARMA equations.

Theorem 1.1.5. *Let* $\{X_t.\ t \in \mathbb{Z}\}$ *be an* ARMA(p, q) *process for which the polynomials* $\phi(\cdot)$ *and* $\theta(\cdot)$ *have no common zeroes. Then* $\{X_t\}$ *is causal if and only if* $\phi(z) \neq 0$ *for all* $z \in \mathbb{C}$ *such that* $|z| \leq 1$. *The coefficients* $\{\psi_j\}$ *in* (1.1.4) *are determined by the relation*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z), \qquad |z| \leq 1. \tag{1.1.5}$$

The proof is given in [12, Section 3.1], whereas the numerical calculation of the coefficients $\{\psi_j\}$ is discussed [12, Section 3.3].

Remark 1.1.6. *The first part of the proof of Theorem 1.1.5 shows that if* $\{X_t\}$ *is a stationary solution of the ARMA equations with* $\phi(z) \neq 0$ *for* $|z| \leq 1$, *then we must have* $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, *where* $\{\psi_j\}$ *is defined by* (1.1.5). *Conversely, if* $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, *then* $\phi(B)X_t = \phi(B)\psi(B)Z_t = \theta(B)Z_t$. *Thus the process* $\{\psi(B)Z_t\}$ *is the* unique *stationary solution of the ARMA equations if* $\phi(z) \neq 0$ *for* $|z| \leq 1$.

1.1.2 The Spectral Representation Theorem on $\mathbb{Z}$

Now, we report the statement of the Spectral Representation Theorem for stationary processes and the sketch of the proof.

Theorem 1.1.7 (The Spectral Representation Theorem on $\mathbb{Z}$, see [12])**.** *If $\{X_t,\ t \in \mathbb{Z}\}$ is a stationary sequence with mean zero and spectral distribution function $F(\cdot)$, then there exists an orthogonal-increment process $\{Z(\lambda),\ -\pi \leq \lambda \leq \pi\}$ such that*

$$\mathbb{E}|Z(\lambda) - Z(-\pi)|^2 = F(\lambda), \qquad -\pi \leq \lambda \leq \pi,$$

and

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda), \qquad \text{with probability one.} \tag{1.1.6}$$

The right-hand side is a stochastic integral with respect to an orthogonal-increment process, a precise definition of which is given, for instance, in [12, Chapter 4].
The proof of the representation (1.1.6) will be achieved by defining a certain isomorphism $\mathcal{I}$ between the closed subspaces $\overline{\text{span}\{X_t,\ t \in \mathbb{Z}\}}$ of $L^2(\Omega, \mathbb{P})$ and $\overline{\text{span}\{\exp(it\cdot),\ t \in \mathbb{Z}\}}$ of $L^2([-\pi,\pi], dF)$. This isomorphism, defined as a linear extension of the mapping $X_t \mapsto e^{it\cdot}$, will provide a link between random variables in the "time domain" and functions on $[-\pi,\pi]$ in the "frequency domain". In particular,

$$\langle X_t, X_s \rangle_{L^2(\Omega,\mathbb{P})} := \mathbb{E}[X_t X_s] = \int_{-\pi}^{\pi} e^{it\lambda} e^{-is\lambda} dF(\lambda),$$

and one also finds that

$$X_t = \mathcal{I}^{-1}(e^{it\cdot}) = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda),$$

taking $Z(\lambda) = \mathcal{I}^{-1}(\mathbb{1}_{[-\pi,\lambda)})$. Note in particular that the linear space of functions $\exp(it\cdot)$ is dense in $L^2([-\pi,\pi], dF)$ and the indicator function of the subsets of $[-\pi,\pi]$ always belongs to $L^2([-\pi,\pi], dF)$ if $F(\cdot)$ is the distribution function of a nonatomic measure on $[-\pi,\pi]$. The complete proof can be found, for example, in [12, Chapter 4].

Example 1.1.8. *Under the same assumptions of Theorem 1.1.5, a causal* ARMA(p,q) *process has spectral density*

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \qquad \pi \leq \lambda \leq \pi,$$

and the following spectral representation holds

$$X_t = \int_{-\pi}^{\pi} e^{it\lambda} \frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})} dW(\lambda),$$

where the orthogonal-increment process $\{W(\lambda),\ -\pi \leq \lambda \leq \pi\}$ is such that

$$Z_t = \int_{-\pi}^{\pi} e^{it\lambda} dW(\lambda)$$

(see, e.g., [12, Theorem 4.4.2 and Theorem 4.10.1]).

1.2 Harmonic Analysis on the Sphere

This section includes some well-established results concerning harmonic analysis on the two-dimensional sphere $\mathbb{S}^2$, with special attention to isotropic random fields defined on it for which a Spectral Representation Theorem holds. The reader is referred for further details to [2, 43] and the references therein. These concepts lay the foundation for our analysis on spherical random fields which show a temporal dependence, the main focus of this dissertation.

Let $L^2(\mathbb{S}^2) := L^2(\mathbb{S}^2, dx)$ be the space of square-integrable complex-valued functions over the unit sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$ [1] with respect to the uniform Lebesgue measure, since now on denoted by dx. When endowed with the inner product

$$\langle f, g\rangle_{L^2(\mathbb{S}^2)} = \int_{\mathbb{S}^2} f(x)\overline{g(x)}dx,$$

$L^2(\mathbb{S}^2)$ is a separable Hilbert space; the Spectral Theorem for compact self-adjoint operators (see Section 1.4) then entails that $L^2(\mathbb{S}^2)$ can be decomposed into the direct sum of orthogonal spaces spanned by eigenfunctions of the corresponding Laplacian. More precisely,

$$L^2(\mathbb{S}^2) = \bigoplus_{\ell=0}^{\infty} \mathcal{Y}_\ell,$$

where $\mathcal{Y}_\ell$ is spanned by the eigenfunctions of the spherical Laplacian associated with the eigenvalue $-\ell(\ell+1)$. These eigenfunctions are called spherical harmonics and they satisfy

$$\Delta_{\mathbb{S}^2} f_\ell = -\ell(\ell+1) f_\ell, \qquad \ell \geq 0,$$

where, as usual,

$$\Delta_{\mathbb{S}^2} := \frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta}\frac{\partial}{\partial\varphi^2}.$$

A standard orthonormal basis for the eigenspace $\mathcal{Y}_\ell$ is chosen to be $\{Y_{\ell,m},\ m = -\ell, \dots, \ell\}$, and then any $f \in L^2(\mathbb{S}^2)$ admits a Fourier series representation of the form

$$f(x) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell,m}(x), \qquad \text{in the } L^2\text{-sense}.$$

An explicit expression for $Y_{\ell,m} : \mathbb{S}^2 \to \mathbb{C}$ can be given, with some abuse of notation, in spherical coordinates by (see [43, page 64])

$$Y_{\ell,m}(\vartheta,\varphi) = \begin{cases} \sqrt{\dfrac{2\ell+1}{4\pi}\dfrac{(\ell-m)!}{(\ell+m)!}} P_{\ell,m}(\cos\vartheta)\exp(im\varphi) & m \geq 0, \\ (-1)^m \overline{Y_{\ell,-m}}(\vartheta,\varphi) & m < 0, \end{cases}$$

where $\vartheta \in [0,\pi]$ and $\varphi \in [0, 2\pi)$ are the colatitude and longitude respectively, and $P_{\ell,m} : [-1,1] \to \mathbb{R}$ is the associated Legendre function of degree ℓ and order m, which is defined in terms of the ℓ-th Legendre polynomial $P_\ell : [-1,1] \to \mathbb{R}$, i.e.,

$$P_{\ell,m}(u) = (-1)^m (1-u^2)^{m/2} \frac{d^m}{du^m} P_\ell(u),$$

[1] We shall use $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ to denote the Euclidean norm and inner product on $\mathbb{R}^3$ respectively.

$$P_\ell(u)=\frac{1}{2^\ell \ell!}\frac{d^\ell}{du^\ell}(u^2-1)^\ell,$$

for $u\in[-1,1]$ (see also [1, Chapter 8]). Here, we display the first few:

$$Y_{0,0}(\vartheta,\varphi)=\sqrt{\frac{1}{4\pi}}$$
$$Y_{1,0}(\vartheta,\varphi)=\sqrt{\frac{3}{4\pi}}\cos\vartheta$$
$$Y_{1,1}(\vartheta,\varphi)=\sqrt{\frac{3}{8\pi}}\sin\vartheta e^{i\varphi}$$
$$Y_{2,0}(\vartheta,\varphi)=\sqrt{\frac{5}{16\pi}}(3\cos^2\vartheta-1)$$
$$Y_{2,1}(\vartheta,\varphi)=\sqrt{\frac{15}{8\pi}}\sin\vartheta\cos\vartheta e^{i\varphi}$$
$$Y_{2,2}(\vartheta,\varphi)=\sqrt{\frac{15}{32\pi}}\sin^2\vartheta e^{i2\varphi}.$$

Remark 1.2.1. *Legendre polynomials are orthogonal over $[-1,1]$, i.e.,*

$$\int_{-1}^{-1}P_\ell(u)P_{\ell'}(u)du=\frac{2}{2\ell+1}\delta_\ell^{\ell'},$$

here δ_a^b is the Kronecker delta function (see [67]). Moreover, the sequence of Legendre polynomials forms an orthogonal basis for the L^2 space of real-valued functions over $[-1,1]$. Note that $P_\ell(1)=1$, for all $\ell\geq 0$.

One of the most fundamental property of spherical harmonics is expressed by the following *addition formula*: for any $x,y\in\mathbb{S}^2$,

$$\sum_{m=-\ell}^{\ell}Y_{\ell,m}(x)\overline{Y_{\ell,m}}(y)=\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle); \tag{1.2.1}$$

for a proof (based on group representation theory) we refer to [43, Chapter 3]. As a consequence, a *duplication property* is satisfied, i.e.,

$$\int_{\mathbb{S}^2}\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle)\frac{2\ell+1}{4\pi}P_{\ell'}(\langle y,z\rangle)dy=\frac{2\ell+1}{4\pi}P_\ell(\langle x,z\rangle)\delta_\ell^{\ell'}. \tag{1.2.2}$$

Remark 1.2.2. *We recall that an orthonormal basis for $L^2(\mathbb{S}^2)$ restricted to real-valued functions can be defined by setting*

$$\widetilde{Y}_{\ell,m}(\vartheta,\varphi)=\begin{cases}\sqrt{2}(-1)^m\operatorname{Re}Y_{\ell,m}(\vartheta,\varphi) & m>0,\\ Y_{\ell,0}(\vartheta,\varphi) & m=0,\\ \sqrt{2}(-1)^m\operatorname{Im}Y_{\ell,|m|}(\vartheta,\varphi) & m<0.\end{cases}$$

These functions have the same orthonormality properties as the complex ones above, and also the addition formula (1.2.1) holds. Note that the central *spherical harmonics*

$$Y_{\ell,0}(\vartheta,\varphi)=\sqrt{\frac{2\ell+1}{4\pi}}P_\ell(\cos\vartheta),\qquad \ell\geq 0,$$

are real-valued functions.

The real basis of spherical harmonics is common in the literature; examples are given in [36, 74, 75]. Throughout this thesis, we will use both the complex and real systems, with the same notation. However, when not specified, we will refer to the former.

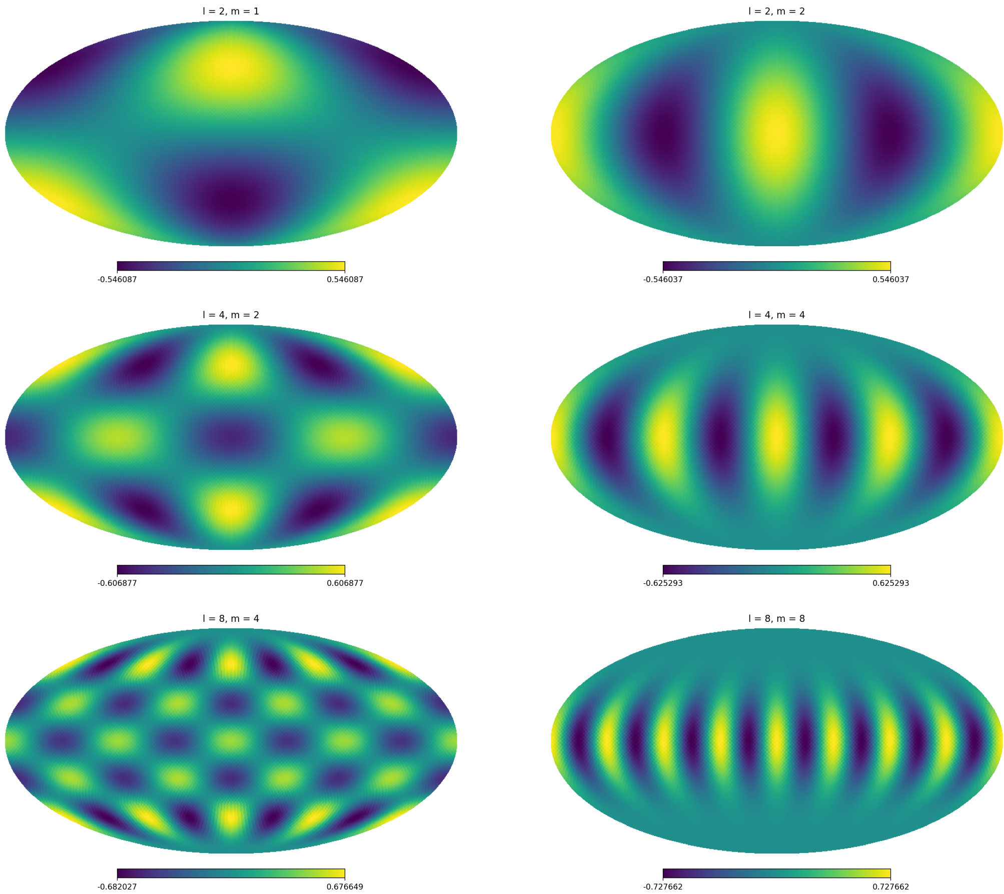


Figure 1.1. Examples of real spherical harmonics, for $\ell = 2, 4, 8,\ m = \ell/2, \ell$.

1.2.1 Isotropic Random Fields on the Sphere

We now consider a random field $\{T(x),\ x \in \mathbb{S}^2\}$ defined on the sphere. This means that $T : \mathbb{S}^2 \times \Omega \to \mathbb{R}$ is a $\mathfrak{B}(\mathbb{S}^2) \times \mathfrak{F}$-measurable mapping, for some appropriate probability space $(\Omega, \mathfrak{F}, \mathbb{P})$; $\mathfrak{B}(\mathbb{S}^2)$ denoting the Borel σ-field on the sphere.
Through this subsection we meanly refer to [43]. In parallel to Section 1.1, we recall the definition of strong and weak isotropy and related second-order properties.

Definition 1.2.3 (see, e.g., [6, 43])**.** *The spherical random field $\{T(x),\ x \in \mathbb{S}^2\}$ is said to be strongly isotropic if, for every $k \in \mathbb{N}$, every $x_1, \ldots, x_k \in \mathbb{S}^2$ and every $g \in SO(3)$ (the special group of rotations in $\mathbb{R}^3$) we have that $(T(x_1), \ldots, T(x_k))'$ and $(T(gx_1), \ldots, T(gx_k))'$ have the same joint distribution.*

For spherical random fields in $L^2(\Omega) := L^2(\Omega, \mathbb{P})$, i.e., with finite variance and therefore a well-defined covariance function $\Gamma(x, y) := \mathrm{Cov}[T(x), T(y)]$, we have the following definition:

Definition 1.2.4 (see, e.g., [43]). *The spherical random field* $\{T(x),\ x \in \mathbb{S}^2\}$ *is said to be isotropic if*

(i) $\mathbb{E}|T(x)|^2 < \infty$*, for all* $x \in \mathbb{S}^2$*,*

(ii) $\mathbb{E}[T(x)] = const$*, for all* $x \in \mathbb{S}^2$*,*

(iii) $\Gamma(x,y) = \Gamma(gx, gy)$*, for all* $x, y \in \mathbb{S}^2,\ g \in SO(3)$*.*

Remark 1.2.5. *As for stationarity, a strongly isotropic field with finite second moment is also (weakly) isotropic and, if the spherical field is Gaussian, the two notions are equivalent.*

Without loss of generality, from now on, we will assume $\mathbb{E}[T(x)] = 0$, uniformly.

Remark 1.2.6. *Note that if* $\{T(x),\ x \in \mathbb{S}^2\}$ *is isotropic, one has that*

$$\mathbb{E}\left[\int_{\mathbb{S}^2} |T(x)|^2 dx\right] = 4\pi \mathbb{E}|T(x_0)|^2 < \infty,$$

for any fixed $x_0 \in \mathbb{S}^2$*. This implies that there exists a* $\mathfrak{F}$*-measurable set* Ω' *of* $\mathbb{P}$*-probability 1 such that, for every* $\omega \in \Omega'$*,* $T(\cdot, \omega)$ *is an element of* $L^2(\mathbb{S}^2)$*. In addition, as pointed out in [44], the field can be shown to be mean-square continuous, meaning that*

$$\lim_{x \to x_0} \mathbb{E}[T(x) - T(x_0)]^2 = 0, \qquad \forall x_0 \in \mathbb{S}^2.$$

By isotropy, we have clearly that

$$\Gamma(x, y) = \Gamma(x', y'),$$

for all pairs $\{(x, y), (x', y')\}$ such that $\langle x, y\rangle = \langle x', y'\rangle$. Hence, the covariance is really a function of the spherical geodesic distance between x and y, i.e.,

$$d_{\mathbb{S}^2}(x, y) = \arccos\langle x, y\rangle;$$

moreover, it is a positive definite continuous function. With some abuse of notation, we will denote with Γ also the "reduced" version, defined on $[-1, 1]$.
I. J. Schoenberg in his seminal paper [64] characterized the class of positive definite continuous function on $[-1, 1]$; see also [8] which offers a natural extension of Schoenberg's Theorem to the product space $[-1, 1] \times G$, for G a locally compact group. The result can be seen as the analogue of the Herglotz's theorem (see Theorem 1.1.3) for stationary time series.

Theorem 1.2.7 (Schoenberg (1942), [64]). *Assume* $\Gamma : [-1, 1] \to \mathbb{R}$ *is a positive definite continuous function. Then there exists a sequence of nonnegative weights* $\{C_\ell,\ \ell \geq 0\}$ *such that for all* $z \in [-1, 1]$ *we have*

$$\Gamma(z) = \sum_{\ell=0}^{\infty} C_\ell \frac{2\ell + 1}{4\pi} P_\ell(z).$$

With a slight abuse of notation, we can also write

$$\Gamma(x, y) = \Gamma(\langle x, y\rangle) = \sum_{\ell=0}^{\infty} C_\ell \frac{2\ell + 1}{4\pi} P_\ell(\langle x, y\rangle). \tag{1.2.3}$$

1.2.2 The Spectral Representation Theorem on $\mathbb{S}^2$

In this subsection we see the Spectral Representation Theorem for isotropic random fields on the sphere: in [43] is given as a special case of the Stochastic Peter-Weyl Theorem; however, we report a different proof (see also [42, 69]), which does not involve the Group Representation Theory and make clearer the affinity with the time series case, since it is based on the construction of a linear isometry between $\overline{\text{span}\,\{T(x),\ x \in \mathbb{S}^2\}}$ and a closed subspace of $L^2(\mathbb{S}^2)$.

Theorem 1.2.8 (The Spectral Representation Theorem on $\mathbb{S}^2$, see [42, 43])**.** *Let $\{T(x),\ x \in \mathbb{S}^2\}$ be a centred isotropic random field. Then, for every $x \in \mathbb{S}^2$,*

$$T(x) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m}(x), \tag{1.2.4}$$

in the $L^2(\Omega)$ sense, that is, for every $x \in \mathbb{S}^2$,

$$\lim_{L\to\infty} \mathbb{E}\left[\left(T(x) - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m}(x)\right)^2\right] = 0.$$

The random coefficients $\{a_{\ell,m},\ \ell \geq 0,\ m = -\ell, \dots, \ell\}$ satisfy

$$\mathbb{E}[a_{\ell,m}\overline{a_{\ell',m'}}] = C_\ell \delta_\ell^{\ell'} \delta_m^{m'}.$$

Remark 1.2.9. *The decomposition (1.2.4) is also shown to hold in the sense of $L^2(\mathbb{S}^2 \times \Omega) := L^2(\mathbb{S}^2 \times \Omega, dx \otimes \mathbb{P})$, that is,*

$$\lim_{L\to\infty} \mathbb{E}\left\| T - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m} \right\|_{L^2(\mathbb{S}^2)}^2 = 0.$$

Indeed, for each $\omega \in \Omega'$ (see Remark 1.2.6),

$$\begin{aligned}\int_{\mathbb{S}^2} \left| T(x,\omega) - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} a_{\ell,m}(\omega) Y_{\ell,m}(x) \right|^2 dx &= \sum_{\ell=L+1}^{\infty} \sum_{m=-\ell}^{\ell} |a_{\ell,m}(\omega)|^2 \\ &\leq \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} |a_{\ell,m}(\omega)|^2 \\ &= \int_{\mathbb{S}^2} |T(x,\omega)|^2 \, dx.\end{aligned}$$

Then, by dominated convergence,

$$\mathbb{E}\left\| T - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m} \right\|^2 \to 0, \qquad L \to \infty. \tag{1.2.5}$$

Proof. Let $\mathbb{T}_0$ be the complex linear space spanned by all finite linear combinations of the $T(x)$'s,

$$\mathbb{T}_0 := \left\{ \sum_{j=1}^{n} c_j T(x_j) : \; n \in \mathbb{N}, \; c_j \in \mathbb{C}, \; x_j \in \mathbb{S}^2 \right\} \subset L^2(\Omega), \tag{1.2.6}$$

and $\mathbb{T}$ be the closure of $\mathbb{T}_0$ in $L^2(\Omega)$. Now, define the linear operator $\mathcal{J}$ by linear extension of the mapping

$$T(x) \mapsto \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle);$$

note that $\mathcal{J} : \mathbb{M}_0 \to L^2(\mathbb{S}^2)$ since

$$\begin{aligned}
&\left\| \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle) \right\|^2_{L^2(\mathbb{S}^2)} \\
&= \int_{\mathbb{S}^2} \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, y \rangle) \sum_{\ell'=0}^{\infty} \frac{2\ell'+1}{4\pi} \sqrt{C_{\ell'}} P_{\ell'}(\langle x, y \rangle) dy \\
&= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell < \infty,
\end{aligned}$$

where for the last equality we have used the duplication property (1.2.2) and the fact that $P_\ell(1) = 1$. Then, we have

$$\begin{aligned}
&\langle \mathcal{J}T(x), \mathcal{J}T(y) \rangle_{L^2(\mathbb{S}^2)} \\
&= \left\langle \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle), \sum_{\ell'=0}^{\infty} \frac{2\ell'+1}{4\pi} \sqrt{C_{\ell'}} P_{\ell'}(\langle y, \cdot \rangle) \right\rangle_{L^2(\mathbb{S}^2)} \\
&= \int_{\mathbb{S}^2} \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, z \rangle) \sum_{\ell'=0}^{\infty} \frac{2\ell'+1}{4\pi} \sqrt{C_{\ell'}} P_{\ell'}(\langle y, z \rangle) \, dz \\
&= \sum_{\ell\ell'} \sqrt{C_\ell} \sqrt{C_{\ell'}} \int_{\mathbb{S}^2} \frac{2\ell+1}{4\pi} P_\ell(\langle x, z \rangle) \frac{2\ell'+1}{4\pi} P_{\ell'}(\langle z, y \rangle) \, dz \\
&= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell P_\ell(\langle x, y \rangle),
\end{aligned}$$

again by the duplication property; thus,

$$\langle \mathcal{J}T(x), \mathcal{J}T(y) \rangle_{L^2(\mathbb{S}^2)} = \mathbb{E}[T(x)T(y)],$$

and, hence, $\mathcal{J}$ is well defined and it is a linear isometry. We extend its domain to $\mathbb{T}$, so that the image space of the extension is made by the closure of the span of all functions which have the form

$$\sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle). \tag{1.2.7}$$

Now, we define $a_{\ell,m} := \int_{\mathbb{S}^2} T(x)\overline{Y_{\ell,m}}(x)\, dx$ and we verify that it is well defined as an element of $\mathbb{T}$. To this aim, we consider the sequence

$$a_{\ell,m}(j) = \sum_{k\in N_j} T(x_{jk})\overline{Y_{\ell,m}}(x_{jk})\mu(V_{jk}),$$

where $\{V_{jk}\}$ is a family of (exhaustive and disjoint) Voronoi cells, so that there exist spherical caps and constants $0 < c < c'$ such that (see, e.g., [5])

$$B_{\epsilon/2}(x_{jk}) \subset V_{jk} \subset B_{\epsilon}(x_{jk}),\ c2^{-j} \leq \epsilon \leq c'2^{-j},\ \text{for all } j,k.$$

By considering refining partitions $\{V_{j'k}\}$, we have that

$$\begin{aligned}
&\mathbb{E}|a_{\ell,m}(j) - a_{\ell,m}(j')|^2\\
=&\mathbb{E}\left[\sum_{k\in N_j}\left\{T(x_{jk})\overline{Y_{\ell,m}}(x_{jk}) - T(x_{j'k})\overline{Y_{\ell,m}}(x_{j'k})\right\}\mu(V_{jk})\right]^2\\
=&\mathbb{E}\left[\sum_{k\in N_j}\left\{\left(T(x_{jk}) - T(x_{j'k})\right)\overline{Y_{\ell,m}}(x_{jk}) + T(x_{j'k})\left(\overline{Y_{\ell,m}}(x_{jk}) - \overline{Y_{\ell,m}}(x_{j'k})\right)\right\}\mu(V_{jk})\right]^2\\
\leq&2\sum_{k\in N_j}\sum_{k'\in N_j}\mathbb{E}\left[\left(T(x_{jk}) - T(x_{j'k})\right)\left(T(x_{jk'}) - T(x_{j'k'})\right)\overline{Y_{\ell,m}}(x_{jk})\overline{Y_{\ell,m}}(x_{jk'})\mu(V_{jk})\mu(V_{jk'})\right]\\
+&2\sum_{k\in N_j}\sum_{k'\in N_j}\mathbb{E}\left[T(x_{j'k})T(x_{j'k'})\left(\overline{Y_{\ell,m}}(x_{jk}) - \overline{Y_{\ell,m}}(x_{j'k})\right)\left(\overline{Y_{\ell,m}}(x_{jk'}) - \overline{Y_{\ell,m}}(x_{j'k'})\right)\mu(V_{jk})\mu(V_{jk'})\right].
\end{aligned}$$

The first summand is bounded by

$$\begin{aligned}
&2\sup_k \mathbb{E}\left|T(x_{jk}) - T(x_{j'k})\right|^2\left\{\sum_{k\in N_j}\overline{Y_{\ell,m}}(x_{jk})\mu(V_{jk})\right\}^2\\
&\leq 2\sup_k \mathbb{E}|T(x_{jk}) - T(x_{jk})|^2 \sup_k\left|\overline{Y_{\ell,m}}(x_{jk})\right|^2\left\{\sum_{k\in N_j}\mu(V_{jk})\right\}^2.
\end{aligned}$$

Since the V_{jk}'s form a partition of $\mathbb{S}^2$, the sum $\sum_{k\in N_j}\mu(V_{jk})$ is smaller than the surface of the sphere and so it is finite. It follows from the mean-square continuity of $\{T(x),\ x\in s\}$ that the first term converges to zero as $j\to\infty$; whereas, the second is bounded by

$$\begin{aligned}
&2\mathbb{E}|T(x)|^2\left\{\sum_{k\in N_j}\left(\overline{Y_{\ell,m}}(x_{jk}) - \overline{Y_{\ell,m}}(x_{j'k})\right)\mu(V_{jk})\right\}^2\\
\leq&2\mathbb{E}|T(x)|^2\sup_k\left|\overline{Y_{\ell,m}}(x_{jk}) - \overline{Y_{\ell,m}}(x_{j'k})\right|^2\left\{\sum_{k\in N_j}\mu(V_{jk})\right\}^2,
\end{aligned}$$

which converges to zero. Thus, $\{a_{\ell m}(j)\}$ is a Cauchy sequence, as claimed.
Note that

$$\mathcal{J}a_{\ell,m} = \mathcal{J}\left[\int_{\mathbb{S}^2} T(x)\overline{Y_{\ell,m}}(x)dx\right] = \int_{\mathbb{S}^2}\mathcal{J}T(x)\,\overline{Y_{\ell,m}}(x)dx$$

$$= \int_{\mathbb{S}^2} \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle) \overline{Y_{\ell,m}}(x) dx; \tag{1.2.8}$$

then by the addition formula (1.2.1) and orthogonality property $\int_{\mathbb{S}^2} Y_{\ell',m'}(y)\overline{Y_{\ell,m}}(y) = \delta_\ell^{\ell'} \delta_m^{m'}$, (1.2.8) is equal to $\sqrt{C_\ell}\, \overline{Y_{\ell,m}}$. Therefore, by the isometry property,

$$\mathbb{E}[a_{\ell,m}\overline{a_{\ell'm'}}] = \left\langle \sqrt{C_\ell}\, \overline{Y_{\ell,m}}, \sqrt{C_\ell}\, \overline{Y_{\ell',m'}} \right\rangle_{L^2(\mathbb{S}^2)} = C_\ell \delta_\ell^{\ell'} \delta_m^{m'}.$$

Moreover, since $\mathcal{J}$ is a linear isometry, it is injective and its inverse (restricted to image space) $\mathcal{J}^{-1}$ is well defined. It follows immediately that

$$\begin{aligned} T(x) &= \mathcal{J}^{-1}\left[\sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} \sqrt{C_\ell} P_\ell(\langle x, \cdot \rangle)\right] = \mathcal{J}^{-1}\left[\sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell,m}(x)\overline{Y_{\ell,m}}(\cdot)\sqrt{C_\ell}\right] \\ &= \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell,m}(x)\mathcal{J}^{-1}\left[\sqrt{C_\ell}\, \overline{Y_{\ell,m}}(\cdot)\right] = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell,m}(x) a_{\ell,m}, \end{aligned}$$

as claimed.

□

Remark 1.2.10. *There are two important differences with time series analysis (see also [42]):*

- *The Spectral Representation Theorem on $\mathbb{S}^2$ involves a series rather than an integral: the reason is that $\mathbb{S}^2$ is compact, while $\mathbb{Z}$ is not. This is closely related to group-theoretic results, in particular to the fact that representations of compact groups are countable, while representations of noncompact groups, such as $\mathbb{Z}$, are uncountably many. Note that the sphere by itself is not a group, but it can be realized as quotient space $\mathbb{S}^2 = SO(3)/SO(2)$, see [43] for more details.*

- *In the time series case there is a single deterministic component $\exp(it\lambda)$ corresponding to each frequency λ, while in the spherical case there are $2\ell+1$ spherical harmonics corresponding to a single multipole ℓ (i.e., spatial frequency). Again, this has a simple explanation in terms of groups: indeed, stationary processes on $\mathbb{Z}$ enjoy some form of invariance with respect to the action of a commutative group ($X_t \to X_{t+h}$), while isotropy implies invariance in distribution with respect to the action of the noncommutative group $SO(3)$. It is a standard result of group representation theory that noncommutative groups have multiple representations of the same dimension, and this leads to the $2\ell+1$ spherical harmonics at the same multipole ℓ. Note that the complex exponentials satisfy $\frac{\partial^2}{\partial t^2}\exp(it\lambda) = -\lambda^2 \exp(it\lambda)$, in perfect analogy with spherical harmonics.*

1.3 Stein-Malliavin Normal Approximations

With Stein-Malliavin approach [49] we refer to an exhaustive theory based on the combination of two probabilistic techniques, namely, the *Malliavin calculus of* *variations* and *Stein's method* for probabilistic approximations, see also [54, 58, 53, 48, 51, 50]. The aim is to provide estimates of the distance between the laws of two random objects, with a focus on normal approximations, as well as the corresponding Central Limit Theorems (CLTs), i.e., convergence results displaying a Gaussian limit. In particular, a crucial role is played by Hermite polynomial and the elements of the so-called *Gaussian Wiener chaos.* We briefly discuss these topics in the next sections to finally end up with the so-called *Fourth Moment Theorem* (FMT) [49, Chapter 5]; this powerful result will be applied in Chapter 3 to processes which exhibit an integral representation as the one in Example 1.1.8, in order to obtain a Quantitative Central Limit Theorem (QCLT), in Wasserstein metric. For a comprehensive treatise, we refer to [49].

The Central Limit Theorem (with all its variants) is by all means among the most important results in probability and statistics, since it allows to apply methods that work for normal distributions to many problems involving other types of distributions. For instance, if Y_n is some statistical estimator with unknown distribution, which satisfies a central limit theorem, then a Gaussian likelihood may be appropriate, or approximated confidence intervals can be computed. However, it is not possible to quantify directly the error one makes when replacing the actual law with its asymptotics, because of the lack of information on the rate of convergence to the limiting Gaussian distribution.

Example 1.3.1 (see, also, [42]). *Take $Z \stackrel{d}{=} \mathcal{N}(0,1)$ and consider the two sequences*

$$Y_n = Z + \frac{Z^2}{\exp(n)}, \quad W_n = Z + \frac{10^3 \times Z^2}{\log\log(n+1)}, \quad n \in \mathbb{N}.$$

While for both sequences we have Y_n, $W_n \xrightarrow{d} \mathcal{N}(0,1)$, for a fixed n their distributions will be completely different and Y_n will be "closer" to Z than W_n.

It is therefore natural to try to extend the Central Limit Theorem by measuring a suitable distance between probability laws, and investigating the rate of convergence to zero of such distance.
Here, we quickly recall the notion of Wasserstein distance, which, as anticipated, will be used in Chapter 3 to produce a quantitative version of the Central Limit Theorem for some functional estimators. See [49] for other examples of probability metrics and properties.

Definition 1.3.2 (Wasserstein Distance). *Fix an integer $d \geq 1$ and let $\mathcal{G}$ be the set of all functions $g : \mathbb{R}^d \to \mathbb{R}$, such that*

$$\|g\|_{\text{Lip}} \leq 1, \qquad \|g\|_{\text{Lip}} := \sup_{\substack{x \neq y \\ x,y \in \mathbb{R}^d}} \frac{|g(x) - g(y)|}{\|x - y\|},$$

with $\|\cdot\|$ the usual Euclidean norm on $\mathbb{R}^d$. Let X, Y be random variables with values in $\mathbb{R}^d$ such that $\mathbb{E}|g(X)| < \infty$, $\mathbb{E}|g(Y)| < \infty$ for every $g \in \mathcal{G}$. The Wasserstein distance between the laws of X and Y is given by the quantity

$$d_W(X,Y) = \sup_{g \in \mathcal{G}} |\mathbb{E}[g(X)] - \mathbb{E}[g(Y)]| . \tag{1.3.1}$$

1.3.1 Hermite Polynomials and Diagram Formulae

We first introduce Hermite polynomials which turn out to be a very useful tool when computing moments and cumulants of Gaussian random variables.

Definition 1.3.3 (see, e.g, [43]). *The sequence of Hermite polynomials* $\{H_q(\cdot),\ q \ge 0\}$ *on* $\mathbb{R}$*, is defined via the following relations:* $H_0(\cdot) \equiv 1$*, and for* $q \ge 1$*,*

$$H_q(x) = (-1)^q e^{x^2/2} \frac{d^q}{dx^q} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Recall that the sequence $\left\{(q!)^{-1/2} H_q(\cdot),\ q \ge 0\right\}$ is an orthonormal basis of the space $L^2(\mathbb{R}, (2\pi)^{-1/2} e^{-x^2/2} dx)$ (see [49, Proposition 1.4.1]) and any finite variance transform of a standard Gaussian random variable X has a representation in terms of Hermite polynomials (see [49, Example 2.2.6]), that is, for F such that $\mathbb{E}[F(X)]^2 < \infty$,

$$F(X) = \sum_{q=0}^{\infty} J_q(F) \frac{H_q(X)}{q!}, \qquad J_q(F) := \mathbb{E}\left[F(X) H_q(X)\right]. \tag{1.3.2}$$

Several relevant other properties can be deduced from the following formula, valid for every $t, x \in \mathbb{R}$,

$$\exp\left(tx - \frac{t^2}{2}\right) = \sum_{q=0}^{\infty} \frac{t^q}{q!} H_q(x);$$

for instance, the recursive formulas

$$\frac{d}{dx} H_q(x) = q H_{q-1}(x), \qquad q \ge 1,$$
$$H_{q+1}(x) = x H_q(x) - q H_{q-1}(x), \qquad q \ge 1,$$

see again [49, Proposition 1.4.1]. Hence, one can easily verify that the few first Hermite polynomials have the following expression $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$ and $H_4(x) = x^4 - 6x^2 + 3$.
The next statement provides a well-known combinatorial description of the moments and cumulants associated with Hermite transformations of (possibly correlated) Gaussian random variables. See [57, Chapters 2 - 4] for a self-contained presentation using integer partitions and Möbius inversion formulae. We refer also to [43] for definition of cumulants and notation.

Proposition 1.3.4 (Diagram Formulae for Hermite Polynomials, [43]). *Let* $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ *be a centred Gaussian vector, with* $\gamma_{ij} = \mathbb{E}[Z_i Z_j]$*,* $i, j \in \{1, \dots, p\}$*. Let* $H_{l_1}, \dots, H_{l_p}$ *be Hermite polynomials of degrees* $l_1, \dots, l_p$ (≥ 1) *respectively. Let* $\Gamma_{\overline{F}}(l_1, \dots, l_p)$ *(resp.* $\Gamma_{\overline{F}}^{C}(l_1, \dots, l_p)$*) be the collection of all diagrams with no flat edges (resp. connected diagrams with no flat edges) of order* $l_1, \dots, l_p$*. Then,*

$$\mathbb{E}\left[\prod_{j=1}^{p} H_{l_j}(Z_j)\right] = \sum_{G \in \Gamma_{\overline{F}}(l_1, \dots, l_p)} \prod_{i=1}^{p} \prod_{j=i+1}^{p} \gamma_{ij}^{\eta_{ij}(G)}, \tag{1.3.3}$$

$$\mathrm{Cum}[H_{l_1}, \dots, H_{l_p}] = \sum_{G \in \Gamma_{\overline{F}}^{C}(l_1, \dots, l_p)} \prod_{i=1}^{p} \prod_{j=i+1}^{p} \gamma_{ij}^{\eta_{ij}(G)}, \tag{1.3.4}$$

where, for each diagram G, $\eta_{ij}(G)$ is the exact number of edges connecting one vertex of the i-th row to one vertex of the j-th row of the diagram G.

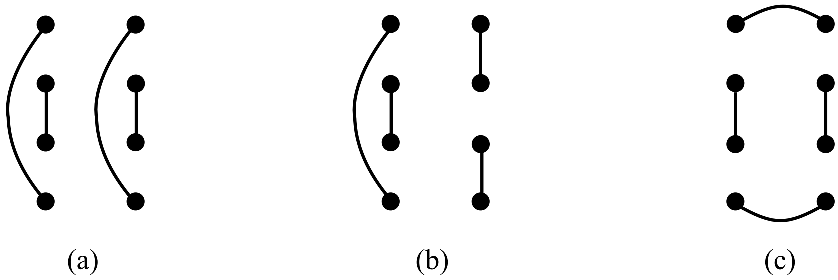


Figure 1.2. Example of three different diagrams: (a) non-flat and not connected diagram; (b) non-flat and connected diagram; (c) not connected diagram with two flat edges.

The Diagram Formulae will be often applied in Chapter 3, specifically to compute fourth moments and cumulants of the examined estimators.

1.3.2 Wiener Chaos and Quantitative Central Limit Theorem

The notion of *Wiener Chaos* plays a role analogous to that of the Hermite polynomials $\{H_q(\cdot),\ q \geq 0\}$ for the one-dimensional Gaussian distribution; again, we fully refer to [49] for more details.
Let $\mathbb{H}$ be a real separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$.

Definition 1.3.5 (see, e.g., Section 2.1 in [49])**.** *An isonormal Gaussian process over $\mathbb{H}$ is a collection $X = \{X(f),\ f \in \mathbb{H}\}$ of jointly centred Gaussian random variables defined on some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, such that $\mathbb{E}[X(f)X(g)] = \langle f, g \rangle_{\mathbb{H}}$ for every $f, g \in \mathbb{H}$.*

In this section, we shall assume that $\mathscr{F}$ is generated by X. Let us now introduce the Wiener chaoses. For each $q \geq 0$, we define $\mathcal{H}_q$ be the closure in $L^2(\Omega)$ of the linear subspace generated by random variables of the form $H_q(X(f)),\ f \in \mathbb{H},\ \|f\|_{\mathbb{H}} = 1$. The space $\mathcal{H}_q$ is the q-th Wiener chaos of X; in particular, $\mathcal{H}_0 = \mathbb{R}$ and $\mathcal{H}_1 = X$.
One has the crucial result

$$L^2(\Omega) = \bigoplus_{q=0}^{\infty} \mathcal{H}_q,$$

see [49, Theorem 2.2.4], which generalizes to an infinite-dimensional setting the case (1.3.2); this means that every random variable $F \in L^2(\Omega)$ admits a unique expansion of the type

$$F = \sum_{q=0}^{\infty} J_q(F), \tag{1.3.5}$$

where $J_q : L^2(\Omega) \to \mathcal{H}_q$ is the orthogonal projection operator; note that $J_0 = \mathbb{E}[F]$.

Now, consider the measure space $(A, \mathfrak{A}, \mu)$, where the set A equipped with the associated Borel σ-field $\mathfrak{A}$ is a Polish space[2], and the measure μ is positive, σ-finite and non-atomic; then, take $\mathbb{H}$ equal to $L^2(A, \mathfrak{A}, \mu)$ with the standard inner product $\langle f, g\rangle_{\mathbb{H}} = \int_A f(a)g(a)d\mu(a)$. For every $f \in \mathbb{H}$, the isonormal Gaussian process

$$X(f) = \int_A f(a)dW(a) \tag{1.3.6}$$

is defined as the Wiener-Itô integral of f with respect to the Gaussian family $W = \{W(B): \; B \in \mathfrak{A}, \; \mu(B) < \infty\}$ such that, for every $B, C \in \mathfrak{A}$ of finite μ-measure, $\mathbb{E}[W(B)W(C)] = \mu(B \cap C)$, see [49, Example 2.1.4].

Remark 1.3.6. *Compare Equation (1.3.6) with the spectral representations in Example 1.1.8. Both $\{X_t\}$ and $\{Z_t\}$ are defined with respect to the same random measure $W(\cdot)$.*

Let $\mathbb{H}^{\otimes q}$ and $\mathbb{H}^{\odot q}$ be the q-th tensor product and the q-th symmetric tensor product of $\mathbb{H}$, respectively; namely, $\mathbb{H}^{\otimes q} = L^2(A^q, \mathfrak{A}^q, \mu^q)$ and $\mathbb{H}^{\odot q} = L^2_s(A^q, \mathfrak{A}^q, \mu^q)$, where L^2_s denotes the space of square-integrable and symmetric functions. For $x_1, x_2, \ldots, x_q \in A$, $f \in \mathbb{H}$, $f_q \in \mathbb{H}^{\odot}$, we define

$$f^{\otimes q}(x_1, x_2, \ldots, x_q) := f(x_1)f(x_2)\cdots f(x_q),$$

and

$$I_q(f_q) := \int_{A^q} f_q(a_1, a_2, \ldots, a_q)dW(a_1)dW(a_2)\cdots dW(a_q),$$

where the right-hand side is the *multiple Wiener-Itô integral* of order $q \geq 1$, of f_q with respect to the Gaussian measure W, see also [49, Exercise 2.7.6]. Then, for $f \in \mathbb{H}$ such that $\|f\|_{\mathbb{H}} = 1$, we have

$$H_q(X(f)) = I_q(f^{\otimes q}), \qquad q \geq 1,$$

see [49, Theorem 2.7.7]. As a consequence, the linear operator I_q provides an isometry from $\mathbb{H}^{\odot q}$ (equipped with the modified norm $\frac{1}{\sqrt{q!}}\|\cdot\|_{\mathbb{H}^{\odot q}}$) onto the q-th Wiener chaos $\mathcal{H}_q$ of X (equipped with the $L^2(\Omega)$ norm). For $q = 0$, we set $I_0(c) = c \in \mathbb{R}$ and the relation in (1.3.5) becomes

$$F = \sum_{q=0}^{\infty} I_q(f_q), \tag{1.3.7}$$

where $f_0 = \mathbb{E}[F]$, and the kernels $f_q \in \mathbb{H}^{\odot q}$, $q \geq 1$, are uniquely determined (see [49, Corollary 2.7.8]). In other words, F in (1.3.7) can be seen as a series of (multiple) stochastic integrals.

For every $p, q \geq 1$, every $f \in \mathbb{H}^{\otimes p}$, every $g \in \mathbb{H}^{\otimes q}$ and every $r \in \{1, \ldots, p \wedge q\}$, the so-called contraction of f and g of order r is the element $f \otimes_r g \in \mathbb{H}^{\otimes p+q-2r}$ defined as

$$f \otimes_r g(x_1, \ldots, x_{p+q-2r}) =$$

[2] A Polish space is a separable completely metrizable topological space; that is, a space homeomorphic to a complete metric space that has a countable dense subset.

$$\int_{A^r} f(a_1,\dots,a_r,x_1,\dots,x_{p-r})g(a_1,\dots,a_r,x_{p-r+1},\dots,x_{p+q-2r})d\mu(a_1)\dots d\mu(a_r).$$

For $p=q=r$, we have $f\otimes_r g=\langle f,g\rangle_{\mathbb{H}^{\otimes r}}$; if $r=0$, then $f\otimes_0 g=f\otimes g$. Denoting by $f\widetilde{\otimes}_r g$ the canonical symmetrization of $f\otimes_r g$, the following multiplication formula holds

$$I_p(f)I_q(g)=\sum_{r=0}^{p\wedge q} r!\binom{p}{r}\binom{q}{r} I_{p+q-2r}(f\widetilde{\otimes}_r g)$$

for $f\in\mathbb{H}^{\odot p}$ and $g\in\mathbb{H}^{\odot q}$, see [49, Theorem 2.7.10].
Finally, we report the Fourth Moment Theorem, which allows us to establish Quantitative Central Limit Theorems.

Theorem 1.3.7 (Fourth Moment Theorem, Theorem 5.2.7 in [49])**.** *Let $\{F_n,\ n\geq 1\}$, be a sequence of random variables belonging to the q-th chaos of X, for some fixed integer $q\geq 1$. Then, for $Z\overset{d}{=}\mathcal{N}(0,1)$,*

$$d_W\left(\frac{F_n}{\sqrt{\mathbb{V}[F_n]}},Z\right)\leq\sqrt{\frac{2q-2}{3\pi q}\frac{\operatorname{Cum}[F_n]}{\mathbb{V}[F_n]^2}}.$$

Remark 1.3.8. *This result is systematically stronger than the so-called* method of moments and cumulants*, which is the most popular tool used in the proof of central limit theorems for functional of Gaussian fields. Basically, it requires proving that all the moments (or cumulants) of F_n converge to those of a standard Gaussian random variable. The Fourth Moment Theorem is, instead, a "simplified" version for sequences of chaotic random variables with possibly different orders, since one has just to study the limit of the fourth moment (or cumulant). More importantly, it gives as a result, not only the simple convergence in distribution, but also an explicit bound on the probability metric and hence, in this sense, a Quantitative Central Limit Theorem for F_n. See, also, Chapter 6 of [49] for a multivariate version.*

In several recent works, the FMT has been applied to obtain QCLTs for specific random functionals. In particular, it was discovered to be a powerful tool to study the geometry of random fields on the sphere (and in general on Riemannian manifolds). See, for example, [46, 71, 70, 15, 16] for the sphere, [45] for the torus, and [52, 59] for the plane.

1.4 Short Background on Operator Theory

This section contains some background notions and notations from operator theory; we mainly refer to [32].
Let $\mathbb{X}_1$ and $\mathbb{X}_2$ be Banach spaces, i.e., complete normed linear spaces, with norms $\|\cdot\|_{\mathbb{X}_i}$, $i=1,2$, and $\mathscr{T}$ a linear transformation that maps from $\mathbb{X}_1$ into $\mathbb{X}_2$. We recall that the linear transformation $\mathscr{T}$ is *bounded* (or *continuous*) is there exists a finite constant $C>0$ such that

$$\|\mathscr{T}x\|_{\mathbb{X}_2}\leq C\|x\|_{\mathbb{X}_1}.$$

Then, it is defined the *operator norm*

$$\|\mathscr{T}\|_{\text{op}} = \sup_{x \in \mathbb{X}_1, \|x\|_{\mathbb{X}_1}=1} \|\mathscr{T}x\|_{\mathbb{X}_2}, \tag{1.4.1}$$

and its *adjoint operator*, that is, the unique element $\mathscr{T}^*$ satisfying

$$\langle \mathscr{T}x_1, x_2 \rangle_{\mathbb{X}_2} = \langle x_1, \mathscr{T}^* x_2 \rangle_{\mathbb{X}_1},$$

for all $x_1 \in \mathbb{X}_1, x_2 \in \mathbb{X}_2$; when $\mathscr{T}^* = \mathscr{T}$, $\mathscr{T}$ is called *self-adjoint*.
Moreover, $\mathscr{T}$ is *compact* if for any bounded sequence $\{x_n\} \in \mathbb{X}_1$, $\{\mathscr{T}x_n\}$ contains a convergent subsequence in $\mathbb{X}_2$. Note that compact linear transformations are necessarily bounded, hence are referred to as compact operators.
Let us focus on linear and bounded operators $\mathscr{T} : \mathbb{H} \to \mathbb{H}$, where $\mathbb{H}$ is some separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \|$. When working with Hilbert spaces, this type of operators can be thought as the extension of the concept of matrix (acting on a finite-dimensional vector space) to infinite-dimensional spaces. Indeed, in this setting, eigenvalues and eigenvectors (or eigenfunctions when $\mathbb{H}$ is a function space) are defined analogously as the scalar λ and nonzero element e of $\mathbb{H}$, respectively, which satisfies

$$\mathscr{T}e = \lambda e.$$

Within the class of compact operators on separable Hilbert spaces, there are the so-called Hilbert-Schmidt and trace class (or nuclear) operators, that arises in our work ahead and, more in general, are pervasive throughout statistics. Specifically in functional data analysis, of special interest are integral operators, for which a key result is the celebrated Mercer's Theorem (see, for instance, [32, Theorem 4.6.5]) that uses the eigenvalue-eigenvector decomposition on an integral operator to obtain a corresponding series expansion for the operator's kernel.
Very briefly, let $\{e_i, i \in \mathbb{N}\}$ be a complete orthonormal system for $\mathbb{H}$, $\mathscr{T}$ is called Hilbert-Schmidt if

$$\|\mathscr{T}\|_{\text{HS}} := \left(\sum_{i=1}^{\infty} \|\mathscr{T}e_i\|^2 \right)^{1/2} < \infty;$$

trace class if

$$\|\mathscr{T}\|_{\text{TR}} := \sum_{i=1}^{\infty} \langle (\mathscr{T}^*\mathscr{T})^{1/2} e_i, e_i \rangle < \infty,$$

where $\mathscr{T}^*$ is the adjoint of $\mathscr{T}$. $\|\mathscr{T}\|_{\text{HS}}$ and $\|\mathscr{T}\|_{\text{TR}}$ are respectively the Hilbert-Schmidt and trace (or nuclear) norms of $\mathscr{T}$ and, by their definitions, it can be deduced that trace class operators are also Hilbert-Schmidt [32, Theorem 4.5.2].
If we restrict our attention to compact self-adjoint operators, we end up with a fundamental result of functional analysis, which asserts that every compact self-adjoint operator may be *diagonalized* in some suitable basis.

Theorem 1.4.1 (Theorem 6.11 in [11]). *Let $\mathbb{H}$ be a separable Hilbert space and let $\mathscr{T} : \mathbb{H} \to \mathbb{H}$ be a compact self-adjoint operator. Then there exists a Hilbert basis composed of eigenvectors of $\mathscr{T}$.*

The proof is given in [11, Section 6.4].
As a consequence, the two norms have a more explicit expression in terms of the corresponding eigenvalues $\{\lambda_i, i \in \mathbb{N}\}$, that is,

$$\|\mathscr{T}\|_{\mathrm{HS}} = \sum_{i=1}^{\infty} |\lambda_i|^2 \quad \text{and} \quad \|\mathscr{T}\|_{\mathrm{TR}} = \sum_{i=1}^{\infty} |\lambda_i|;$$

while the operator norm (1.4.1) can be written as

$$\|\mathscr{T}\|_{\mathrm{op}} = \max_{i \in \mathbb{N}} |\lambda_i|.$$

Chapter 2

Spherical Random Fields with Temporal Dependence

In this chapter we combine the standard notions of stationary processes and random fields defined on the sphere, to analyze harmonic properties of time-varying spherical random fields. As mentioned earlier, harmonic analysis on the sphere has already been studied extensively and proved to be a valid tool to perform statistical analysis (see, for instance, [23, 36]). This naturally leads to investigate possible extensions in a time-dependent framework, due to also the growing necessity to model sequences of potentially dependent spherical data in many areas, such as Cosmology, Geophysics, and also Medical Imaging. All the results can be framed within the context of functional time series analysis.

This chapter is principally divided into two main parts. The first section provides some results on a double spectral representation, both in (discrete) time and space, of time-dependent spherical random fields, while the other section focuses on properties of a specific class of sphere-cross-time random fields which can be also interpreted as functional autoregressive processes (see [10]) taking values in $L^2(\mathbb{S}^2)$.

2.1 Double Spectral Representation

This section is devoted to the spectral representation of functional time series which are essentially sequences of spherical random fields. The main purpose is to study these objects, trying to simultaneously capture the surface structure (spatial component) as well as the dynamics in time (temporal component); what in [55] is called *within/between curve dynamics*. Indeed, we extended the work [55], for dependent random functions on the interval $[0,1]$, to the case of the sphere.
We start with very general assumptions, and then we gradually add some structure to the objects we are treating.

Notation. Only in this section, we will denote with $L^2(\mathbb{S}^2;\mathbb{C}) := L^2(\mathbb{S}^2, dx\,;\mathbb{C})$ the Hilbert space of square-integrable complex-valued functions on $\mathbb{S}^2$ endowed with usual the inner product $\langle f,g\rangle_{L^2(\mathbb{S}^2;\mathbb{C})} = \int_{\mathbb{S}^2} f(x)\overline{g(x)}dx$. $\|\cdot\|_{L^2(\mathbb{S}^2;\mathbb{C})}$ will be the norm induced by $\langle\cdot,\cdot\rangle_{L^2(\mathbb{S}^2;\mathbb{C})}$; when it does not cause confusion, we will omit the subscript

$L^2(\mathbb{S}^2;\mathbb{C})$. Moreover, the restriction of $L^2(\mathbb{S}^2;\mathbb{C})$ to real-valued functions will be denoted by $L^2(\mathbb{S}^2;\mathbb{R})$.
For a real- or complex-valued function f defined on a set D, we define $\|f\|_\infty := \sup_{x\in D}|f(x)|$. Recall that $\|\mathscr{T}\|_{\mathrm{TR}}$ is the trace (or nuclear) norm of the operator $\mathscr{T}$, see Section 1.4.

2.1.1 A Minimal Set of Assumptions

Let T be a random element of $L^2(\mathbb{S}^2;\mathbb{R})$ defined on a probability space $(\Omega,\mathfrak{F},\mathbb{P})$, i.e., a measurable mapping $T:\Omega\to L^2(\mathbb{S}^2;\mathbb{R})$, and such that $\mathbb{E}\|T\|^2<\infty$. Then, for each $\omega\in\Omega$,

$$\left\| T(\omega) - \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell} a_{\ell,m}(\omega) Y_{\ell,m}\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad L\to\infty,$$

where $a_{\ell,m} := \langle T_t, Y_{\ell,m}\rangle$. Following a similar argument as Remark 1.2.9, we also deduce that

$$\mathbb{E}\left\| T - \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell} a_{\ell,m} Y_{\ell,m}\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad L\to\infty.$$

Since $\mathbb{E}\|T\|^2<\infty$, the mean element of T and the covariance operator $\mathscr{R}: L^2(\mathbb{S}^2;\mathbb{C})\to L^2(\mathbb{S}^2;\mathbb{C})$ are both well-defined as Bochner integrals (see [32, Chapter 7]),

$$\begin{aligned} m &= \mathbb{E}\left[T\right] := \int_\Omega T\, d\mathbb{P},\\ \mathscr{R} &= \mathbb{E}\left[(T-m)\otimes(T-m)\right] := \int_\Omega (T-m)\otimes(T-m)\, d\mathbb{P}, \end{aligned}$$

where for $u,v\in L^2(\mathbb{S}^2;\mathbb{C})$ the tensor product $u\otimes v$ is defined to be the mapping that takes any element $f\in L^2(\mathbb{S}^2;\mathbb{C})$ to $u\langle f,v\rangle\in L^2(\mathbb{S}^2;\mathbb{C})$.

Remark 2.1.1. *The Bochner integral extends the concept of Lebesgue integration to integration over a Banach space (see [32] for definition and properties). Given a function f on a measure space $(E,\mathfrak{E},\mu)$ that takes on values in a Banach space $\mathbb{X}$ (with norm $\|\cdot\|$), the Bochner integral of f over E is defined as*

$$\int_E f d\mu = \lim_{n\to\infty}\int_E f_n d\mu,$$

where $\{f_n, n\in\mathbb{N}\}$ is a sequence of simple functions (in the sense of [32, Definition 2.6.1]) such that $\lim_{n\to\infty}\int_E\|f_n-f\|d\mu=0$. Clearly, by triangle inequality,

$$\left\|\int_E f d\mu\right\| \le \int_E \|f\| d\mu.$$

In the specific case of random elements, the measure space is $(\Omega,\mathfrak{F},\mathbb{P})$, whereas the Banach space is $L^2(\mathbb{S}^2;\mathbb{C})$ with norm $\|\cdot\|_{L^2(\mathbb{S}^2;\mathbb{C})}$.

Here, we recall two basic properties of m and $\mathscr{R}$:

$$\mathbb{E}\langle T, f\rangle = \langle m, f\rangle, \tag{2.1.1}$$
$$\mathbb{E}\left[\langle T-m, g\rangle\overline{\langle T-m, f\rangle}\right] = \langle \mathscr{R}f, g\rangle, \tag{2.1.2}$$

$f, g \in L^2(\mathbb{S}^2;\mathbb{C})$, and $\langle \mathscr{R}f, g\rangle = \langle f, \mathscr{R}g\rangle$, that is, $\mathscr{R}$ is self-adjoint. Without loss of generality, we shall assume $m = 0$.
Now consider a sequence $\{T_t,\ t \in \mathbb{Z}\}$ of zero-mean random elements of $L^2(\mathbb{S}^2;\mathbb{R})$ defined on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and such that, for each $t \in \mathbb{Z}$, $\mathbb{E}\|T_t\|^2 < \infty$; then we can also define as Bochner integrals the autocovariance operators

$$\mathscr{R}_{t,s} = \mathbb{E}\left[T_t \otimes T_s\right] := \int_\Omega T_t \otimes T_s \, d\mathbb{P}, \qquad t, s \in \mathbb{Z},$$

which satisfies

$$\mathbb{E}\left[\langle T_t, g\rangle\overline{\langle T_s, f\rangle}\right] = \langle \mathscr{R}_{t,s}f, g\rangle, \tag{2.1.3}$$

$f, g \in L^2(\mathbb{S}^2;\mathbb{C})$, and $\langle \mathscr{R}_{t,s}f, g\rangle = \langle f, \mathscr{R}_{s,t}\, g\rangle$, that is, the adjoint of $\mathscr{R}_{t,s}$ is $\mathscr{R}_{s,t}$.

Remark 2.1.2. *Properties such as those in Equations* (2.1.1), (2.1.2), *and* (2.1.3) *are proved in [32]. However, the authors restrict their attention to real Hilbert spaces. Also in [55] the autocovariance operators are defined on the restriction of* $L^2([0,1])$ *to real-valued functions. Here, even if we work with random elements of* $L^2(\mathbb{S}^2;\mathbb{R})$, *their spectral representation is in terms of the standard complex basis of spherical harmonics* $\{Y_{\ell,m}\}$, *hence we shall define the* $\mathscr{R}_{t,s}$*'s on* $L^2(\mathbb{S}^2;\mathbb{C})$.

It is then natural to extend the definition of a stationary functional time series of $L^2(\mathbb{S}^2;\mathbb{R})$.

Definition 2.1.3 (see, for instance, Definition 2.4 in [10])**.** *We say that the sequence* $\{T_t,\ t \in \mathbb{Z}\}$ *of zero-mean random elements of* $L^2(\mathbb{S}^2;\mathbb{R})$ *is stationary if* $\mathbb{E}\|T_t\|^2 < \infty$, *for all* $t \in \mathbb{Z}$, *and*

$$\mathscr{R}_{t,s} = \mathscr{R}_{t+h,s+h}, \qquad \text{for all } t, s, h \in \mathbb{Z}.$$

Thus, in this case, as for standard time series, we can simplify the notation as follows:

$$\mathscr{R}_t \equiv \mathscr{R}_{t,0} = \mathbb{E}\left[T_t \otimes T_0\right], \qquad t \in \mathbb{Z}. \tag{2.1.4}$$

Remark 2.1.4. *Note that Definition 2.1.3 can be generalized to sequences of random elements of any separable Hilbert space, since the mean element* m *and the autocovariance operators* $\mathscr{R}_{t,s}$ *are, in any case, well-defined (see [32, Chapters 2 and 7]); an example is given in [56, 55] for the Hilbert space of square-integrable functions on* $[0,1]$. *In even broader terms, [10] provides the definition for sequences of random elements of separable Banach spaces.*

A very first double spectral representation for stationary sequences of spherical random elements is given by the following theorem:

Theorem 2.1.5. *Let $\{T_t,\ t \in \mathbb{Z}\}$ be a sequence of stationary zero-mean random elements of $L^2(\mathbb{S}^2;\mathbb{R})$ such that, for every $t \in \mathbb{Z}$, $\mathbb{E}\|T_t\|^2 < \infty$; then it holds*

$$\mathbb{E}\left\| T_t - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \int_{-\pi}^{\pi} e^{i\lambda t} d\alpha_{\ell,m}(\lambda) Y_{\ell,m} \right\|_{L^2(\mathbb{S}^2;\mathbb{C})}^2 \to 0, \qquad L \to \infty,$$

where $\{\alpha_{\ell,m}(\lambda),\ -\pi \le \lambda \le \pi\}$ is a complex-valued orthogonal increment process, that is,

$$\mathbb{E}(\alpha_{\ell,m}(\lambda_1) - \alpha_{\ell,m}(\lambda_2))\overline{(\alpha_{\ell,m}(\lambda_3) - \alpha_{\ell,m}(\lambda_4))} = 0,$$

for $\lambda_1 > \lambda_2 \ge \lambda_3 > \lambda_4$, and the stochastic integral involved can be understood as a Riemann-Stieltjes limit, in the sense that

$$\mathbb{E}\left| \langle T_t, Y_{\ell,m}\rangle - \sum_{j=1}^{J} e^{i\lambda_j t} \left[\alpha_{\ell,m}(\lambda_{j+1}) - \alpha_{\ell,m}(\lambda_j)\right] \right|^2 \to 0, \qquad J \to \infty,$$

where $-\pi = \lambda_1 < \cdots < \lambda_{J+1} = \pi$ and $\max_{j=1,\ldots,J} |\lambda_{j+1} - \lambda_j| \to 0$ as $J \to \infty$.

Proof. Define $a_{\ell,m}(t) := \langle T_t, Y_{\ell,m}\rangle$. For every fixed (ℓ, m), $\{a_{\ell,m}(t),\ t \in \mathbb{Z}\}$ forms a zero-mean complex-valued stationary sequence, i.e.,

$$\mathbb{E}[a_{\ell,m}(t)] = 0;$$
$$\mathbb{E}|a_{\ell,m}(t)|^2 < \infty;$$
$$\mathbb{E}[a_{\ell,m}(t)\overline{a_{\ell,m}}(s)] = C_{\ell,m}(t-s).$$

Indeed, from Equation (2.1.1), we have

$$\mathbb{E}[a_{\ell,m}(t)] = \mathbb{E}\langle T_t, Y_{\ell,m}\rangle = \langle 0, Y_{\ell,m}\rangle = 0.$$

Moreover,

$$\mathbb{E}|a_{\ell,m}(t)|^2 \le \mathbb{E}\|T_t\|^2 < \infty,$$

and, by Equation (2.1.3),

$$\mathbb{E}[a_{\ell,m}(t)\overline{a_{\ell,m}}(s)] = \mathbb{E}\left[\langle T_t, Y_{\ell,m}\rangle \overline{\langle T_s, Y_{\ell,m}\rangle}\right] = \langle \mathscr{R}_{t-s} Y_{\ell,m}, Y_{\ell,m}\rangle.$$

Therefore, as a result of the Spectral Theorem 1.1.7, the following representation holds

$$a_{\ell,m}(t) = \int_{-\pi}^{\pi} e^{i\lambda t} d\alpha_{\ell,m}(\lambda), \qquad \text{a.s.},$$

where $\{\alpha_{\ell,m}(\lambda),\ -\pi \le \lambda \le \pi\}$ is an orthogonal increment process, and the stochastic integral involved can be understood as a Riemann-Stieltjes limit, in the sense that

$$\mathbb{E}\left| a_{\ell,m}(t) - \sum_{j=1}^{J} e^{i\lambda_j t} \left[\alpha_{\ell,m}(\lambda_{j+1}) - \alpha_{\ell,m}(\lambda_j)\right] \right|^2 \to 0, \qquad J \to \infty,$$

where $-\pi = \lambda_1 < \cdots < \lambda_{J+1} = \pi$ and $\max_{j=1,\ldots,J} |\lambda_{j+1} - \lambda_j| \to 0$ as $J \to \infty$. As a consequence, for every $t \in \mathbb{Z}$,

$$\mathbb{E}\left\| T_t - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \int_{-\pi}^{\pi} e^{i\lambda t} d\alpha_{\ell,m}(\lambda) Y_{\ell,m} \right\|_{L^2(\mathbb{S}^2;\mathbb{C})}^2 \to 0, \qquad L \to \infty.$$

□

2.1.2 Adding Mean-Square Continuity

Now, consider the collection of random variables $\{T(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$. For every fixed $t \in \mathbb{Z}$, $\{T(x,t),\ x \in \mathbb{S}^2\}$ is a spherical random field as defined in Section 1.2.1; recall that we are implicitly assuming measurability with respect to the product σ-field $\mathfrak{B}(\mathbb{S}^2) \times \mathfrak{F}$. We name $\{T(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ *space-time spherical random field*. For simplicity, we will assume that $\mathbb{E}[T(x,t)] = 0$, for all $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$. As usual in the context of functional data analysis, we model the mapping $T_t : \omega \mapsto T(\cdot, t, \omega)$ as a random element of the separable Hilbert space of functions $L^2(\mathbb{S}^2; \mathbb{R})$.

Remark 2.1.6. *Recall that if $\{T(x),\ x \in \mathbb{S}^2\}$ is jointly measurable and $T(x,\omega) \in L^2(\mathbb{S}^2; \mathbb{R})$ for each ω, then the mapping $\omega \mapsto T(\cdot,\omega)$ is a random element of $L^2(\mathbb{S}^2; \mathbb{R})$ (see [32, Theorem 7.4.1]).*

If $\mathbb{E}\|T_t\|^2 < \infty$, by Fubini's theorem $\mathbb{E}|T(x,t)|^2 < \infty$ almost everywhere. Then, the autocovariance kernel $r_{t,s}(\cdot,\cdot) : (x,y) \mapsto \mathbb{E}\left[T(x,t)T(y,s)\right]$ is in $L^2(\mathbb{S}^2 \times \mathbb{S}^2)$, and the corresponding operator $\mathcal{R}_{t,s} : L^2(\mathbb{S}^2; \mathbb{C}) \to L^2(\mathbb{S}^2; \mathbb{C})$ induced by right integration

$$(\mathcal{R}_{t,s} f)(\cdot) := \int_{\mathbb{S}^2} r_{t,s}(\cdot, y) f(y) dy$$

coincides with the autocovariance operator $\mathscr{R}_{t,s}$ in (2.1.3), see also [56]. We shall use the notation $\mathscr{R}_{t,s}$ for both. For a stationary sequence, clearly we can define $r_t(\cdot,\cdot) \equiv r_{t,0}(\cdot,\cdot)$ associated with (2.1.4).
The following conditions are the equivalent on the sphere of Conditions 1.1 in [55]. We shall use these conditions to prove first a Functional Cramér Representation Theorem which involves a $L^2(\mathbb{S}^2; \mathbb{C})$-valued orthogonal increment process, and then to obtain a double spectral representation with respect to both space and time. We stress that here, as in [55], it is not assumed any other prior structural properties for the stationary sequence (e.g., linearity or Gaussianity).

Condition 2.1.7 (see also Conditions 1.1 in [55])**.** *The zero-mean space-time spherical random field $\{T(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is such that, for every $t \in \mathbb{Z}$,*

- *(i) the mapping $T_t : \omega \mapsto T(\cdot,t,\omega)$ is a random element of $L^2(\mathbb{S}^2; \mathbb{R})$ and the sequence $\{T_t,\ t \in \mathbb{Z}\}$ is stationary;*
- *(ii) the autocovariance kernel $r_t(\cdot,\cdot) : (x,y) \mapsto \mathbb{E}\left[T(x,t)T(y,0)\right]$ is continuous on $\mathbb{S}^2 \times \mathbb{S}^2$.*

Furthermore, we assume that the autocovariance kernels and the associated autocovariance operators satisfies respectively:

$$\sum_{t \in \mathbb{Z}} \|r_t\|_\infty < \infty, \qquad \sum_{t \in \mathbb{Z}} \|\mathscr{R}_t\|_{\mathrm{TR}} < \infty. \tag{2.1.5}$$

Remark 2.1.8. *The compactness of $\mathbb{S}^2$ has the consequence that $r_t(\cdot,\cdot)$ is also uniformly continuous. This continuity is translated to the image of $\mathscr{R}_t$, that is, for each $h \in L^2(\mathbb{S}^2; \mathbb{C})$, $(\mathscr{R}_t h)(\cdot)$ is uniformly continuous [32, Lemma 4.6.1]. Moreover, notice* *that if $r_0(\cdot,\cdot)$ is continuous on $\mathbb{S}^2 \times \mathbb{S}^2$, then each random field $\{T(x,t),\ x \in \mathbb{S}^2\}$, $t \in \mathbb{Z}$, is mean-square continuous, that is,*

$$\lim_{x \to x_0} \mathbb{E}[T(x,t) - T(x_0,t)]^2 = 0, \qquad \forall x_0 \in \mathbb{S}^2. \tag{2.1.6}$$

Indeed, as

$$\mathbb{E}[T(x,t) - T(x_0,t)]^2 = r_0(x,x) + r_0(x_0,x_0) - 2r_0(x,x_0),$$

the continuity of $r_0(\cdot,\cdot)$ implies (2.1.6).

In [56] there is an extensive discussion on the role of Condition 2.1.7, and at what cost it may be weakened. However, in this context, it is possible to define the spectral density kernel at frequency $\lambda \in \mathbb{R}$,

$$f_\lambda(\cdot,\cdot) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} e^{-i\lambda t} r_t(\cdot,\cdot),$$

where the convergence is in $\|\cdot\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}$ and $\|\cdot\|_\infty$. It is uniformly bounded and also uniformly continuous in λ with respect to $\|\cdot\|_{L^2(\mathbb{S}^2\times\mathbb{S}^2)}$ and $\|\cdot\|_\infty$. For each λ, $f_\lambda(\cdot,\cdot)$ is continuous on $\mathbb{S}^2 \times \mathbb{S}^2$ and

$$f_{-\lambda}(x,y) = \overline{f_\lambda(x,y)} = f_\lambda(y,x).$$

Moreover, for all $t \in \mathbb{Z},\ x,y \in \mathbb{S}^2$, the following *inversion formula* holds

$$\int_{-\pi}^{\pi} f_\lambda(x,y) e^{i\lambda t} d\lambda = r_t(x,y). \tag{2.1.7}$$

The spectral density operator $\mathscr{F}_\lambda : L^2(\mathbb{S}^2;\mathbb{C}) \to L^2(\mathbb{S}^2;\mathbb{C})$, the operator induced by the spectral density kernel through right-integration, is self-adjoint and nonnegative definite for all $\lambda \in \mathbb{R}$. $\mathscr{F}_\lambda$ is also 2π-periodic with respect to λ, trace class, since $\|\mathscr{F}_\lambda\|_{\mathrm{TR}} \leq \sum_{t\in\mathbb{Z}} \|\mathscr{R}_t\|_{\mathrm{TR}} < \infty$, $\lambda \mapsto \|\mathscr{F}_\lambda\|_{\mathrm{TR}}$ is uniformly continuous and

$$\|\mathscr{F}_\lambda\|_{\mathrm{TR}} = \int_{\mathbb{S}^2} f_\lambda(x,x) dx.$$

The reader is referred to [56] for proofs of these assertions.

Remark 2.1.9. *The assumptions in Equation (2.1.5) are strictly related to the concept of short-range dependent stationary processes; indeed, stationary processes which exhibit short-range dependence are those with absolutely summable covariance and, hence, bounded and continuous spectral density, e.g., stationary ARMA processes. For functional time series, this translates into an "absolutely summable" covariance operator and a "bounded and continuous" spectral density operator, that is, their nuclear norms are, respectively, absolutely summable, and bounded and continuous.*

The following theorem is the analogue of Theorem 2.1 in [55].

Theorem 2.1.10. *(Spherical Functional Cramér Representation, see also [55]) Under Condition 2.1.7, T_t admits the representation*

$$T_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_\lambda, \qquad \text{a.s. in } L^2(\mathbb{S}^2), \tag{2.1.8}$$

where, for fixed λ, Z_λ is a random element of $L^2(\mathbb{S}^2;\mathbb{C})$ with $\mathbb{E}\|Z_\lambda\|^2 = \int_{-\pi}^{\lambda} \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu$, and the process $\{Z_\lambda, -\pi \leq \lambda \leq \pi\}$ has orthogonal increments:

$$\mathbb{E}\langle Z_{\lambda_1} - Z_{\lambda_2}, Z_{\lambda_3} - Z_{\lambda_4}\rangle = 0, \qquad \lambda_1 > \lambda_2 \geq \lambda_3 > \lambda_4. \tag{2.1.9}$$

The representation (2.1.8) is called the Cramér representation of T_t, and the stochastic integral involved can be understood as a Riemann-Stieltjes limit, in the sense that

$$\mathbb{E}\left\| T_t - \sum_{j=1}^{J} e^{i\lambda_j t}(Z_{\lambda_{j+1}} - Z_{\lambda_j})\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad J \to \infty,$$

where $-\pi = \lambda_1 < \cdots < \lambda_{J+1} = \pi$ and $\max_{j=1,\ldots,J} |\lambda_{j+1} - \lambda_j| \to 0$ as $J \to \infty$.

Proof. The proof follows the same lines of [55]. Let $\mathbb{H}$ be the Hilbert space of $L^2(\mathbb{S}^2;\mathbb{C})$-valued random elements with finite second moment[1] and $\mathbb{M}_0$ be the complex linear space spanned by all finite linear combinations of the T_t's,

$$\mathbb{M}_0 := \left\{ \sum_{j=1}^{n} b_j T_j : \; n \in \mathbb{N}, \; b_j \in \mathbb{C}, \; t_j \in \mathbb{Z} \right\} \subset \mathbb{H}.$$

Let $e_t : \nu \mapsto e^{it\nu}$, which belongs to the (complex) Hilbert space $L^2([-\pi,\pi], \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu)$ endowed with the standard inner product

$$\int_{-\pi}^{\pi} f(\nu)\overline{g(\nu)} \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu, \qquad f, g \in L^2([-\pi,\pi], \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu),$$

$\|\mathscr{F}_\nu\|_{\mathrm{TR}}$ being the nuclear norm of the spectral density operator. Now, define the linear operator E by linear extension of the mapping $T_t \mapsto e_t$. E is well defined and a linear isometry; in particular, the inversion formula (2.1.7) gives

$$\langle T_t, T_s\rangle_{\mathbb{H}} = \mathbb{E}\left[\int_{\mathbb{S}^2} T(x,t)T(x,s)dx\right] = \int_{-\pi}^{\pi} e^{i(t-s)\lambda} \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu.$$

Then, we extend its domain to $\mathbb{M}$, the closure of $\mathbb{M}_0$ in $\mathbb{H}$ (see [55] for further details); the extension has a well-defined inverse $E^{-1} : L^2([-\pi,\pi], \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu) \to \mathbb{M}$. For any $\omega \in (-\pi,\pi]$, we define $Z_\omega = E^{-1}(\mathbb{1}_{[-\pi,\omega)}) \in \mathbb{M}$ and $Z_{-\pi} \equiv 0$. By the isometry property,

$$\langle Z_\omega, Z_\beta\rangle_{\mathbb{H}} = \langle E^{-1}\mathbb{1}_{[-\pi,\omega)}, E^{-1}\mathbb{1}_{[-\pi,\beta)}\rangle_{\mathbb{H}} = \int_{-\pi}^{\min\{\omega,\beta\}} \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu. \tag{2.1.10}$$

[1] $f \in \mathbb{H}$ is such that $\mathbb{E}\|f\|^2_{\mathbb{H}} < \infty$. The associated inner product is defined as

$$\langle f, g\rangle_{\mathbb{H}} = \mathbb{E}\langle f, g\rangle_{L^2(\mathbb{S}^2;\mathbb{C})}, \quad \text{for } f, g \in \mathbb{H}$$

.

Hence, $\omega \mapsto Z_\omega$ is an orthogonal increment process.
The proof follows with definition of an operator ζ as extension of the mapping

$$\sum_{j=1}^{n} g_j \mathbb{1}_{[\omega_j,\omega_{j+1})} \mapsto \sum_{j=1}^{n} g_j (Z_{\omega_{j+1}} - Z_{\omega_j}).$$

The operator ζ is, by (2.1.10), an isomorphism with domain $L^2([-\pi,\pi], \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu)$, and in addition $\zeta = E^{-1}$. This in turn implies $T_t = E^{-1}(e_t) = \zeta(e_t)$. If g is cadlag with a finite number of jumps, then $\zeta(g)$ is in fact the Riemann–Stieltjes integral (in the mean square sense) with respect to the orthogonal increment process Z_ω:

$$\zeta(g) = \int_{-\pi}^{\pi} g(\lambda) dZ_\lambda.$$

In conclusion, $T_t = \int_{-\pi}^{\pi} e^{it\lambda} dZ_\lambda$, as claimed. □

Now, define the operator

$$\begin{aligned} Y_{\ell,m} \otimes Y_{\ell,m} &: L^2(\mathbb{S}^2;\mathbb{C}) \to L^2(\mathbb{S}^2;\mathbb{C}) \\ &: f \mapsto \langle f, Y_{\ell,m} \rangle Y_{\ell,m}. \end{aligned}$$

We are going to establish a double spectral representation result, by showing the relation between the orthogonal increment process $\{\alpha_{\ell,m}(\lambda), -\pi \le \lambda \le \pi\}$ (see Theorem 2.1.5) and $\{Z_\lambda, -\pi \le \lambda \le \pi\}$. It is worth to notice that, under Condition (2.1.7), all the results presented in [55] can be easily extended to our framework, including the so-called Cramér–Karhunen–Loève Representation. Such a representation decomposes the space-time spherical random field into uncorrelated functional frequency components, exploiting an orthonormal basis for $L^2(\mathbb{S}^2;\mathbb{C})$ made up of eigenfunctions of the spectral density operator $\mathscr{F}_\lambda$. However, in the anisotropic case, these eigenfunctions are unknown and have to be estimated.
The next result does not go through the eigenvalue-eigenfunction decomposition of $\mathscr{F}_\lambda$, but is is based on the standard orthonormal basis of spherical harmonics.

Theorem 2.1.11 (Double Spectral Representation)**.** *Under Condition 2.1.7, for every $t \in \mathbb{Z}$, it holds that*

$$\mathbb{E} \left\| T_t - \int_{-\pi}^{\pi} e^{it\lambda} \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} Y_{\ell,m} \otimes Y_{\ell,m} dZ_\lambda \right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad L \to \infty, \qquad (2.1.11)$$

with $\{Z_\lambda, -\pi \le \lambda \le \pi\}$ as defined in Theorem 2.1.10.

Remark 2.1.12. *Equation (2.1.11) cannot be interpreted as a proper Cramér–Karhunen–Loève decomposition, since the random components at different multipoles are correlated. It result not to be useful when making inference, but here it is still reported for completeness.*

Proof. Recall from Theorem 2.1.5 that

$$\mathbb{E} \left\| T_t - \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} \int_{-\pi}^{\pi} e^{i\lambda t} d\alpha_{\ell,m}(\lambda) Y_{\ell,m} \right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad L \to \infty.$$

First we prove that $\alpha_{\ell,m}(\lambda) \overset{\text{a.s.}}{=} \langle Z_\lambda, Y_{\ell,m}\rangle$.
For a fixed λ, $\alpha_{\ell,m}(\lambda) \in \overline{\operatorname{span}\{a_{\ell,m}(t),\ t\in\mathbb{Z}\}} = \overline{\operatorname{span}\{\langle T_t, Y_{\ell,m}\rangle,\ t\in\mathbb{Z}\}} \subset L^2(\Omega)$.
Indeed, from [12], we know that there exist a sequence $\{\alpha_j\}_{j\in\mathbb{Z}} \subset \mathbb{C}$ such that

$$\mathbb{E}\left|\alpha_{\ell,m}(\lambda) - \sum_{|j|\le k} \alpha_j \langle T_{t_j}, Y_{\ell,m}\rangle\right|^2 \to 0, \qquad k\to\infty.$$

The sequence is given by

$$\alpha_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathbb{1}_{[-\pi,\lambda)}(\nu) e^{-ij\nu} d\nu, \qquad j\in\mathbb{Z}. \tag{2.1.12}$$

Now,

$$\begin{aligned}\mathbb{E}\left|\langle Z_\lambda, Y_{\ell,m}\rangle - \sum_{|j|\le k}\alpha_j\langle T_{t_j}, Y_{\ell,m}\rangle\right|^2 &= \mathbb{E}\left|\left\langle Z_\lambda - \sum_{|j|\le k}\alpha_j T_{t_j}, Y_{\ell,m}\right\rangle\right|^2 \\ &\le \mathbb{E}\left\|Z_\lambda - \sum_{|j|\le k}\alpha_j T_{t_j}\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})},\end{aligned}$$

by Cauchy-Schwartz inequality e orthonormality of the $Y_{\ell,m}$'s.
We just need to prove that

$$\mathbb{E}\left\|Z_\lambda - \sum_{|j|\le k}\alpha_j T_{t_j}\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} \to 0, \qquad k\to\infty.$$

Recall that $\{\alpha_j,\ j\in\mathbb{Z}\}$ as defined in (2.1.12) represent the Fourier coefficients of the indicator function $\mathbb{1}_{[-\pi,\lambda)}(\cdot)$. Then, its k-th order Fourier series approximation is given by

$$h_k(\cdot) = \sum_{|j|\le k} a_j e^{ik\cdot},$$

and $\sum_{|j|\le k}\alpha_j T_{t_j} = E^{-1}(h_k)$, where E is the isomorphism of Theorem 2.1.8. Since $\|\mathscr{F}_\lambda\|_{\mathrm{TR}} \le const$ uniformly over λ by assumption, it holds that

$$\int_{-\pi}^{\pi} |h_k(\nu) - \mathbb{1}_{[-\pi,\lambda)}(\nu)|^2 \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu \to 0, \qquad k\to\infty.$$

By continuity of E, we conclude that

$$E^{-1}(h_k) \to E^{-1}(\mathbb{1}_{[-\pi,\lambda)}) = Z_\lambda, \qquad k\to\infty,$$

in the L^2-sense. Finally, the triangular inequality gives the result.
As a consequence, we have that

$$\int_{-\pi}^{\pi} e^{it\lambda} Y_{\ell,m}\otimes Y_{\ell,m} dZ_\lambda = \int_{-\pi}^{\pi} e^{it\lambda} d\alpha_{\ell,m}(\lambda) Y_{\ell,m}, \qquad \text{a.s. in } L^2,$$

see [55] for a definition of stochastic integrals of operators as the one on the left-hand side. By linearity of the stochastic integral in (2.1.11) (see [55]), we conclude the proof. □

Remark 2.1.13. *By orthonormality of the $Y_{\ell,m}$'s, the approximation error is given by*

$$\mathbb{E}\left\| T_t - \int_{-\pi}^{\pi} e^{it\lambda} \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} Y_{\ell,m} \otimes Y_{\ell,m} dZ_\lambda \right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} = \sum_{\ell>L} \sum_{m=-\ell}^{\ell} C_{\ell,m}(0),$$

see also [55, Remark 3.10].

2.1.3 Adding Isotropy

Now, we strengthen our conditions, introducing joint isotropy-stationarity (see also [20]). To this purpose, we give the following definition, that will be crucial for the rest of the thesis.

Definition 2.1.14. *The space-time spherical random field $\{T(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is said to be isotropic stationary if*

(i) $\mathbb{E}|T(x,t)|^2 < \infty$, *for all* $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$,

(ii) $\mathbb{E}[T(x,t)] = const$, *for all* $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}$,

(iii) the covariance function Γ on $(\mathbb{S}^2 \times \mathbb{Z})^2$ is such that

$$\Gamma(x,t,y,s) = \Gamma(gx, t+h, gy, s+h),$$

for all $x, y \in \mathbb{S}^2,\ g \in SO(3),\ t,s,h \in \mathbb{Z}$.

Then, the new assumptions become:

Condition 2.1.15. *The zero-mean space-time spherical random field $\{T(x,t),$ $(x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is isotropic stationary with autocovariance kernels $r_t(\cdot,\cdot) \equiv \Gamma(\cdot,t,\cdot,0)$, $t \in \mathbb{Z}$.*
Furthermore, we assume that the autocovariance kernels and the associated autocovariance operators satisfies respectively:

$$\sum_{t\in\mathbb{Z}} \|r_t\|_\infty < \infty, \qquad \sum_{t\in\mathbb{Z}} \|\mathscr{R}_t\|_{\mathrm{TR}} < \infty. \tag{2.1.13}$$

Remark 2.1.16. *In this setup, most of the conditions given in [55] are satisfied. The first part of Condition 2.1.15 implies that there exists T_t random element of $L^2(\mathbb{S}^2;\mathbb{R})$ such that $T(\cdot,t) = T_t$ $\mathbb{P}$-a.s., and clearly $\mathbb{E}\|T_t\|^2 < \infty$. Moreover, continuity of all kernels $r_t(\cdot,\cdot)$, $t \in \mathbb{Z}$, follows from mean-square continuity of $\{T(x,t),\ x \in \mathbb{S}^2\}$, $t \in \mathbb{Z}$, (which is in turn consequence of isotropy, see [43]). Write*

$$r_t(x,y) - r_t(x_0,y_0) = (r_t(x,y) - r_t(x_0,y)) + (r_t(x_0,y) - r_t(x_0,y_0)).$$

The Cauchy-Schwartz inequality then gives

$$|r_t(x,y) - r_t(x_0,y)| \le \left(\mathbb{E}[T(y,0)]^2\right)^{1/2} \left(\mathbb{E}[T(x,t) - T(x_0,t)]^2\right)^{1/2}$$

and

$$|r_t(x_0,y) - r_t(x_0,y_0)| \le \left(\mathbb{E}[T(x_0,t)]^2\right)^{1/2} \left(\mathbb{E}[T(y,0) - T(y_0,0)]^2\right)^{1/2},$$

so that continuity of $r_t(\cdot,\cdot)$ follows immediately from (2.1.6). Hence, all the previous results can be directly applied to T_t.

These stronger conditions allows to apply directly theorems from [55], since we have an explicit eigenvalue-eigenfunction decomposition of the spectral density operator in terms of spherical harmonics. Exploiting joint isotropy-stationarity of the space-time spherical random field, we can obtain a neat expression for some quantities of interest, such as the autocovariance and the spectral density kernels.
First of all, the sequence of zero-mean random coefficients satisfies

$$\mathbb{E}[a_{\ell,m}(t)\overline{a_{\ell',m'}}(s)] = C_\ell(t-s)\delta_\ell^{\ell'}\delta_m^{m'}, \qquad t,s \in \mathbb{Z};$$

and the covariance kernel is shown to have a spectral decomposition in terms of Legendre polynomials, i.e.,

$$r_t(x,y) = \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_\ell(t) P_\ell(\langle x,y\rangle), \tag{2.1.14}$$

for all $x,y \in \mathbb{S}^2$, $t \in \mathbb{Z}$.

Remark 2.1.17. *Following the works [64] and [26], in [8] the authors give a mathematical characterization of covariance functions for isotropic stationary random fields over $\mathbb{S}^2 \times \mathbb{R}$. In [20] the regularity properties of such covariance functions have been investigated for the case where a double Karhunen–Loève expansion holds. Examples of random fields satisfying this decomposition are found in the Appendix of [62].*

As a consequence of (2.1.14), the orthonormal basis $\{Y_{\ell,m},\ \ell \geq 0,\ m = -\ell,\dots,\ell\}$ satisfies

$$\begin{aligned}(\mathscr{F}_\lambda Y_{\ell,m})(\cdot) &= \int_{\mathbb{S}^2} f_\lambda(\cdot,y) Y_{\ell,m}(y)dy \\ &= \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-it\lambda} \int_{\mathbb{S}^2} r_t(\cdot,y)Y_{\ell,m}(y)dy \\ &= \left[\frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-it\lambda} C_\ell(t)\right] Y_{\ell,m}(\cdot).\end{aligned}$$

Hence, $f_\ell(\lambda) = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-it\lambda}C_\ell(t)$ is the eigenvalue associated with the eigenvector $Y_{\ell,m}$ and corresponds to the spectral density of the time series $\{a_{\ell,m}(t),\ t \in \mathbb{Z}\}$, for $m = -\ell,\dots,\ell$. This implies

$$\begin{aligned} f_\lambda(x,y) &= \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} f_\ell(\lambda) P_\ell(\langle x,y\rangle), \\ \|\mathscr{F}_\lambda\|_{\mathrm{TR}} &= \sum_{\ell=0}^{\infty} (2\ell+1) f_\ell(\lambda),\end{aligned}$$

for all $x,y \in \mathbb{S}^2$, $-\pi \leq \lambda \leq \pi$.
Finally, in the isotropic case, it is also possible to show that the Cramér–Karhunen–Loève Representation, with respect to the orthogonal increment process $\{Z_\lambda, -\pi \leq \lambda \leq \pi\}$, holds in the $L^2(\Omega)$ sense for almost every $x \in \mathbb{S}^2$.

Theorem 2.1.18 (Spherical Cramér–Karhunen–Loève Representation)**.** *Under Condition 2.1.15, for every $t \in \mathbb{Z}$ and almost every $x \in \mathbb{S}^2$,*

$$\mathbb{E}\left|\int_{-\pi}^{\pi} e^{it\lambda} dZ_\lambda(x) - \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\int_{-\pi}^{\pi} e^{it\lambda} d\alpha_{\ell,m}(\lambda) Y_{\ell,m}(x)\right|^2 \to 0, \qquad L \to \infty,$$

with $\{Z_\lambda, -\pi \leq \lambda \leq \pi\}$ as defined in Theorem 2.1.10.

Proof. Under isotropy, it holds that

$$\mathbb{E}\left|\int_{-\pi}^{\pi} e^{it\lambda} dW_\lambda(x) - \sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell}\int_{-\pi}^{\pi} e^{it\lambda} d\alpha_{\ell,m}(\lambda) Y_{\ell,m}(x)\right|^2 \to 0, \qquad L \to \infty,$$

for every $x \in \mathbb{S}^2$, where $\{W_\lambda(x),\ -\pi \leq \lambda \leq \pi\}$ is such that

$$\mathbb{E}\left|W_\lambda(x) - \sum_{|j|\leq k} \alpha_j T_{t_j}(x)\right|^2 \to 0, \qquad k \to \infty,$$

for every $x \in \mathbb{S}^2$. We want to prove that, for every $\lambda \in [-\pi, \pi]$, $W_\lambda = Z_\lambda$ almost surely, almost everywhere.
First note that, thanks to the standard isomorphism,

$$\mathbb{E}\left|W_\lambda(x) - \sum_{|j|\leq k} \alpha_j T_{t_j}(x)\right|^2 = \int_{-\pi}^{\pi}\left|\mathbb{1}_{[-\pi,\lambda)}(\nu) - \sum_{|j|\leq k}\alpha_j e^{it_j\nu}\right|^2 f_\lambda(x,x)d\nu,$$

and,

$$\int_{\mathbb{S}^2}\left(\int_{-\pi}^{\pi}\left|\mathbb{1}_{[-\pi,\lambda)}(\nu) - \sum_{|j|\leq k}\alpha_j e^{it_j\nu}\right|^2 f_\lambda(x,x)d\nu\right)dx$$
$$= \int_{-\pi}^{\pi}\left|\mathbb{1}_{[-\pi,\lambda)}(\nu) - \sum_{|j|\leq k}\alpha_j e^{it_j\nu}\right|^2 \|\mathscr{F}_\nu\|_{\mathrm{TR}} d\nu < \infty.$$

Then, by the Dominated Convergence Theorem,

$$\lim_{k\to\infty}\mathbb{E}\left\|W_\lambda - \sum_{|j|\leq k}\alpha_j T_{t_j}\right\|^2_{L^2(\mathbb{S}^2;\mathbb{C})} = \int_{\mathbb{S}^2}\lim_{k\to\infty}\mathbb{E}\left|W_\lambda(x) - \sum_{|j|\leq k}\alpha_j T_{t_j}(x)\right| dx = 0.$$

□

2.2 Spherical Functional Autoregressions

In this section, we introduce a particular class of space-time spherical random fields, i.e., what we call *spherical functional autoregressions*. The main purpose here is to study the existence and uniqueness of an isotropic stationary solution of the functional autoregressive equation, see also [10, Chapter 5].
As usual in the context of autoregressive processes, we start with the definition of a *spherical white noise*.

Definition 2.2.1 (Spherical White Noise). *The collection of random variables* $\{Z(x,t),(x,t)\in\mathbb{S}^2\times\mathbb{Z}\}$ *is said to be a spherical white noise if:*

(i) for every fixed $t\in\mathbb{Z}$, $\{Z(x,t),\ x\in\mathbb{S}^2\}$ *is a zero-mean isotropic random field, with covariance function*

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

$\{C_{\ell;Z}\}$ *denoting as usual the angular power spectrum of* $Z(\cdot,t)$*;*

(ii) for every $t\neq s$*, the random fields* $\{Z(x,t),\ x\in\mathbb{S}^2\}$ *and* $\{Z(x,s),\ x\in\mathbb{S}^2\}$ *are independent.*

Remark 2.2.2. *Note that we are writing the spherical white noise as a collection of random variables defined on every pair* $(x,t)\in\mathbb{S}^2\times\mathbb{Z}$*. Alternatively, following [10, page 72], one could give the definition in terms of random elements of a separable Hilbert space (in our case, corresponding to* $L^2(\mathbb{S}^2)$*). The two approaches are equivalent here, because throughout this section and the next chapters we will always be dealing with jointly-measurable mean-square continuous random fields.*

Definition 2.2.3. *A spherical isotropic kernel operator is an application* $\Phi: L^2(\mathbb{S}^2)\to L^2(\mathbb{S}^2)$ *which satisfies*

$$(\Phi f)(x)=\int_{\mathbb{S}^2}k(\langle x,y\rangle)f(y)dy,\quad x\in\mathbb{S}^2,$$

for some continuous $k:[-1,1]\to\mathbb{R}$.

The following representation holds in the L^2-sense for the kernel associated with Φ:

$$k(\langle x,y\rangle)=\sum_{\ell=0}^{\infty}\phi_\ell\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle). \tag{2.2.1}$$

The coefficients $\{\phi_\ell,\ \ell\geq 0\}$ corresponds to the eigenvalues of the operator Φ and the associated eigenfunctions are the family of spherical harmonics $\{Y_{\ell,m}\}$, yielding

$$\Phi Y_{\ell,m}=\phi_\ell Y_{\ell,m},$$

Thus, it holds $\sum_\ell(2\ell+1)\phi_\ell^2<\infty$, and hence this operator is Hilbert-Schmidt (see, e.g., [32]). In the next chapters, we shall also consider trace class operators, namely, such that $\sum_\ell(2\ell+1)|\phi_\ell|<\infty$, for which the representation (2.2.1) holds pointwise for every $x,y\in\mathbb{S}^2$.
Now, we focus on a space-time spherical random field $\{T(x,t),(x,t)\in\mathbb{S}^2\times\mathbb{Z}\}$, as defined in Section 2.1.2, for which it holds almost surely $T(\cdot,t)\in L^2(\mathbb{S}^2;\mathbb{R})$, $t\in\mathbb{Z}$.

Definition 2.2.4. $\{T(x,t),(x,t)\in\mathbb{S}^2\times\mathbb{Z}\}$ *is called Spherical Autoregressive Process of order* p *(written* SPHAR(p)*) if there exist* p *isotropic kernel operators* $\{\Phi_1,\dots,\Phi_p\}$ *and a spherical white noise* $\{Z(x,t),(x,t)\in\mathbb{S}^2\times\mathbb{Z}\}$ *such that*

$$T(x,t)-(\Phi_1T(\cdot,t-1))(x)-\cdots-(\Phi_pT(\cdot,t-p))(x)-Z(x,t)=0, \tag{2.2.2}$$

for all $(x,t)\in\mathbb{S}^2\times\mathbb{Z}$*, the equality holding both in the* $L^2(\Omega)$ *and in the* $L^2(\mathbb{S}^2\times\Omega)$ *sense.*

Remark 2.2.5. *It should be noted that the solution process is defined pointwise, i.e., for each (x,t) there exists a random variable defined on $(\Omega,\Im,\mathbb{P})$ such that the identity (2.2.2) holds.*

Let us define the eigenvalues $\{\phi_{\ell;j},\ \ell \geq 0,\ j=1,\dots,p\}$, which satisfy

$$\Phi_j Y_{\ell,m} = \phi_{\ell;j} Y_{\ell,m}, \quad \text{and} \quad k_j(\langle x,y\rangle) = \sum_{\ell=0}^{\infty} \phi_{\ell;j} \frac{2\ell+1}{4\pi} P_\ell(\langle x,y\rangle).$$

Hence, for any $t \in \mathbb{Z}$,

$$\Phi_j T(\cdot,t-j) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \phi_{\ell;j} a_{\ell,m}(t-j) Y_{\ell,m}, \quad \text{a.s. in } L^2(\mathbb{S}^2),$$

that is, $\Phi_j T(\cdot,t-j)$ admits a spectral representation in terms of spherical harmonics with coefficients $\{\phi_{\ell;j} a_{\ell m}(t-j),\ \ell\geq 0,\ m=-\ell,\dots,\ell\}$. Likewise, we obtain

$$a_{\ell,m}(t) = \phi_{\ell;1} a_{\ell,m}(t-1) + \dots + \phi_{\ell;p} a_{\ell,m}(t-p) + a_{\ell,m;Z}(t); \tag{2.2.3}$$

to ensure identifiability, we assume that there exists at least an ℓ such that $\phi_{\ell;p} \neq 0$, so that $\Pr\{\Phi_p T(\cdot,t)\neq 0\} > 0,\ t\in\mathbb{Z}$, see again [10].
Now, define the polynomials $\phi_\ell : \mathbb{C}\to\mathbb{C},\ \ell \geq 0$, such that

$$\phi_\ell(z) = 1 - \phi_{\ell;1} z - \dots - \phi_{\ell;p} z^p, \tag{2.2.4}$$

see Section 1.1.

Condition 2.2.6. *The sequence of polynomials (2.2.4) is such that*

$$|z| \leq 1 \ \Rightarrow\ \phi_\ell(z) \neq 0.$$

More explicitly, there are no roots in the unit disk, for all $\ell \geq 0$.

Remark 2.2.7. *This condition, together with the summability of $\left\{\phi^2_{\ell;j}\right\}$, ensures that the smallest root taken among all non-degenerate polynomials is bounded away from one. Indeed, if $\xi_{\ell;1},\dots,\xi_{\ell;r_\ell}$ are the distinct roots of the d_ℓ-degree polynomial (2.2.4), $1\leq d_\ell \leq p$, then*

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

uniformly over ℓ. Equivalently, there exists $\delta > 0$ such that

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

Example 2.2.8 (SPHAR(1)). *The family of random variables $\{T(x,t),\ (x,t)\in \mathbb{S}^2\times\mathbb{Z}\}$ is a spherical autoregressive process of order one if for all pairs $(x,t)\in\mathbb{S}^2\times\mathbb{Z}$ it satisfies*

$$T(x,t) = (\Phi_1 T(\cdot,t-1))(x) + Z(x,t); \tag{2.2.5}$$

in this case, Condition 2.2.6 simply becomes $|\phi_\ell| < 1$ or $|\phi_\ell| \leq \frac{1}{1+\delta}$, for all $\ell \geq 0$.

For all integer $\ell \geq 0$, it is standard to show that there exist real-valued sequences $\{\psi_{j;\ell},\ j \geq 0\}$ such that

$$\begin{cases} \psi_{j;\ell} = 1 & j = 0, \\ \psi_{j;\ell} - \sum_{0<k\leq j} \phi_{\ell;j}\psi_{j-k;\ell} = 0 & 1 < j < p, \\ \psi_{j;\ell} - \sum_{0<k\leq p} \phi_{\ell;j}\psi_{j-k;\ell} = 0 & j \geq p. \end{cases} \tag{2.2.6}$$

Excluding the degenerate cases, the sequences can be written explicitly as

$$\psi_{j;\ell} = \sum_{u=1}^{r_\ell} \sum_{v=1}^{s_{\ell;u}} c_{u,v;\ell} j^v \xi_{\ell;u}^{-j},$$

where $\left\{\xi_{\ell;u}^{-j} : u = 1, \ldots, r_\ell\right\}$ denotes the distinct roots of $\phi_\ell(z)$, each of them of multiplicity $s_{\ell;u}$, so that $\sum_{u=1}^{r_\ell} s_{\ell;u} = d_\ell \leq p$, for all ℓ; the constants $\{c_{u,v;\ell}\}$ are determined by the initial conditions (2.2.6), see [12, Section 3.3].
The proof of the following statement is given already in [10] for the simplest case of order one processes, but here we construct explicitly the solution with a slightly different argument for completeness.

Theorem 2.2.9. *Under Condition 2.2.6, the unique isotropic stationary solution to* (2.2.2) *is given by*

$$T(x,t) = \lim_{k\to\infty} T_k(x,t), \qquad T_k(x,t) = \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} \sum_{j=0}^{k} \psi_{j;\ell} a_{\ell m;Z}(t-j) Y_{\ell m}(x), \tag{2.2.7}$$

in the $L^2(\Omega)$ and $L^2(\mathbb{S}^2 \times \Omega)$ sense. The coefficients $\{\psi_{j;\ell}\}$ are determined by the relation

$$\psi_\ell(z) = \sum_{j=0}^{\infty} \psi_{j;\ell} z^j = 1/\phi_\ell(z), \qquad |z| \leq 1, \tag{2.2.8}$$

see also Equation (1.1.5).

Remark 2.2.10. *Notice that the isotropic stationary solutions of the* SPHAR(1) *equation* (2.2.5) *take the form*

$$T(\cdot,t) = \sum_{j=0}^{\infty} \Phi_1^j Z(\cdot,t),$$

and Condition 2.2.6 is satisfied if and only if the operator norm

$$\|\Phi_1\|_{\text{op}} := \max_{\ell\geq 0} |\phi_\ell| < 1,$$

see also [10, Section 3.4].

The proof is composed of two steps. First we show that the $T(x,t) = \lim_{k\to\infty} T_k(x,t)$ is a solution of the SPHAR(p) equation (2.2.2); and then we prove that any isotropic stationary solution of (2.2.2) takes the form (2.2.7).

Proof. First note that, under Condition 2.2.6, for any $\ell \geq 0$, $1/\phi_\ell(z)$ has a power series expansion, that is,

$$1/\phi_\ell(z) = \sum_{j=0}^{\infty} \psi_{j;\ell} z^j = \psi_\ell(z), \qquad |z| \leq 1,$$

and

$$\sum_{j=0}^{\infty} |\psi_{j;\ell}| < \infty,$$

see [12, Proof of Theorem 3.1.1, page 85].
Now, let us show first that the sequence $\{T_k\}$ is Cauchy. Indeed we have, for $k' > k$, $L_{k'} > L_k$,

$$\begin{aligned} T_{k'}(x,t) - T_k(x,t) &= \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} \sum_{j=k}^{k'} \psi_{j;\ell} a_{\ell,m;Z}(t-j) Y_{\ell,m}(x) \\ &+ \sum_{\ell=L_k+1}^{L_{k'}} \sum_{m=-\ell}^{\ell} \sum_{j=0}^{k'} \psi_{j;\ell} a_{\ell,m;Z}(t-j) Y_{\ell,m}(x) \end{aligned}$$

and, therefore,

$$\begin{aligned} &\mathbb{E}\left|T_{k'}(x,t) - T_k(x,t)\right|^2 \\ =& \sum_{\ell=0}^{L_k} \frac{2\ell+1}{4\pi} C_{\ell;Z} \sum_{j=k}^{k'} |\psi_{j;\ell}|^2 + \sum_{\ell=L_k+1}^{L_{k'}} \frac{2\ell+1}{4\pi} C_{\ell;Z} \sum_{j=0}^{k'} |\psi_{j;\ell}|^2 \\ \leq& \sum_{\ell=0}^{\infty} \frac{2\ell+1}{4\pi} C_{\ell;Z} \sum_{j=k}^{\infty} |\psi_{j;\ell}|^2 + \sum_{\ell=L_k+1}^{\infty} \frac{2\ell+1}{4\pi} C_{\ell;Z} \sum_{j=0}^{\infty} |\psi_{j;\ell}|^2. \end{aligned} \tag{2.2.9}$$

For $\ell \geq 0$, consider the stationary process

$$X_\ell(t) = \sum_{j=0}^{\infty} \psi_{j;\ell} \varepsilon(t-j), \qquad t \in \mathbb{Z};$$

here we take $\{\varepsilon(t),\ t \in \mathbb{Z}\}$ to be a white noise sequence with variance identically equal to one. The spectral density of $\{X_\ell(t),\ t \in \mathbb{Z}\}$ is given by (see [12])

$$\begin{aligned} f_\ell(\lambda) &= \frac{1}{2\pi} \left| \sum_{j=0}^{\infty} \psi_{j;\ell} \exp(i\lambda j) \right|^2 = \frac{1}{2\pi} |\psi_\ell(e^{i\lambda})|^2 = \frac{1}{2\pi} \frac{1}{|\phi_\ell(e^{i\lambda})|^2}, \\ \psi_\ell(e^{i\lambda}) &:= \sum_{j=0}^{\infty} \psi_{j;\ell} \exp(i\lambda j), \quad \phi_\ell(e^{i\lambda}) = 1 - \phi_{\ell;1} e^{i\lambda} - \cdots - \phi_{\ell;p} e^{ip\lambda}. \end{aligned}$$

Now, recall the identity

$$\mathbb{V}[X_\ell(t)] = \sum_{j=0}^{\infty} |\psi_{j;\ell}|^2 = \int_{-\pi}^{\pi} f_\ell(\lambda) d\lambda,$$

whence

$$\sum_{j=0}^{\infty} |\psi_{j;\ell}|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\psi_\ell(e^{i\lambda})|^2 \, d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|\phi_\ell(e^{i\lambda})|^2} d\lambda.$$

Moreover, under Condition 2.2.6, for the non-degenerate polynomials it holds that

$$|\phi_\ell(e^{i\lambda})| = \prod_{u=1}^{r_\ell} |1 - \xi_{\ell;u}^{-1} e^{i\lambda}|^{s_{\ell;u}} \geq \prod_{u=1}^{r_\ell} (1 - |\xi_{\ell;u}^{-1}|)^{s_{\ell;u}} \geq (1 - \xi_*^{-1})^p > 0 \ ,$$

see Remark 2.2.7; hence, as a consequence,

$$\sum_{j=0}^{\infty} |\psi_{j;\ell}|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{|\phi_\ell(e^{i\lambda})|^2} d\lambda \leq \left(\frac{\xi_*}{\xi_* - 1}\right)^{2p} < const,$$

uniformly over ℓ, and

$$\sum_{\ell=0}^{\infty}\sum_{j=0}^{\infty} |\psi_{j;\ell}|^2 \frac{2\ell+1}{4\pi} C_{\ell;Z} < const < \infty.$$

Then, by the Dominated Convergence Theorem, we have

$$\lim_{k\to\infty} \sum_{\ell=0}^{\infty}\sum_{j=k}^{\infty} |\psi_{j;\ell}|^2 \frac{2\ell+1}{4\pi} C_{\ell;Z} = \sum_{\ell=0}^{\infty} \left\{ \lim_{k\to\infty} \sum_{j=k}^{\infty} |\psi_{j;\ell}|^2 \right\} \frac{2\ell+1}{4\pi} C_{\ell;Z} = 0,$$

and (2.2.9) $\to 0$ as $k \to \infty$, so that $\{T_k\}$ is indeed a Cauchy sequence. The proof that it satisfies (2.2.2) is standard; we have that

$$\begin{aligned}
&\left\| T(x,t) - \sum_{j=1}^{p} (\Phi_j T(\cdot, t-j))(x) \right\|_{L^2(\Omega)} \\
&= \lim_{k\to\infty} \left\| T_k(x,t) - \sum_{j=1}^{p} (\Phi_j T_k(\cdot, t-j))(x) - Z(x,t) \right\|_{L^2(\Omega)} \\
&= \lim_{k\to\infty} \left\| \sum_{j=1}^{p}\sum_{\ell=1}^{L_k}\sum_{m=-\ell}^{\ell}\sum_{h=k-j+1}^{k} \phi_{\ell;j}\psi_{h;\ell} a_{\ell,m;Z}(t-j-h) Y_{\ell,m}(x) \right\|_{L^2(\Omega)} \\
&\leq \sum_{j=1}^{p} \lim_{k\to\infty} \left\| \sum_{\ell=1}^{L_k}\sum_{m=-\ell}^{\ell}\sum_{h=k-j+1}^{k} \phi_{\ell;j}\psi_{h;\ell} a_{\ell,m;Z}(t-j-h) Y_{\ell,m}(x) \right\|_{L^2(\Omega)} ,
\end{aligned}$$

which again is easily shown to be zero by $\lim_{k\to\infty} \sum_{h=k-j+1}^{\infty} |\psi_{h;\ell}|^2 = 0$ and Dominated Convergence Theorem. The argument involving the $L^2(\mathbb{S}^2 \times \Omega)$ limit is analogous.

To complete the proof, we need to show that if $\{U(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is an isotropic stationary solution of (2.2.2), then we must have

$$U(x,t) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\sum_{j=0}^{\infty} \psi_{j;\ell} a_{\ell,m;Z}(t-j) Y_{\ell,m}(x)$$

in $L^2(\Omega \times \mathbb{S}^2)$ and $L^2(\Omega)$. If $\{U(x,t),(x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is an isotropic stationary solution of (2.2.2), then $a_{\ell,m;U}(t) = \int_{\mathbb{S}^2} U(x,t)Y_{\ell,m}(x)dx$ is a stationary solution of the standard AR(p) equation and, under Condition 2.2.6,

$$a_{\ell,m;U}(t) = \sum_{j=0}^{\infty} \psi_{j;\ell} a_{\ell,m;Z}(t-j), \qquad \text{in } L^2(\Omega).$$

Then, by stationarity, $\mathbb{E}|a_{\ell,m;U}(t)|^2 = C_{\ell;U}$ and

$$\left\| \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} a_{\ell,m;U}(t)Y_{\ell,m} - \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} \sum_{j=0}^{k} \psi_{j;\ell} a_{\ell,m;Z}(t-j)Y_{\ell,m} \right\|_{L^2(\Omega\times\mathbb{S}^2)}$$
$$= \sum_{\ell=0}^{L_k} \sum_{j=k+1}^{\infty} |\psi_{j;\ell}|^2 (2\ell+1) C_{\ell;U},$$

which goes to zero as $k \to \infty$. Hence, by triangular inequality,

$$\left\| U(\cdot,t) - \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} \sum_{j=0}^{k} \psi_{j;\ell} a_{\ell,m;Z}(t-j)Y_{\ell,m} \right\|_{L^2(\Omega\times\mathbb{S}^2)} \to 0, \qquad k \to \infty.$$

The same result holds in the sense of convergence in $L^2(\Omega)$, for every fixed pair (x,t). Indeed, we have

$$\mathbb{E}\left| \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} a_{\ell,m;U}(t)Y_{\ell,m}(x) - \sum_{\ell=0}^{L_k} \sum_{m=-\ell}^{\ell} \sum_{j=0}^{k} \psi_{j;\ell} a_{\ell,m;Z}(t-j)Y_{\ell,m}(x) \right|^2$$
$$= \sum_{\ell=0}^{L_k} \sum_{j=k+1}^{\infty} |\psi_{j;\ell}|^2 \frac{2\ell+1}{4\pi} C_{\ell;U}.$$

□

In the chapters to follow, we are going to introduce two estimation procedures for the spherical autoregressive kernels $\{k_j : j = 1, \dots, p\}$ and investigate asymptotic properties of the corresponding nonparametric estimators. Specifically, in Chapter 3, we focus on the solutions of a functional L^2-minimization problem, while, in Chapter 4, we add a convex penalty term to study LASSO-type estimators under sparsity assumptions.

Chapter 3

Asymptotics for Spherical Functional Autoregressions[1]

3.1 Introduction

Our purpose in this chapter is to investigate a class of space-time processes, which can be viewed as functional autoregressions taking values in $L^2(\mathbb{S}^2)$; we refer to [10] for a general textbook analysis of functional autoregressions taking values in Hilbert spaces, and [3, 31, 56, 55] for a very partial list of some important recent references. Dealing with functional spherical autoregressions ensures some very convenient simplifications; in particular, we exploit the analytic properties of the standard orthonormal basis of $L^2(\mathbb{S}^2)$ and some natural isotropy requirements to obtain neat expressions for the autoregressive operators, which are then estimated by a form of frequency-domain least squares. For our estimators, we are able to establish rates of consistency (in L^2 and L^∞ norms) and a quantitative version of the Central Limit Theorem, in Wasserstein distance. In particular, we derive explicit bounds for the rate of convergence to the limiting Gaussian distribution by means of the rich machinery of Stein-Malliavin methods (see [49] and Section 1.3); to the best of our knowledge, this is the first Quantitative Central Limit Theorem established in the framework of functional-valued stationary processes. Under stronger regularity conditions, we are able to establish a weak convergence result for the kernel estimators; our results are then illustrated by simulations.
The plan of this chapter is then as follows: in Section 3.2 we recall our basic model; we show how, under isotropy, the model enjoys a number of symmetry properties which greatly simplify our approach. The main results are then collected in Section 3.3, where we investigate rates of convergence and the Quantitative Central Limit Theorem; we consider also weak convergence in $C_p\left([-1,1]\right)$, under stronger regularity conditions for the autoregressive kernels. Large parts of the proofs and many auxiliary lemmas, some of possible independent interest, are collected in Sections 3.4 and in the Appendix (Section 3.6). Finally, Section 3.5 provides numerical estimates on the behaviour of our procedures.

[1]This chapter is partially based on the preprint *Asymptotics for Spherical Functional Autoregressions* [18], written jointly with Domenico Marinucci, accepted for publication in *The Annals of Statistics*.

Notation. Throughout this chapter we consider the real-valued basis of spherical harmonics already defined in Section 1.2, and therefore the random coefficients will be always real-valued random variables for all (ℓ, m).
In the sequel, given any two positive sequences $\{a_k,\ k \in \mathbb{N}\}$, $\{b_k,\ k \in \mathbb{N}\}$, we shall write $a_k \sim b_k$ if $\exists c_1, c_2 > 0$ such that $c_1 b_k \leq a_k \leq c_2 b_k, \forall k \in \mathbb{N}$. In addition, we will denote with *const* a positive real constant, which may change from line to line; also, we use $\|\cdot\|_{L^2(\mathbb{S}^2)}$ for the usual L^2 norm on the sphere, $\Lambda_{\min}(A)$ and $\Lambda_{\max}(A)$ for the minimum and maximum eigenvalues of the matrix A, respectively, $\|A\|_{\text{op}}$ for the operator norm of A, i.e., $\|A\|_{\text{op}} = \sqrt{\lambda_{\max}(A'A)}$, and $\text{Tr}(A)$ for the trace of A.

3.2 Background and Assumptions

We briefly recall our model of interest, already introduced in Chapter 2. As usual, by space-time spherical random field we mean a collection of random variables $\{T(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ such that, for every $t \in \mathbb{Z}$, the mapping $(x, \omega) \mapsto T(x, t, \omega)$ is $\mathfrak{B}(\mathbb{S}^2) \times \mathfrak{F}$-measurable, for some probability space $(\Omega, \mathfrak{F}, \mathbb{P})$.
In particular, we consider (zero-mean) isotropic stationary random fields (see Definition 2.1.14), which are also Gaussian. In this case, of course, weak isotropy-stationarity entails strong isotropy-stationarity, i.e., the law of $T(g\cdot, \cdot + \tau)$ is the same as the law of $T(\cdot, \cdot)$, in the sense of processes, for all $g \in SO(3)$ and $\tau \in \mathbb{Z}$.
Thus, recall that, for any fixed $t \in \mathbb{Z}$, there exists a random element T_t such that $T(\cdot, t) = T_t$ $\mathbb{P}$-a.s. (see Remark 2.1.16) and

$$T(x,t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell,m}(t) Y_{\ell,m}(x), \tag{3.2.1}$$

where $\{Y_{\ell,m}(\cdot),\ \ell \geq 0,\ m = -\ell, \dots, \ell\}$ is a standard basis of *real* spherical harmonics (see Remark 1.2.2) and $\{a_{\ell,m}(t),\ \ell \geq 0,\ m = -\ell, \dots, \ell\}$ are (zero-mean) random coefficients which satisfy

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

The sequence $\{C_\ell(0),\ \ell \geq 0\}$ corresponds to the angular power spectrum of the spherical field at a given time point, for which we will simply write $\{C_\ell\}$. Also recall that, for fixed $t, s \in \mathbb{Z}$, the covariance function $\Gamma(x,t,y,s)$ is easily shown to have a spectral decomposition in terms of Legendre polynomials (Schoenberg's Theorem 1.2.7, see also [8]), i.e., for every $(x,t), (y,s) \in \mathbb{S}^2 \times \mathbb{Z}$,

$$\Gamma(x,t,y,s) = \sum_{\ell=0}^{\infty} C_\ell(t-s) \frac{2\ell+1}{4\pi} P_\ell(\langle x, y \rangle).$$

Consider now a Gaussian spherical white noise $\{Z(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ (see Definition 2.2.1), i.e., a sequence of independent and identically distributed Gaussian isotropic spherical random fields with the same covariance function

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

Moreover, assume that the collection of random variables $\{T(x,t),(x,t)\in\mathbb{S}^2\times\mathbb{Z}\}$ satisfies the autoregressive equation

$$T(x,t)-(\Phi_1 T(\cdot,t-1))(x)-\cdots-(\Phi_p T(\cdot,t-p))(x)-Z(x,t)=0, \tag{3.2.2}$$

for all $(x,t)\in\mathbb{S}^2\times\mathbb{Z}$, where $\{\Phi_j : j=1,\ldots,p\}$ are p integral operators (see Definition 2.2.3) associated with p continuous isotropic kernel $\{k_j : j=1,\ldots,p\}$. Hence, for each $j=1,\ldots,p$, it holds that

$$k_j(\langle x,y\rangle)=\sum_{\ell=0}^{\infty}\phi_{\ell;j}\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle),$$

where $\{\phi_{\ell;j},\ell\geq 0\}$ are the eigenvalues of Φ_j and, for any $(x,t)\in\mathbb{S}^2\times\mathbb{Z}$, we have

$$(\Phi_j T(\cdot,t-j))(x)=\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}\phi_{\ell;j}a_{\ell,m}(t-j)Y_{\ell,m}(x), \qquad \text{in } L^2(\Omega).$$

Furthermore,

$$a_{\ell,m}(t)=\phi_{\ell;1}a_{\ell,m}(t-1)+\cdots+\phi_{\ell;p}a_{\ell,m}(t-p)+a_{\ell,m;Z}(t),$$

and it is well-defined the sequence of associated polynomials $\phi_\ell:\mathbb{C}\to\mathbb{C},\ \ell\geq 0$, such that

$$\phi_\ell(z)=1-\phi_{\ell;1}z-\cdots-\phi_{\ell;p}z^p. \tag{3.2.3}$$

Condition 3.2.1. *The sequence of polynomials* (3.2.3) *is such that*

$$|z|<1+\delta \ \Rightarrow\ \phi_\ell(z)\neq 0,$$

for some $\delta>0$. More explicitly, there are no roots in a δ-enlargement of the unit disk, for all $\ell\geq 0$.

Remark 3.2.2. *Recall that, under Condition 3.2.1, Equation* (3.2.2) *admits a unique isotropic stationary solution; the proof is given in Section 2.2, along the same lines as in [10].*

Remark 3.2.3. *The autocovariance function of an isotropic stationary* SPHAR(1) *process (see Example 2.2.8) is easily seen to be given by (writing $\tau=t-s$)*

$$\Gamma(x,t,y,s)=\sum_{\ell=0}^{\infty}C_\ell(\tau)\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle)=\sum_{\ell=0}^{\infty}\frac{\phi_\ell^{|\tau|}C_{\ell;Z}}{1-\phi_\ell^2}\frac{2\ell+1}{4\pi}P_\ell(\langle x,y\rangle).$$

It is easy hence to envisage a number of parametric models for sphere-time covariances; for instance, a simple proposal is

$$\begin{aligned}
&\phi_\ell=G\times\{|\ell-\ell^*|+1\}^{-\alpha_\phi}, && \textit{with } \ell^*\geq 0,\ \alpha_\phi>2,\ 0<G<1,\\
&C_{\ell;Z}=G_Z(1+\ell)^{-\alpha_Z}, && \textit{with } \alpha_Z>2.
\end{aligned}$$

Here, the parameters α_Z,α_ϕ control, respectively, the smoothness of the innovation process and the regularity of the autoregressive kernel (see [37]); the positive integer

ℓ^* can be seen as a sort of "characteristic scale", where the power of the kernel is concentrated. More generally, we can take $\phi_\ell = G(\ell; \alpha_1, \ldots, \alpha_q)$, where $\alpha_1, \ldots, \alpha_q$ are fixed parameters and G is any function such that

$$\sup_\ell |G(\ell; \alpha_1, \ldots, \alpha_q)| < 1 \text{ and } \sum_\ell (2\ell + 1) |G(\ell; \alpha_1, \ldots, \alpha_q)| < \infty,$$

uniformly over all values of $(\alpha_1, \ldots, \alpha_q)$.

Condition 3.2.4 (Identifiability)**.** *The Gaussian spherical white noise process $\{Z(x,t),\ (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ is such that $C_{\ell;Z} > 0$, for all $\ell \geq 0$.*

Remark 3.2.5. *The previous condition is an identifiability assumption; indeed, it is simple to verify from our arguments below that for $C_{\ell;Z} = 0$ the component of the kernel corresponding to the ℓ-th multipole is not observable, i.e., the* AR(p) *process has the same distribution whatever the values of $\{\phi_{\ell;j}, j = 1, \ldots, p\}$. It is possible, however, to estimate the "sufficient" version of the kernel, i.e., its projection on the relevant subspace, such that $C_{\ell,Z} > 0$. The extension is straightforward and we avoid it just for brevity and notational simplicity. Of course, as a consequence we have that*

$$\int_{\mathbb{S}^2 \times \mathbb{S}^2} \Gamma_Z(x,y) f(x) f(y) dx dy > 0, \qquad \forall f(\cdot) \in L^2(\mathbb{S}^2),\ f(\cdot) \neq 0.$$

3.3 Main Results

Throughout this chapter, we shall assume to be able to observe the projections of the fields on the orthonormal basis $\{Y_{\ell m}\}$, i.e., we assume to observe

$$a_{\ell,m}(t) := \int_{\mathbb{S}^2} T(x,t) Y_{\ell,m}(x) dx, \qquad t = 1, \ldots, n.$$

The estimator we shall focus on is a form of least squares regression on an increasing subset of the real orthonormal system $\{Y_{\ell,m}\}$; more precisely, we shall define $k(\cdot) := (k_1(\cdot), \cdots, k_p(\cdot))'$ for the vector of nuclear kernels, a growing sequence of integers L_N, $L_N \to \infty$ as $N \to \infty$; and a vector of estimators

$$\widehat{k}_N(\cdot) := (\widehat{k}_{1;N}(\cdot), \ldots, \widehat{k}_{p;N}(\cdot))' = \operatorname*{arg\,min}_{k(\cdot) \in \mathcal{P}_N^p} \sum_{t=1}^{N} \left\| T_{t+p} - \sum_{j=1}^{p} \Phi_j T_{t+p-j} \right\|_{L^2(\mathbb{S}^2)}^2, \tag{3.3.1}$$

where $N := n - p$, $N > p$, and $\mathcal{P}_N^p$ is the Cartesian product of p copies of

$$\mathcal{P}_N = \operatorname{span} \left\{ \frac{2\ell + 1}{4\pi} P_\ell(\cdot),\ \ell \leq L_N \right\}.$$

As common in the autoregressive context, we drop the first p observations when computing our estimators, in order to avoid initialization issues. We shall write $\mathcal{L}_N(\cdot)$ for the function $\mathcal{L}_N : [-1,1] \to \mathbb{R}$,

$$\mathcal{L}_N(z) = \sum_{\ell=0}^{L_N} \frac{2\ell + 1}{16\pi^2} P_\ell^2(z), \qquad z \in [-1,1]. \tag{3.3.2}$$

Note that

$$\mathcal{L}_N(1) = \mathcal{L}_N(-1) = \sum_{\ell=0}^{L_N} \frac{2\ell+1}{16\pi^2} = \frac{(L_N+1)^2}{16\pi^2};$$

on the other hand, for $z \in (-1,1)$ we have the identity (see [67, 28])

$$\sum_{\ell=0}^{L_N} \frac{2\ell+1}{16\pi^2} P_\ell^2(z) = \frac{L_N+1}{16\pi^2}\left[P'_{L_N+1}(z)P_{L_N}(z) - P'_{L_N}(z)P_{L_N+1}(z)\right];$$

it is then possible to show that (see Lemma 3.6.4 in the Appendix)

$$\mathcal{L}_N(z) \simeq \frac{2L_N}{\pi\sqrt{1-z^2}}, \qquad \text{as } L_N \to \infty, \tag{3.3.3}$$

where $\simeq$ indicates that the ratio of left- and right-hand sides converges to unity. For our results, we need slightly stronger assumptions on the "high frequency" behaviour of the kernels $k_j(\cdot)$. More precisely, we shall introduce the following:

Condition 3.3.1 (Smoothness)**.** *For all $j = 1, \dots, p$, there exists positive constants β_j, γ_j such that*

$$|\phi_{\ell;j}| \le \frac{\gamma_j}{\ell^{\beta_j}}, \qquad \beta_j > 1, \ \ell > 0. \tag{3.3.4}$$

We let $\beta_* = \min_{j\in\{1,\dots,p\}} \beta_j$. We shall say that this condition is satisfied in the strong *sense if $\beta_j > 2, \ j = 1, \dots, p$.*

Remark 3.3.2. *It is readily seen that Condition 3.3.1 leads to Hilbert-Schmidt operators, since it implies $\sum_\ell (2\ell+1)\phi^2_{\ell;j} < \infty$, $j = 1, \dots, p$; whereas the strong version Condition 3.3.1 is specific for nuclear operators, since it entails $\sum_\ell (2\ell+1)|\phi_{\ell;j}| < \infty$, $j = 1, \dots, p$, see again [32].*

Remark 3.3.3. *Condition 3.3.1 is interpretable in terms of the regularity of each kernel $k_j(\cdot)$. Indeed, in [37] it is shown that*

$$\sum_{\ell=0}^{\infty} |\phi_{\ell;j}|^2 \frac{2\ell+1}{4\pi}(1+\ell^{2\eta}) < \infty$$

implies integrability of the first η derivatives of $k_j(\cdot)$, i.e., $k_j(\cdot)$ belongs to the Sobolev space $W_{1,\eta}$.

Our first result refers to the asymptotic consistency of the kernel estimators that we just introduced.

Theorem 3.3.4 (Consistency)**.** *Consider $\widehat{k}_N(\cdot)$ in Equation (3.3.1). Under Conditions 3.2.1, 3.2.4 and 3.3.1, for $L_N \sim N^d$, $0 < d < 1$, we have that*

$$\mathbb{E}\left[\int_{-1}^{1} \left\|\widehat{k}_N(z) - k(z)\right\|^2 dz\right] = \mathcal{O}\left(N^{d-1} + N^{2d(1-\beta_*)}\right). \tag{3.3.5}$$

Moreover, under Conditions 3.2.1, 3.2.4 and 3.3.1 (in the strong *sense), for $L_N \sim N^d$, $0 < d < \frac{1}{3}$,*

$$\mathbb{E}\left[\sup_{z\in[-1,1]} \left\|\widehat{k}_N(z) - k(z)\right\|\right] = \mathcal{O}\left(N^{(3d-1)/2} + N^{d(2-\beta_*)}\right).$$

Remark 3.3.5 (Optimal choice of d)**.** *The optimal choice of d, in terms of the best convergence rates, is given by $d^* = \frac{1}{2\beta_*-1}$, leading to the exponents $\frac{2-2\beta_*}{2\beta_*-1}$ and $\frac{2-\beta_*}{2\beta_*-1}$, respectively. Heuristically, the result can be explained as follows: larger values of β_* entail higher regularity/smoothness properties of the kernels to be estimated; as usual in nonparametric estimation, more regular functions can be estimated with better convergence rates, as the bias term is controlled more efficiently. Indeed, for $d = d^*$ and $\beta_* \to \infty$, the mean squared error approximates the parametric rate $1/N$, as expected.*

Remark 3.3.6 (Plug-in estimates)**.** *For applications to empirical data, the optimal rate can be implemented by means of plug-in techniques, i.e., estimating (under additional regularity conditions) the value of the parameter β_* by means of first step-estimators of the coefficients $\{\phi_{\ell,j}\}$. Let us sketch the main ideas for this approach, omitting some details for brevity. Consider for simplicity the* SPHAR(1) *case, and let us make Condition 3.3.1 stronger by assuming that*

$$|\phi_\ell| = \frac{\gamma}{\ell^\beta} + o\left(\frac{1}{\ell^\beta}\right), \qquad \text{some } \gamma > 0,\ \beta > 1,\ \forall \ell > 0.$$

Consider the estimator

$$\widehat{\phi}_{\ell,N} := \frac{\sum_t a_{\ell,m}(t-1)a_{\ell,m}(t)}{\sum_t a_{\ell,m}^2(t-1)}, \qquad \ell = 0,1,2,\dots,$$

from which we can now build the pseudo log-regression model

$$\log \widehat{\phi}_{\ell,N}^2 = \log \frac{\widehat{\phi}_{\ell,N}^2}{\gamma^2 \ell^{-2\beta}} + \log\left(\gamma^2 \ell^{-2\beta}\right) = \log\left(\gamma^2\right) - 2\beta \log \ell + v_\ell$$
$$v_\ell := \log \frac{\widehat{\phi}_{\ell,N}^2}{\gamma^2 \ell^{-2\beta}}, \qquad \ell = 0,1,2,\dots,$$

where the "regression residuals" $\{v_\ell\}$ are independent over ℓ, with asymptotically mean zero and bounded variance as $N \to \infty$. It is then possible to study the asymptotic consistency of the OLS-like estimator (see also [63] for the related log-periodogram estimator)

$$\widehat{\beta}_N := -\frac{\sum_\ell \left\{\log \ell \times \log \widehat{\phi}_{\ell,N}^2\right\}}{2\sum_\ell \{\log \ell\}^2}.$$

The optimal rates can then be consistently estimated by means of the plug-in estimates $\widehat{d}_N^* = \frac{1}{2\widehat{\beta}_N - 1}$.
A more rigorous and complete investigation on these issues is currently in preparation and is not reported here for brevity's sake.

Our second result refers to a Quantitative Central Limit Theorem for the kernel estimators. Consider $\widehat{k}_N(\cdot)$ in Equation (3.3.1) and, for any $m \in \mathbb{N}$, any $z_1, \dots, z_m \in (-1,1)$, $z_1 \neq \dots \neq z_m$, define the $mp \times 1$ vectors

$$K_N = K_N(z_1, z_2, \dots, z_m) := \begin{pmatrix} \sqrt{\frac{N}{\mathcal{L}_N(z_1)}} \left(\widehat{k}_N(z_1) - k(z_1)\right) \\ \vdots \\ \sqrt{\frac{N}{\mathcal{L}_N(z_m)}} \left(\widehat{k}_N(z_m) - k(z_m)\right) \end{pmatrix}, \quad Z \overset{d}{=} \mathcal{N}_{mp}\left(0_{mp}, I_{mp}\right).$$

Theorem 3.3.7. *Under Conditions 3.2.1, 3.2.4 and 3.3.1 (in the strong sense), for* $L_N \sim N^d$, $d > \frac{1}{2\beta_* - 2}$, *we have that*

$$d_W(Z, K_N) = \mathcal{O}\left(N^{-1/2} + N^{1/2+d(1-\beta_*)} + N^{-d}\log N\right).$$

An immediate Corollary is the following.

Corollary 3.3.8. *Under the same Conditions and notation as in Theorem 3.3.7, for any fixed* $z \in [-1, 1]$, *we have that*

$$\sqrt{\frac{N}{\mathcal{L}_N(z)}}\left(\widehat{k}_N(z) - k(z)\right) \to \mathcal{N}_p\left(0_p, I_p\right), \qquad N \to \infty.$$

Remark 3.3.9. *As usual, the values of* d *that guarantee asymptotic normality do not minimize the mean squared error; in fact, we have that* $d^* = \frac{1}{2\beta_*-1} < \frac{1}{2\beta_*-2}$, *which is the minimal value of* d *for Theorem 3.3.7 to hold. Indeed, asymptotic Gaussianity requires undersmoothing, i.e., a value of* d *which makes the asymptotic bias negligible, rather than of the same order as the variance. Once again the rate can be taken to approach* $N^{-1/2}$ *for* $\beta_* \to \infty$.

For our third and final result, we need to strengthen the conditions on the regularity of the autoregressive kernels.

Condition 3.3.10. *The kernel* $k_j(\cdot)$ *admits a finite expansion in the Legendre basis, i.e., there exist an (arbitrary large but finite) integer* $L > 0$ *such that*

$$\int_{-1}^{1} k_j(z) P_\ell(z) dz = 0, \qquad \text{for all } j = 1, \ldots, p \text{ and } \ell > L.$$

Condition 3.3.10 clearly implies that there exist finite integers $L_1, \ldots, L_p \leq L$ such that

$$k_j(z) = \sum_{\ell=0}^{L_j} \frac{2\ell+1}{4\pi} \phi_{\ell;j} P_\ell(z), \qquad z \in [-1, 1],\ j = 1, \ldots, p;$$

we also need to introduce, for $\ell = 0, 1, 2, \ldots$, the $p \times p$ autocovariance matrix

$$\Gamma_\ell := \begin{pmatrix} C_\ell & C_\ell(1) & \cdots & C_\ell(p-1) \\ C_\ell(1) & C_\ell & \cdots & C_\ell(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_\ell(p-1) & C_\ell(p-2) & \cdots & C_\ell \end{pmatrix},$$

and we shall write $W_p(\cdot)$ for the zero-mean, p-dimensional Gaussian process with covariance function

$$\Gamma_k(z, z') = \sum_{\ell=0}^{L} C_{\ell;Z} \Gamma_\ell^{-1} \frac{2\ell+1}{16\pi^2} P_\ell(z) P_\ell(z').$$

We are now able to state our last theorem.

Theorem 3.3.11. *Under Conditions 3.2.1, 3.2.4 and 3.3.10, we have that*

$$\sqrt{N}\left(\widehat{k}_N(\cdot) - k(\cdot)\right) \Longrightarrow W_p(\cdot)\,, \qquad N \to \infty,$$

where $\Longrightarrow$ denotes weak convergence in $C_p([-1,1])$ (the space of continuous functions from $[-1,1]$ to $\mathbb{R}^p$, with the standard uniform metric).

Remark 3.3.12. *At first sight, it may look surprising that the weak convergence for the estimators in Theorem 3.3.11 occurs at a faster rate $\sqrt{N}$ than the convergence in finite-dimensional distributions of Theorem 3.3.7. This comparison, however, is misleading; indeed, in Theorem 3.3.7 we are not assuming the expansion of the kernels to be finite, and therefore we need to include a growing number of multipoles L_N, to ensure that bias terms are asymptotically negligible. On the other hand, note that weak convergence cannot hold under the conditions of Theorem 3.3.7, as the limiting finite dimensional distributions correspond to Gaussian independent random variables for any choice of fixed points $(z_1, \dots, z_m)$: no Gaussian process with measurable trajectories can have these finite-dimensional distributions. The limiting distribution is characterized by the nuisance parameters $\{C_\ell, C_\ell(1), \dots, C_\ell(p-1), C_{\ell;Z}\}$; for brevity's sake, estimation of these parameters is deferred to future work.*

3.4 Proofs of the Main Results

We now present the main arguments of our proofs, which are based on a number of technical results collected in the Appendix (Section 3.6). For $\ell = 0, 1, 2, \dots,$ it is convenient to introduce the $N(2\ell+1)$-dimensional vectors

$$Y_{\ell;N} := (a_{\ell,-\ell}(p+1), \dots, a_{\ell,\ell}(p+1), \dots, a_{\ell,\ell}(n))',$$
$$\varepsilon_{\ell;N} := (a_{\ell,-\ell;Z}(p+1), \dots, a_{\ell,\ell;Z}(p+1), \dots, a_{\ell,\ell;Z}(n))';$$

moreover, let us consider the $N(2\ell+1) \times p$ matrix

$$X_{\ell;N} := \{Y_{\ell;N-1} : Y_{\ell;N-2} : \dots : Y_{\ell;N-p}\}\,,$$

where

$$Y_{\ell;N-j} := (a_{\ell,-\ell}(p+1-j), \dots, a_{\ell,\ell}(p+1-j), \dots, a_{\ell,\ell}(n-j))', \quad j = 1, \dots, p.$$

We start from the proof of the consistency results.

Proof (Theorem 3.3.4). It is easy to see that we have

$$\widehat{k}_N(\cdot) = \underset{k(\cdot)\,\in\,\mathcal{P}_N^p}{\arg\min} \sum_{t=p+1}^{n} \left\| T_t - \sum_{j=1}^{p} \Phi_j T_{t-j} \right\|^2_{L^2(\mathbb{S}^2)}$$
$$= \sum_{\ell=0}^{L_N} \widehat{\boldsymbol{\phi}}_{\ell;N} \frac{2\ell+1}{4\pi} P_\ell(\cdot),$$

where

$$\widehat{\boldsymbol{\phi}}_{\ell;N} := (\widehat{\phi}_{\ell;N}(1), \dots, \widehat{\phi}_{\ell;N}(p))'$$

$$= \arg\min_{\phi_\ell \in \mathbb{R}^p} \sum_{t=p+1}^{n} \sum_{m=-\ell}^{\ell} \left(a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right)^2 .$$

Now, let $r_N(z)$ be the difference between the kernel and its truncated version

$$k_N(\cdot) = \sum_{\ell=0}^{L_N} \boldsymbol{\phi}_\ell \frac{2\ell+1}{4\pi} P_\ell(z),$$

i.e.,

$$r_N(z) = k(z) - k_N(z) = \sum_{\ell=L_N+1}^{\infty} \boldsymbol{\phi}_\ell \frac{2\ell+1}{4\pi} P_\ell(z),$$

where the equality holds in the L^2-sense. Then,

$$\mathbb{E}\left[\int_{-1}^{1} \left\| \widehat{k}_N(z) - k(z) \right\|^2 dz\right] = \mathbb{E}\left[\int_{-1}^{1} \left\| \widehat{k}_N(z) - k_N(z) \right\|^2 dz\right] + \int_{-1}^{1} \|r_N(z)\|^2 dz, \tag{3.4.1}$$

since $\mathbb{E}\left[\int_{-1}^{1} \left\langle \widehat{k}_N(z) - k_N(z), r_N(z) \right\rangle dz\right] = 0$, from orthogonality of Legendre polynomials.

Now notice that

$$\begin{aligned}
\int_{-1}^{1} \left\| \widehat{k}_N(z) - k_N(z) \right\|^2 dz &= \sum_{\ell=0}^{L_N} \sum_{\ell'=0}^{L_N} \left\langle \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell, \widehat{\boldsymbol{\phi}}_{\ell';N} - \boldsymbol{\phi}_{\ell'} \right\rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \int_{-1}^{1} P_\ell(z) P_{\ell'}(z) dz \\
&= \sum_{\ell=0}^{L_N} \sum_{\ell'=0}^{L_N} \left\langle \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell, \widehat{\boldsymbol{\phi}}_{\ell';N} - \boldsymbol{\phi}_{\ell'} \right\rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \frac{2}{2\ell+1} \delta_\ell^{\ell'} \\
&= \sum_{\ell=0}^{L_N} \left\| \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell \right\|^2 \frac{2\ell+1}{8\pi^2}.
\end{aligned}$$

Then, from Lemma 3.6.2 in the Appendix,

$$\mathbb{E}\left[\int_{-1}^{1} \left\| \widehat{k}_N(z) - k_N(z) \right\|^2 dz\right] = \sum_{\ell=0}^{L_N} \mathbb{E} \left\| \widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell \right\|^2 \frac{2\ell+1}{8\pi^2} \leq const \frac{L_N+1}{N}.$$

On the other hand,

$$\begin{aligned}
\int_{-1}^{1} \|r_N(z)\|^2 dz &= \sum_{\ell=L_N+1}^{\infty} \sum_{\ell'=L_N+1}^{\infty} \langle \boldsymbol{\phi}_\ell, \boldsymbol{\phi}_{\ell'} \rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \int_{-1}^{1} P_\ell(z) P_{\ell'}(z) dz \\
&= \sum_{\ell=L_N+1}^{\infty} \sum_{\ell'=L_N+1}^{\infty} \langle \boldsymbol{\phi}_\ell, \boldsymbol{\phi}_{\ell'} \rangle \frac{2\ell+1}{4\pi} \frac{2\ell'+1}{4\pi} \frac{2}{2\ell+1} \delta_\ell^{\ell'} \\
&= \sum_{\ell=L_N+1}^{\infty} \|\boldsymbol{\phi}_\ell\|^2 \frac{2\ell+1}{8\pi^2}.
\end{aligned}$$

Therefore, under Condition 3.3.1 and for $L_N \sim N^d$, $0 < d < 1$, we have

$$\int_{-1}^{1} \|r_N(z)\|^2 dz = \mathcal{O}\left(N^{2d(1-\beta_*)}\right),$$

and
$$\mathbb{E}\left[\int_{-1}^{1}\left\|\widehat{k}_N(z)-k(z)\right\|^2 dz\right]=\mathcal{O}\left(N^{d-1}+N^{2d(1-\beta_*)}\right),$$
where $\beta_*=\min_{j\in\{1,\dots,p\}}\beta_j$, as claimed.
Under the strong version of Condition 3.3.1, each kernel $k_j(\cdot)$ is defined for all $z\in[-1,1]$ as the pointwise limit of its expansion in terms of Legendre polynomials and
$$\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\widehat{k}_N(z)-k(z)\right\|\right]\leq\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\widehat{k}_N(z)-k_N(z)\right\|\right]+\sup_{z\in[-1,1]}\|r_N(z)\|,$$
by the triangle inequality. Hence, for the first component we have
$$\begin{aligned}\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\sum_{\ell=0}^{L_N}\left(\widehat{\phi}_{\ell;N}-\phi_\ell\right)\frac{2\ell+1}{4\pi}P_\ell(z)\right\|\right]&\leq\sum_{\ell=0}^{L_N}\mathbb{E}\left\|\widehat{\phi}_{\ell;N}-\phi_\ell\right\|\frac{2\ell+1}{4\pi}\\&\leq const\sum_{\ell=0}^{L_N}\frac{\sqrt{2\ell+1}}{\sqrt{N}}\\&\leq const\frac{(L_N+1)^{3/2}}{\sqrt{N}},\end{aligned}$$
again in view of Lemma 3.6.2 in the Appendix and the Cauchy-Schwartz inequality. On the other hand,
$$\sup_{z\in[-1,1]}\|r_N(z)\|\leq\sum_{\ell=L_N+1}^{\infty}\|\phi_\ell\|\frac{2\ell+1}{4\pi}.$$
Therefore, again under the strong version of Condition 3.3.1 and for $L_N\sim N^d$, $0<d<\frac{1}{3}$, we have
$$\sup_{z\in[-1,1]}\|r_N(z)\|=\mathcal{O}\left(N^{d(2-\beta_*)}\right)$$
and thus
$$\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\widehat{k}_N(z)-k(z)\right\|\right]=\mathcal{O}\left(N^{(3d-1)/2}+N^{d(2-\beta_*)}\right).$$
as claimed. □

We are now in the position to establish the Quantitative Central Limit Theorem.

Proof (Theorem 3.3.7). Let us recall that the minimizing estimator takes the form
$$\begin{aligned}\widehat{k}_N(\cdot)&=\arg\min_{k(\cdot)\in\mathcal{P}_N^p}\sum_{t=p+1}^{n}\left\|T_t-\sum_{j=1}^{p}\Phi_jT_{t-j}\right\|^2_{L^2(\mathbb{S}^2)}\\&=\sum_{\ell=0}^{L_N}\widehat{\phi}_{\ell;N}\frac{2\ell+1}{4\pi}P_\ell(\cdot),\end{aligned}$$

where

$$\widehat{\boldsymbol{\phi}}_{\ell;N} = \underset{\boldsymbol{\phi}_\ell \in \mathbb{R}^p}{\arg\min} \sum_{t=p+1}^{n} \sum_{m=-\ell}^{\ell} \left(a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right)^2$$
$$= (X'_{\ell;N} X_{\ell;N})^{-1} X'_{\ell;N} Y_{\ell;N} \; = \boldsymbol{\phi}_\ell + (X'_{\ell;N} X_{\ell;N})^{-1} X'_{\ell;N} \boldsymbol{\varepsilon}_{\ell;N}.$$

We shall introduce some more notation:

$$A_{\ell;N} := \frac{1}{C_\ell N(2\ell+1)} X'_{\ell;N} X_{\ell;N}, \qquad \Sigma_\ell := \mathbb{E}[A_{\ell;N}] = \frac{\Gamma_\ell}{C_\ell},$$

and

$$B_{\ell;N} := \frac{1}{C_\ell \sqrt{N(2\ell+1)}} X'_{\ell;N} \boldsymbol{\varepsilon}_{\ell;N}.$$

Therefore

$$\sqrt{N(2\ell+1)} \left(\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell \right) = A_{\ell;N}^{-1} B_{\ell;N}.$$

Heuristically, the proof of the Quantitative Central Limit Theorem can be described as follows: in order to be able to exploit Stein-Malliavin techniques, we need to deal with variables belonging to some q-th order chaos; now the ratio above does not fulfill this requirement, because $A_{\ell;N}^{-1}$ is a random quantity which does not belong to any $\mathcal{H}_q$. On the other hand, componentwise we have $B_{\ell;N} \in \mathcal{H}_2$, for each ℓ. We shall then show that it is possible to replace $A_{\ell;N}^{-1}$ by its (deterministic) probability limit Σ_ℓ^{-1}, without affecting asymptotic results; because our kernel estimators will be written as linear combinations of $\widehat{\boldsymbol{\phi}}_{\ell;N}$, the proof can be completed by a careful investigation of multivariate fourth-order cumulants.

Let us now make the previous argument rigorous. Let K_N and U_N be two mp-dimensional random vectors, defined as

$$K_N := \begin{pmatrix} \sqrt{\frac{N}{\mathcal{L}_N(z_1)}} \left(\widehat{k}_N(z_1) - k(z_1) \right) \\ \vdots \\ \sqrt{\frac{N}{\mathcal{L}_N(z_m)}} \left(\widehat{k}_N(z_m) - k(z_m) \right) \end{pmatrix},$$

and

$$U_N = \begin{pmatrix} U_N(z_1) \\ \vdots \\ U_N(z_m) \end{pmatrix} := \begin{pmatrix} \frac{1}{\sqrt{\mathcal{L}_N(z_1)}} \sum_{\ell=0}^{L_N} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_1) \\ \vdots \\ \frac{1}{\sqrt{\mathcal{L}_N(z_m)}} \sum_{\ell=0}^{L_N} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_m) \end{pmatrix}.$$

In particular, $\mathbb{E}[U_N] = 0_{mp}$ and $\mathbb{E}[U_N U'_N] = V_N$, where V_N is a block matrix whose generic ij-th block, $i, j \in \{1, \dots, m\}$, is given by

$$V_N(i,j) = \mathbb{E}[U_N(z_i) U'_N(z_j)]$$
$$= \frac{1}{\sqrt{\mathcal{L}_N(z_i)}} \frac{1}{\sqrt{\mathcal{L}_N(z_j)}} \sum_{\ell=0}^{L_N} \frac{C_{\ell;Z}}{C_\ell} \Sigma_\ell^{-1} \frac{2\ell+1}{16\pi^2} P_\ell(z_i) P_\ell(z_j).$$

Now, consider $Z \overset{d}{=} \mathcal{N}_{mp}(0_{mp}, I_{mp})$ and $Z_N \overset{d}{=} \mathcal{N}_{mp}(0_{mp}, V_N)$. Applying the triangle inequality twice, it follows that

$$\begin{aligned} d_W(Z, K_N) &\leq d_W(Z, U_N) + d_W(U_N, K_N) \\ &\leq d_W(Z, Z_N) + d_W(Z_N, U_N) + d_W(U_N, K_N). \end{aligned}$$

From [49, Equation 6.4.2, page 126], we have

$$d_W(Z, Z_N) \leq \sqrt{mp}\, \min\{\|V_N^{-1}\|_{\text{op}} \|V_N\|_{\text{op}}^{1/2}, 1\} \|V_N - I_{mp}\|_{\text{HS}},$$

where $\|A\|_{\text{HS}} = \sqrt{\text{Tr}(A'A)}$, and we observe that

$$\|V_N - I_{mp}\|_{\text{HS}} \leq mp \|V_N - I_{mp}\|_\infty = \mathcal{O}\left(N^{-d} \log N\right), \tag{3.4.2}$$

from Lemmas 3.6.3 and 3.6.4 in the Appendix. Indeed, for every $i \in \{1, \ldots, m\}$,

$$\begin{aligned} \|V_N(i,i) - I_p\|_{\text{HS}} &\leq \frac{const}{L_N + 1} \sum_{\ell=0}^{L_N} \left\| \frac{C_{\ell;Z}}{C_\ell} \Sigma_\ell^{-1} - I_p \right\|_\infty (2\ell + 1) \\ &\leq \frac{const}{L_N + 1}; \end{aligned}$$

the logarithmic term comes from Equation (3.6.8) in the Appendix Lemma 3.6.4. Equation (3.4.2) entails that $V_N \to I_{mp}$, thus we have $\|V_N^{-1}\|_{\text{op}} \|V_N\|_{\text{op}}^{1/2} \to 1$, as $N \to \infty$, and

$$d_W(Z, Z_N) = \mathcal{O}\left(N^{-d} \log N\right). \tag{3.4.3}$$

Let us recall again from [49, page 122] (second point of Theorem 6.2.2) that

$$d_W(Z_N, U_N) \leq \sqrt{mp} \|V_N^{-1}\|_{\text{op}} \|V_N\|_{\text{op}}^{1/2} m(U_N),$$

where

$$m(U_N) = 2mp \sum_{i=1}^m \sum_{j=1}^p \sqrt{\text{Cum}_4 \left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}} \sum_{\ell=0}^{L_N} \tilde{b}_{\ell;N}(j) \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_i) \right]},$$

$\tilde{b}_{\ell;N}(j)$ being the j-th element of $\Sigma_\ell^{-1} B_{\ell;N}$. Moreover, for the j-th element of $\Sigma_\ell^{-1} B_{\ell;N}$ we have

$$\text{Cum}_4 \left[\tilde{b}_{\ell;N}(j) \right] = \frac{6}{N(2\ell+1)} \left(\frac{C_{\ell;Z}}{C_\ell} s_\ell(j,j) \right)^2,$$

see Equation (3.6.4) in Lemma 3.6.1. In addition,

$$\begin{aligned} &\text{Cum}_4 \left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}} \sum_{\ell=0}^{L_N} \tilde{b}_{\ell;N}(j) \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z_i) \right] \\ =& \frac{1}{\mathcal{L}_N^2(z_i)} \sum_{\ell=0}^{L_N} \text{Cum}_4 \left[\tilde{b}_{\ell;N}(j) \right] \frac{(2\ell+1)^2}{(4\pi)^4} P_\ell^4(z_i), \end{aligned}$$

in view of the independence across different multipoles ℓ. Therefore

$$\begin{aligned}
&\mathrm{Cum}_4\left[\frac{1}{\sqrt{\mathcal{L}_N(z_i)}}\sum_{\ell=0}^{L_N}\tilde{b}_{\ell;N}(j)\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z_i)\right]\\
=&\frac{6}{N\mathcal{L}_N^2(z_i)}\sum_{\ell=0}^{L_N}\left(\frac{C_{\ell;Z}}{C_\ell}s_\ell(j,j)\right)^2\frac{2\ell+1}{(4\pi)^4}P_\ell^4(z_i)\\
\leq&\frac{6}{N\mathcal{L}_N^2(z_i)}\sum_{\ell=0}^{L_N}\left[\frac{C_{\ell;Z}}{C_\ell}\mathrm{Tr}(\Sigma_\ell^{-1})\right]^2\frac{2\ell+1}{(4\pi)^4}P_\ell^4(z_i)\\
\leq&\frac{const}{N(L_N+1)^2}\sum_{\ell=0}^{L_N}(2\ell+1)P_\ell^4(z_i).
\end{aligned}$$

Thus, we have

$$m(U_N)\leq const\frac{m^2p^2}{L_N+1}\sqrt{\frac{\log N}{N}},$$

and

$$d_W(Z_N,U_N)=\mathcal{O}\left(N^{-(d+1/2)}(\log N)^{1/2}\right). \tag{3.4.4}$$

Now, consider the decomposition

$$\begin{aligned}
\sqrt{\frac{N}{\mathcal{L}_N(z)}}\left(\widehat{k}_N(z)-k(z)\right)=&\frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}\sqrt{N(2\ell+1)}\left(\widehat{\phi}_{\ell;N}-\phi_\ell\right)\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z)\\
&-\sqrt{\frac{N}{\mathcal{L}_N(z)}}\sum_{\ell=L_N+1}^{\infty}\phi_\ell\frac{2\ell+1}{4\pi}P_\ell(z)\\
=&\frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}\Sigma_\ell^{-1}B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z)\\
&+\frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z)\\
&-\sqrt{\frac{N}{\mathcal{L}_N(z)}}\sum_{\ell=L_N+1}^{\infty}\phi_\ell\frac{2\ell+1}{4\pi}P_\ell(z).
\end{aligned}$$

Without loss of generality, we shall focus on the case $m=1$; the more general argument is basically identical, with a slightly more cumbersome notation. For $z\in(-1,1)$,

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z)\right\|\\
\leq&\frac{const}{\sqrt{L_N+1}}\sum_{\ell=0}^{L_N}\left\|[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}\right\|\sqrt{2\ell+1}|P_\ell(z)|,
\end{aligned}$$

and then,

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{\mathcal{L}_N(z)}}\sum_{\ell=0}^{L_N}[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_\ell(z)\right\|\right]$$

$$
\begin{aligned}
&\leq \frac{const}{\sqrt{L_N+1}} \sum_{\ell=0}^{L_N} \mathbb{E} \left\| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \right\| \sqrt{2\ell+1} |P_\ell(z)| \\
&\leq \frac{const}{\sqrt{L_N+1}} \sum_{\ell=0}^{L_N} \frac{1}{\sqrt{N(2\ell+1)}} \sqrt{2\ell+1} |P_\ell(z)| \\
&= \mathcal{O}\left(\frac{1}{\sqrt{N}}\right),
\end{aligned} \tag{3.4.5}
$$

where for the second inequality we have exploited the Appendix Lemma 3.6.2, while for the last step the Hilb's asymptotics (3.6.11) in the Appendix (see, also, [67, 75]). Likewise,

$$
\begin{aligned}
\left\| \sqrt{\frac{N}{\mathcal{L}_N(z)}} \sum_{\ell=L_N+1}^{\infty} \phi_\ell \frac{2\ell+1}{4\pi} P_\ell(z) \right\| &\leq const \sqrt{\frac{N}{L_N+1}} \sum_{\ell=L_N+1}^{\infty} \|\phi_\ell\| (2\ell+1) |P_\ell(z)| \\
&\leq const \sqrt{\frac{N}{L_N+1}} \sum_{\ell=L_N+1}^{\infty} \|\phi_\ell\| \sqrt{2\ell+1} \\
&= \mathcal{O}\left(\frac{1}{N^{d(\beta_*-1)-1/2}}\right).
\end{aligned} \tag{3.4.6}
$$

From Equations (3.4.5) and (3.4.6),

$$
d_W(U_N, K_N) = \mathcal{O}\left(N^{-1/2} + N^{1/2+d(1-\beta_*)}\right) . \tag{3.4.7}
$$

In the end, combining Equations (3.4.3), (3.4.4) and (3.4.7), it holds that

$$
d_W(Z, K_N) = \mathcal{O}\left(N^{-1/2} + N^{1/2+d(1-\beta_*)}\right).
$$

Note that the constant in this bound may depend on the choice of m and $z_1, \dots, z_m$. □

We can now give the proof of the third (and final) result.

Proof (Theorem 3.3.11). Under Condition 3.3.10, we have that, for $z \in [-1, 1]$,

$$
\begin{aligned}
\sqrt{N}\left(\widehat{k}_N(z) - k(z)\right) &= \sum_{\ell=0}^{L} \sqrt{N(2\ell+1)} \left(\widehat{\phi}_{\ell;N} - \phi_\ell\right) \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \\
&= \sum_{\ell=0}^{L} A_{\ell;N}^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \\
&= \sum_{\ell=0}^{L} \Sigma_\ell^{-1} B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \\
&\quad + \sum_{\ell=0}^{L} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z).
\end{aligned} \tag{3.4.8}
$$

Then,

$$
\sup_{z \in [-1,1]} \left\| \sum_{\ell=0}^{L} [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \frac{\sqrt{2\ell+1}}{4\pi} P_\ell(z) \right\| \leq \sum_{\ell=0}^{L} \left\| [A_{\ell;N}^{-1} - \Sigma_\ell^{-1}] B_{\ell;N} \right\| \frac{\sqrt{2\ell+1}}{4\pi},
$$

and hence

$$\mathbb{E}\left[\sup_{z\in[-1,1]}\left\|\sum_{\ell=0}^{L}[A_{\ell;N}^{-1}-\Sigma_{\ell}^{-1}]B_{\ell;N}\frac{\sqrt{2\ell+1}}{4\pi}P_{\ell}(z)\right\|\right]$$
$$\leq\sum_{\ell=0}^{L}\mathbb{E}\left\|[A_{\ell;N}^{-1}-\Sigma_{\ell}^{-1}]B_{\ell;N}\right\|\frac{\sqrt{2\ell+1}}{4\pi}\to 0,\qquad N\to\infty,$$

in view of the Appendix Lemma 3.6.2. Then the second part of the sum in (3.4.8) goes to zero in probability. Since the sum (over ℓ) has independent components, we just need to prove that, for each $\ell=0,1,2,\dots,L$, $\{B_{\ell;N}P_{\ell}(\cdot)\}$ forms a *tight* sequence. Using the tightness criterion given in [9, Equation 13.14, page 143], it is sufficient to show that, for $z_1\leq z\leq z_2$,

$$\begin{aligned}&\mathbb{E}\|B_{\ell;N}P_{\ell}(z)-B_{\ell;N}P_{\ell}(z_1)\|\|B_{\ell;N}P_{\ell}(z_2)-B_{\ell;N}P_{\ell}(z)\|\\&=|P_{\ell}(z)-P_{\ell}(z_1)||P_{\ell}(z_2)-P_{\ell}(z)|\mathbb{E}\|B_{\ell;N}\|^2\\&\leq p\frac{C_{\ell;Z}}{C_{\ell}}Q_{\ell}^2|z-z_1||z_2-z|\\&\leq p\frac{C_{\ell;Z}}{C_{\ell}}Q_{\ell}^2(z_2-z_1)^2.\end{aligned}$$

Convergence of the finite-dimensional distributions is standard and we omit the details, which are close to those given in the proofs of the previous Theorem. Thus the sequence converges weakly to a zero-mean multivariate Gaussian process with covariance function $\Gamma_{k_L}(z,z')=\sum_{\ell=0}^{L}C_{\ell;Z}\Gamma_{\ell}^{-1}\frac{2\ell+1}{16\pi^2}P_{\ell}(z)P_{\ell}(z')$. □

3.5 Some Numerical Evidence

In this section, we present some short numerical results to illustrate the models and methods that we discussed in this chapter.

We stress first that random fields on the sphere cross time can be very conveniently generated by combining the general features of Python with the HEALPix software (see [27] and `https://healpix.sourceforge.io`). More precisely, HEALPix (which stands for *Hierarchical Equal Area and iso-Latitude Pixelation*) is a multi-purpose computer software package for a high resolution numerical analysis of functions on the sphere, based on a clever tessellation scheme: the spherical surface is hierarchically partitioned into curvilinear quadrilaterals of equal area (at a given resolution), distributed on lines of constant latitude, as suggested in the name. In particular, we shall make use of `healpy`, which is a Python package based on the HEALPix C++ library. HEALPix was developed to efficiently process Cosmic Microwave Background data from cosmological experiments (like *Planck*, [60]), but it is now used in many other branches of Astrophysics and applied sciences.

In short, HEALPix allows to create spherical maps according to the spectral representation (1.2.4) (see Section 1.2.1), accepting in input either an array of random coefficients $\{a_{\ell,m}\}$, or the angular power spectrum $\{C_{\ell}\}$, by means of the routines `alm2map` and `synfast`: in the latter case, random $\{a_{\ell,m}\}$ are generated according to a Gaussian zero mean distribution with variance $\{C_{\ell}\}$. The routine is extremely efficient and allows to generate maps of resolution up to a few thousands multipoles in a matter of seconds on a standard laptop computer.

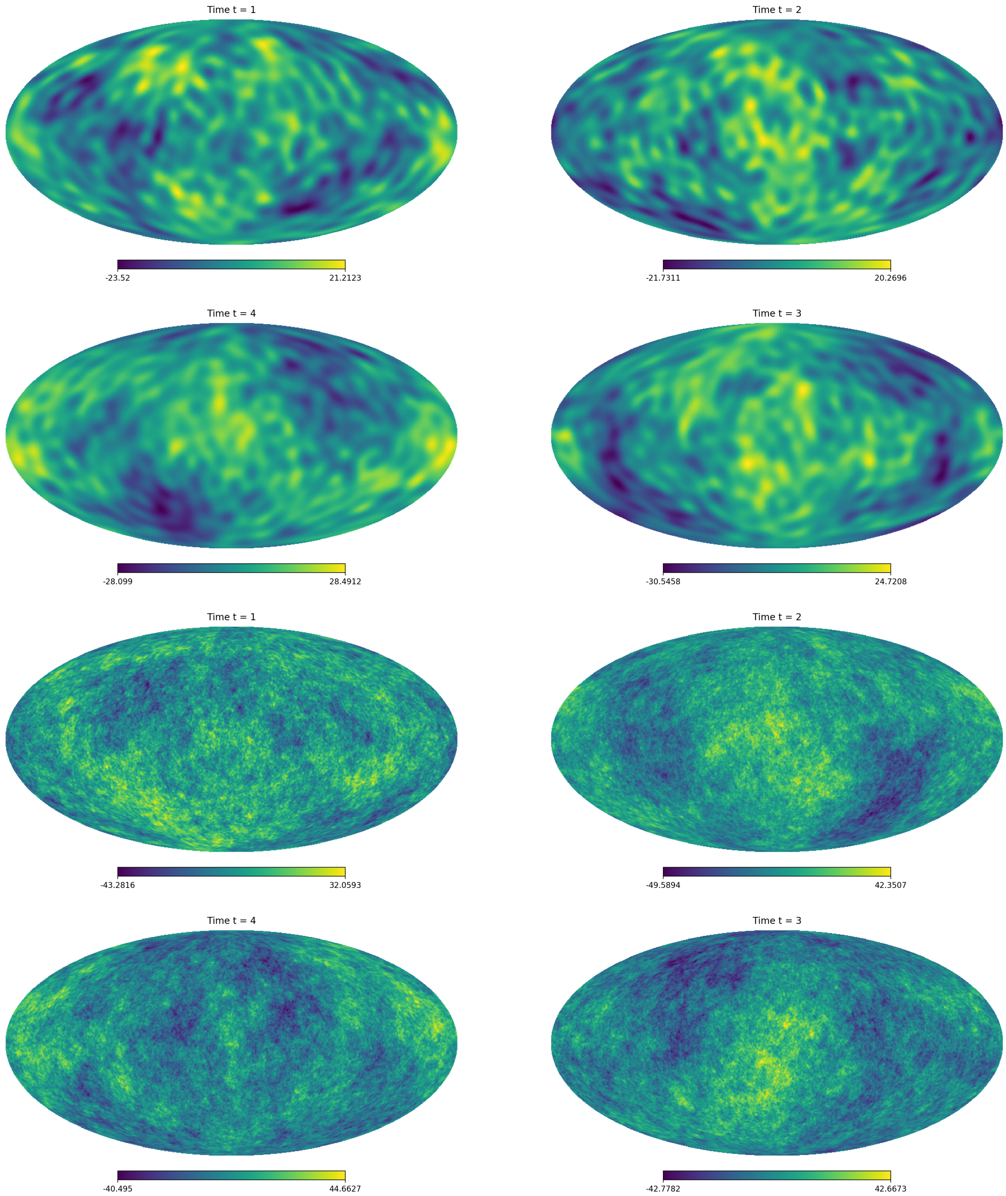


Figure 3.1. Two realizations of sphere-cross-time random fields at times $t = 1, 2, 3, 4$ (clockwise). Upper panel: maximum resolution $L_{\max} = 30$. Lower panel: maximum resolution $L_{\max} = 200$.

In our case, however, we need random fields where the random spherical harmonics coefficients have themselves a temporal dependence structure. For this reason, we implemented a simple routine in Python, to simulate Gaussian $\{a_{\ell,m}(t)\}$ processes, each following an AR(p) dependence structure. These random harmonic coefficients are then uploaded into HEALPix, to generate maps such as those that are given in Figure 3.1. In particular, in these two cases we fixed $L_{\max} = \max(\ell) = 30, 200$, respectively. Then we generated $\{a_{\ell,m}(t)\}$ according to stationary AR(1) processes,

with parameters $\phi_\ell \simeq const \times \ell^{-3}$; similarly, we took here $C_{\ell;Z} \simeq const \times \ell^{-2}$. In Figure 3.1, we report the realization for the first 4 periods, simply for illustrative purposes.
We are now in the position to use simulations to validate the previous results. In our first Tables 3.1-3.3, we report for $B = 1000$ Monte Carlo replications the values of the "variance" and "bias" terms, i.e., the first and second summand in the mean squared error equation (3.4.1); the second term is actually deterministic, and it is reported to illustrate the approximation one obtains by cutting the expansion to a finite multipole value. In the third column, we report, the actual (squared) L^2 error. On the left-hand side, we fix the number of multipoles to be exploited in the reconstruction of the kernel; on the right-hand side, we consider a sort of "oracle" estimator, where the number of multipoles grows with the optimal rate $N^{\frac{1}{2\beta_*-1}}$. As before, we took $C_{\ell;Z} \simeq const \times \ell^{-2}$, $\phi_\ell \simeq const \times \ell^{-\beta}$ for $\beta = 2, 2.5, 3$; for $N = 100, 300, 700$ the left-hand side uses $L_N \sim N^{0.6}$, while the right-hand side takes $L_N \sim N^{\frac{1}{2\beta_*-1}}$, as explained above.
We note how the estimators perform very efficiently, and show the errors scale approximately as N^α, where $\alpha \approx \frac{2-2\beta_*}{2\beta_*-1}$, as predicted by our computations, see Remark 3.3.5. In particular, Figure 3.2 shows the behaviour of the L^2 error, as a function of N. For $\beta_* = 2, 2.5, 3$, the empirical mean squared error is computed over a grid of N which ranges from 50 to 1000 in steps of 50. The green lines represent respectively the curves

$$y = \exp(-4.28)\, x^{-0.667}, \quad y = \exp(-3.7)\, x^{-0.75}, \quad y = \exp(-3.7)\, x^{-0.80}. \tag{3.5.1}$$

As explained earlier, the exponents match our theoretical results, whereas the multiplicative constants have been chosen by a least squares fit.

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00082	0.00006	0.00088	100	0.00041	0.00023	0.00065
300	0.00057	0.00001	0.00059	300	0.00022	0.00010	0.00031
700	0.00041	0.00001	0.00042	700	0.00012	0.00005	0.00018

Table 3.1. L^2 errors obtained with $\beta_* = 2$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right).

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00081	0.00007	0.00088	100	0.00063	0.00014	0.00077
300	0.00056	0.00001	0.00057	300	0.00029	0.00006	0.00035
700	0.00041	0.00000	0.00041	700	0.00016	0.00003	0.00019

Table 3.2. L^2 errors obtained with $\beta_* = 2.5$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right).

N	Variance	Bias	MSE	N	Variance	Bias	MSE
100	0.00082	0.00001	0.00084	100	0.00041	0.00021	0.00062
300	0.00058	0.00000	0.00058	300	0.00021	0.00004	0.00025
700	0.00041	0.00000	0.00041	700	0.00009	0.00004	0.00013

Table 3.3. L^2 errors obtained with $\beta_* = 3$; $L_N \sim N^{0.6}$ (left) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$ (right).

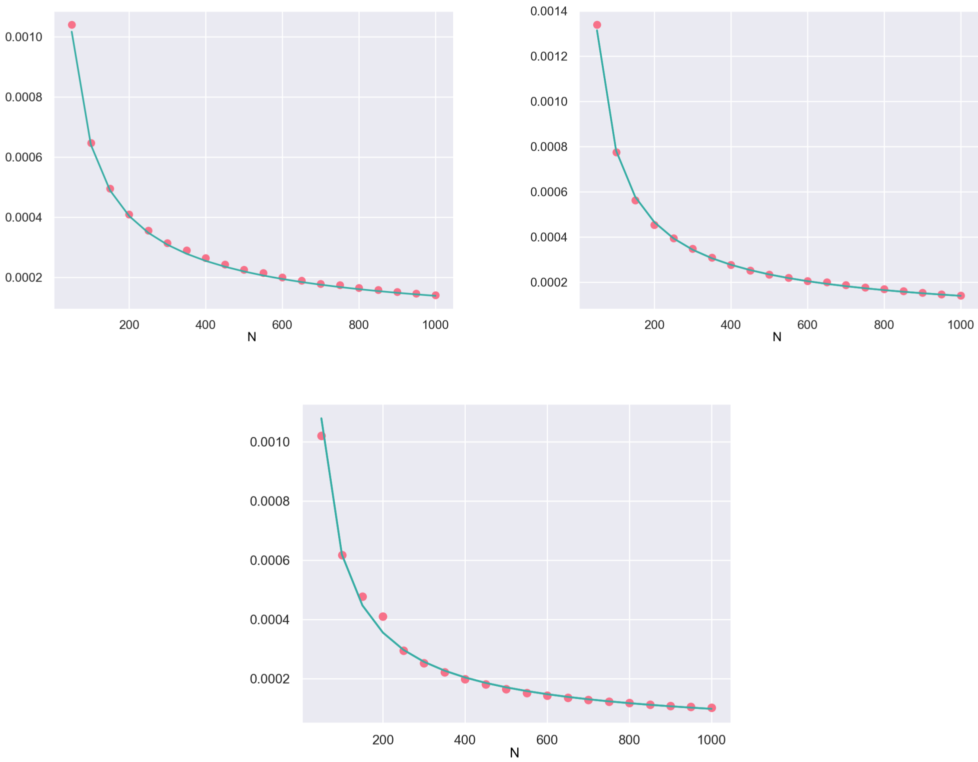


Figure 3.2. L^2 errors (dots) over a grid of N, for $\beta_* = 2, 2.5, 3$ (clockwise) and $L_N \sim N^{\frac{1}{2\beta_*-1}}$. The green lines represent the (calibrated) theoretical upper bounds in Equation (3.5.1).

We can now focus quickly on the main result of this chapter, dealing with the Quantitative Central Limit Theorem, in Wasserstein distance; the latter is computed following the Python routine (`scipy.stats.wasserstein_distance`). We consider again a model where the autoregressive parameter and the angular power spectra are exactly the same as in the previous settings, in particular, taking $\beta = 3$ and $d = 0.5$, up to integer approximations; we fix $L_{\max} = 1000$ for the number of components under the null hypothesis. Under these circumstances, we evaluate (univariate) Wasserstein distances for the kernel estimators at $m = 9$ different locations, performing $B = 10000$ Monte Carlo replications.

In our simulations, we took a number of time-domain observations ranging from $N = 100$ to $N = 1000$ in steps of 100; it should be noted that huge sample sizes are quite common when dealing with sphere-cross-time data, see, e.g., the NCEP/NCAR reanalysis datasets [34] for atmospheric research. In Table 3.4 we report for brevity a subset of these results, while the full sample is considered in Figure 3.3.

Again, we note that simulations track closely the theoretical predictions. More precisely, by our theoretical upper bound, we expect the Wasserstein distance $d_W(\cdot,\cdot)$ to decay faster than $N^{-0.5}$ (up to logarithmic factors) in the setting of Table 3.4, in good agreeement with simulations. To help visualize this behaviour, we report in Figure 3.3 the decay of numerically estimated Wasserstein distances for $K_N(z)$ (see Theorem 3.3.7) considered for three different values $z = -0.5, 0, 0.5$, for N in steps of 100 ranging from 100 to 1000; in blue, we reproduce also the expected upper bound, of order $\log N \times N^{-0.5}$. It is evident that the realized values are well controlled by the theoretical bound, with the exception of the smallest samples.

$N \backslash z$	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8
100	0.52	0.15	0.70	0.19	0.69	0.66	0.42	1.13	0.79
500	0.04	0.10	0.11	0.11	0.10	0.08	0.08	0.13	0.26
1000	0.03	0.03	0.03	0.03	0.04	0.04	0.04	0.03	0.07

Table 3.4. Wasserstein distances obtained with $\beta_* = 3$ and $L_N \sim N^{0.5}$.

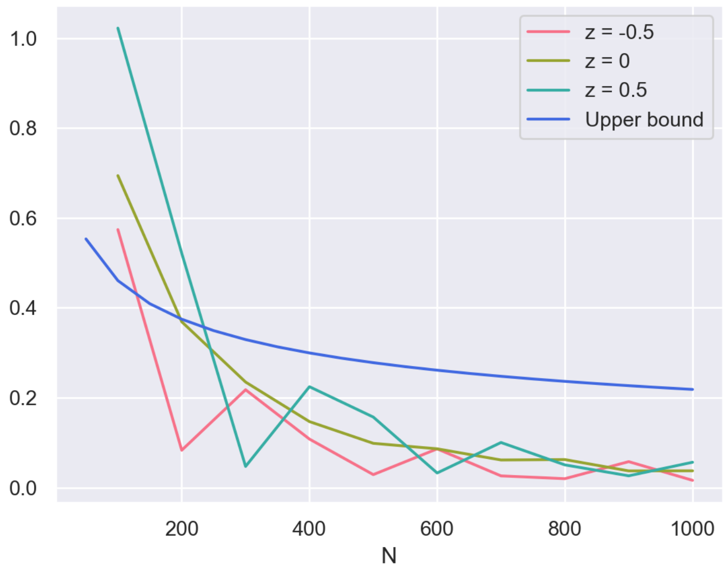


Figure 3.3. Wasserstein distances for $z = -0.5, 0, 0.5$ and theoretical upper bound $\log N \times N^{-0.5}$.

Remark 3.5.1. *Although the setting considered in this chapter is mainly theoretical, we believe that the models and procedures introduced here have plenty of potential for important applications. A possible dataset which is in our view amenable to SPHAR modeling is the NCEP reanalysis catalogue (see [34]), which provides the near-surface air temperature of the planet Earth over a grid of* 94×192 *unique spatial locations with a time span of 50 years (starting in 1948), sampled every day; overall, then, there are publicly available* 18048×18250 *space-cross-time observations. Clearly for temperature (and, more generally, climate) variables we cannot expect isotropy to hold exactly, due to the presence of features which depend on the location on the surface of the Earth; our idea, however, is that these anisotropic components can be estimated and removed in a preliminary stage of the analysis, just like trend and cyclical components are usually subtracted from time series data before standard ARMA models are implemented (see [12, Section 1.4]). These topics will be the object of a future, more applied work.*

3.6 Appendix

Throughout this appendix, we assume that Conditions 3.2.1 and 3.2.4 hold. Under these assumptions the proof that Equation 3.2.2 admits a unique stationary and isotropic solution can be given along the same lines as in [10] and it is omitted for brevity's sake; see Chapter 2 for more discussion and details. Note that, under these two Conditions, the variance C_ℓ can be written in terms of the coefficients $\phi_{\ell;j}$, $j = 1, \dots, p$, the autocorrelations $\rho_\ell(j) = C_\ell(j)/C_\ell$, $j = 1, \dots, p$, and the error variance $C_{\ell;Z}$, namely,

$$C_\ell = \frac{C_{\ell;Z}}{1 - \phi_{\ell;1}\rho_\ell(1) - \cdots - \phi_{\ell;p}\rho_\ell(p)} > 0, \qquad \ell \geq 0.$$

Hence,

$$0 < \frac{C_{\ell;Z}}{C_\ell} = 1 - \phi_{\ell;1}\rho_\ell(1) - \cdots - \phi_{\ell;p}\rho_\ell(p),$$

and there exists a positive constant ϕ^* such that, uniformly over ℓ,

$$\sum_{j=1}^{p} \phi_{\ell;j}\rho_\ell(j) \leq \phi^* < 1. \tag{3.6.1}$$

Recall that $C_{\ell;Z}/C_\ell$ and $C_\ell/C_{\ell;Z}$ are (in absolute value) bounded by positive constants since both converge to 1 as $\ell \to \infty$. Now, we denote with $g_\ell(\lambda)$ the correlation spectral density

$$\begin{aligned} g_\ell(\lambda) := \frac{f_\ell(\lambda)}{C_\ell} &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho_\ell(\tau) e^{i\lambda\tau} \\ &= \frac{1}{2\pi} \frac{1 - \phi_{\ell;1}\rho_\ell(1) - \cdots - \phi_{\ell;p}\rho_\ell(p)}{|1 - \phi_{\ell;1}e^{i\lambda} - \cdots - \phi_{\ell;p}e^{i\lambda p}|^2}, \qquad \lambda \in [-\pi, \pi], \end{aligned}$$

where $\rho_\ell(\cdot) := C_\ell(\cdot)/C_\ell$ is the autocorrelation function, and we recall that Σ_ℓ is the $p \times p$ matrix of autocorrelations, with ij-th element $\rho_\ell(i - j)$. Since $g_\ell(\cdot)$ is a continuous symmetric function on $[-\pi, \pi]$, it follows that (see [76])

$$2\pi \underline{g_\ell} \leq \lambda_{\min}(\Sigma_\ell) \leq \lambda_{\max}(\Sigma_\ell) \leq 2\pi \overline{g_\ell}, \tag{3.6.2}$$

where $\underline{g_\ell}$ and $\overline{g_\ell}$ are the minimum and maximum of $g_\ell(\cdot)$ in $[-\pi, \pi]$, respectively; $\lambda_{\min}(\Sigma_\ell)$ and $\lambda_{\max}(\Sigma_\ell)$ are the minimum and maximum eigenvalues of Σ_ℓ, respectively. Moreover, because we assumed $g_\ell(\lambda) > 0$, $\forall \lambda \in [-\pi, \pi]$, from (3.6.2) we conclude that the minimum eigenvalue is strictly positive (and hence bounded away from zero) and Σ_ℓ is positive definite (and then invertible). Since Σ_ℓ is a $p \times p$ real symmetric positive definite matrix, then

$$\|\Sigma_\ell\|_{\text{op}} = \lambda_{\max}(\Sigma_\ell) \leq \text{Tr}(\Sigma_\ell) = p, \quad \|\Sigma_\ell^{-1}\|_{\text{op}} = \frac{1}{\lambda_{\min}(\Sigma_\ell)} \leq \frac{1}{2\pi \underline{g_\ell}},$$

$$\text{Tr}(\Sigma_\ell^{-1}) \leq p \|\Sigma_\ell^{-1}\|_{\text{op}} \leq \frac{p}{2\pi \underline{g_\ell}},$$

where $\|A\|_{\text{op}} = \sqrt{\lambda_{\max}(A'A)}$, and $\text{Tr}(A)$ is the trace of A. In addition,

$$\frac{1}{2\pi \underline{g_\ell}} = \max_{\lambda \in [-\pi,\pi]} \frac{1}{2\pi g_\ell(\lambda)} = \max_{\lambda \in [-\pi,\pi]} \frac{|1 - \phi_{\ell;1} e^{i\lambda} - \cdots - \phi_{\ell;p} e^{i\lambda p}|^2}{1 - \phi_{\ell;1}\rho_\ell(1) - \cdots - \phi_{\ell;p}\rho_\ell(p)}$$
$$\leq \frac{const}{1 - \sum_{j=1}^{p} \phi_{\ell;j}\rho_\ell(j)},$$

since

$$|1 - \phi_{\ell;1} e^{i\lambda} - \cdots - \phi_{\ell;p} e^{i\lambda p}| \leq 1 + \sum_{j=1}^{p} |\phi_{\ell;j}| \leq const.$$

Then, from Equation (3.6.1), we can conclude that, uniformly over ℓ,

$$\frac{1}{2\pi \underline{g_\ell}} \leq C, \quad \text{some } C > 0.$$

The first result below will be exploited to prove convergence in probability of the denominator for our estimators, while the second part gives the fourth-cumulant bound which is crucial for Stein-Malliavin arguments. We recall here for convenience the equalities

$$\sqrt{N(2\ell+1)}\left(\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell\right) = A_{\ell;N}^{-1} B_{\ell;N},$$

where

$$A_{\ell;N} = \frac{1}{C_\ell N(2\ell+1)} X_{\ell;N}' X_{\ell;N}, \qquad \Sigma_\ell = \mathbb{E} A_{\ell;N},$$

and

$$B_{\ell;N} = \frac{1}{C_\ell \sqrt{N(2\ell+1)}} X_{\ell;N}' \boldsymbol{\varepsilon}_{\ell;N}.$$

Lemma 3.6.1. *For any integers $\ell \geq 0$, $N > p$, there exists $M > 0$ such that*

$$\mathbb{E}\left[a_{\ell;N}(i,j) - \mathbb{E}[a_{\ell;N}(i,j)]\right]^2 \leq \frac{M}{N(2\ell+1)}, \qquad i,j \in \{1,\ldots,p\}, \tag{3.6.3}$$

and

$$\text{Cum}_4\left[\widetilde{b}_{\ell;N}(i)\right] = \frac{6}{N(2\ell+1)}\left(\frac{C_{\ell;Z}}{C_\ell} s_\ell(i,i)\right)^2, \tag{3.6.4}$$

where $\widetilde{b}_{\ell;N}(i) = \sum s_\ell(i,j) b_{\ell;N}(j)$ is the i-th element of the p-dimensional vector $\widetilde{B}_{\ell;N} = \Sigma_\ell^{-1} B_{\ell;N}$, and $s_\ell(i,j)$ are the elements of the inverse matrix Σ_ℓ^{-1}.

The following result shows that replacing $A_{\ell;N}$ by its expected value Σ_ℓ in the definition of the OLS-like estimator $\widehat{\phi}_{\ell;N}$ does not have any asymptotic effect, as $N \to \infty$.

Lemma 3.6.2. *For any integers $\ell \geq 0$ and $N > 7 + p$, there exists generic positive constants such that*

$$\mathbb{E}\left\|\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell\right\|^2 \leq \frac{const}{N(2\ell+1)}, \tag{3.6.5}$$

and

$$\mathbb{E}\left\|\sqrt{N(2\ell+1)}\left(\widehat{\boldsymbol{\phi}}_{\ell;N} - \boldsymbol{\phi}_\ell\right) - \Sigma_\ell^{-1} B_{\ell;N}\right\| \leq \frac{const}{\sqrt{N(2\ell+1)}}. \tag{3.6.6}$$

The following results entail that $\lim_{N\to\infty} V_N = I_{mp}$; actually the two lemmas below allow to obtain a uniform rate of convergence.

Lemma 3.6.3. *If* $\|\phi_\ell\| \le \frac{\gamma}{\ell^\beta},\ \beta > 1,\ \ell > 0$,

$$\left\| \frac{C_{\ell;Z}}{C_\ell} \Sigma_\ell^{-1} - I_p \right\|_\infty = \mathcal{O}\left(\frac{1}{\ell^\beta}\right).$$

The next result is technical; given the huge amount of work which has taken place on Legendre polynomials, we expect that the statement could be known already, but we failed to locate a reference and therefore we report a full proof for the sake of completeness.

Lemma 3.6.4. *Let* $z = \cos\theta,\ \theta \in (0,\pi)$,

$$\lim_{L\to\infty} \frac{1}{L+1} \sum_{\ell=0}^{L} (2\ell+1) P_\ell^2(\cos\theta) = \frac{2}{\pi \sin\theta}; \tag{3.6.7}$$

on the other hand, for $\theta, \theta' \in (0,\pi),\ \theta \neq \theta'$, *as* $L \to \infty$,

$$\frac{1}{L+1} \sum_{\ell=0}^{L} (2\ell+1) P_\ell(\cos\theta) P_\ell(\cos\theta') = O\left(\frac{\log L}{L}\right). \tag{3.6.8}$$

We can now start with the proof of these Lemmas.

Proof (Lemma 3.6.1). Let us start by observing that the ij-th element of $A_{\ell;N}$, denoted by $a_{\ell;N}(i,j)$, has expected value

$$\begin{aligned} \mathbb{E}\left[a_{\ell;N}(i,j)\right] &= \mathbb{E}\left[\frac{1}{N(2\ell+1)C_\ell} \sum_{t=p+1}^{n} \sum_{m=-\ell}^{\ell} a_{\ell,m}(t-i) a_{\ell,m}(t-j)\right] \\ &= \frac{1}{N(2\ell+1)C_\ell} \sum_{t=p+1}^{n} \sum_{m=-\ell}^{\ell} \mathbb{E}\left[a_{\ell,m}(t-i) a_{\ell,m}(t-j)\right] \\ &= \rho_\ell(i-j). \end{aligned}$$

Now, we have

$$\mathbb{E}\left[a_{\ell;N}(i,j) - \mathbb{E}[a_{\ell;N}(i,j)]\right]^2 = \sum_{tt'} \sum_{mm'} \frac{\mathrm{Cov}[a_{\ell,m}(t_1-i)a_{\ell,m}(t_1-j), a_\ell(t_2-i)a_{\ell,m}(t_2-j)]}{N^2(2\ell+1)^2 C_\ell^2}$$

$$\overset{\tau=t_1-t_2}{=} \frac{1}{N(2\ell+1)} \sum_{\tau=1-N}^{N-1} \left(1 - \frac{|\tau|}{N}\right) \left[\left(\frac{C_\ell(\tau)}{C_\ell}\right)^2 + \frac{C_\ell(\tau+i-j)}{C_\ell} \frac{C_\ell(\tau+j-i)}{C_\ell}\right].$$

Now observe $\rho_\ell^2(\cdot) = (C_\ell(\cdot)/C_\ell)^2$, the squared autocorrelation function of the process, is nonnegative and summable; that is, there exists $\rho_\ell^* \in \mathbb{R}^+$ so that $\sum_{\tau=-\infty}^{+\infty} \rho_\ell^2(\tau) = \rho_\ell^* < \infty$, and

$$\sum_{\tau=-\infty}^{\infty} |\rho_\ell(\tau+i-j)\rho_\ell(\tau+j-i)| \le \sum_{\tau=-\infty}^{+\infty} \rho_\ell^2(\tau),$$

in view of the Cauchy-Schwartz inequality. Thus, it holds that

$$\begin{aligned}
&\sum_{\tau=1-N}^{N-1}\left(1-\frac{|\tau|}{N}\right)\left[\left(\frac{C_\ell(\tau)}{C_\ell}\right)^2+\frac{C_\ell(\tau+i-j)}{C_\ell}\frac{C_\ell(\tau+j-i)}{C_\ell}\right]\\
&\leq \sum_{\tau=1-N}^{N-1}\rho_\ell^2(\tau)+\sum_{\tau=1-N}^{N-1}|\rho_\ell(\tau+i-j)\rho_\ell(\tau+j-i)|\\
&\leq \sum_{\tau=-\infty}^{+\infty}\rho_\ell^2(\tau)+\sum_{\tau=-\infty}^{+\infty}\rho_\ell^2(\tau)=2\rho_\ell^*.
\end{aligned}$$

On the other hand,

$$\rho_\ell^*=2\pi\int_{-\pi}^{\pi}[g_\ell(\lambda)]^2d\lambda,$$

and

$$\begin{aligned}
g_\ell(\lambda)&=\frac{1}{2\pi}\frac{1-\phi_{\ell;1}\rho_\ell(1)-\cdots-\phi_{\ell;p}\rho_\ell(p)}{|1-\phi_{\ell;1}e^{i\lambda}-\cdots-\phi_{\ell;p}e^{i\lambda p}|^2}\\
&\leq\frac{1}{2\pi}\frac{const}{\left(1-\xi_*^{-1}\right)^{2p}},
\end{aligned}$$

since

$$1-\sum_{j=1}^{p}\phi_{\ell;j}\rho_\ell(j)\leq 1+\sum_{j=1}^{p}|\phi_{\ell;j}|\leq const,$$

and

$$|1-\phi_{\ell;1}e^{i\lambda}-\cdots-\phi_{\ell;p}e^{i\lambda p}|\geq(1-\xi_*^{-1})^p>0,$$

see also Chapter 2 (Remark 2.2.7). Then, $\rho_\ell^*\leq const$, uniformly over ℓ.
In conclusion, uniformly over ℓ and N,

$$\mathbb{E}\left[a_{\ell;N}(i,j)-\mathbb{E}[a_{\ell;N}(i,j)]\right]^2\leq\frac{M}{N(2\ell+1)},\qquad i,j\in\{1,\dots,p\},$$

$M>0$.
Let us now focus on the elements of $\widetilde{B}_{\ell;N}=\Sigma_\ell^{-1}B_{\ell;N}$; they are given by

$$\widetilde{b}_{\ell;N}(i)=\sum_{j=1}^{p}s_\ell(i,j)b_{\ell;N}(j),\qquad i=1,\dots,p.$$

These elements can be shown to satisfy the following properties:

(i) $\mathbb{E}\left[\widetilde{b}_{\ell;N}(j)\right]=\sum_{j=1}^{p}s_\ell(i,j)\mathbb{E}[b_{\ell;N}(j)]=0$;

(ii) $\mathbb{E}\left[\widetilde{b}_{\ell;N}(i)\widetilde{b}_{\ell;N}(j)\right]=s_\ell(i,j)\frac{C_{\ell;Z}}{C_\ell}$, since

$$\mathbb{E}[\Sigma_\ell^{-1}B_{\ell;N}(\Sigma_\ell^{-1}B_{\ell;N})']=\Sigma_\ell^{-1}\mathbb{E}[B_{\ell;N}B_{\ell;N}']\Sigma_\ell^{-1}=\frac{C_{\ell;Z}}{C_\ell}\Sigma_\ell^{-1},$$

and because

$$\begin{aligned}\mathbb{E}[b_{\ell;N}(i)b_{\ell;N}(j)] &= \frac{1}{C_\ell^2}\frac{1}{N(2\ell+1)}\sum_{tt'}\sum_{mm'}\mathbb{E}[a_{\ell,m}(t-i)a_{\ell,m;Z}(t)a_{\ell,m}(t'-j)a_{\ell,m;Z}(t')] \\ &= \frac{1}{C_\ell^2}\frac{1}{N(2\ell+1)}\sum_{tt'}\sum_{m}\mathbb{E}\left[a_{\ell,m}(t-i)a_{\ell,m;Z}(t)a_{\ell,m}(t'-j)a_{\ell,m;Z}(t')\right] \\ &= \frac{1}{C_\ell^2}\frac{1}{N(2\ell+1)}\sum_{tm}C_\ell(i-j)C_{\ell;Z} \\ &= \frac{C_\ell(i-j)C_{\ell;Z}}{C_\ell^2}.\end{aligned}$$

(iii) $\mathrm{Cum}_4\left[\widetilde{b}_{\ell;N}(i)\right] = \frac{6}{N(2\ell+1)}\left(s_\ell(i,i)\frac{C_{\ell;Z}}{C_\ell}\right)^2$.

To compute $\mathrm{Cum}_4\left[\widetilde{b}_{\ell;N}(i)\right]$ we use once again the multilinearity property of cumulants, the real expansion and the diagram formula, so that we obtain:

$$\mathrm{Cum}_4[\widetilde{b}_{\ell;N}(i)] = \sum_{j_1j_2j_3j_4} s_\ell(i,j_1)s_\ell(i,j_2)s_\ell(i,j_3)s_\ell(i,j_4)\mathrm{Cum}[b_{\ell;N}(j_1),b_{\ell;N}(j_2),b_{\ell;N}(j_3),b_{\ell;N}(j_4)],$$

with $\mathrm{Cum}[b_{\ell;N}(j_1),b_{\ell;N}(j_2),b_{\ell;N}(j_3),b_{\ell;N}(j_4)] = \mathrm{Cum}(j_1,j_2,j_3,j_4)$ given by

$$\begin{aligned}&\mathrm{Cum}(j_1,j_2,j_3,j_4) = \frac{1}{C_\ell^4}\frac{1}{N^2(2\ell+1)^2} \\ &\times\sum_{tm}\mathrm{Cum}\left[a_{\ell,m}(t-j_1)a_{\ell,m;Z}(t),a_{\ell,m}(t-j_2)a_{\ell,m;Z}(t),a_{\ell,m}(t-j_3)a_{\ell,m;Z}(t),a_{\ell,m}(t-j_4)a_{\ell,m;Z}(t)\right] \\ &= \frac{1}{N(2\ell+1)}\left[2\frac{C_\ell(j_1-j_2)}{C_\ell}\frac{C_\ell(j_3-j_4)}{C_\ell}\left(\frac{C_{\ell;Z}}{C_\ell}\right)^2\right. \\ &+2\frac{C_\ell(j_1-j_3)}{C_\ell}\frac{C_\ell(j_2-j_4)}{C_\ell}\left(\frac{C_{\ell;Z}}{C_\ell}\right)^2 \\ &\left.+2\frac{C_\ell(j_1-j_4)}{C_\ell}\frac{C_\ell(j_2-j_3)}{C_\ell}\left(\frac{C_{\ell;Z}}{C_\ell}\right)^2\right].\end{aligned}$$

Hence,

$$\mathrm{Cum}_4\left[\widetilde{b}_{\ell;N}(i)\right] = \frac{6}{N(2\ell+1)}\left(s_\ell(i,i)\frac{C_{\ell;Z}}{C_\ell}\right)^2,$$

as claimed. □

Proof (Lemma 3.6.2). First, rewrite

$$\sqrt{N(2\ell+1)}\left(\widehat{\boldsymbol{\phi}}_{\ell;N}-\boldsymbol{\phi}_\ell\right) = \Sigma_\ell^{-1}B_{\ell;N} + [A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}.$$

Since

$$\begin{aligned}\left\|\sqrt{N(2\ell+1)}\left(\widehat{\phi}_{\ell;N}-\phi_\ell\right)-\Sigma_\ell^{-1}B_{\ell;N}\right\| &= \|[A_{\ell;N}^{-1}-\Sigma_\ell^{-1}]B_{\ell;N}\| \\ &= \|[I_p-\Sigma_\ell^{-1}A_{\ell;N}]A_{\ell;N}^{-1}B_{\ell;N}\|\end{aligned}$$

$$\leq \|I_p - \Sigma_\ell^{-1} A_{\ell;N}\|_{\text{op}} \|A_{\ell;N}^{-1}\|_{\text{op}} \|B_{\ell;N}\|,$$

we have

$$\begin{aligned}\mathbb{E}\|[A_{\ell;N}^{-1} - \Sigma_\ell^{-1}]B_{\ell;N}\| &\leq \left(\mathbb{E}\|A_{\ell;N}^{-1}\|_{\text{op}}^2 \|B_{\ell;N}\|^2\right)^{1/2} \left(\mathbb{E}\|I_p - \Sigma_\ell^{-1} A_{\ell;N}\|_{\text{op}}^2\right)^{1/2} \\ &\leq \left(\mathbb{E}\|A_{\ell;N}^{-1}\|_{\text{op}}^4\right)^{1/4} \left(\mathbb{E}\|B_{\ell;N}\|^4\right)^{1/4} \left(\mathbb{E}\|I_p - \Sigma_\ell^{-1} A_{\ell;N}\|_{\text{op}}^2\right)^{1/2},\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}\|B_{\ell;N}\|^4 &= \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbb{E}\left[b_{\ell;N}^2(i) b_{\ell;N}^2(j)\right] \\ &\leq \sum_{i=1}^{p}\sum_{j=1}^{p} \left(\mathbb{E}\left[b_{\ell;N}^4(i)\right]\right)^{1/2} \left(\mathbb{E}\left[b_{\ell;N}^4(j)\right]\right)^{1/2} \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} \left[\frac{6}{N(2\ell+1)}\left(\frac{C_{\ell;Z}}{C_\ell}\right)^2 + 3\left(\frac{C_{\ell;Z}}{C_\ell}\right)^2\right] \\ &< p^2\left(\frac{24}{N(2\ell+1)} + 12\right),\end{aligned}$$

and, from (3.6.3),

$$\begin{aligned}\mathbb{E}\|I_p - \Sigma_\ell^{-1} A_{\ell;N}\|_{\text{op}}^2 &\leq \|\Sigma_\ell^{-1}\|_{\text{op}}^2 \mathbb{E}\|\Sigma_\ell - A_{\ell;N}\|_{\text{op}}^2 \\ &\leq const \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbb{E}\left[a_{\ell;N}(i,j) - \mathbb{E}[a_{\ell;N}(i,j)]\right]^2 \\ &\leq \frac{const}{N(2\ell+1)}.\end{aligned}$$

By definition,

$$\|A_{\ell;N}^{-1}\|_{\text{op}} = N(2\ell+1)\frac{C_\ell}{\lambda_{\min}(X'_{\ell;N}X_{\ell;N})}.$$

Since $X'_{\ell;N}X_{\ell;N}$ is a real symmetric $p \times p$ matrix,

$$\lambda_{\min}(X'_{\ell;N}X_{\ell;N}) = \min_{\|\boldsymbol{\gamma}\|=1} \boldsymbol{\gamma}' X'_{\ell;N}X_{\ell;N}\boldsymbol{\gamma}.$$

$X'_{\ell;N}X_{\ell;N}$ can be seen as the sum of $2\ell+1$ independent matrix, i.e.,

$$X'_{\ell;N}X_{\ell;N} = \sum_{m=-\ell}^{\ell} X'_{\ell,m;N}X_{\ell,m;N},$$

where $X_{\ell,m;N}$ is a $N \times p$ matrix, defined by (recalling that $n = N + p$)

$$X_{\ell,m;N} = \begin{pmatrix} a_{\ell,m}(p) & a_{\ell,m}(p+1) & \cdots & a_{\ell,m}(n-1) \\ \vdots & \vdots & \vdots & \vdots \\ a_{\ell,m}(1) & a_{\ell,m}(2) & \cdots & a_{\ell,m}(n-p) \end{pmatrix}.$$

Then,

$$\lambda_{\min}(X'_{\ell;N}X_{\ell;N}) = \min_{\|\gamma\|=1} \gamma' \left[\sum_{m=-\ell}^{\ell} X'_{\ell,m;N}X_{\ell,m;N} \right] \gamma$$
$$= \min_{\|\gamma\|=1} \sum_{m=-\ell}^{\ell} \gamma' X'_{\ell,m;N}X_{\ell,m;N}\gamma. \tag{3.6.9}$$

Now recall that Σ_ℓ is the $p \times p$ matrix of autocorrelations; similarly we define $\Sigma_{\ell;N}$ as the $N \times N$ matrix of autocorrelations. Both are invertible since we assumed that the spectral density

$$g_\ell(\lambda) = \frac{1}{2\pi} \frac{1 - \phi_{\ell;1}\rho_\ell(1) - \cdots - \phi_{\ell;p}\rho_\ell(p)}{|1 - \phi_{\ell;1}e^{i\lambda} - \cdots - \phi_{\ell;p}e^{i\lambda p}|^2}, \qquad \lambda \in [-\pi, \pi],$$

is a continuous positive function.
$X_{\ell,m;N}$ is a zero-mean Gaussian matrix with $\mathbb{E}[X_{\ell,m;N}X'_{\ell,m;N}] = pC_\ell\Sigma_{\ell;N}$ and $\mathbb{E}[X'_{\ell,m;N}X_{\ell,m;N}] = NC_\ell\Sigma_\ell$, therefore it can be written as $X_{\ell,m;N} = (C_\ell\Sigma_{\ell;N})^{1/2}Z_{\ell,m;N}$, where $Z_{\ell,m;N}$ is a zero-mean Gaussian matrix with independent rows. If $\Sigma_{\ell;N} = P\Lambda P'$, where P is an orthogonal matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues, then

$$\begin{aligned}
\gamma' X'_{\ell,m;N}X_{\ell,m;N}\gamma &= C_\ell\gamma' Z'_{\ell,m;N}\Sigma_{\ell;N}Z_{\ell,m;N}\gamma \\
&= C_\ell\gamma' Z'_{\ell,m;N}P\Lambda P' Z_{\ell,m;N}\gamma \\
&\geq C_\ell\lambda_{\min}(\Sigma_{\ell;N})\gamma' Z'_{\ell,m;N}PP'Z_{\ell,m;N}\gamma \\
&= C_\ell\lambda_{\min}(\Sigma_{\ell;N})\gamma' Z'_{\ell,m;N}Z_{\ell,m;N}\gamma,
\end{aligned}$$

where $Z'_{\ell,m;N}Z_{\ell,m;N}$ is a Wishart random matrix with N degrees of freedom. The same argument applies to all $2\ell + 1$ components of (3.6.9), so that

$$\begin{aligned}
\frac{\lambda_{\min}(X'_{\ell;N}X_{\ell;N})}{C_\ell} &\geq \lambda_{\min}(\Sigma_{\ell;N}) \min_{\|\gamma\|=1} \sum_{m=-\ell}^{\ell} \gamma' Z'_{\ell,m;N}Z_{\ell,m;N}\gamma \\
&= \lambda_{\min}(\Sigma_{\ell;N}) \min_{\|\gamma\|=1} \gamma' \left[\sum_{m=-\ell}^{\ell} Z'_{\ell,m;N}Z_{\ell,m;N} \right] \gamma \\
&= \lambda_{\min}(\Sigma_{\ell;N}) \min_{\|\gamma\|=1} \gamma' Z'_{\ell;N}Z_{\ell;N}\gamma' \\
&= \lambda_{\min}(\Sigma_{\ell;N})\lambda_{\min}(Z'_{\ell;N}Z_{\ell;N}).
\end{aligned} \tag{3.6.10}$$

The summation in (3.6.10) includes $2\ell + 1$ independent Wishart random matrix each with N degrees of freedom and Σ_ℓ as scale matrix, then $Z'_{\ell;N}Z_{\ell;N}$ is a Wishart random matrix with $N(2\ell + 1)$ degrees of freedom and Σ_ℓ as scale matrix, and $\lambda_{\min}(Z'_{\ell;N}Z_{\ell;N})$ its minimum eigenvalue. Furthermore, this result guarantees the invertibility of the matrix $X'_{\ell;N}X_{\ell;N}$.
By the standard inequality on trace and operator norms for matrices, we obtain that

$$\mathbb{E}\|A^{-1}_{\ell;N}\|^4_{\text{op}} \leq \frac{N^4(2\ell + 1)^4}{\lambda^4_{\min}(\Sigma_{\ell;N})} \mathbb{E}\|(Z'_{\ell;N}Z_{\ell;N})^{-1}\|^4_{\text{op}}$$

$$\leq \frac{N^4(2\ell+1)^4}{\lambda^4_{\min}(\Sigma_{\ell;N})}\mathbb{E}\left[\mathrm{Tr}((Z'_{\ell;N}Z_{\ell;N})^{-1})\right]^4$$
$$\leq \frac{N^4(2\ell+1)^4}{(2\pi \underline{g_\ell})^4}\mathbb{E}\left[\mathrm{Tr}((Z'_{\ell;N}Z_{\ell;N})^{-1})\right]^4.$$

For $N(2\ell+1) > 7+p$ the fourth moment of the trace of an inverse Wishart matrix is given in [47]:

$$\begin{aligned} u_4(\eta)\mathbb{E}\left[\mathrm{Tr}((Z'_{\ell;N}Z_{\ell;N})^{-1})\right]^4 &= 48(5\eta-3)\mathrm{Tr}(\Sigma_\ell^{-4}) \\ &+ 128\eta(\eta-2)\mathrm{Tr}(\Sigma_\ell^{-3})\mathrm{Tr}(\Sigma_\ell^{-1}) \\ &+ 12(2\eta^2-5\eta+9)(\mathrm{Tr}(\Sigma_\ell^{-2}))^2 \\ &+ 12(4\eta^3-12\eta^2+3\eta+3)\mathrm{Tr}(\Sigma_\ell^{-2})(\mathrm{Tr}(\Sigma_\ell^{-1}))^2 \\ &+ (\eta+1)(2\eta-3)(4\eta^2-12\eta+1)(\mathrm{Tr}(\Sigma_\ell^{-1}))^4, \end{aligned}$$

where $\eta = \frac{N(2\ell+1)}{2} - \frac{p+1}{2}$, and

$$u_4(\eta) = 2^4\eta(\eta-1)(\eta-2)(\eta-3)(2\eta-1)(\eta+1)(2\eta+1)(2\eta+3).$$

If $\lambda_{\ell;1}, \dots, \lambda_{\ell;p}$ are the eigenvalues of Σ_ℓ, we have

$$0 < \mathrm{Tr}(\Sigma_\ell^{-k}) = \sum_{j=1}^p \left(\frac{1}{\lambda_{\ell;j}}\right)^k \leq \left(\sum_{j=1}^p \frac{1}{\lambda_{\ell;j}}\right)^k = (\mathrm{Tr}(\Sigma_\ell^{-1}))^k,$$

$k \geq 1$. Then, for $2\eta > 7+p$,

$$\begin{aligned} u_4(\eta)\mathbb{E}\left[\mathrm{Tr}((Z'_{\ell;N}Z_{\ell;N})^{-1})\right]^4 &\leq (\mathrm{Tr}(\Sigma_\ell^{-1}))^4(8\eta^4+20\eta^3+10\eta^2-5\eta-3) \\ &= (\mathrm{Tr}(\Sigma_\ell^{-1}))^4(2\eta-1)(\eta+1)(2\eta+1)(2\eta+3), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\left[\mathrm{Tr}((Z'_{\ell;N}Z_{\ell;N})^{-1})\right]^4 &\leq \frac{(\mathrm{Tr}(\Sigma_\ell^{-1}))^4}{2^4\eta(\eta-1)(\eta-2)(\eta-3)} \\ &= \frac{(\mathrm{Tr}(\Sigma_\ell^{-1}))^4}{\prod_{k=1}^4(N(2\ell+1)-p+1-2k))}. \end{aligned}$$

In addition,

$$\begin{aligned} \frac{N^4(2\ell+1)^4}{\prod_{k=1}^4\left(N(2\ell+1)-p+1-2k\right)} &= \frac{1}{\prod_{k=1}^4\left(1-\frac{p-1+2k}{N(2\ell+1)}\right)} \\ &\leq \frac{1}{\left(1-\frac{p+7}{N(2\ell+1)}\right)^4} \\ &\leq \frac{1}{\left(1-\frac{p+7}{p+8}\right)^4}, \end{aligned}$$

for every $\ell \geq 0$ and $N > 7 + p$. Thus, (3.6.6) holds.
The second part of this Lemma follows easily, indeed

$$\begin{aligned}
\mathbb{E}\left\|\widehat{\phi}_{\ell;N} - \phi_\ell\right\|^2 &= \frac{1}{N(2\ell+1)}\mathbb{E}\left\|\sqrt{N(2\ell+1)}(\widehat{\phi}_{\ell;N} - \phi_\ell)\right\|^2 \\
&= \frac{1}{N(2\ell+1)}\mathbb{E}\left\|A^{-1}_{\ell;N}B_{\ell;N}\right\|^2 \\
&\leq \frac{1}{N(2\ell+1)}\left(\mathbb{E}\left\|A^{-1}_{\ell;N}\right\|^4\right)^{1/2}\left(\mathbb{E}\left\|B_{\ell;N}\right\|^4\right)^{1/2} \\
&\leq \frac{const}{N(2\ell+1)},
\end{aligned}$$

in view of the bounds that we just established on the fourth-moments of the norms of $A^{-1}_{\ell,N}$ and $B_{\ell,N}$. □

Proof (Lemma 3.6.3). We first need to prove that $\lim_{\ell\to\infty}\Sigma_\ell = I_p$, where we recall that Σ_ℓ is the matrix of autocorrelations $\rho_\ell(i-j)$. For $i = j$, $\rho_\ell(i-j) = 1$, for all ℓ; on the other hand, for $i \neq j$,

$$\rho_\ell(i-j) = \phi_{\ell;1}\rho_\ell(i-j-1) + \cdots + \phi_{\ell;p}\rho_\ell(i-j-p),$$

and

$$|\rho_\ell(i-j)| \leq \sum_{k=1}^{p}|\phi_{\ell;k}| \to 0, \qquad \ell \to \infty.$$

For $\ell > 0$,

$$\left\|\frac{C_{\ell;Z}}{C_\ell}\Sigma_\ell^{-1} - I_p\right\|_\infty \leq \left\|I_p - \frac{C_\ell}{C_{\ell;Z}}\Sigma_\ell\right\|_\infty\left\|\frac{C_{\ell;Z}}{C_\ell}\Sigma_\ell^{-1}\right\|_\infty \leq const\left|\frac{C_{\ell;Z}}{C_\ell}\right|\left\|I_p - \frac{C_\ell}{C_{\ell;Z}}\Sigma_\ell\right\|_\infty.$$

Moreover, since

$$|\rho_\ell(i-j)| \leq p\|\phi_\ell\| \leq \frac{p\gamma}{\ell^\beta}, \qquad i \neq j,$$

and

$$\left|1 - \frac{C_\ell}{C_{\ell;Z}}\right| = \left|\frac{C_\ell}{C_{\ell;Z}}\right|\left|\sum_{j=1}^{p}\phi_{\ell;j}\rho_\ell(j)\right| \leq \left|\frac{C_\ell}{C_{\ell;Z}}\right|p\|\phi_\ell\| \leq \left|\frac{C_\ell}{C_{\ell;Z}}\right|\frac{p\gamma}{\ell^\beta},$$

we have

$$\left\|\frac{C_{\ell;Z}}{C_\ell}\Sigma_\ell^{-1} - I_p\right\|_\infty \leq \frac{const}{\ell^\beta},$$

as claimed. □

The last proof is for the technical Lemma on summation of squared Legendre polynomials.

Proof (Lemma 3.6.4). For $\ell \geq 1$, by Hilb's asymptotics (see [67, 75]), it holds that

$$P_\ell(\cos\theta) = \sqrt{\frac{2}{\pi\ell\sin\theta}}\sin(\ell\theta+\alpha) + \mathcal{O}\left(\ell^{-3/2}\right), \qquad 0 < \theta < \pi, \tag{3.6.11}$$

with $\alpha = \frac{\theta}{2} + \frac{\pi}{4}$. Then,

$$(2\ell+1)P_\ell^2(\cos\theta) = (2\ell+1)\left(\sqrt{\frac{2}{\pi\ell\sin\theta}}\sin\left(\ell\theta+\alpha\right) + \mathcal{O}\left(\ell^{-3/2}\right)\right)^2$$
$$= \frac{4}{\pi\sin\theta}\sin^2\left(\ell\theta+\alpha\right) + \mathcal{O}\left(\ell^{-1}\right), \qquad 0 < \theta < \pi.$$

In view of the standard identities

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i},$$

and

$$\sum_{k=0}^{n-1} e^{ixk} = \frac{1 - e^{ixn}}{1 - e^{ix}}, \qquad x \neq 0,$$

we have

$$\sum_{\ell=1}^{L}\sin^2(\ell\theta+\alpha) = \sum_{\ell=1}^{L}\left(\frac{e^{i(\ell\theta+\alpha)} - e^{-i(\ell\theta+\alpha)}}{2i}\right)^2$$
$$= -\frac{1}{4}\sum_{\ell=1}^{L}\left[e^{i2(\ell\theta+\alpha)} + e^{-i2(\ell\theta+\alpha)} - 2\right]$$
$$= -\frac{e^{i2(\theta+\alpha)}}{4}\left(\frac{1 - e^{i2\theta L}}{1 - e^{i2\theta}}\right) - \frac{e^{-i2(\theta+\alpha)}}{4}\left(\frac{1 - e^{-i2\theta L}}{1 - e^{-i2\theta}}\right) + \frac{1}{2}(L+1),$$

hence,

$$\lim_{L\to\infty}\frac{1}{L+1}\sum_{\ell=1}^{L}\sin^2(\ell\theta+\alpha) = \frac{1}{2}.$$

Also, it holds that if $\lim_{k\to\infty} a_k = A$, $|A| < \infty$, then $\lim_{k\to\infty}\frac{1}{n}\sum_{k=1}^{n} a_k = A$. As a consequence,

$$\lim_{L\to\infty}\frac{1}{L+1}\sum_{\ell=0}^{L}(2\ell+1)P_\ell^2(\cos\theta) = \frac{2}{\pi\sin\theta}, \qquad \theta \in (0,\pi).$$

Likewise, for $\theta, \theta' \in (0,\pi),\ \theta \neq \theta'$,

$$(2\ell+1)P_\ell(\cos\theta)P_\ell(\cos\theta') = \frac{4}{\pi\sqrt{\sin\theta\sin\theta'}}\sin\left(\ell\theta+\alpha\right)\sin\left(\ell\theta'+\alpha'\right) + \mathcal{O}\left(\ell^{-1}\right),$$

and

$$\sum_{\ell=1}^{L}\sin(\ell\theta+\alpha)\sin(\ell\theta'+\alpha') = \sum_{\ell=1}^{L}\left(\frac{e^{i(\ell\theta+\alpha)} - e^{-i(\ell\theta+\alpha)}}{2i}\right)\left(\frac{e^{i(\ell\theta'+\alpha')} - e^{-i(\ell\theta'+\alpha')}}{2i}\right)$$
$$= -\frac{e^{i(\theta+\theta'+\alpha+\alpha')}}{4}\left(\frac{1 - e^{i(\theta+\theta')L}}{1 - e^{i(\theta+\theta')}}\right) - \frac{e^{-i(\theta+\theta'+\alpha+\alpha')}}{4}\left(\frac{1 - e^{-i(\theta+\theta')L}}{1 - e^{-i(\theta+\theta')}}\right)$$
$$+ \frac{e^{i(\theta-\theta'+\alpha-\alpha')}}{4}\left(\frac{1 - e^{i(\theta-\theta')L}}{1 - e^{i(\theta-\theta')}}\right) + \frac{e^{-i(\theta-\theta'+\alpha-\alpha')}}{4}\left(\frac{1 - e^{-i(\theta-\theta')L}}{1 - e^{-i(\theta-\theta')}}\right),$$

hence,

$$\frac{1}{L+1}\sum_{\ell=1}^{L}\sin(\ell\theta+\alpha)\sin(\ell\theta'+\alpha')=\mathcal{O}\left(\frac{1}{L}\right).$$

In addition, since $\sum_{\ell=1}^{L}\ell^{-1}=\mathcal{O}(\log L)$, we can then conclude that

$$\frac{1}{L+1}\sum_{\ell=0}^{L}(2\ell+1)P_\ell(\cos\theta)P_\ell(\cos\theta')=\mathcal{O}\left(\frac{\log L}{L}\right),\qquad \theta,\theta'\in(0,\pi),\ \theta\neq\theta'.$$

as $L\to\infty$. □

Remark 3.6.5. *Note that (3.6.8) does not converge pointwise if θ or $\theta'=0$; for instance, for $\theta=\theta'=0$ we have $\frac{1}{L+1}\sum_{\ell=0}^{L}(2\ell+1)=L+1$, whereas for $\theta\neq 0$, $\theta'=0$ (3.6.4) oscillates among given constants.*

Chapter 4

LASSO for Spherical Functional Autoregressions[1]

4.1 Introduction

This chapter is concerned with the same class of space-time fields introduced in Chapter 2, i.e., functional autoregressive processes defined over $L^2(\mathbb{S}^2)$ (see also [10] and the references therein). More precisely, we recall that the SPHAR(p) equation is given by

$$T(x,t) = \sum_{j=1}^{p} (\Phi_j T(\cdot, t-j))(x) + Z(x,t) , \qquad (x,t) \in \mathbb{S}^2 \times \mathbb{Z} \tag{4.1.1}$$

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

$$(\Phi_j f)(x) = \int_{\mathbb{S}^2} k_j(\langle x, y \rangle) f(y)\, dy, \quad f \in L^2(\mathbb{S}^2),$$

and, as noted earlier, $k_j : [-1,1] \to \mathbb{R}$ is the corresponding autoregressive kernel, that we assume to be continuous. Furthermore, each k_j has an harmonic expansion

$$k_j(\langle x, y \rangle) = \sum_{\ell \geq 0} \phi_{\ell;j} \frac{2\ell+1}{4\pi} P_\ell(\langle x, y \rangle), \tag{4.1.2}$$

where $P_\ell : [-1,1] \to \mathbb{R}$ denotes as usual 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. Once again, as a standard consequence of the so-called duplication property for spherical harmonics (see Section 1.2), we can write

$$a_{\ell,m}(t) = \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) + a_{\ell,m;Z}(t). \tag{4.1.3}$$

In the previous chapter (see also [18]), estimators for the kernels $\{k_j : \ j = 1, \ldots, p\}$ have been defined according to a functional L^2-minimization criterion, exploiting

[1] This chapter is partially based on the preprint *LASSO for Spherical Functional Autoregressions* [17], written jointly with Claudio Durastanti and Anna Vidotto, submitted for publication.

their spectral decomposition (4.1.2); consistency, quantitative central limit theorem, and weak convergence results have then been established under some additional regularity conditions.

Here, we approach the estimation of the autoregressive kernels under a sparsity assumption, namely, for any $j = 1, \dots, p$, we assume that only a few of the components of k_j in (4.1.2) are nonzero (see Definition 4.1.1 below). It is now well-known that assuming sparsity conditions can lead to considerable advantages in estimation problems (see, for example, [30]). Indeed, on one hand, the proper identification of the null components allows one to reduce the number of predictor variables, preserving accuracy; on the other hand, sparsity enhances computational efficiency. LASSO - or ℓ_1-regularized - regression, introduced in the statistical literature by [68], is among the most popular penalization techniques to estimate sparse models. As well known, it corrects the L^2-loss for sparse models by adding a convex penalty term. 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 [29, 73] and references therein). The connections between LASSO, ridge regression, best subset selection and other ℓ_q-based penalization methods have been widely investigated; further links between LASSO and other nonparametric statistical techniques, such as soft and hard thresholding, have been widely investigated, for instance, in [13, 30].

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 [7], 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 [21, 66] 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 [7], 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 [12, 41] as well as in [10] for the functional case (see also Section 4.2). In turn, these deviation bounds are instrumental to establish concentration properties of the estimators, and so-called *oracle inequalities*.

The aim of this chapter is to define and study LASSO-type estimators for spherical autoregressive kernels under sparsity assumptions. In line with [18], 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 [72]. It is important to stress that, given the nonparametric nature of the model (4.1.1), we are dealing with a *functional* penalized regression problem, hence stepping away from the framework of VAR(p) processes, where estimators assume a vectorial form. More specifically, our oracle inequalities will involve functions rather than scalar or vectorial parameters (see Section 4.1.1). Exploiting the harmonic expansion for the spherical autoregressions (4.1.1) and the isotropy assumption on $\{k_j : j = 1, \dots, p\}$ in (4.1.2), together with an extension of the concept of stability measure introduced in [7], we will be able to establish concentration properties in functional norms for the autoregressive kernels (see Section 4.3). Moreover, the sparsity enforcement properties of LASSO procedures will allow to avoid overfitting and to select asymptotically the most relevant multipole components in the functional autoregressions.

4.1.1 Background and Main Results

Let us preliminarily introduce some standard notation, which will be instrumental to state our main findings. For real valued 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, which does not depend on the model parameters, such that $a_n \geq c\, b_n$, for all $n \in \mathbb{N}$. For a vector $v \in \mathbb{R}^d$, $\|v\|_q$ denotes the ℓ_q-norm of v,

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

for $0 < q < \infty$, $q = 0$ and $q = \infty$ respectively. Unless stated otherwise, for the sake of the simplicity, $\|\cdot\|$ denotes the ℓ_2-norm of v. Let $f : [-1,1] \to \mathbb{R}^p$, $f \in L^q([-1,1], dz)$, where dz is the Lebesgue measure over $[-1,1]$. Then, for $1 \leq q < \infty$, the L^q-norm of f is given by

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

Analogously, the L^∞-norm of f is given by $\|f\|_{L^\infty} = \sup_{z\in[-1,1]} \|f(z)\|$. Finally, recall that 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

$$\|\mathscr{T}\|_{\mathrm{HS}} = \sum_{i=1}^{\infty} |\lambda_i|^2; \quad \|\mathscr{T}\|_{\mathrm{TR}} = \sum_{i=1}^{\infty} |\lambda_i|,$$

where $\{\lambda_i\}_{i\in\mathbb{N}}$ are the eigenvalues of $\mathscr{T}$ (see, for example, [32]).
We provide now the definition of sparsity index, which can be understood as a more rigorous characterization of sparsity.

Definition 4.1.1 (Sparsity index)**.** *For any $\ell \geq 0$, $\boldsymbol{\phi}_\ell = (\phi_{\ell;1}, \dots, \phi_{\ell;p})$ is a q_ℓ-sparse vector if*

$$\|\boldsymbol{\phi}_\ell\|_0 = q_\ell,$$

where q_ℓ is the ℓ-th sparsity index, which satisfies $1 \leq q_\ell \leq p$. We call $\{q_\ell : \ell \geq 0\}$ the sparsity set.

Remark 4.1.2. *Following [10, 18], to ensure identifiability we assume that there exists at least one $\ell \geq 0$ such that $\phi_{\ell;p} \neq 0$, so that $\mathbb{P}(\Phi_p T(\cdot,t) \neq 0) > 0$, for all $t \in \mathbb{Z}$. As a consequence, for some $\ell \geq 0$, we can have $\boldsymbol{\phi}_\ell = 0$ and hence $\|\boldsymbol{\phi}_\ell\|_0 = q_\ell = 0$; however, $q = \max_{\ell\geq 0} q_\ell \geq 1$.*

Set, as usual, the sequence of polynomials $\phi_\ell : \mathbb{C} \to \mathbb{C}$, $\ell \geq 0$, so that

$$\phi_\ell(z) = 1 - \phi_{\ell;1} z - \dots - \phi_{\ell;p} z^p. \tag{4.1.4}$$

In line with the previous chapter, we must introduce some conditions on the model, essential to achieve our results.

Condition 4.1.3 (Identifiability). *Let $\{Z(x,t), (x,t) \in \mathbb{S}^2 \times \mathbb{Z}\}$ be the Gaussian spherical white noise used in (4.1.1). Then it holds that*

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

for any $f \in L^2(\mathbb{S}^2)$ such that $f(\cdot) \neq 0$. Equivalently $C_{\ell;Z} > 0$, for all $\ell \geq 0$.

Condition 4.1.4 (Stationarity). *The sequence of polynomials (4.1.4) is such that*

$$|z| < 1 + \delta \;\Rightarrow\; \phi_\ell(z) \neq 0,$$

for some $\delta > 0$. More explicitly, there are no roots in a δ-enlargement of the unit disk, for all $\ell \geq 0$.

Condition 4.1.5 (Smoothness). *For all $j = 1, \ldots, p$, we have that*

$$\|\Phi_j\|_{\mathrm{TR}} = \sum_{\ell=0}^{\infty} (2\ell+1)|\phi_{\ell;j}| < \infty, \tag{4.1.5}$$

that is, Φ_j is a nuclear operator, see again [32].

From now on, for any $\ell \geq 0$, we assume that we are able to observe the harmonic coefficients $\{a_{\ell,m}(t) : m = -\ell, \ldots, \ell\}$ over a finite set of times $\{1, \ldots, n\}$. The vector of functions

$$k := (k_1, \ldots, k_p)$$

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

$$\widehat{k}_N^{\text{lasso}} = \operatorname*{argmin}_{k \in \mathcal{P}_N^p} \sum_{t=p+1}^{n} \left\| T(\cdot,t) - \sum_{j=1}^{p} \Phi_j T(\cdot, t-j) \right\|_{L^2(\mathbb{S}^2)}^2 + \lambda \sum_{j=1}^{p} \|\Phi_j\|_{\mathrm{TR}}, \tag{4.1.6}$$

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

$$\operatorname{span}\left\{ \frac{2\ell+1}{4\pi} P_\ell(\cdot) : \ell = 0, \ldots, L_N - 1 \right\}, \tag{4.1.7}$$

where the integer $L_N > 0$ is the truncation level, which corresponds to the frequency of the highest component in (4.1.2) estimated by (4.1.6), see Section 4.2.1 for a detailed discussion. Let us furthermore define

$$k_{j;N}(\langle x,y \rangle) := \sum_{\ell=0}^{L_N-1} \phi_{\ell;j} \frac{2\ell+1}{4\pi} P_\ell(\langle x,y \rangle) \tag{4.1.8}$$

and accordingly $k_N := (k_{1;N}, \ldots, k_{p;N})$.

Our main result is extensively stated in Theorem 4.3.12 and can be compactly formulated as follows.

Theorem 4.1.6. *Consider the estimation problem (4.1.6), assume that Conditions 4.1.3 and 4.1.4 hold, and suppose that*

$$N \succeq b_1 \, q \log\left(pL_N\right),$$

where $q = \max_{\ell \geq 0} q_\ell$*. Then, for any penalty parameter* $\lambda = \lambda_N \geq b_2 \sqrt{\frac{\log(pL_N)}{N}}$*, the solution* $\widehat{k}_N^{\text{lasso}}$ *of (4.1.6) satisfies*

$$\mathbb{P}\left(\left\| \widehat{k}_N^{\text{lasso}} - k \right\|_{L^2}^2 \leq \frac{18}{\pi^2} \lambda_N^2 \sum_{\ell=0}^{L_N-1} \frac{q_\ell}{\alpha_\ell^2} (2\ell+1) + \left\| k - k_N \right\|_{L^2}^2 \right) \geq 1 - c_1 e^{-c_2 \log(pL_N)},$$

where the sequence $\{\alpha_\ell : \ell = 0, \ldots, L_N - 1\}$ *is defined in (4.3.12), and* $c_1, c_2 > 0$ *are absolute constants. Moreover, under the additional Condition 4.1.5, it holds that*

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

Remark 4.1.7. *The constants* b_1 *and* b_2 *depend on the model. In particular,* $b_1 = \max\left\{\omega^2, 1\right\}$ *and* $b_2 = 4\mathcal{F}$*, where* $\mathcal{F}, \omega > 0$ *are tightly connected to the stability measure introduced in Section 4.3.1. The reader is referred to Section 4.3.2 for further details and comments.*

Our findings provide upper bounds for the L^2- and the L^∞-distances between the LASSO-type estimator $\widehat{k}_N^{\text{lasso}}$ and the true 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 k with $\widehat{k}_N^{\text{lasso}}$. The second one arises because $\widehat{k}_N^{\text{lasso}}$ provides an estimation of 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 errors, respectively (see [72]). 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 is arbitrarily close to 1. Of course, with an appropriate choice of L_N, both the upper bounds converge to 0 and their probabilities converge to 1, as $N \to \infty$; as a consequence, our result can be read in terms of asymptotic consistency.

4.1.2 Plan of the Chapter

The chapter is organized as follows. In Section 4.2, we present the LASSO estimators for spherical autoregressive kernels under sparsity assumptions. Section 4.3 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 4.4 we briefly show the performance of the LASSO estimators under sparsity assumptions. Finally, Section 4.5 collects the proofs.

4.2 LASSO Estimation on the Sphere

Here we present ℓ_1-regularized estimators for spherical autoregressive kernels under sparsity assumptions. More specifically, merging the techniques based on the stability measure presented in [7] with the properties of Gaussian isotropic stationary

SPHAR(p) processes enables the construction of functional estimators for the kernels $\{k_j : j = 1, \ldots, p\}$.

4.2.1 The Construction of the Estimator

As seen in Section 4.1, 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.

Definition 4.2.1 (LASSO Estimator)**.** *The LASSO estimator for the vector of kernels k is defined by*

$$\widehat{k}_N^{\text{lasso}} := \underset{k \in \mathcal{P}_N^p}{\operatorname{argmin}} \sum_{t=p+1}^{n} \left\| T(\cdot, t) - \sum_{j=1}^{p} \Phi_j T(\cdot, t-j) \right\|_{L^2(\mathbb{S}^2)}^2 + \lambda \sum_{j=1}^{p} \|\Phi_j\|_{\mathrm{TR}}. \tag{4.2.1}$$

Arguing as in the previous chapter leads to

$$\begin{aligned} Z(x,t) &= T(x,t) - \sum_{j=1}^{p} \left(\Phi_j T(\cdot, t-j)\right)(x) \\ &= \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left(a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right) Y_{\ell,m}(x), \end{aligned}$$

where $\{Y_{\ell,m}(\cdot),\ \ell \geq 0,\ m = -\ell, \ldots, \ell\}$ is a standard basis of *real* spherical harmonics (see Remark 1.2.2) and, for fixed (ℓ, m), the random coefficients $\{a_{\ell,m;Z}(t),\ t \in \mathbb{Z}\}$ associated with the Gaussian spherical white noise can be seen as the residuals of the one-dimensional autoregressive process $\{a_{\ell,m}(t),\ t \in \mathbb{Z}\}$.
For any $\ell \geq 0$, the p-dimensional vector of regressors is given by

$$\boldsymbol{\phi}_\ell = \left(\phi_{\ell;1}, \ldots, \phi_{\ell;p}\right)'.$$

As a consequence,

$$\sum_{t=p+1}^{n} \left\| T(\cdot, t) - \sum_{j=1}^{p} \Phi_j T(\cdot, t-j) \right\|_{L^2(\mathbb{S}^2)}^2 = \sum_{t=p+1}^{n} \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \left| a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right|^2.$$

Let us fix a truncation level L_N, thus depending on the number of observations N. We can also define the *truncated residual sum of squares* as

$$\begin{aligned} S(\boldsymbol{\phi}_0, \ldots, \boldsymbol{\phi}_{L-1}) &= \sum_{t=p+1}^{n} \sum_{\ell=0}^{L_N-1} \sum_{m=-\ell}^{\ell} \left| a_{\ell,m}(t) - \sum_{j=1}^{p} \phi_{\ell;j} a_{\ell,m}(t-j) \right|^2 \\ &= \sum_{\ell=0}^{L_N-1} S_\ell(\boldsymbol{\phi}_\ell), \end{aligned} \tag{4.2.2}$$

where

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

Using (4.2.2), we obtain that

$$\begin{aligned}\widehat{k}_N^{\text{lasso}} &= \underset{k\in\mathcal{P}_N^p}{\operatorname{argmin}}\, S\left(\boldsymbol{\phi}_0,\ldots,\boldsymbol{\phi}_{L_N-1}\right) + \lambda \sum_{j=1}^{p} \|\Phi_j\|_{\text{TR}} \\ &= \underset{k\in\mathcal{P}_N^p}{\operatorname{argmin}} \sum_{\ell=0}^{L_N-1} \left[S_\ell(\boldsymbol{\phi}_\ell) + \lambda\left(2\ell+1\right)\|\boldsymbol{\phi}_\ell\|_1\right] \\ &= \sum_{\ell=0}^{L_N-1} \widehat{\boldsymbol{\phi}}_{\ell;N}^{\text{lasso}} \frac{2\ell+1}{4\pi} P_\ell,\end{aligned}$$

where $\mathcal{P}_N^p$ is given by (4.1.7), and

$$\begin{aligned}\widehat{\boldsymbol{\phi}}_{\ell;N}^{\text{lasso}} &= \underset{\boldsymbol{\phi}_\ell\in\mathbb{R}^p}{\operatorname{argmin}}\, S_\ell(\boldsymbol{\phi}_\ell) + \lambda(2\ell+1)\|\boldsymbol{\phi}_\ell\|_1 \\ &= \underset{\boldsymbol{\phi}_\ell\in\mathbb{R}^p}{\operatorname{argmin}} \frac{1}{N(2\ell+1)} S_\ell(\boldsymbol{\phi}_\ell) + \frac{\lambda}{N}\|\boldsymbol{\phi}_\ell\|_1 . \end{aligned} \tag{4.2.3}$$

Remark 4.2.2. *Note that the penalization procedure* (4.2.1) *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* $a_{\ell.}$ *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* j*-th component of* $\widehat{k}_N^{\text{lasso}}$*. The reader is referred for further discussions to [14], where it is shown that not all the* ℓ_1*-penalized problems have solutions which are isotropic, and to [38] for sparsity enforcing procedures in the case of isotropic spherical random fields.*

An alternative form for the ℓ_1-penalized problem given by (4.2.3) can be introduced as follows. First, we define the $N\left(2\ell+1\right)$-dimensional vectors

$$\begin{aligned}&Y_{\ell;N} = \left(a_{\ell,-\ell}\left(n\right),\ldots,a_{\ell,-\ell}\left(p+1\right),\ldots,a_{\ell,\ell}\left(p+1\right)\right)',\\ &Y_{\ell;N}\left(h\right) = \left(a_{\ell,-\ell}\left(n-h\right),\ldots,a_{\ell,-\ell}\left(p+1-h\right),\ldots,a_{\ell,\ell}\left(p+1-h\right)\right)', \quad h=1,\ldots,p,\\ &E_{\ell;N} = \left(a_{\ell,-\ell;Z}\left(n\right),\ldots,a_{\ell,-\ell;Z}\left(p+1\right),\ldots a_{\ell,\ell;Z}(p+1)\right),\end{aligned}$$

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

$$X_{\ell;N} = \left[Y_{\ell;N}\left(1\right),\ldots,Y_{\ell;N}\left(p\right)\right] \tag{4.2.4}$$

Note that the LASSO problem (4.2.1) reduces to

$$\sum_{\ell=0}^{L_N-1} \underset{\boldsymbol{\phi}_\ell\in\mathbb{R}^p}{\operatorname{argmin}} \left[\frac{1}{N(2\ell+1)}\|Y_{\ell;N} - X_{\ell;N}\boldsymbol{\phi}_\ell\|_2^2 + \frac{\lambda}{N}\|\boldsymbol{\phi}_\ell\|_1\right]\frac{2\ell+1}{4\pi}P_\ell.$$

Fixed $\ell=0,\ldots,L_N-1$, we define the covariance matrix Γ_ℓ, namely, the $p\times p$ matrix with generic ij-th element $C_\ell\left(i-j\right)$. We can use (4.2.4) to define its unbiased estimator

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

Let us now consider the product $X'_{\ell;N}E_{\ell;N}/N(2\ell+1)$. Observe that $E_{\ell;N}$ 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 [13] and the references therein), associated with the multipole ℓ. Furthermore, observe that the ℓ-th empirical process corresponds to the following sum

$$\frac{X'_{\ell;N}E_{\ell;N}}{N\left(2\ell+1\right)}=\frac{X'_{\ell;N}Y_{\ell;N}}{N\left(2\ell+1\right)}-\frac{X'_{\ell;N}X_{\ell;N}}{N\left(2\ell+1\right)}\boldsymbol{\phi}_{\ell}=\widehat{\gamma}_{\ell;N}-\widehat{\Gamma}_{\ell;N}\boldsymbol{\phi}_{\ell}\,, \tag{4.2.5}$$

where

$$\widehat{\gamma}_{\ell;N}=\frac{X'_{\ell;N}Y_{\ell;N}}{N\left(2\ell+1\right)}\,.$$

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

4.3 Main Results

In this section, we present the main results of this chapter, which consist of properties for the LASSO-type estimator k given in (4.2.1). First of all, in Section 4.3.1 we introduce the concept of stability measure, a powerful tool, firstly proposed in the LASSO framework by [7], to obtain some bounds on the concentration of the sample covariances and the empirical processes around their expected values. Then, in Section 4.3.2 we will follow the standard scheme of LASSO-techniques, see [30, 13], to establish, as it is usually done, a *basic inequality*, a *deviation condition* and a *compatibility condition*. Finally, in Section 4.3.3 we present our main theorem, showing the so-called *oracle inequalities* for $\widehat{k}_N^{\mathrm{lasso}}$.

4.3.1 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 correlation of the components of the process. Several proposals aiming to represent the stability of a given process have been suggested in the literature over the years (see, for example [77, 65, 35]), mostly involving set of mixing conditions in order to assess for how long in time the dependence between the components is effective. The stability measure considered here is in line with the one defined by [7].

First of all, recall that $\{a_{\ell,m}\left(t\right), t\in\mathbb{Z}\}$ can be read as a real-valued autoregressive process of order p. Under standard stationarity assumptions (see [12, page 123]), we can define its spectral density as

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

which is bounded and continuous (see also [18]). Upper and lower extrema of the spectral density over the unit circle are hence given by

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

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 [7], 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 p-dimensional process

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

with spectral density and corresponding stability measure

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

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 $j_1 < \ldots < j_r$. Then, we define

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

where $(\tilde{a}_{\ell,m}(t))_i$ is the i-th component of $\{\tilde{a}_{\ell,m}(t) : t \in \mathbb{Z}\}$. This subprocess has spectral density $\tilde{f}^J_\ell(\nu)$. We can finally introduce the associated band-limited, global and observed stability measures, respectively

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

Notice that $\mathcal{M}\left(\tilde{f}_\ell\right) = \mathcal{M}\left(\tilde{f}_\ell, p\right)$, while, for the sake of completeness, we define $\mathcal{M}\left(\tilde{f}_\ell, r\right) = \mathcal{M}(\tilde{f}_\ell)$, for all $r > p$. Moreover, it can be shown that

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

The following quantities are also well-defined

$$\mu_{\min;\ell} := \min_{z \in \mathbb{C}:|z|=1} |\phi_\ell(z)|^2, \quad \mu_{\max;\ell} := \max_{z \in \mathbb{C}:|z|=1} |\phi_\ell(z)|^2, \tag{4.3.1}$$

for every $\ell \geq 0$.

Remark 4.3.1. *Note that,*

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

see also Chapter 2 (Remark 2.2.7).

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 analyse our regression problem. Note that, $\{a_{\ell,m}(t) : t \in \mathbb{Z}\}$, $m = -\ell, \ldots, \ell$, can be seen as a tool to provide an alternative notation for the empirical process (4.2.5). Indeed the h-th component of $X'_{\ell;N}E_{\ell;N}/N(2\ell+1)$ is given by

$$\frac{Y_{\ell;N}(h)' E_{\ell;N}}{N(2\ell+1)} = \frac{1}{N(2\ell+1)}\sum_{m=-\ell}^{\ell}\sum_{t=p+1}^{n} a_{\ell,m}(t-h)\,a_{\ell,m;Z}(t)\,. \tag{4.3.2}$$

Proposition 4.3.2 (Deviation bounds)**.** *There exists a constant $c>0$ such that for 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*

$$\mathbb{P}\left(\left|v'\left(\widehat{\Gamma}_{\ell;N}-\Gamma_{\ell}\right)v\right| > 2\pi\mathcal{M}\left(\widetilde{f}_{\ell}, r\right)\eta\right) \leq 2e^{-c\,N(2\ell+1)\min\left\{\eta^{2},\eta\right\}}, \tag{4.3.3}$$
$$\mathbb{P}\left(\left|u'\left(\widehat{\Gamma}_{\ell;N}-\Gamma_{\ell}\right)v\right| > 6\pi\mathcal{M}\left(\widetilde{f}_{\ell}, 2r\right)\eta\right) \leq 6e^{-cN(2\ell+1)\min\left\{\eta^{2},\eta\right\}}. \tag{4.3.4}$$

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

$$\mathbb{P}\left(\left|\left(\widehat{\Gamma}_{\ell;N}-\Gamma_{\ell}\right)_{ij}\right| > 6\pi\mathcal{M}(\widetilde{f}_{\ell},2)\eta\right) \leq 6e^{-c\,N(2\ell+1)\min\left\{\eta^{2},\eta\right\}}\,. \tag{4.3.5}$$

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

$$\mathbb{P}\left(\left|\frac{Y'_{\ell;N}(h)\,E_{\ell;N}}{N(2\ell+1)}\right| > 2\pi\,C_{\ell;Z}\left(1+\frac{1+\mu_{\max;\ell}}{\mu_{\min;\ell}}\right)\eta\right) \leq 6e^{-cN(2\ell+1)\min\left\{\eta^{2},\eta\right\}}, \tag{4.3.6}$$

where $\mu_{\max;\ell}$ and $\mu_{\min;\ell}$ are defined by (4.3.1).

Remark 4.3.3. *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 [7] 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 [7] is that, in our framework, we use observations from a group of $2\ell+1$ stationary processes, namely, $\{a_{\ell,m}(t) : t \in \mathbb{Z}\}$, $m = -\ell, \ldots, \ell$, to estimate the same covariance matrix Γ_ℓ, exploiting the isotropy of the field.
Similarly to [7], (4.3.3)-(4.3.6) quantify how the underlying estimators concentrate around their expected values. In particular, (4.3.3) will be used to verify the

compatibility condition (see Proposition 4.3.6), while (4.3.6) will be used to prove the deviation condition (see Proposition 4.3.8). 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, analogously to [7], the ℓ-th 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.

4.3.2 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 4.3.4 (Basic Inequality)**.** *Consider the estimation problem* (4.2.1)*. For any $\ell = 0, \dots, L_N - 1$, set $v_\ell := \widehat{\boldsymbol{\phi}}^{\text{lasso}}_{\ell;N} - \boldsymbol{\phi}_\ell$. The following basic inequality holds*

$$v_\ell' \widehat{\Gamma}_{\ell;N} v_\ell \leq \frac{2 v_\ell' X_{\ell;N}' E_{\ell;N}}{N(2\ell+1)} + \frac{\lambda}{N} \left[\|\boldsymbol{\phi}_\ell\|_1 - \|\boldsymbol{\phi}_\ell + v_\ell\|_1 \right] \tag{4.3.7}$$

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 $X_{\ell;N}' E_{\ell;N} / N\,(2\ell+1)$. The second one is deterministic and its value depends on the penalty parameter λ, the number of observations N, and the chosen linear model itself.

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

$$\frac{2 v_\ell' X_{\ell;N}' E_{\ell;N}}{N(2\ell+1)}\,, \quad \ell = 0, \dots, L_N - 1\,,$$

when conditioning 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 4.3.5. *In the setting previously described, let*

$$\mathscr{S}_N = \bigcap_{\ell=0}^{L_N-1} \left\{ \left\| \widehat{\gamma}_{\ell;N} - \widehat{\Gamma}_{\ell;N} \boldsymbol{\phi}_\ell \right\|_\infty \leq \mathcal{F}_N \sqrt{\frac{\log p L_N}{N}} \right\}, \tag{4.3.8}$$

where $\mathcal{F}_N$ is a deterministic function depending only on the parameters $(\boldsymbol{\phi}_0, \dots, \boldsymbol{\phi}_{L_N-1})$ ad noise variances $(C_{0;Z}, \dots, C_{L_N-1;Z})$. 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 4.3.6 (Deviation condition)**.** *Consider the regression problem* (4.1.1) *and the proposed estimator* $\widehat{k}_N^{\text{lasso}}$ *described in* (4.2.1)*. Assume that Conditions 4.1.3 and 4.1.4 hold. There exist some absolute constants* $c_0, c_1, c_2 > 0$ *such that, if we define*

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

and if $N \succeq \log pL_N$*, then*

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1}\left\|\widehat{\gamma}_{\ell;N}-\widehat{\Gamma}_{\ell;N}\phi_\ell\right\|_\infty \leq \mathcal{F}_N\sqrt{\frac{\log pL_N}{N}}\right) \geq 1-c_1e^{-c_2\log(pL_N)}.$$

Remark 4.3.7. *Observe that:*

(i) the results presented in Proposition 4.3.6 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 [7];

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

$$\log\left(pL_N\right) = o\left(N\right).$$

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 $\left\{\left\|X_{\ell;N}\left(\widehat{\phi}_{\ell;N}^{\text{lasso}}-\phi_\ell\right)\right\|_2^2\right\}$ and the estimation errors $\left\{\left\|\widehat{\phi}_{\ell;N}^{\text{lasso}}-\phi_\ell\right\|_2^2\right\}$ 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 $\vartheta \in \mathbb{R}^d$,

$$\vartheta' A\vartheta \geq \alpha\left\|\vartheta\right\|_2^2 - \tau\left\|\vartheta\right\|_1^2. \tag{4.3.9}$$

The next result gives some sufficient conditions in order to have

$$\widehat{\Gamma}_{\ell;N} \sim RE(\alpha,\tau),$$

for some α and τ, with high probability.

Proposition 4.3.8 (Compatibility condition)**.** *Consider the estimation problem* (4.2.1) *and assume that Conditions 4.1.3 and 4.1.4 hold. Define* $q_N = \max_{\ell<L_N} q_\ell$*. There exist some absolute constants* $c_1, c_2, c_3 > 0$ *such that, if*

$$N \succeq \max\left\{\omega_N^2, 1\right\} q_N \log\left(pL_N\right), \quad \textit{with} \quad \omega_N = c_3\max_{\ell<L_N}\frac{\mu_{\max;\ell}}{\mu_{\min;\ell}}, \tag{4.3.10}$$

then

$$\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1}\left\{\widehat{\Gamma}_{\ell;N}\sim RE(\alpha_\ell,\tau_\ell)\right\}\right)\geq 1-c_1\,e^{-c_2\,N\,\min\{\omega_N^{-2},1\}}, \tag{4.3.11}$$

with

$$\alpha_\ell=\frac{C_{\ell;Z}}{2\,\mu_{\max;\ell}},\quad \text{and}\quad \tau_\ell=\alpha_\ell\,\max\left\{\omega_N^2,1\right\}\frac{\log(pL_N)}{N}. \tag{4.3.12}$$

Remark 4.3.9. *The results presented in Proposition 4.3.8 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 [7].

Remark 4.3.10. *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 (4.3.10) ensures that, with high probability, the minimum (sample) eigenvalue of the matrix $\widehat{\Gamma}_{\ell;N}$ is bounded away from zero.*

Remark 4.3.11. *Note that $0\leq\mathcal{F}_N\leq\mathcal{F}$, where $\mathcal{F}:=\max_{\ell\geq 0}\mathcal{F}(\phi_\ell,C_{\ell;Z})$ exists finite. Indeed,*

$$\mathcal{F}(\phi_\ell,C_{\ell;Z})=c_0\left[C_{\ell;Z}\left(1+\frac{1+\mu_{\max;\ell}}{\mu_{\min;\ell}}\right)\right]\to 0,\qquad \ell\to\infty,$$

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 4.3.1. Similarly, $0\leq\omega_N\leq\omega$, 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 chapter can be stated using $\mathcal{F},\omega,q$ instead of $\mathcal{F}_N,\omega_N,q_N$. Without loss of generality, we can assume $q_N\geq 1$.

4.3.3 Oracle Inequalities

Oracle inequalities are used to estimate the accuracy of the $\widehat{k}_N^{\text{lasso}}$. Observe that, in general, $\widehat{k}_N^{\text{lasso}}$ depends on the penalty parameter λ, according to (4.2.1). As a consequence, given a proper choice of λ, oracle inequalities produce upper bounds for the estimation error with high probability. Such upper bounds are characterized by a multiplicative factor $\log(pL_N)$; roughly speaking, this factor is the cost for not knowing explicitly the set of nonzero coefficients.

Theorem 4.3.12. *Consider the estimation problem (4.2.1) and assume that Conditions 4.1.3 and 4.1.4 hold. Moreover, suppose that, for any $\ell=0,\dots,L_N-1$, $\widehat{\Gamma}_{\ell;N}\sim RE(\alpha_\ell,\tau_\ell)$ with $q_\ell\tau_\ell\leq\alpha_\ell/32$ and that $(\widehat{\Gamma}_{\ell;N},\widehat{\gamma}_{\ell;N})$ satisfies the deviation condition almost surely, that is,*

$$\left\|\widehat{\gamma}_{\ell;N}-\widehat{\Gamma}_{\ell;N}\phi_\ell\right\|_\infty\leq\mathcal{F}_N\sqrt{\frac{\log(pL_N)}{N}}\quad a.s. \tag{4.3.13}$$

Then, for any $\lambda_N \geq 4\mathcal{F}_N\sqrt{\log(pL_N)/N}$*, where* $\mathcal{F}_N := \max_{\ell<L_N} \mathcal{F}\left(\phi_\ell, C_{\ell;Z}\right)$*, any solution* $\widehat{k}_N^{\text{lasso}}$ *of* (4.2.1) *satisfies*

$$\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_{L^2}^2 \leq \frac{18}{\pi^2}\lambda_N^2 \sum_{\ell=0}^{L_N-1} \frac{q_\ell}{\alpha_\ell^2}(2\ell+1) + \left\|k - k_N\right\|_{L^2}^2; \tag{4.3.14}$$

moreover, under the additional Condition 4.1.5, it holds that

$$\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_{L^\infty} \leq \frac{3}{\pi}\lambda_N \sum_{\ell=0}^{L_N-1} \frac{\sqrt{q_\ell}}{\alpha_\ell}(2\ell+1) + \left\|k - k_N\right\|_{L^\infty}. \tag{4.3.15}$$

Rates of Convergence. To discuss the possible rates of convergence in (4.3.14)-(4.3.15), let us choose $\lambda_N = 4\mathcal{F}_N\sqrt{\log(pL_N)/N}$ and $L_N \sim N^d$. Moreover, we impose some semiparametric structure to the set $\{\phi_{\ell;j} : \ell \geq 0\}$, that is,

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

where $\beta_j > 1$ and $G_j > 0$ (see also Chapter 3). Note that, since we are looking at the asymptotic behaviour as $N \to \infty$, the sufficient condition (4.3.10) automatically holds and the coefficients $\alpha_\ell \sim C_{\ell;Z}$. In particular, a standard assumption for the behaviour of the power spectrum of a spherical white noise is $C_{\ell;Z} \sim \ell^{-\alpha_*}$, with $\alpha_* > 2$, see [43]. As a consequence, in this framework, we have that

$$\begin{aligned}\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_{L^2}^2 &\leq \frac{18}{\pi^2}\lambda_N^2 \sum_{\ell=0}^{L_N-1} \frac{q_\ell}{\alpha_\ell^2}(2\ell+1) + \sum_{\ell=L_N}^{\infty} \|\phi_\ell\|_2^2 \frac{2\ell+1}{8\pi^2} \\ &\leq \mathit{const}\left[\frac{\log N}{N}\sum_{\ell=0}^{L_N-1} \ell^{2\alpha_*}(2\ell+1) + \sum_{\ell=L_N}^{\infty} \ell^{-2\beta_*}(2\ell+1)\right] \\ &= O\left(\log N\, N^{2d(\alpha_*+1)-1} + N^{2d(1-\beta_*)}\right),\end{aligned}$$

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

$$\begin{aligned}\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_{L^\infty} &\leq \frac{3}{\pi}\lambda_N \sum_{\ell=0}^{L_N-1} \frac{\sqrt{q_\ell}}{\alpha_\ell}(2\ell+1) + \sum_{\ell=L_N}^{\infty} \|\phi_\ell\|_2 \frac{2\ell+1}{4\pi} \\ &= O\left((\log N)^{1/2} N^{d(\alpha_*+2)-1/2} + N^{d(2-\beta_*)}\right)\end{aligned}$$

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

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

We stress that the parameter α_* can be estimated via a Whittle-like procedure, see [22, 23] and Remark 3.3.6.

4.4 Some Numerical Results

In this Section we will describe the numerical implementation to support the results provided along the chapter. More specifically, here we briefly discuss the performance of the LASSO estimator $\widehat{k}_N^{\text{lasso}}$ under sparsity assumptions.
Fixed $N = 300$ and $L_N = L = 50$, we are concerned with the empirical evaluation of the L^2-risk of $\widehat{k}_N^{\text{lasso}}$ for the penalty parameters $\lambda_i = 10^{i-6}$, $i = 1, \ldots, 6$, in comparison with the one of the non-penalized estimator described in Chapter 3 (see also [18]), 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 four different case studies, all belonging to the class of SPHAR(2) processes, so that (4.1.1) becomes

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

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 second case, the random field T_2 is characterized by less sparsity; in particular, the non-null coefficients are

$$\phi_{30;1} = -0.72, \quad \phi_{31;1} = 0.31, \quad \phi_{32;1} = 0.85, \quad \phi_{2,2} = 0.25, \phi_{3,2} = -0.87, \quad \phi_{5,2} = -0.98.$$

In the third case, the random field T_3 is not sparse, even if the eigenvalues are taken to be relevant only on the first 20 multipoles, that is, $\phi_{\ell,j} \propto \ell^{-2}$ for $\ell \geq 20$ and $j = 1, 2$. Finally, in the fourth case T_4, all the multipoles are assumed to be relevant.

MSE	T_1	T_2	T_3	T_4
λ_0	0.00218	0.00212	0.00189	0.00143
λ_1	0.00214	0.00208	0.00185	0.00144
λ_2	0.00181	0.00175	0.00153	0.00149
λ_3	0.00083	0.00087	0.00071	0.00650
λ_4	0.00032	0.00478	0.00190	0.46580
λ_5	0.00015	0.15363	0.10620	7.32961
λ_6	0.00049	0.53786	1.57303	15.18901

Table 4.1. Values of the mean squared error (MSE) for the four case studies $T_1, \ldots, T_4$, by varying the penalty parameter λ. Note that λ_0 corresponds to the non-penalized estimation.

Table 4.1 collects the values of the empirical mean squared error (MSE) associated with the four models of interest. In particular, for $B = 1000$ replications, we have

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

where $\{z_1, \ldots, z_G\}$, $G = 2000$, is an equally spaced grid over $[-1, 1]$. As expected, LASSO-type estimators provide smaller MSE when we consider highly sparse models (T_1 and T_2). Regarding T_3, the situation does not really change because, after the

20-th multipole, the values of the eigenvalues are very small. On the contrary, for the model T_4, the penalized estimator always performs poorly when λ grows. Note that for $\lambda = \lambda_1$ the penalization is very small and, as a consequence, the penalized and non-penalized estimators are almost equivalent.

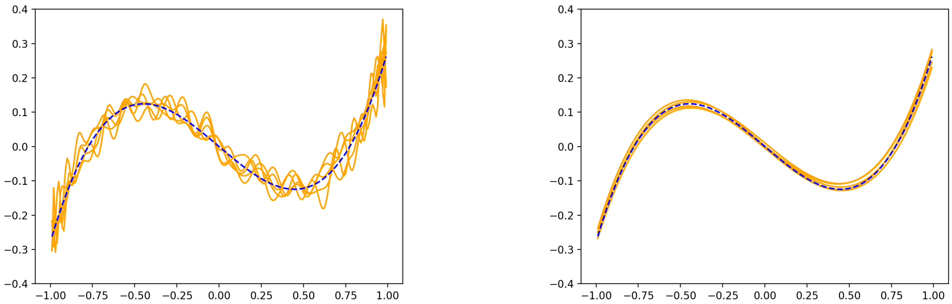


Figure 4.1. Five estimates (orange lines) of k_2 for the model T_1, with $\lambda = \lambda_5$. The true k_2 is the blue dashed line.

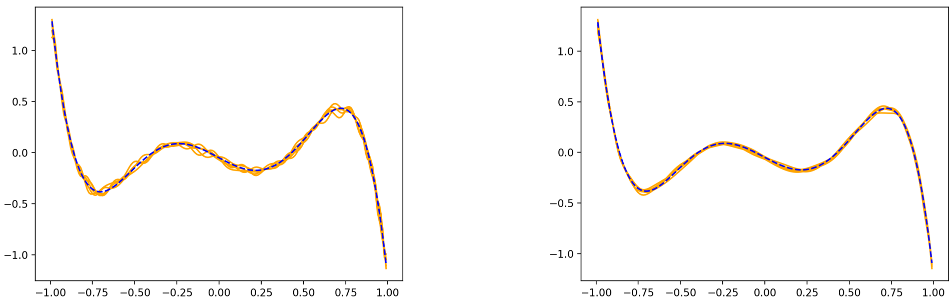


Figure 4.2. Five estimates (orange lines) of k_2 for the model T_2, with $\lambda = \lambda_3$. The true k_2 is the blue dashed line.

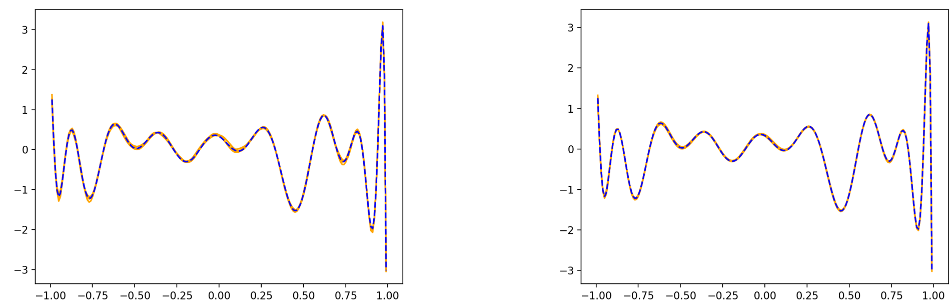


Figure 4.3. Five estimates (orange lines) of k_2 for the model T_3, with $\lambda = \lambda_3$. The true k_2 is the blue dashed line.

Figures 4.1-4.3 illustrate the functional forms of the estimated kernel, compared to the real one. 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 for T_1 and T_2 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. For the model T_3, which is non-sparse but has few relevant multipoles, the difference between the two functional estimations is not significant.

4.5 Proofs

In the present section we prove the bounds showed in Section 4.3.2 as well as our main theorem. Many of the proofs have arguments which are broadly similar to those given for related results in [7].

Proof of Proposition 4.3.2. Let us define $J = \operatorname{supp}(v) = \{j_1, \dots, j_r\} \subset \{1, \dots, p\}$, $r \geq 1$, and

$$W_{\ell,J} = X_{\ell;N} v = \sum_{j \in J} v_j Y_{\ell;N}(j)\,.$$

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

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

As a consequence, $\|Q_{\ell,J}\|_{op} = \|B_{\ell,J}\|_{op} \leq 2\pi \mathcal{M}(\widetilde{f}_\ell, r)$ (see [7, Proposition 2.4]) and (4.3.3) is proved. To prove (4.3.4), 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 (4.3.3) separately on each of the three terms on the right. The element-wise deviation bound (4.3.5) is obtained by choosing $u = e_i$, $v = e_j$.

Let us now prove (4.3.6). Recall (4.3.2); the following decomposition holds

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

Implementing (4.3.3) for $v = e_h$, we have

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

which implies

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

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 (4.3.3) (setting $v \in \mathbb{R}$, $v = 1$, and obviously $r = 1$), one can show that

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

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

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

Moreover,

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

which implies, $\mathcal{M}(f_{\ell;T+Z}) \leq \mathcal{M}(f_\ell) + \mathcal{M}(f_{\ell;Z}) + \mathcal{M}(f_{\ell;(T,Z)})$, 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 $\{(a_{\ell,m}(t-h), a_{\ell,m;Z}(t))\,,\, t \in \mathbb{Z}\}$.
Now, using the obvious implications

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

we have that

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

Following the last steps of the proof of Proposition 2.4 in [7], we obtain

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

which finally implies (4.3.6).

□

Proof of Proposition 4.3.4. Since $\widehat{\phi}^{\text{lasso}}_{\ell;N}$ is the solution of the minimization problem (4.2.3), we have that

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

which, using the definition of $Y_{\ell;N}$, becomes

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

Now, we have

$$\frac{\left\| X_{\ell;N}\phi_\ell - X_{\ell;N}\widehat{\phi}^{\text{lasso}}_{\ell;N} \right\|_2^2}{N(2\ell+1)} + \frac{\| E_{\ell;N} \|_2^2}{N(2\ell+1)} - 2\frac{\left(\widehat{\phi}^{\text{lasso}}_{\ell;N} - \phi_\ell\right)' X'_{\ell;N}E_{\ell;N}}{N(2\ell+1)} + \frac{\lambda}{N}\left\| \widehat{\phi}^{\text{lasso}}_{\ell;N} \right\|_1$$

$$\leq \frac{\| E_{\ell;N} \|_2^2}{T} + \frac{\lambda}{N}\|\phi_\ell\|_1$$

and finally, using the notation v_ℓ,

$$\frac{v'_\ell X'_{\ell;N}X_{\ell;N}v_\ell}{N(2\ell+1)} - 2\frac{v'_\ell X'_{\ell;N}E_{\ell;N}}{N(2\ell+1)} + \frac{\lambda}{N}\left\| \widehat{\phi}^{\text{lasso}}_{\ell;N} \right\|_1 \leq \frac{\lambda}{N}\|\phi_\ell\|_1 .$$

□

Proof of Proposition 4.3.6. First of all, we have

$$\left\| \widehat{\gamma}_{\ell;N} - \widehat{\Gamma}_{\ell;N}\phi_\ell \right\|_\infty = \frac{1}{N(2\ell+1)}\left\| X'_{\ell;N}E_{\ell;N} \right\|_\infty = \max_{1\leq h\leq p}\left| \frac{Y'_{\ell;N}(h)E_{\ell;N}}{N(2\ell+1)} \right| .$$

Now, using (4.3.6), we obtain that, for any $\eta \geq 0$ and $c > 0$,

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

Thus it follows that

$$\mathbb{P}\left(\max_{1\leq h\leq p}\left| \frac{Y'_{\ell;N}(h)E_{\ell;N}}{N(2\ell+1)} \right| > C_{\ell;Z}\left(1 + \frac{1+\mu_{max;\ell}}{\mu_{\min;\ell}}\right)\eta \right)$$

$$\leq 6\,p\,\exp\left(-c\,N(2\ell+1)\,\min\left\{\eta,\eta^2\right\}\right) .$$

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

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

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

$$\begin{aligned}
&\mathbb{P}\left(\bigcap_{\ell=0}^{L_N-1}\left\|\widehat{\gamma}_{\ell;N}-\widehat{\Gamma}_{\ell;N}\phi_\ell\right\|_\infty \leq \mathcal{F}_N\sqrt{\frac{\log pL_N}{N}}\right)\\
&=\mathbb{P}\left(\max_{\ell<L_N}\left\|\widehat{\gamma}_{\ell;N}-\widehat{\Gamma}_{\ell;N}\phi_\ell\right\|_\infty \leq \mathcal{F}_N\sqrt{\frac{\log pL_N}{N}}\right)\\
&=\mathbb{P}\left(\max_{\ell<L_N}\max_{1\leq h\leq p}\left|Y'_{\ell;N}(h)E_{\ell;N}\right| \leq \mathcal{F}_N\sqrt{\frac{\log pL_N}{N}}\right)\\
&\geq 1-6\,p\,L_N\,\exp\left(-c\,N\,\min\left\{c_0\sqrt{\frac{\log pL_N}{N}},c_0^2\,\frac{\log pL_N}{N}\right\}\right)\\
&\geq 1-6\,p\,L_N\,\exp\left(-c\,N\,\min\left\{c_0,c_0^2\right\}\,\frac{\log pL_N}{N}\right)\\
&=1-6\,p\,L_N\,e^{-c\,\min\left\{c_0,c_0^2\right\}\,\log pL_N}\\
&=1-6\,e^{-\left(c\,\min\left\{c_0,c_0^2\right\}-1\right)\log pL_N}\,,
\end{aligned}$$

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 $c_2 > 0$. □

Proof of Proposition 4.3.8. Our goal is to prove (4.3.11), which can be rewritten as follows

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

where $A_\ell = \left\{v'_\ell\widehat{\Gamma}_{\ell;N}v_\ell \geq \alpha_\ell\left\|v_\ell\right\|_2^2-\tau_\ell\left\|v_\ell\right\|_1^2,\ \forall\, v_\ell\in\mathbb{R}^p\right\}$.
We start from Equation (4.3.3), and considering that

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

we have

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

Using Lemma F.2 in the supplementary material of [7] yields

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

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

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

to obtain

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

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

and we apply Lemma 12 of the supplementary material of [40] to get

$$\mathbb{P}\left(\left|v_\ell' \left(\widehat{\Gamma}_{\ell;N} - \Gamma_\ell\right) v_\ell\right| \leq \frac{1}{2} \frac{C_{\ell;Z}}{\mu_{\max;\ell}} \left\{\|v_\ell\|_2^2 + \frac{1}{s}\|v_\ell\|_1^2\right\},\ \forall\, v_\ell \in \mathbb{R}^p\right)$$

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

Moreover, it holds that

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

which implies

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

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

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

$$\mathbb{P}\left(v_\ell'\widehat{\Gamma}_{\ell;N}v_\ell \geq \frac{1}{2}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_2^2 - \frac{1}{2s}\frac{C_{\ell;Z}}{\mu_{\max;\ell}}\|v_\ell\|_1^2,\ \forall\, v_\ell \in \mathbb{R}^p\right)$$

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

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

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

which is the desired conclusion, once we set

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

□

Proof of Theorem 4.3.12. Set $v_\ell = \widehat{\boldsymbol{\phi}}^{\text{lasso}}_{\ell;N} - \boldsymbol{\phi}_\ell$. Using the basic inequality (4.3.7) and the deviation condition (4.3.13), we obtain that, almost surely,

$$\begin{aligned} v_\ell' \widehat{\Gamma}_{\ell;N} v_\ell &\leq 2 v_\ell' \left(\widehat{\gamma}_{\ell;N} - \widehat{\Gamma}_{\ell;N} \boldsymbol{\phi}_\ell \right) + \lambda_N \left(\|\boldsymbol{\phi}_\ell\|_1 - \|\boldsymbol{\phi}_\ell + v_\ell\|_1 \right) \\ &\leq 2 \|v_\ell\|_1 \left\| \widehat{\gamma}_{\ell;N} - \widehat{\Gamma}_{\ell;N} \boldsymbol{\phi}_\ell \right\|_\infty + \lambda_N \left(\|\boldsymbol{\phi}_\ell\|_1 - \|\boldsymbol{\phi}_\ell + v_\ell\|_1 \right) \\ &\leq 2 \|v_\ell\|_1 \mathcal{F}_N \sqrt{\frac{\log(pL_N)}{N}} + \lambda_N \left(\|\boldsymbol{\phi}_\ell\|_1 - \|\boldsymbol{\phi}_\ell + v_\ell\|_1 \right) . \end{aligned} \tag{4.5.1}$$

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

$$\begin{aligned} \|\boldsymbol{\phi}_\ell + v_\ell\|_1 &= \|\boldsymbol{\phi}_{\ell,J} + v_{\ell,J}\|_1 + \|v_{\ell,J^c}\|_1 \\ &\geq \|\boldsymbol{\phi}_{\ell,J}\|_1 - \|v_{\ell,J}\|_1 + \|v_{\ell,J^c}\|_1 , \end{aligned}$$

which implies

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

Having explicitly required that $\lambda_N \geq 4\mathcal{F}_N \sqrt{\log(pL_N)/N}$, Equation (4.5.1) becomes

$$\begin{aligned} 0 \leq v_\ell' \widehat{\Gamma}_{\ell;N} v_\ell &\leq \frac{\lambda_N}{2} \|v_\ell\|_1 + \lambda_N \left(\|v_{\ell,J}\|_1 - \|v_{\ell,J^c}\|_1 \right) \\ &= \frac{\lambda_N}{2} \left(\|v_{\ell,J}\|_1 + \|v_{\ell,J^c}\|_1 \right) + \lambda_N \left(\|v_{\ell,J}\|_1 - \|v_{\ell,J^c}\|_1 \right) \\ &= \frac{3\lambda_N}{2} \|v_{\ell,J}\|_1 - \frac{\lambda_N}{2} \|v_{\ell,J^c}\|_1 \leq \frac{3}{2} \lambda_N \|v_\ell\|_1 . \end{aligned} \tag{4.5.2}$$

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

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

from Cauchy-Schwartz inequality.
Now we use this property into the (RE) inequality (4.3.9), keeping in mind that we specifically required that $q_\ell \tau_\ell \leq \alpha_\ell / 32$, and we obtain

$$\begin{aligned} v_\ell' \widehat{\Gamma}_{\ell;N} v_\ell &\geq \alpha_\ell \|v_\ell\|_2^2 - \tau_\ell \|v_\ell\|_1^2 \geq \alpha_\ell \|v_\ell\|_2^2 - 16 q_\ell \tau_\ell \|v_\ell\|_2^2 \\ &\geq \alpha_\ell \|v_\ell\|_2^2 - \frac{\alpha_\ell}{2} \|v_\ell\|_2^2 \geq \frac{\alpha_\ell}{2} \|v_\ell\|_2^2 . \end{aligned} \tag{4.5.3}$$

Hence, combining Equation (4.5.2) and (4.5.3), 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

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

As a consequence,

$$\|v_\ell\|_2 \leq 12\sqrt{q_\ell}\,\frac{\lambda_N}{\alpha_\ell}\,, \tag{4.5.4}$$

$$\|v_\ell\|_1 \leq 4\sqrt{q_\ell}\,\|v_\ell\|_2 \leq 48\,q_\ell\,\frac{\lambda_N}{\alpha_\ell}\,, \tag{4.5.5}$$

$$v_\ell'\widehat{\Gamma}_{\ell;N}v_\ell \leq \frac{3}{2}\lambda_N\,\|v_\ell\|_1 \leq 72\,\frac{\lambda_N^2}{\alpha_\ell}\,. \tag{4.5.6}$$

Following similar arguments as in the proof of Theorem 3.3.4, previous chapter, it is readily seen that

$$\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_2^2 = \sum_{\ell=0}^{L_N-1}\|v_\ell\|_2^2\,\frac{2\ell+1}{8\pi^2} + \left\|k - k_N\right\|_{L^2}^2,$$

by the orthonormality of Legendre polynomials (see (1.2.2)), and that

$$\left\|\widehat{k}_N^{\text{lasso}} - k\right\|_{L^\infty} \leq \sum_{\ell=0}^{L_N-1}\|v_\ell\|_2\,\frac{2\ell+1}{4\pi} + \left\|k - k_N\right\|_{L^\infty},$$

by the triangle inequality. Then, using 4.5.4, the proof is concluded. □

Appendix A

HEALPix Tutorial

A.1 Installation Procedure

`healpy` is a Python package based on the Hierarchical Equal Area isoLatitude Pixelization (HEALPix) scheme, see [27]. It depends on the HEALPix C++ and cfitsio C libraries, whose source code is included with `healpy` and is built automatically during its installation. Only Linux and MAC OS X are supported, not Windows.
Here we just report the procedure for the binary installation with `conda`, which is also recommended by the package developers.
Conda forge provides a conda channel with a pre-compiled version of `healpy` for Linux 64-bit and MAC OS X platforms; to install it in Anaconda, just open your terminal and write the following lines:

```
conda config --add channels conda-forge
conda install healpy
```

see https://healpy.readthedocs.io/en/latest/install.html for other types of installation.

A.2 Packages

First, import the Python packages necessary to run the commands that will follow.

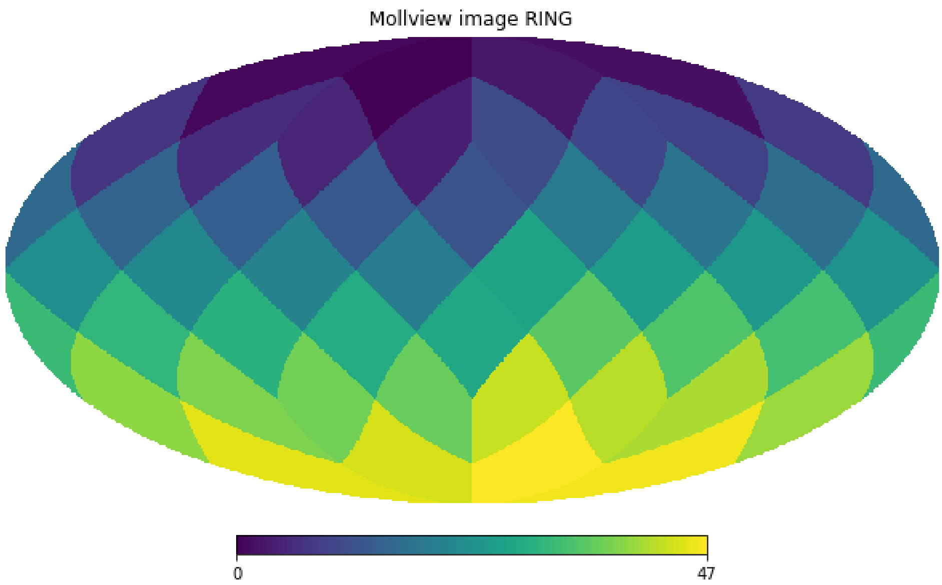
```
import healpy as hp

# import the module for simulating AR(p) processes
import statsmodels.api as sm

import numpy as np
import math as math
import matplotlib.pyplot as plt
from matplotlib import cm

%matplotlib inline
```

Note: The `inline` backend of `maptlotlib` is added to display the plots next to the code. For this reason it is recommended to use an IPython shell or a Jupyter notebook.

A.3 Basics on HEALPix Maps

Maps are simply numpy arrays of N_{pix} elements, where each element refers to a location in the sky as defined by the HEALPix pixelization schemes (see the HEALPix website). The same pixels in the map can be ordered in 2 ways, either RING or NESTED. Here, we focus on RING ordering since in this case spherical harmonics transforms are easy to implement.
In the RING scheme, pixels are numbered moving down from the North to the South Pole along rings of constant latitude:

```
N_pix = 48

m = np.arange(N_pix)
hp.mollview(m, title="Mollview image RING")
```



However, NESTED ordering is displayed by passing the `nest = True` argument to the `healpy.visufunc.mollview` function, as well as the most `healpy` routines.
Recall that the spherical coordinates are the colatitude ϑ, 0 at the North Pole, $\pi/2$ at the equator, and π at the South Pole and the longitude ϕ between 0 and 2π eastward. In a standard Mollweide projection (Galactic coordinates), $\phi = 0$ is at the center and increases eastward toward the left of the map.
Now, we are going to work with real maps to understand their structure. The next line will automatically execute in Terminal the bash script `healpy_get_wmap_maps.sh` (which should be available in the path; if not, install the latest version of `healpy`), and download the higher resolution WMAP data into the current directory.

```
!healpy_get_wmap_maps.sh
```

According to HEALPix convention, data are stored in FITS format, which is the most commonly used digital file format in astronomy.
The function `healpy.fitsfunc.read_map` read a HEALPix map from a FITS file.

```
wmap_map_I = hp.read_map("wmap_band_iqumap_r9_7yr_W_v4.fits")
```

```
NSIDE = 512
ORDERING = NESTED in fits file
INDXSCHM = IMPLICIT
Ordering converted to RING
```

By default, `read_map` loads the first column and convert the input map to RING ordering, even if it was stored as NESTED. Its resolution, and hence the number of pixels it contains, is defined through the N_{side} parameter, which is generally a power of 2:

```
N_side = hp.get_nside(wmap_map_I)
print(N_side)
```

```
512
```

For a given N_{side}, a HEALPix map contains $12N_{\text{side}}^2$ pixels placed on $4N_{\text{side}} - 1$ iso-latitude rings.

```
N_pix = hp.nside2npix(N_side)
print(N_pix)
```

```
3145728
```

A.4 Simulations

Now, we are ready to simulate a map. Thus, we first need a power spectrum $\{C_\ell\}$. Note that we create a numpy array by setting

$$C_\ell = \frac{1}{\ell(\ell+1)}, \qquad 0 < \ell \leq 3000;$$

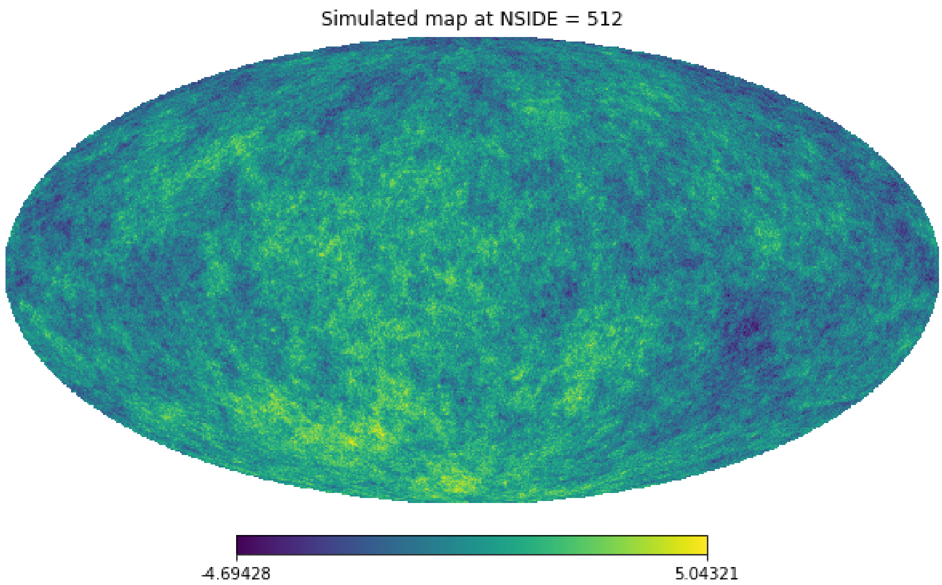
of course this model would not be acceptable for the full range of multipoles, because the resulting field would not have finite variance. However, this is not an issue here because we are considering only a finite range of values (otherwise, we can implicitly assume that, for greater ℓ's, the sequence decreases at a faster rate).

```
L = 3000

cl = [1]
cl[1:L] = [1 / (l * (l + 1)) for l in range(1, L + 1)]

cl = np.asarray(cl)

# Writes the array cl into a healpix file named cl.fits
# hp.write_cl("cl.fits", cl, overwrite = True)
```

The desired output is obtained by first simulating the spherical random coefficients $\{a_{\ell,m}\}$ up to the maximum multipole $L_{\max} = 1000$ (`healpy.sphtfunc.synalm`), and then by creating a map at $N_{\text{side}} = 512$ (`healpy.sphtfunc.alm2map`). The keyword `new = True` returns a numpy array of $\{a_{\ell,m}\}$ sorted according the new HEALPix order, that is, $(a_{0,0}, a_{1,0}, a_{2,0}, \dots, a_{1,1}, a_{2,1}, \dots a_{L_{\max},L_{\max}})$.

```
L_max = 1000
N_side = 512
```

```
alm = hp.synalm(cl, lmax = L_max, new = True)

field = hp.alm2map(alm, nside = N_side)
hp.mollview(field, title = "Simulated map at NSIDE = %i" %N_side)
```




```
# This command is equivalent to the previous one.
# field, alm = hp.synfast(cl, nside = N_side, lmax = L_max, alm =
    True, new = True)
```

`write_map` writes the output map to disk in FITS format:

```
hp.write_map("my_map.fits", field, overwrite=True)
```

A.5 Direct and Inverse Spherical Harmonic Transforms

As anticipated, `healpy` provides bindings to the C++ HEALPIX library for performing spherical harmonic transforms. It is then possible to extract the spherical harmonic coefficients $\{a_{\ell,m}\}$ from a map, and vice versa, using the routines `healpy.sphtfunc.alm2map` and `healpy.sphtfunc.map2alm`.
In practice, given the locations $x_0, \ldots, x_{N_{\text{pix}}-1}$ where the signal is observed, it is possible to compute

$$\widehat{a}_{\ell,m} = \frac{4\pi}{N_{\text{pix}}} \sum_{p=0}^{N_{\text{pix}}-1} T(x_p)\overline{Y_{\ell,m}}(x_p) \qquad \text{and} \qquad \widehat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |\widehat{a}_{\ell,m}|^2.$$

```
alm_hat = hp.map2alm(field, lmax = L_max)
cl_hat = hp.alm2cl(alm_hat)
```

As a rule, you can use $L_{\max} \approx 2N_{\text{side}}$, that is, the maximum multipole for which the $a_{\ell,m}$'s are computed with high precision – e.g., $L_{\max} \approx 1000$ for $N_{\text{side}} = 512$. For higher multipoles, the precision is gradually lost until $L_{\max} = 3N_{\text{side}}$. For more information see the HEALPix primer.

A.6 Simulating Spherical Functional Autoregressions

Here, we provide the code to simulate from a spherical functional autoregressive process of order p (see [18]). The function `sphar_p` collects the simulated harmonic coefficients $\{a_{\ell,m}(t)\}$ in a list of length `L`; the `l`-th element of the list a numpy array of shape `(n_series, l+1)`.

```
def sphar_p(L, n_series, phi, Cz, p):
    data = [0]*L

    for l in range(0, L):

        arp = np.insert(-phi[l], 0, 1)
        ma1 = np.array([1])
        # Both the AR and MA components
        # should include the coefficient on the zero-lag.
        # The AR parameters should have
        # the opposite sign of what you might expect.

        al0 = sm.tsa.arma_generate_sample(arp, ma1,
                            nsample = n_series,
                            sigma = math.sqrt(Cz[l]))
        data[l] = al0 + 1j*0

        if l!=0:
            for m in range(0,l):

                Re_alm = sm.tsa.arma_generate_sample(arp, ma1,
                                    nsample = n_series,
                                    sigma = math.sqrt(Cz[l]/2))
                Im_alm = sm.tsa.arma_generate_sample(arp, ma1,
                                    nsample = n_series,
                                    sigma = math.sqrt(Cz[l]/2))
                data[l] = np.column_stack((data[l],
                                    Re_alm + 1j* Im_alm))

    return data
```

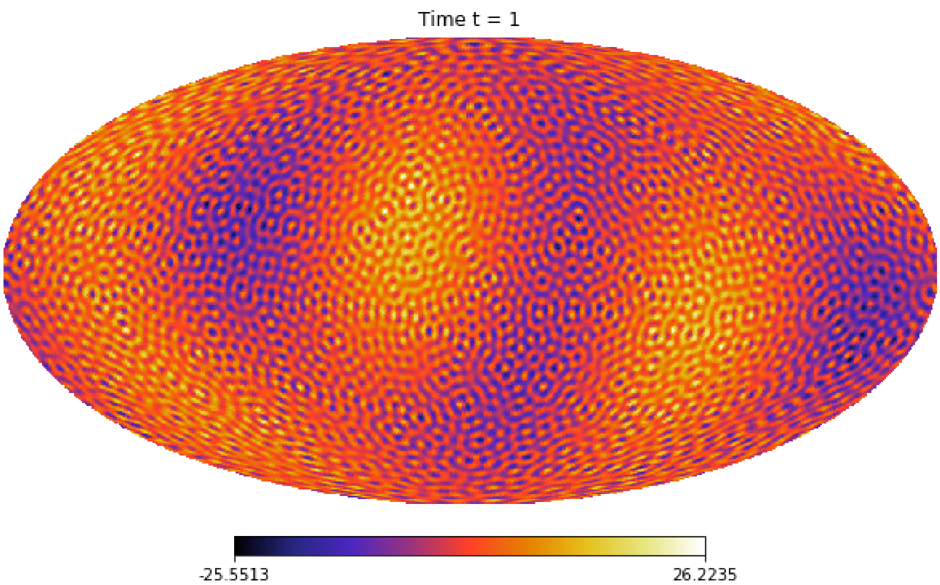
The example below shows simulation from a SPHAR(1) process for a particular choice of the sequence of parameters $\{\phi_\ell\}$. The variances of the non-null components (multipoles $\ell = 3, 90$) are set to be approximately equals.

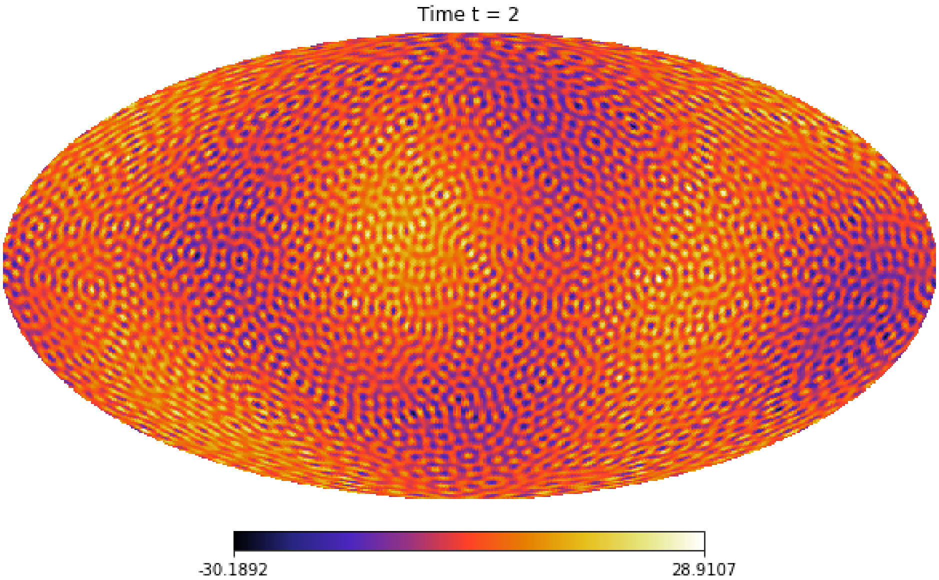
```
L = 100
n_series = 50

phi = [0]*L
phi[3] = 0.95
phi[90] = -0.95

Cz = [0]*L
Cz[3] = 25
Cz[90] = 1
```

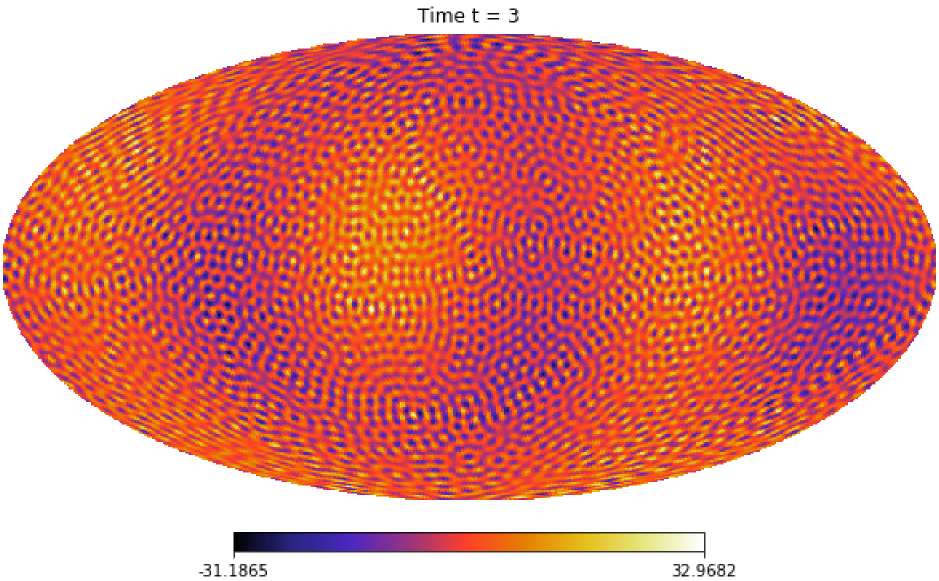
```
data = sphar_p(L, n_series, phi, Cz, p = 1)
```

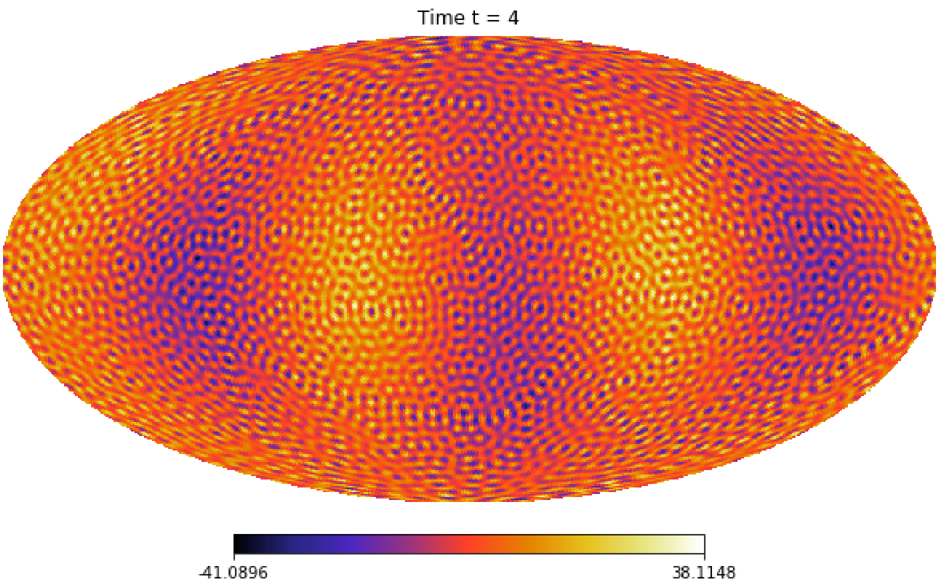
Once obtained the output, we can plot the spherical maps $T(\cdot, t)$ for given time points, say $t \in \{t_{\min}, \ldots, t_{\max} - 1\}$. The harmonic coefficients are sorted according to new HEALPix order.

```
# Colour palettes
cool_cmap = cm.CMRmap
cool_cmap.set_under("w") # sets background to white

t_min = 1
t_max = 5

for t in range(t_min, t_max):
    At = [item[(t,)] for item in data]

    n_iter = int(L*(L+1)/2)
    At_sort = np.zeros(n_iter,dtype=np.complex_)

    k=0
    for j in range(0, L):
        for i in range(0, L - j):
            if j==0 and i==0:
                At_sort[k] = At[0]

            else:
                k=k+1
                At_sort[k] = At[i+j][j]

    field = hp.alm2map(At_sort, lmax = L-1, nside = 128)
    hp.mollview(field, cmap = cool_cmap, title = 'Time t = %i' %t)
```




Time t = 2
-30.1892
28.9107


Time t = 3
-31.1865
32.9682

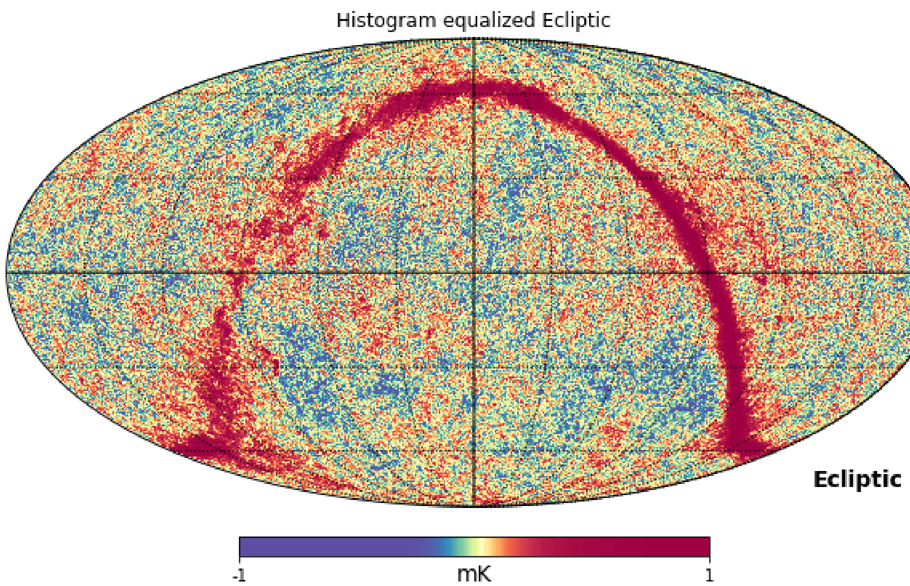

Time t = 4
-41.0896
38.1148

If we focus our attention to a single big blue spot, what immediately can be noticed is that it is preserved over time. Contrary, the small regions change rapidly their color. This is because we simulate the fields setting a coefficient $\phi_\ell = 0.95$ (strongly positive correlation) for a low frequency $\ell = 3$ (large scale) and $\phi_\ell = -0.95$ (strongly negative correlation) for a higher frequency $\ell = 90$ (small scale).

A.7 Visualization

The Mollweide projection introduced above with `mollview` is the most common visualization tool for HEALPix maps. The function contains many arguments which allow to customize the visualization through graphical parameter. For instance,

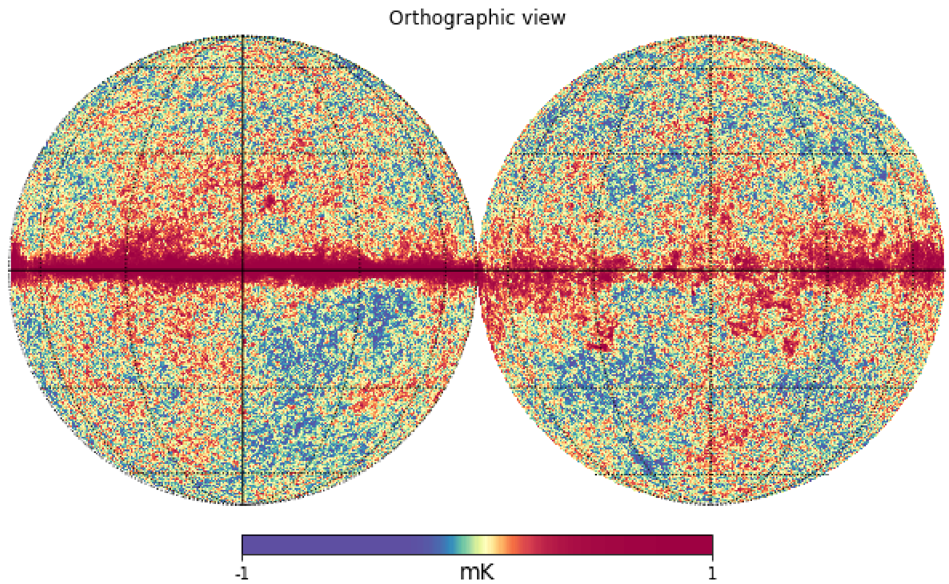
```
cool_cmap = cm.Spectral_r
cool_cmap.set_under("w") # sets background to white

# print(cm.cmap_d.keys())

hp.mollview(
    wmap_map_I,
    coord = ["G", "E"],
    title = "Histogram equalized Ecliptic",
    unit = "mK",
    norm = "hist",
    min = -1,
    max = 1,
    cmap = cool_cmap)
hp.graticule()
```



`coord = ["G", "E"]` rotate the map from the first (Galactic) to the second (Ecliptic) coordinate system; `norm = 'hist'` sets a histogram equalized color scale from `min = -1` to `max = 1`; `graticule` adds meridians and parallels. `print(cm.cmap_d.keys())` shows all the possible colour palettes.

However, there exists other types of visualization functions; an example is given by `healpy.visufunc.orthview` (see also the healpy documentation):

```
hp.orthview(
    wmap_map_I,
    unit = "mK",
    norm = "hist",
    min = -1,
    max = 1,
    cmap = cool_cmap)
hp.graticule()
```



References

[1] Abramowitz, M. and Stegun, I. A. *Handbook of Mathematical Functions, with Formulas Graphs and Mathematical Tables.* Dover (1964).

[2] Atkinson, K. and Han, W. *Spherical Harmonics and Approximations on the Unit Sphere: an Introduction.* Springer (2012).

[3] Aue, A. and Van Delft, A. Testing for stationarity of functional time series in the frequency domain. *The Annals of Statistics*, **in press** (2019+). arXiv preprint arXiv:1701.01741.

[4] Baldi, P., Kerkyacharian, G., Marinucci, D., and Picard, D. Asymptotics for spherical needlets. *The Annals of Statistics*, **37** (2009), 1150.

[5] Baldi, P., Kerkyacharian, G., Marinucci, D., and Picard, D. Subsampling needlet coefficients on the sphere. *Bernoulli*, **15** (2009), 438.

[6] Baldi, P. and Trapani, S. Fourier coefficients of invariant random fields on homogeneous spaces of compact groups. *Annales de l'Institut Henri Poincaré Poincaré Probabilités et Statistiques*, **51** (2015), 648.

[7] Basu, S. and Michailidis, G. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, **43** (2015), 1535.

[8] Berg, C. and Porcu, E. From Schoenberg coefficients to Schoenberg functions. *Constructive Approximation*, **45** (2017), 217.

[9] Billingsley, P. *Convergence of Probability Measures.* John Wiley & Sons, second edn. (1999).

[10] Bosq, D. *Linear Processes in Function Spaces. Theory and Applications.* Springer-Verlag (2000).

[11] Brezis, H. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Springer-Verlag (2011).

[12] Brockwell, P. J. and Davis, R. A. *Time Series: Theory and Methods.* Springer Series in Statistics. Springer-Verlag, 2nd edn. (1991).

[13] Bühlmann, P. and van de Geer, S. *Statistics for High-Dimensional Data.* Springer (2011).

[14] Cammarota, V. and Marinucci, D. The stochastic properties of ℓ^1-regularized spherical Gaussian fields. *Applied and Computational Harmonic Analysis*, **38** (2015), 262.

[15] Cammarota, V. and Marinucci, D. A quantitative central limit theorem for the Euler-Poincaré characteristic of random spherical eigenfunctions. *The Annals of Probability*, **46** (2018), 3188.

[16] Cammarota, V. and Marinucci, D. On the correlation of critical points and angular trispectrum for random spherical harmonics. *arXiv preprint arXiv:1907.05810*, (2019).

[17] Caponera, A., Durastanti, C., and Vidotto, A. Lasso estimation for spherical autoregressive processes. *arXiv preprint arXiv:1911.11470*, (2019). Submitted for publication.

[18] Caponera, A. and Marinucci, D. Asymptotics for spherical functional autoregressions. *The Annals of Statistics*, **in press** (2020+). arXiv preprint arXiv:1907.05802.

[19] Cheng, D., Cammarota, V., Fantaye, Y., Marinucci, D., and Schwartzman, A. Multiple testing of local maxima for detection of peaks on the (celestial) sphere. *Bernoulli*, **26** (2019), 31.

[20] Clarke De la Cerda, J., Alegría, A., and Porcu, E. Regularity properties and simulations of Gaussian random fields on the sphere cross time. *Electronic Journal of Statistics*, **12** (2018), 399.

[21] Davis, R. A., Zang, P., and Zheng, T. Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, **25** (2016), 1077.

[22] Durastanti, C., Lan, X., and Marinucci, D. Needlet-Whittle estimates on the unit sphere. *Electronic Journal of Statistics*, **7** (2013), 597.

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

[24] Fan, M., Paul, D., Lee, T., and Matsuo, T. A multi-resolution model for non-Gaussian random fields on a sphere with application to ionospheric electrostatic potentials. *The Annals of Applied Statistics*, **1** (2018).

[25] Fan, M., Paul, D., Lee, T. C., and Matsuo, T. Modeling tangential vector fields on a sphere. *Journal of the American Statistical Association*, **113** (2018), 1625.

[26] Gneiting, T. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19** (2013), 1327.

[27] Gorski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelman, M. HEALPix – a framework for high resolution discretization, and fast analysis of data distributed on the sphere. *Astrophysical Journal*, **622** (2005), 759.

[28] Gradshteyn, I. S. and Ryzhik, I. M. *Table of Integrals, Series, and Products*. Elsevier Academic Press, 8th edn. (2014).

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

[30] Hastie, T., Tibshirani, R., and Wainwright, M. *Statistical Learning with Sparsity: the Lasso and Generalizations.* CRC Press (2015).

[31] Hörmann, S., Kokoszka, P., and Nisol, G. Testing for periodicity in functional time series. *The Annals of Statistics*, **46** (2018), 2960.

[32] Hsing, T. and Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* John Wiley & Sons (2015).

[33] Jun, M. Matérn-based nonstationary cross-covariance models for global processes. *Journal of Multivariate Analysis*, **128** (2014), 134.

[34] Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Zhu, Y., et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77** (1996), 437.

[35] Kalousis, A., Prados, J., and Hilario, M. Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowledge and Information Systems*, **12** (2007), 95.

[36] Kim, P. T. and Koo, J.-Y. Optimal spherical deconvolution. *Journal of Multivariate Analysis*, **80** (2002), 21.

[37] Lang, A. and Schwab, C. Isotropic Gaussian random fields on the sphere: Regularity, fast simulation and stochastic partial differential equations. *The Annals of Applied Probability*, **25** (2015), 3047.

[38] Le Gia, Q., Sloan, I., Womersley, R., and Wang, Y. G. Sparse isotropic regularization for spherical harmonic representations of random fields on the sphere. *Applied and Computational Harmonic Analysis*, **in press** (2019+). arXiv preprint arXiv:1801.03212.

[39] Leonenko, N. N., Taqqu, M. S., and Terdik, G. H. Estimation of the covariance function of Gaussian isotropic random fields on spheres, related Rosenblatt-type distributions and the cosmic variance problem. *Electronic Journal of Statistics*, **12** (2018), 3114.

[40] Loh, P.-L. and Wainwright, M. J. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, **40** (2012), 1637.

[41] Lütkepohl, H. *New Introduction to Multiple Time Series Analysis.* Springer (2005).

[42] Marinucci, D. Lecture notes on spherical random fields. Finnish School in Probability and Statistics (2015).

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

[44] Marinucci, D. and Peccati, G. Mean-square continuity on homogeneous spaces of compact groups. *Electronic Communications in Probability*, **18** (2013).

[45] Marinucci, D., Peccati, G., Rossi, M., and Wigman, I. Non-universality of nodal length distribution for arithmetic random waves. *Geometry and Functional Analysis*, **3** (2016), 926.

[46] Marinucci, D., Rossi, M., and Wigman, I. The asymptotic equivalence of the sample trispectrum and the nodal length for random spherical harmonics. *Annales de l'Institut Henri Poincaré Poincaré Probabilités et Statistiques*, **56** (2020), 374.

[47] Matsumoto, S. General moments of the inverse real Wishart distribution and orthogonal Weingarten functions. *Journal of Theoretical Probability*, **25** (2012), 798.

[48] Nourdin, I. and Peccati, G. Stein's method on Wiener chaos. *Probability Theory and Related Fields*, **145** (2009), 75.

[49] Nourdin, I. and Peccati, G. *Normal Approximations with Malliavin Calculus: from Stein's Method to Universality.* Cambridge University Press (2012).

[50] Nourdin, I., Peccati, G., and Reinert, G. Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos. *The Annals of Probability*, **38** (2010), 1947.

[51] Nourdin, I., Peccati, G., and Reveillac, A. Multivariate normal approximation using Stein's method and Malliavin calculus. *Annales de l'Institut Henri Poincaré Poincaré Probabilités et Statistiques*, **46** (2010), 45.

[52] Nourdin, I., Peccati, G., and Rossi, M. Nodal statistics of planar random waves. *Communications in Mathematical Physics*, **369** (2019), 99.

[53] Nualart, D. and Ortiz-Latorre, S. Central limit theorems for multiple stochastic integrals and Malliavin calculus. *Stochastic Processes and their Applications*, **118** (2008), 614.

[54] Nualart, D. and Peccati, G. Central limit theorems for sequences of multiple stochastic integrals. *The Annals of Probability*, **33** (2005), 177.

[55] Panaretos, V. M. and Tavakoli, S. Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, **123** (2013), 2779.

[56] Panaretos, V. M. and Tavakoli, S. Fourier analysis of stationary time series in function space. *The Annals of Statistics*, **41** (2013), 568.

[57] Peccati, G. and Taqqu, M. S. *Wiener Chaos: Moments, Cumulants and Diagrams.* Springer-Verlag (2010).

[58] Peccati, G. and Tudor, C. A. Gaussian limits for vector-valued multiple stochastic integrals. In *Séminaire de Probabilités XXXVIII*, pp. 247–262. Springer (2005).

[59] Peccati, G. and Vidotto, A. Gaussian random measures generated by Berry's nodal sets. *Journal of Statistical Physics*, **178** (2020), 996.

[60] Planck Collaboration. Planck 2015 results-I. Overview of products and scientific results. *Astronomy & Astrophysics*, **594** (2016), A1.

[61] Porcu, E., Alegría, A., and Furrer, R. Modeling temporally evolving and spatially globally dependent data. *International Statistical Review*, **86** (2018), 344.

[62] Porcu, E., Bevilacqua, M., and Genton, M. Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, **111** (2016), 888.

[63] Robinson, P. M. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23** (1995), 1048.

[64] Schoenberg, I. J. Positive definite functions on spheres. *Duke Mathematical Journal*, **9** (1942), 96.

[65] Sodin, M. and Tsirelson, B. Random complex zeroes, I. Asymptotic normality. *Israel Journal of Mathematics*, **144** (2004), 125.

[66] Song, S. and Bickel, P. J. Large vector auto regressions. *arXiv preprint arXiv:1106.3915*, (2011).

[67] Szego, G. *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications, 4th edn. (1975).

[68] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, **58** (1996), 267.

[69] Todino, A. P. *Local Geometry of Random Spherical Harmonics*. Ph.D. thesis, Gran Sasso Science Institute (2018).

[70] Todino, A. P. Nodal lengths in shrinking domains for random eigenfunctions on $\mathbb{S}^2$. *arXiv preprint arXiv:1807.11787*, (2018).

[71] Todino, A. P. A quantitative central limit theorem for the excursion area of random spherical harmonics over subdomains of $\mathbb{S}^2$. *Journal of Mathematical Physics*, **60** (2019), 023505.

[72] Tsybakov, A. *Introduction to Nonparametric Estimation*. Springer (2009).

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

[74] Wigman, I. On the distribution of the nodal sets of random spherical harmonics. *Journal of Mathematical Physics*, **50** (2009), 01352.

[75] Wigman, I. Fluctuations of the nodal length of random spherical harmonics. *Communications in Mathematical Physics*, **298** (2010), 787.

[76] Xiao, H. and Wu, W. B. Covariance matrix estimation for stationary time series. *The Annals of Statistics*, **40** (2012), 466.

[77] Yu, B. Stability. *Bernoulli*, **19** (2013), 1484.