Normal approximations for wavelet coefficients on spherical Poisson fields

Claudio Durastanti\textsuperscript{a}, Domenico Marinucci\textsuperscript{a,∗}, Giovanni Peccati\textsuperscript{b}

\textsuperscript{a} Department of Mathematics, University of Rome Tor Vergata, Italy
\textsuperscript{b} Unité de Recherche en Mathématiques, Luxembourg University, Luxembourg

\textbf{A B S T R A C T}

We compute explicit upper bounds on the distance between the law of a multivariate Gaussian distribution and the joint law of wavelet/needlet coefficients based on a homogeneous spherical Poisson field. In particular, we develop some results from Peccati and Zheng (2010) \cite{42}, based on Malliavin calculus and Stein’s methods, to assess the rate of convergence to Gaussianity for a triangular array of needlet coefficients with growing dimensions. Our results are motivated by astrophysical and cosmological applications, in particular related to the search for point sources in Cosmic Rays data.

© 2013 Elsevier Inc. All rights reserved.

\section{1. Introduction}

The aim of this paper is to establish multidimensional normal approximation results for vectors of random variables having the form of wavelet coefficients integrated with respect to a Poisson measure on the unit sphere. The specificity of our analysis is that we require the dimension of such vectors to grow to infinity. Our techniques are based on recently obtained bounds for the normal approximation of functionals of general Poisson measures (see \cite{40,42}), as well as on the use of the localization properties of wavelet systems on the sphere (see \cite{36}, as well as the recent monograph \cite{30}). A large part of the paper is devoted to the explicit determination of the above quoted bounds in terms of dimension.

\subsection{1.1. Motivation and overview}

A classical problem in asymptotic statistics is the assessment of the speed of convergence to Gaussianity (that is, the computation of explicit Berry–Esseen bounds) for parametric and nonparametric estimation procedures—for recent references connected to the main topic of the present paper, see for instance \cite{16,29,54}. In this area, an important novel development is given by the derivation of effective Berry–Esseen bounds by means of the combination of two probabilistic techniques, namely the Malliavin calculus of variations and the Stein’s method for probabilistic approximations. The monograph \cite{6} is the standard modern reference for Stein’s method, whereas \cite{38} provides an exhaustive discussion of the use of Malliavin calculus for proving normal approximation results on a Gaussian space. The fact that one can use Malliavin calculus to deduce normal approximation bounds (in total variation) for functionals of Gaussian fields was first exploited in \cite{37}—where one can find several quantitative versions of the “fourth moment theorem” for chaotic random variables proved in \cite{39}. Lower bounds can also be computed, entailing that the rates of convergence provided by these techniques are sharp in many instances—see again \cite{38}.

∗ Corresponding author.
E-mail address: marinucc@mat.uniroma2.it (D. Marinucci).

0022-247X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmaa.2013.06.028
In a recent series of contributions, the interaction between Stein’s method and Malliavin calculus has been further exploited for dealing with the normal approximation of functionals of a general Poisson random measure. The most general abstract results appear in [40] (for one-dimensional normal approximations) and [42] (for normal approximations in arbitrary dimensions). These findings have recently found a wide range of applications in the field of stochastic geometry—see [25,26,34,28,47] for a sample of geometric applications, as well as the webpage

http://www.iecn.u-nancy.fr/~nourdin/steinmalliavin.htm

for a constantly updated resource on the subject.

The purpose of this paper is to apply and extend the main findings of [40,42] in order to study the multidimensional normal approximation of the elements of the first Wiener chaos of a given Poisson measure. Our main goal is to deduce bounds that are well-adapted to deal with applications where the dimension of a given statistic increases with the number of observations. This is a framework which arises naturally in many relevant fields of modern statistical analysis; in particular, our principal motivation originates from the implementation of observations. This is a framework which arises naturally in many relevant fields of modern statistical analysis; in particular, our principal motivation originates from the implementation of wavelet systems on the sphere. In these circumstances, when more and more data become available, a higher number of wavelet coefficients is evaluated, as it is customarily the case when considering, for instance, thresholding nonparametric estimators. We shall hence be concerned with sequences of Poisson fields, whose intensity grows monotonically. We then exploit the wavelet localization properties to establish bounds that grow linearly with the number of functionals considered; we are then able to provide explicit recipes, for instance, for the number of joint testing procedures that can be simultaneously entertained ensuring that the Gaussian approximation may still be shown to hold, in a suitable sense.

1.2. Main contributions

Consider a sequence of i.i.d. random variables \( \{X_i : i \geq 1\} \) with values in the unit sphere \( S^2 \), and define \( \{\psi_{jk}\} \) to be the collection of the spherical needlets associated with a certain constant \( B > 1 \), see Section 3.1 for more details and discussion. Write also \( \sigma_{jk}^2 = E[|\psi_{jk}(X_i)|^2] \) and \( b_{jk} = E[\psi_{jk}(X_i)] \), and consider an independent (possibly inhomogeneous) Poisson process \( \{N_t : t \geq 0\} \) on the real line such that \( E[N_t] = R(t) \to \infty \), as \( t \to \infty \). Formally, our principal aim is to establish conditions on the sequences \( \{j(n) : n \geq 1\} \), \( \{R(n) : n \geq 1\} \) and \( \{d(n) : n \geq 1\} \) ensuring that the distribution of the centered \( d(n) \)-dimensional vector

\[
Y_n = (Y_{n,1}, \ldots, Y_{n,d(n)})
\]

is asymptotically close, in the sense of some smooth distance denoted \( d_2 \) (see Definition 2.6), to the law of a \( d(n) \)-dimensional Gaussian vector, say \( Z_n \), with centered and independent components having unit variance. The use of a smooth distance allows one to deduce minimal conditions for this kind of asymptotic Gaussianity. The crucial point is that we allow the dimension \( d(n) \) to grow to infinity, so that our results require to explicitly assess the dependence of each bound on the dimension. We shall perform our tasks through the following main steps: (i) Proposition 4.1 deals with one-dimensional normal approximations, (ii) Proposition 5.4 deals with normal approximations in a fixed dimension, and finally (iii) in Theorem 5.5 we deduce a bound that is well-adapted to the case \( d(n) \to \infty \). More precisely, Theorem 5.5 contains an upper bound linear in \( d(n) \), that is, an estimate of the type

\[
d_2(Y_n, Z_n) \leq C(n) \times d(n).
\]

It will be shown in Corollary 5.6, that the sequence \( C(n) \) can be chosen to be

\[
O\left(1/\sqrt{R(n)}B^{-2j(n)}\right);
\]

as discussed below in Remark 4.3, \( R(n) \times B^{-2j(n)} \) can be viewed as a measure of the “effective sample size” for the components of \( Y_n \).

1.3. About de-Poissonization

Our results can be used in order to deduce the asymptotic normality of de-Poissonized linear statistics with growing dimension. To illustrate this point, assume that the random variables \( X_i \) are uniformly distributed on the sphere. Then, it is well known that \( b_{jk} = 0 \), whenever \( j > 1 \). In this framework, when \( j(n) > 1 \) for every \( n, R(n) = n \) and \( d(n)/n^{1/4} \to 0 \), the conditions implying that \( Y_n \) is asymptotically close to Gaussian, automatically ensure that the law of the de-Poissonized vector

\[
Y'_n = (Y'_{n,1}, \ldots, Y'_{n,d(n)}) = \frac{1}{\sqrt{n}} \left( \sum_{k=1}^{n} \frac{\psi_{j(n)k_1}(X_i)}{\sigma_{j(n)k_1}}, \ldots, \sum_{k=1}^{n} \frac{\psi_{j(n)k_{d(n)}}(X_i)}{\sigma_{j(n)k_{d(n)}}} \right)
\]

is also asymptotically close to Gaussian. The reason for this phenomenon is nested in the statement of the forthcoming (elementary) Lemma 1.1 (see also [9] for similar computations).
Lemma 1.1. Assume that $R(n) = n$, that the $X_i$'s are uniformly distributed on the sphere, and that $j(n) > 1$ for every $n$. Then, there exists a universal constant $M$ such that, for every $n$ and every Lipschitz function $\varphi : \mathbb{R}^{d(n)} \to \mathbb{R}$, the following estimate holds:

$$|E[\varphi(Y'_n)] - E[\varphi(Y_n)]| \leq M \|\varphi\|_{\text{Lip}} \frac{d(n)}{n^{1/4}}.$$ 

Proof. Fix $l = 1, \ldots, d(n)$, and write $\beta_l(x) = \frac{\varphi_{i_l}(x)}{\sigma_{i_l}}$, in such a way that $E[\beta_l(X_1)^2] = 1$. One has that

$$E[(Y'_{nl} - Y_{nl})^2] = 2(1 - \alpha_n),$$

where

$$\alpha_n = \frac{1}{n} \sum_{m=0}^{n} \frac{e^{-n}n^m}{m!} (n \land m) = 1 - \frac{e^{-n}n^n}{n^n}.$$ 

This gives the estimate

$$E[|Y'_{nl} - Y_{nl}|] \leq \sqrt{E[(Y'_{nl} - Y_{nl})^2]} \leq \sqrt{2} \frac{e^{-n}n^n}{n^n},$$

so that the conclusion follows from an application of Stirling’s formula and of the Lipschitz property of $\varphi$. \qed

Remark 1.2. (i) Lemma 1.1 implies that one can obtain an inequality similar to (1.2) for $Y'_n$, that is:

$$d_2(Y'_n, Z_n) \leq \left( C(n) + \frac{M}{n^{1/4}} \right) \times d(n).$$

(ii) With some extra work, one can obtain estimates similar to those in Lemma 1.1 also when the constants $b_{j(n)k}$ are possibly different from zero. This point, that requires some lengthy technical considerations, falls slightly outside the scope of this paper and will be pursued in full generality elsewhere.

(iii) In [5], Bentkus proved the following (yet unsurpassed) bound. Assume that $\{X_i : i \geq 1\}$ is a collection of i.i.d. $d$-dimensional vectors, such that $X_1$ is centered and with covariance equal to the identity matrix. Set $S_n = n^{-1/2}(X_1 + \cdots + X_n)$, $n \geq 1$ and let $Z$ be a $d$-dimensional centered Gaussian vector with i.i.d. components having unit variance. Then, for every convex set $C \subset \mathbb{R}^d$

$$|E[1_C(S_n)] - E[1_C(Z)]| \leq d^{1/4} \frac{400\beta}{\sqrt{n}},$$

where $\beta = E[\|X_1\|_{\text{Lip}}^3]$. It is unclear whether one can effectively use this bound in order to investigate the asymptotic Gaussianity of sequences of random vectors of the type (1.1)–(1.3), in particular because, for a fixed $n$, the components of $Y_n$, $Y'_n$ have in general a non trivial correlation. Note also that a simple application of Jensen inequality shows that $\beta d^{1/4} n^{-1/2} \geq d^{1/4}/n^{1/2}$. However, a direct comparison of Bentkus’ estimates with our “linear” rate in $d$ (see (1.2), as well as Theorem 5.5 below) is unfeasible, due to the differences with our setting, namely concerning the choice of distance, the structure of the considered covariance matrices, the Poissonized environment, and the role of $b^{(n)}$ discussed in Remark 4.3.

(iv) A careful inspection of the proofs of our main results reveals that the findings of this paper have a much more general validity, and in particular can be extended to kernel estimators on compact spaces satisfying mild concentration and equispacing properties (see also [19,20]). In this paper, however, we decided to stick to the presentation on the sphere for definiteness, and to make the connection with applications clearer. Some more general frameworks are discussed briefly at the end of Section 5.

(v) For notational simplicity, throughout this paper we will stick to the case where all the components in our vector statistics are evaluated at the same scale $j(n)$ (see below for more precise definitions and detailed discussion). The relaxation of this assumption to cover multiple scales $(j_1(n), \ldots, j_a(n))$ does not require any new ideas and is not considered here for brevity’s sake.

1.4. Plan

The plan of the paper is as follows: in Section 2 we provide some background material on Stein–Malliavin bounds in the case of Poisson random fields, and we describe a suitable setting for the current paper, entailing sequences of fields with monotonically increasing governing measures. We provide also some new results, ensuring that the Central Limit Theorems we are going to establish are stable, in the classical sense. In Section 3 we recall some background material on the construction of tight wavelet systems on the sphere (see [36,35] for the original references, as well as [30, Chapter 10]) and we explain how to express the corresponding wavelet coefficients in terms of stochastic integrals with respect to a Poisson random measure. We also illustrate shortly some possible statistical applications. In Section 4 we provide our bounds in the one-dimensional case; these are simple results which could have been established by many alternative techniques, but still
they provide some interesting insights into the "effective area of influence" of a single component of the wavelet system. The core of the paper is in Section 5, where the bound is provided in the multidimensional case, allowing in particular for the number of coefficients to be evaluated to grow with the number of observations. This result requires a careful evaluation of the upper bound, which is made possible by the localization properties in real space of the wavelet construction.

2. Poisson random measures and Stein–Malliavin bounds

In order to study the asymptotic behavior of linear functionals of Poisson measures on the sphere $S^2$, we start by recalling the definition of a Poisson random measure—more details, see for instance \([41,21,46,49]\). We work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

**Definition 2.1.** Let $(\Theta, \mathcal{A}, \mu)$ be a $\sigma$-finite measure space such that $\Theta$ is a Polish space and $\mathcal{A}$ is its associated Borel $\sigma$-field. Assume that $\mu$ has no atoms (that is, $\mu(\{x\}) = 0$, for every $x \in \Theta$). A collection of random variables $\{N(A) : A \in \mathcal{A}\}$, taking values in $\mathbb{Z}_+ \cup \{+\infty\}$, is called a Poisson random measure (PRM) on $\Theta$ with intensity measure (or control measure) $\mu$ if the following two properties hold:

1. For every $A \in \mathcal{A}$, $N(A)$ has Poisson distribution with mean $\mu(A)$; 
2. If $A_1, \ldots, A_n \in \mathcal{A}$ are pairwise disjoint, then $N \left( \bigcup_{i=1}^n A_i \right)$ is independent of $N \left( \bigcap_{i=1}^n A_i \right)$, and $\sum_{i=1}^n N(A_i)$ is Poisson distributed with mean $\mu(A_1) + \cdots + \mu(A_n)$.

**Remark 2.2.** (i) In Definition 2.1, a Poisson random variable with parameter $\lambda = \infty$ is implicitly set to be equal to $\infty$.
(ii) Points 1 and 2 in Definition 2.1 imply that, for every $\omega \in \Omega$, the mapping $A \mapsto N(A, \omega)$ is a measure on $\Theta$. Moreover, since $\mu$ is non-atomic, one has that

$$P\left[ N(\{x\}) = 0 \text{ or } 1, \forall x \in \Theta \right] = 1.$$  

(2.4)

**Assumption 2.3.** Our framework for the rest of the paper will be the following special case of Definition 2.1:

(a) We take $\Theta = \mathbb{R}_+ \times S^2$, with $\mathcal{A} = \mathcal{B}(\Theta)$, the class of Borel subsets of $\Theta$.
(b) The symbol $N$ indicates a Poisson random measure on $\Theta$, with homogeneous intensity given by $\mu = \rho \times \nu$, where $\rho$ is some measure on $\mathbb{R}_+$ and $\nu$ is a probability on $S^2$ of the form $\nu(dx) = f(x)dx$, where $f$ is a density on the sphere. We shall assume that $\rho([0]) = 0$ and that the mapping $\rho \mapsto \rho([0, t])$ is strictly increasing and diverging to infinity as $t \to \infty$.

We also adopt the notation

$$R_t := \rho([0, t]), \quad t \geq 0,$$

(2.5)

that is, $t \mapsto R_t$ is the distribution function of $\rho$.

**Remark 2.4.** (i) For a fixed $t > 0$, the mapping

$$A \mapsto N_t(A) := V([0, t] \times A)$$

(2.6)

defines a Poisson random measure on $S^2$, with non-atomic intensity

$$\mu_t(dx) = \rho \cdot \nu(dx) = \rho \cdot f(x)dx.$$  

(2.7)

Throughout this paper, we shall assume $f(x)$ to be bounded and bounded away from zero, e.g.

$$\zeta_1 \leq f(x) \leq \zeta_2, \quad \text{some } \zeta_1, \zeta_2 > 0, \text{ for all } x \in S^2.$$  

(2.8)

(ii) Let $\{X_i = i \geq 1\}$ be a sequence of i.i.d. random variables with values in $S^2$ and common distribution equal to $\nu$. Then, for a fixed $t > 0$, the random measure $A \mapsto N_t(A) = V([0, t] \times A)$ has the same distribution as $A \mapsto \sum_{i=1}^V \delta_{X_i}(A)$, where $\delta_x$ indicates a Dirac mass at $x$, and $V$ is an independent Poisson random variable with parameter $R_t$. This is the so-called binomial representation of a Poisson measure.

(iii) By definition, for every $t_1 < t_2$ one has that a random variable of the type $N_{t_2}(A) - N_{t_1}(A), A \subset S^2$, is independent of the random measure $N_{t_1}$, as defined in (2.6).

(iv) To simplify the discussion, one can assume that $\rho(ds) = \ell(ds)$, where $\ell$ is the Lebesgue measure and $R > 0$, in such a way that $R_t = R \cdot t$.

We will now introduce two distances between laws of random variables taking values in $\mathbb{R}^d$. Both distances define topologies, over the class of probability distributions on $\mathbb{R}^d$, that are strictly stronger than convergence in law. One should observe that, in this paper, the first one (Wasserstein distance) will be only used for random elements with values in $\mathbb{R}$. Given a function $g \in C^1(\mathbb{R}^d)$, we write $\|g\|_{\text{Lip}} = \sup_{x \in \mathbb{R}^d} \|\nabla g(x)\|_{\mathbb{R}^d}$. If $g \in C^2(\mathbb{R}^d)$, we set

$$M_2(g) = \sup_{x \in \mathbb{R}^d} \|\text{Hess } g(x)\|_{\text{op}},$$

where $\|\cdot\|_{\text{op}}$ indicates the operator norm.
**Definition 2.5.** The Wasserstein distance $d_W$, between the laws of two random vectors $X, Y$ with values in $\mathbb{R}^d$ ($d \geq 1$) and such that $E \|X\|_{\mathbb{R}^d}, E \|Y\|_{\mathbb{R}^d} < \infty$, is given by:

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

**Definition 2.6.** The distance $d_2$ between the laws of two random vectors $X, Y$ with values in $\mathbb{R}^d$ ($d \geq 1$), such that $E \|X\|_{\mathbb{R}^d}, E \|Y\|_{\mathbb{R}^d} < \infty$, is given by:

$$d_2(X, Y) = \sup_{g \in \mathcal{H}} |E [g(X)] - E [g(Y)]|,$$

where $\mathcal{H}$ denotes the collection of all functions $g \in C^2(\mathbb{R}^d)$ such that $\|g\|_{\text{Lip}} \leq 1$ and $M_2(g) \leq 1$.

We now present, in a form adapted to our goals, two upper bounds involving random variables living in the so-called first Wiener chaos of $N$. The first bound was proved in [40], and concerns normal approximations in dimension 1 with respect to the Wasserstein distance. The second bound appears in [42], and provides estimates for multidimensional normal approximations with respect to the distance $d_2$. Both bounds are obtained by means of a combination of the Malliavin calculus of variations and the Stein's method for probabilistic approximations.

**Remark 2.7.** (i) Let $f \in L^2(\Theta, \mu) \cap L^1(\Theta, \mu)$. In what follows, we shall use the symbols $N(f)$ and $\hat{N}(f)$, respectively, to denote the Wiener–Itô integrals of $f$ with respect to $N$ and with respect to the compensated Poisson measure

$$N(A) = N(A) - \mu(A), \quad A \in \mathcal{B}(\Theta),$$

where one uses the convention $N(A) - \mu(A) = \infty$ whenever $\mu(A) = \infty$ (recall that $\mu$ is $\sigma$-finite). Note that, for $N(f)$ to be well-defined, one needs that $f \in L^1(\Theta, \mu)$, whereas for the isometry property to hold one clearly needs that $f \in L^2(\Theta, \mu)$. We will also make use of the following isometric property: for every $f, g \in L^2(\Theta, \mu)$,

$$E[N(f)N(g)] = \int_{\Theta} f(x)g(x)\mu(dx).$$

The reader is referred e.g. to [41, Chapter 5] for an introduction to Wiener–Itô integrals.

(ii) For most of this paper, we shall consider Wiener–Itô integrals of functions $f$ having the form $f = [0, t] \times h$, where $t > 0$ and $h \in L^2(S^2, \nu) \cap L^1(S^2, \nu)$. For a function $f$ of this type one simply writes

$$N(h) = N([0, t] \times h) := N_t(h), \quad \text{and} \quad \hat{N}(h) = \hat{N}([0, t] \times h) := \hat{N}_t(h).$$

Observe that this notation is consistent with the one introduced in (2.6). Indeed, it is easily seen that $N_t(h)$ (resp. $\hat{N}_t(h)$) coincide with the Wiener–Itô integral of $h$ with respect to $N_t$ (resp. with respect to the compensated measure $\hat{N}_t = N_t - \mu_t = N_t - R_t \cdot \nu$).

(iii) In view of Remark 2.4-(ii), one also has that, for $h \in L^2(S^2, \nu) \cap L^1(S^2, \nu)$,

$$N_t(h) = \sum_{x \in \supp(N_t)} h(x), \quad \text{and} \quad \hat{N}_t(h) = \sum_{x \in \supp(N_t)} h(x) - \int_{S^2} h(x)\mu_t(dx),$$

with $\mu_t$ defined as in (2.7).

Now write $L^2(\nu) \equiv L^2(S^2, \nu)$ and, for a fixed integer $d \geq 1$, let $Y \sim \mathcal{N}_d(0, C)$, with $C$ positive definite; let also

$$F_t = (F_{t,1}, \ldots, F_{t,d}) = \left(\hat{N}_t(h_{t,1}), \ldots, \hat{N}_t(h_{t,d})\right)$$

be a collection of $d$-dimensional random vectors such that $h_{t,a} \in L^2(\nu)$. We call $I_t$ the covariance matrix of $F_t$, that is,

$$I_t(a, b) = E \left[N_t(h_{t,a}) N_t(h_{t,b})\right] = \langle h_{t,a}, h_{t,b} \rangle_{L^2(S^2, \mu_t)}, \quad a, b = 1, \ldots, d.$$

As usual, $\| \cdot \|_{\text{op}}$ and $\| \cdot \|_{\text{HS}}$ stand, respectively, for the operator and Hilbert–Schmidt norms. The formulation of the following result is pretty close to [41,42].

**Theorem 2.8.** Let the notation and assumptions of this section prevail.

1. Let $h \in L^2(\nu)$, let $Z \sim \mathcal{N}(0, 1)$ and fix $t > 0$. Then, the following bound holds (see (2.7)):

$$d_W(\hat{N}_t(h), Z) \leq \left|1 - \|h\|_{L^2(\nu)}^2\right| + \int_{S^2} |h(z)|^3 \mu_t(dz).$$

(2.13)
2. For a fixed integer \( d \geq 1 \), we have

\[
d_2(F_t, Y) \leq \| C^{-1} \|_{op} \| C \|_{op} \| C - J_t \|_{H.S.} + \frac{\sqrt{2\pi}}{8} \| C^{-1} \|_{op}^2 \| C \|_{op} \sum_{i,j,k=1}^{d} \int_{\mathbb{S}^2} |h_{t,i}(x)| |h_{t,j}(x)| |h_{t,k}(x)| \mu_t(dx), \tag{2.14}
\]

\[
\leq \| C^{-1} \|_{op} \| C \|_{op} \| C - J_t \|_{H.S.} + \frac{d^2 \sqrt{2\pi}}{8} \| C^{-1} \|_{op}^{3/2} \| C \|_{op} \sum_{i=1}^{d} \int_{\mathbb{S}^2} |h_{t,i}(x)|^3 \mu_t(dx). \tag{2.15}
\]

**Remark 2.9.** From the previous theorem, it follows immediately that, if \( \{h_t\} \subset L^2(\nu) \cap L^3(\nu) \) is a collection of kernels verifying, as \( t \to \infty \),

\[
\|h_t\|_{L^2(\mathbb{S}^2, \mu_t)} \to 1 \quad \text{and} \quad \|h_t\|_{L^3(\mathbb{S}^2, \mu_t)} \to 0,
\]

one has the CLT

\[
\hat{N}(h_t) \xrightarrow{Law} Z, \tag{2.17}
\]

and the inequality (2.13) provides an explicit upper bound in the Wasserstein distance. Likewise, if \( F_t(\omega) \to C(\omega, b) \) and \( \int_{\mathbb{S}^2} |h_{t,i}(x)|^3 \mu_t(dx) \to 0 \) as \( t \to \infty \), for \( a, b = 1, \ldots, d \), then \( d_2(F_t, Y) \to 0 \) and \( F_t \) converges in distribution to \( Y \).

**Remark 2.10.** The estimate (2.14) will be used to deduce one of the main multidimensional bounds in the present paper. It is a direct consequence of Theorem 3.3 in [42], where the following relation is proved: for every vector \( (F_1, \ldots, F_d) \) of sufficiently regular centered functionals of \( \hat{N}_t \),

\[
d_2(F, X) \leq \| C^{-1} \|_{op} \| C \|_{op}^{1/2} \sum_{i,j} \mathbb{E} \left[ |C(i,j) - \{DF_i, -DL^{-1}F_j\}|_{L^2(\mu)} \right]^2
\]

\[
+ \frac{\sqrt{2\pi}}{8} \| C^{-1} \|_{op}^{3/2} \| C \|_{op} \int_{\mathbb{S}^2} \mu_t(dz) \mathbb{E} \left[ \left( \sum_{i=1}^{d} |D_2F_i| \right)^2 \left( \sum_{j=1}^{d} |D_2L^{-1}F_j| \right) \right],
\]

where

\[
D_2F(N(\omega)) = F_2(N(\omega)) - F(N(\omega)), \quad \text{a.e.} -\mu(dz)P(d\omega),
\]

and

\[
F_2(N(\omega)) = F_2(N(\omega) + \delta_2),
\]

that is, the random variable \( F_2 \) is obtained by adding to the argument of \( F \) (which is a function of the point measure \( N \), a Dirac mass at \( z \), and \( L^{-1} \) is the so-called pseudo-inverse of the Ornstein–Uhlenbeck operator. The estimate (2.14) is then obtained by observing that, when \( F_t = F_{t,i} = \hat{N}_t(h_{t,i}) \), then \( D_2F_t = -D_2L^{-1}F = h_{t,i}(z) \), in such a way that

\[
\sum_{i,j} \mathbb{E} \left[ |C(i,j) - \{DF_i, -DL^{-1}F_j\}|_{L^2(\mu)} \right]^2 = \| C - K_t \|_{H.S.},
\]

and

\[
\int_{\mathbb{S}^2} \mu_t(dz) \mathbb{E} \left[ \left( \sum_{i=1}^{d} |D_2F_i| \right)^2 \left( \sum_{j=1}^{d} |D_2L^{-1}F_j| \right) \right] = \sum_{i,j,k=1}^{d} \int_{\mathbb{S}^2} |h_{t,i}(x)| |h_{t,j}(x)| |h_{t,k}(x)| \mu_t(dx).
\]

The next statement deals with the interesting fact that the convergence in law implied by Theorem 2.8 is indeed stable, as defined e.g. in the classic Ref. [18, Chapter 4].

**Proposition 2.11.** The central limit theorem described at the end of Point 2 of Theorem 2.8 (and a fortiori the CLT at Point 1 of the same theorem) is stable with respect to \( \sigma(N) \) (the \( \sigma \)-field generated by \( N \)) in the following sense: for every random variable \( X \) that is \( \sigma(N) \)-measurable, one has that

\[
(X, F_t) \xrightarrow{Law} (X, Y),
\]

where \( Y \sim \mathcal{N}_d(0, C) \) is independent of \( N \).
Proof. We just deal with the case \( d = 1 \), the extension to a general \( d \) following from elementary considerations. An approximation argument shows that it is enough to prove the following claim: if \( \tilde{N}(h_n) \) \( (h_n \in L^2(\mu), n \geq 1) \) is a sequence of random variables verifying \( E[|\tilde{N}(h_n)|^2] = E[|h_n|_{L^2(\mu)}^2] \to 1 \) and \( \int_{\mathbb{S}} |h_n|^2 d\mu \to 0 \), then for every fixed \( f \in L^2(\mu) \), the pair \( (\tilde{N}(f), \tilde{N}(h_n)) \) converges in distribution, as \( n \to \infty \), to \( (\tilde{N}(f), Z) \), where \( Z \sim \mathcal{N}(0, 1) \) is independent of \( N \). To see this, we start with the explicit formula (see e.g. [41, formula (5.3.31)]): for every \( \lambda, \gamma \in \mathbb{R} \)

\[
\psi_n(\lambda, \gamma) := E[\exp(i\lambda \tilde{N}(f) + \gamma \tilde{N}(h_n))]
= \exp \left[ \int_{\mathbb{S}} \left[ e^{i\lambda f(x) + \gamma h_n(x)} - 1 - i\lambda f(x) + \gamma h_n(x) \right] \mu(dx) \right].
\]

Our aim is to prove that, under the stated assumptions,

\[
\lim_{n \to \infty} \log(\psi_n(\lambda, \gamma)) = \int_{\mathbb{S}} \left[ e^{i\lambda f(x)} - 1 - i\lambda f(x) \right] \mu(dx) - \frac{\gamma^2}{2}.
\]

Standard computations show that

\[
\left| \log(\psi_n(\lambda, \gamma)) - \int_{\mathbb{S}} \left[ e^{i\lambda f(x)} - 1 - i\lambda f(x) \right] \mu(dx) - \frac{\gamma^2}{2} \right| 
\leq \left| \frac{\gamma^2}{2} - \frac{\gamma^2}{2} \int_{\mathbb{S}} h_n(x)^2 \mu(dx) \right| + |\gamma\lambda| \left| (h_n, f)_{L^2(\mu)} \right| + \frac{|\gamma|^2}{6} \int_{\mathbb{S}} |h_n(x)|^3 \mu(dx).
\]

Since \( \int_{\mathbb{S}} |h_n(x)|^3 \mu(dx) \to 0 \) and the mapping \( n \mapsto \|h_n\|^2_{L^2(\mu)} \) is bounded, one has that \( (h_n, f)_{L^2(\mu)} \to 0 \), and the conclusion follows by using the fact that \( \|h_n\|^2_{L^2(\mu)} \to 1 \) by assumption. \( \square \)

3. Needlet coefficients

3.1. Background: the needlet construction

We now provide an overview of the construction of the set of needlets on the unit sphere. The reader is referred to [30, Chapter 10] for an introduction to this topic. Relevant references on this subject are: the seminal papers [36,35], where needlets have been first defined; [12,13,11,14], among others, for generalizations to homogeneous spaces of compact groups and spin fiber bundles: [3,4,27,32] for the analysis of needlets on spherical Gaussian fields, and [31,45,7,10] for some (among many) applications to cosmological and astrophysical issues; see also [33,48] for other approaches to spherical wavelet construction.

(Spherical harmonics) In Fourier analysis, the set of spherical harmonics

\[
\{Y_{lm} : l \geq 0, m = -l, \ldots, l\}
\]

provides an orthonormal basis for the space of square-integrable functions on the unit sphere \( L^2(\mathbb{S}^2) := L^2(\mathbb{S}^2, dx) \), where \( dx \) stands for the Lebesgue measure on \( \mathbb{S}^2 \) (see for instance [1,23,30,52]). Spherical harmonics are defined as the eigenfunctions of the spherical Laplacian \( \Delta_{S^2} \) corresponding to eigenvalues \(-l(l+1)\), e.g. \( \Delta_{S^2} Y_{lm} = -l(l+1) Y_{lm} \), see again [30,52,53] for analytic expressions and more details and properties. For every \( l \geq 0 \), we define \( \mathcal{K}_l \) as the linear space given by the restriction to the sphere of the polynomials with degree at most \( l \). Plainly, one has that

\[
\mathcal{K}_l = \bigoplus_{k=0}^{l} \text{span} \{Y_{lm} : m = -k, \ldots, k\},
\]

where the direct sum is in the sense of \( L^2(\mathbb{S}^2) \).

(Cubature points) It is well-known that for every integer \( l = 1, 2, \ldots \) there exists a finite set of cubature points \( Q_l \subset \mathbb{S}^2 \), as well as a collection of weights \( \{\lambda_{\eta}\} \), indexed by the elements of \( Q_l \), such that

\[
\forall f \in L^2(\mathbb{S}^2), \quad \int_{\mathbb{S}^2} f(x) dx = \sum_{\eta \in Q_l} \lambda_\eta f(\eta).
\]

Now fix \( B > 1 \), and write \([x]\) to indicate the integer part of a given real \( x \). In what follows, we shall denote by \( \mathcal{X}_j = \{\xi_k\} \) and \( \{\lambda_{jk}\} \), respectively, the set \( Q_{2^B} \) and the associated class of weights. We also write \( K_j = \text{card} (\mathcal{X}_j) \). As proved in [36,35], cubature points and weights can be chosen to satisfy

\[
\lambda_{jk} \approx B^{-2j}, \quad K_j \approx B^{2j},
\]

(3.18)
where by \( a \approx b \), we mean that there exists \( c_1, c_2 > 0 \) such that \( c_1 a \leq b \leq c_2 a \) (see also e.g. [2, 43, 44] and [30, Chapter 10]).

(Spherical needlets) Fix \( B > 1 \) as before, as well as a real-valued mapping \( b \) on \((0, \infty)\). We assume that \( b \) verifies the following properties: (i) the function \( b(\cdot) \) has compact support in \([B^{-1}, B]\) (in such a way that the mapping \( l \mapsto b\left(\frac{l}{B}\right) \) has compact support in \( l \in [B^{-1}, B^{1+}] \)); (ii) for every \( \xi \geq 1 \), \( \sum_{l=0}^{\infty} b^2(\xi B^l) = 1 \) (partition of unit property), and (iii) \( b(\cdot) \in C^\infty(0, \infty) \).

Now, let us introduce the function \( L_t: [-1, 1] \rightarrow \mathbb{R} \) as

\[
L_t(\cos \vartheta) := \frac{2l + 1}{4\pi} P_l(\cos \vartheta), \quad \vartheta \in [0, \pi],
\]

where \( P_l(\cdot) \geq 0 \), \( l \geq 0 \), denotes as usual the set of Legendre polynomials. Note also that \( L_t(\langle x, y \rangle) = \sum_{m=-l}^{l} Y_{lm}(x) Y_{lm}(y) \), where \( \langle \cdot, \cdot \rangle \) denotes Euclidean inner product. The collection of spherical needlets \( \{\psi_{jk}\} \) associated with \( B \) and \( b(\cdot) \), are then defined as a weighted convolution, that is

\[
\psi_{jk}(x) := \sqrt{\lambda_{jk}} \sum_{l \geq 0} b\left(\frac{l}{B}\right) L_t(\langle x, \xi_{jk} \rangle).
\]

(Localization) The properties of \( b \) entail the following quasi-exponential localization property (see [36] or [30, Section 13.3]): for any \( \tau = 1, 2, \ldots \) there exists \( \kappa_\tau > 0 \) such that for any \( x \in \mathbb{S}^2 \),

\[
|\psi_{jk}(x)| \leq \frac{\kappa_\tau B^j}{(1 + B^j \arccos(\langle x, \xi_{jk} \rangle))^{\tau}},
\]

where \( d(x, y) := \arccos(\langle x, y \rangle) \) is the spherical distance. From localization, the following bound can be established on the \( L_p(\mathbb{S}^2) \) norms: for all \( 1 \leq p \leq +\infty \), there exist two positive constants \( q_p \) and \( q'_p \) such that

\[
q_p B^{j(1 - \frac{2}{p})} \leq \|\psi_{jk}\|_{L_p(\mathbb{S}^2)} \leq q'_p B^{j(1 - \frac{2}{p})}.
\]

(Needlets as frames) Finally, the fact that \( b \) is a partition of unit, allows on to deduce the following reconstruction formula (see again [36]): for \( f \in L^2(\mathbb{S}^2) \):

\[
f(x) = \sum_{j=0}^{\infty} \sum_{k=1}^{K_j} \beta_{jk} \psi_{jk}(x),
\]

where the convergence of the series is in \( L^2(\mathbb{S}^2) \), and

\[
\beta_{jk} := \langle f, \psi_{jk} \rangle_{L^2(\mathbb{S}^2)} = \int_{\mathbb{S}^2} f(x) \psi_{jk}(x) \, dx,
\]

represents the so-called needlet coefficient of index \( j, k \).

3.2. Two motivations: density estimates and point sources

The principal aim of this paper is to establish multidimensional asymptotic results for some possibly randomized version of random variables of the type

\[
\widehat{\beta}_{jk} = \widehat{\beta}_{jk}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(X_i), \quad j = 1, 2, \ldots, k = 1, \ldots, K_j,
\]

where the function \( \psi_{jk} \) is defined according to (3.19), and \( \{X_i : i \geq 1\} \) is some adequate sequence of i.i.d. random variables. We may also study the asymptotic behavior, as \( t \to \infty \), of multidimensional object of the type \( \{\widehat{\beta}_{jk}, k = 1, 2, \ldots, K_j(t)\} \), where \( t \mapsto K_j(t) \) is a non-decreasing mapping possibly diverging to infinity, and \( j \) may change with \( t \). In other words, as happens in realistic experimental circumstances, we may decide to focus on a growing number of coefficients as the number of (expected) events increase. Two strong motivations for this analysis, both coming from statistical applications, are detailed below.

(Density estimates) Consider a density function \( f \) on the sphere \( \mathbb{S}^2 \), that is: \( f \) is a mapping from \( \mathbb{S}^2 \) into \( \mathbb{R}_+ \), verifying \( \int_{\mathbb{S}^2} f(x) \, dx = 1 \), where \( dx \) indicates the Lebesgue measure on \( \mathbb{S}^2 \). Let \( \{X_i : i = 1, \ldots, n\} \) be a collection of i.i.d. observations with values in \( \mathbb{S}^2 \) with common distribution given by \( f(x) \, dx \). A classical statistical problem, considered for instance by [22, 24], concerns the estimation of \( f \) by wavelet/needlet thresholding techniques. To this aim, keeping in mind the notation (3.23), one uses [8, 15] the following estimator of \( f \):

\[
\hat{f}(x) = \sum_{jk} \beta_{jk}^{(n)} \psi_{jk}(x), \quad \beta_{jk}^{(n)} := \widehat{\beta}_{jk}^{(n)} \{ |\widehat{\beta}_{jk}^{(n)}| \geq c_n \},
\]
where \( t_n = \sqrt{\log n/\pi} \) and \( c \) is a constant to be determined. Finite-sample approximations on the distributions of \( \hat{\beta}_{jk} \) can then be instrumental for the exact determination of the thresholding value \( c_t \), see e.g. [8,15].

 Searching for point sources The joint distribution of the coefficients \( \{\hat{\beta}_{jk}\} \) (as defined in (3.23)) is required in statistical procedures devised for the research of so-called point sources, again for instance in an astrophysical context (see for instance [51]). The physical issue can be formalized as follows:

- Under the null hypothesis, we are observing a background of cosmic rays governed by a Poisson measure on the sphere \( S^2 \), with the form of the measure \( N_t(t) \) defined in (2.6) for some \( t > 0 \). In particular, \( N_t \) is built from a measure \( N \) verifying Assumption 2.3, and the intensity of \( \mu_t(dx) = E[N_t(dx)] \) is given by the absolutely continuous measure \( R_t \cdot f(x)dx \), where \( R_t > 0 \) and \( f \) is a density on the sphere. This situation corresponds, for instance, to the presence of a diffuse background of cosmological emissions.

- Under the alternative hypothesis, the background of cosmic rays is generated by a Poisson measure of the type:

\[
N^*_t(A) = N_t(A) + \sum_{p=1}^p N^{(p)}_{t} \delta_{\xi_p}(A),
\]

where \( \{\xi_1, \ldots, \xi_p\} \subset S^2 \), each mapping \( t \mapsto N^{(p)}_t \) is an independent Poisson process over \([0, \infty)\) with intensity \( \lambda_p \), and \( \delta_{\xi_p} \) is the Dirac mass at \( \xi_p \). In this case, one has that \( N^*_t \) is a Poisson measure with atomic intensity

\[
\mu^*_t(A) := E[N^*_t(A)] = R_t \int_A f(x)dx + \sum_{p=1}^p \lambda_p t \cdot \delta_{\xi_p}(A).
\]

In this context, the informal expression “searching for point sources” can then be transferred into “testing for \( P = 0 \)” or “jointly testing for \( \lambda_p > 0 \) at \( p = 1, \ldots, P \)”. The number \( P \) and the locations \( \{\xi_1, \ldots, \xi_p\} \) can be in general known or unknown. We refer to [17,50] for astrophysical applications of these ideas.

Remark 3.1. In order to directly apply the findings of [40,42], in what follows we shall focus on a randomized version of (3.23), where \( n \) is replaced by an independent Poisson number whose parameter diverges to infinity. Also, we will prefer a deterministic normalization over a random one. As formally shown in the discussion to follow, the resulting randomized coefficients can be neatly put into the framework of Section 2.

### 3.3. Needlet coefficients as Wiener–Itô integrals

Let \( N \) be a Poisson measure on \( \mathbb{R}_{+} \times S^2 \) satisfying the requirements of Assumption 2.3 (in particular, the intensity of \( N \) has the form \( \rho \otimes \nu \), where \( \nu(dx) = f(x)dx \), for some probability density \( f \) on the sphere, and one writes \( R_t = \rho([0, t]) \), \( t > 0 \). For every \( t > 0 \), let the Poisson measure \( N_t \) on \( S^2 \) be defined as in (2.6). For every \( j \geq 1 \) and every \( k = 1, \ldots, N_j \), consider the function \( \psi_{jk} \) defined in (3.19), and observe that \( \psi_{jk} \) is trivially an element of \( L^2(S^2, \nu) \cap L^1(S^2, \nu) \cap L^1(S^2, \nu) \). We write

\[
\sigma_{jk}^2 := \int_{S^2} \psi_{jk}(x)^2 f(x)dx, \quad b_{jk} := \int_{S^2} \psi_{jk}(x)f(x)dx.
\]

Observe that, if \( f(x) = 1_{S^2} \) (that is, the uniform density on the sphere), then \( b_{jk} = 0 \) for every \( j > 1 \). On the other hand, under (2.8),

\[
\xi_1 \| \psi_{jk} (\cdot) \|_{L^2}^2 \leq \sigma_{jk}^2 \leq \xi_2 \| \psi_{jk} (\cdot) \|_{L^2}^2.
\]

Note that (see (3.21)) the \( L^2 \)-norm of \( \{ \psi_{jk} \} \) is uniformly bounded above and below, and therefore the same is true for \( \{ \sigma_{jk}^2 \} \) (indeed, there exists \( \kappa > 0 \), independent of \( j \) and \( k \), such that \( 0 < \kappa < \| \psi_{jk} \|_{L^2(S^2)}^2 < 1 \)). For every \( t > 0 \) and every \( j, k \), we introduce the kernel

\[
h_{jk}^{(R_t)}(x) = \frac{\psi_{jk}(x)}{\sqrt{R_t} \sigma_{jk}}, \quad x \in S^2,
\]

and write

\[
\tilde{\beta}_{jk}^{(R_t)} = \hat{N}_t \left( h_{jk}^{(R_t)} \right) = \int_{S^2} h_{jk}^{(R_t)}(x) \hat{N}_t(dx) = \sum_{x \in \text{supp}(N_t)} h_{jk}^{(R_t)}(x) - R_t \cdot \int_{S^2} h_{jk}^{(R_t)}(x) \nu(dx).
\]

In view of Remark 2.4-(ii), the random variable \( \tilde{\beta}_{jk}^{(R_t)} \) can always be represented in the form

\[
\tilde{\beta}_{jk}^{(R_t)} = \frac{\left( \sum_{i=1}^{N_j(S^2)} \psi_{jk}(X_i) - R_t b_{jk} \right)}{\sqrt{R_t} \sigma_{jk}}.
\]
where \( \{X_i : i \geq 1\} \) is a sequence of i.i.d. random variables with common distribution \( \nu \), and independent of the centered random variable \( N_t(S^2) \). Moreover, the following relations are immediately checked:

\[
E\beta^{(R)}_j = 0, \quad E[|\beta^{(R)}_j|^2] = 1. \tag{3.27}
\]

**Remark 3.2.** Using the notation (3.23), we have that

\[
\beta^{(R)}_j = \frac{\left( N_t(S^2) \times \beta^{(R)}_j(S^2) - R_j b_j \right)}{\sqrt{R_j \sigma_j}}.
\]

**4. Bounds in dimension one**

We are now going to apply the content of Theorem 2.8-(1) to the random variables \( \beta^{(R)}_j \) introduced in the previous section. In the next statement, we write \( Z \sim \mathcal{N}(0, 1) \) to indicate a centered Gaussian random variable with unit variance. Recall that \( \xi_2 := \sup_{x \in \mathbb{R}^2} |f(x)| \cdot p \geq 1 \) and that the constants \( q_p, q'_p \) have been defined in (3.21).

**Proposition 4.1.** For every \( j, k \) and every \( t > 0 \), one has that

\[
d_W \left( \beta^{(R)}_j, Z \right) \leq \frac{(q'_3)^3 \xi_2 B_j}{\sqrt{R_j \sigma_j}}.
\]

It follows that for any sequence \( (j(n), k(n), t(n)) \), \( \beta^{(R)}(j(n), k(n)) \) converges in distribution to \( Z \), as \( n \to \infty \), provided \( B^{2(n)} = o(R^{2(n)}) \). The convergence is \( \sigma (N) \)-stable, in the sense of Proposition 2.11.

**Proof.** Using (3.25)–(3.26) together with (2.15) and (2.8),

\[
d_W \left( \beta^{(R)}_j, Z \right) \leq \int_{\mathbb{R}^2} \left| \frac{h^{(R)}_j(x)}{\sigma_j} \right|^3 \mu_t(dx) = \frac{R_j}{\sqrt{R_j \sigma_j}} \int_{\mathbb{R}^2} \left| \psi_j(x) \right|^3 f(x) dx \leq \frac{\xi_2}{\sqrt{R_j \sigma_j}} \left\| \psi_j \right\|^3_{L^3(S^2)} \leq \frac{(q'_3)^3 \xi_2 B_j}{\sqrt{R_j \sigma_j}}.
\]

where in the last inequality we use the property (3.21) with \( p = 3 \) to have:

\[
\left\| \psi_j \right\|^3_{L^3(S^2)} \leq (q'_3)^3 B_j^{3/2} \left( 1 - \frac{2}{3} \right) = (q'_3)^3 B_j.
\]

The last part of the statement follows from the fact that the topology induced by the Wasserstein distance (on the class of probability distributions on the real line) is strictly stronger than the topology of the convergence in law. \( \square \)

**Remark 4.2.** For \( f(x) \equiv (4\pi)^{-1} \) we have

\[
\sigma_j^2 = \frac{1}{4\pi} \int_{\mathbb{R}^2} \psi_j^2(x) dx = \left\| \psi_j \right\|^2_{L^2(S^2)},
\]

and more generally, under (2.8),

\[
d_W \left( \beta^{(R)}_j, Z \right) \leq \frac{B_j}{\sqrt{R_j} \sigma_j^{3/2}} \left\| \psi_j \right\|^{3/2}_{L^3(S^2)} \left( 1 - \frac{2}{3} \right) = \gamma(j, k, t). \tag{4.28}
\]

**Remark 4.3.** The previous result can be given the following heuristic interpretation. The factor \( B^{-j} \) can be viewed as the “effective scale” of the wavelet, i.e. it is the radius of the region centered at \( \xi_j \) where the value of the wavelet function is not negligible. Because needlets are isotropic, the “effective area” is of order \( B^{-2j} \). For governing measures with density which is bounded and bounded away from zero, the expected number of observations on a spherical cap of radius \( B^{-j} \) around \( \xi_j \) is hence given by

\[
E \left[ \text{card} \{ X_i : d(X_i, \xi_j) \leq B^{-j} \} \right] \approx R_j \int_{d(x, \xi_j) \leq B^{-j}} f(x) dx,
\]

\[
\zeta_1 B^{-2j} R_j \leq R_j \int_{d(x, \xi_j) \leq B^{-j}} f(x) dx \leq \zeta_2 B^{-2j} R_j,
\]

where \( \{X_i : i \geq 1\} \) is a sequence of i.i.d. random variables with common distribution \( \nu \), and independent of the centered random variable \( N_t(S^2) \). Moreover, the following relations are immediately checked:

\[
E\beta^{(R)}_j = 0, \quad E[|\beta^{(R)}_j|^2] = 1. \tag{3.27}
\]
In order to evaluate this integral, we can for instance follow [36], by splitting the sphere $S^2$ into two regions:

$$\begin{align*}
S_1 &= \{ x \in S^2 : d(x, \xi_{j_1}) > d(\xi_{j_1}, \xi_{j_2}) / 2 \} \\
S_2 &= \{ x \in S^2 : d(x, \xi_{j_1}) > d(\xi_{j_1}, \xi_{j_2}) / 2 \}.
\end{align*}$$

For what concerns the integral on $S_1$, we obtain:

$$\int_{S_1} \frac{1}{(1 + B'd(x, \xi_{j_1}))^{\tau}} \frac{1}{(1 + B'd(x, \xi_{j_2}))^{\tau}} dx \leq \frac{2^\tau}{(1 + B'd(\xi_{j_1}, \xi_{j_2}))^{\tau}} \int_{S_1} \frac{dx}{(1 + B'd(x, \xi_{j_2}))^{\tau}}.$$ 

One also has that

$$\int_{S_1} \frac{dx}{(1 + B'd(x, \xi_{j_2}))^{\tau}} \leq \int_{S_2} \frac{dx}{(1 + B'd(x, \xi_{j_2}))^{\tau}} = 2\pi \int_0^{\varphi} \frac{\sin \vartheta}{(1 + B't)^{\tau}} d\vartheta \leq \frac{2\pi}{B^{\tau}} \int_0^{\infty} \frac{y}{(1 + y)^{\tau}} dy \leq \frac{2\pi}{B^{\tau}} \left[ \int_0^1 y dy + \int_1^{\infty} y^{1-\tau} dy \right]$$

$$\leq \frac{2\pi C}{B^{\tau}}.$$
Because calculations on the region $S_2$ are exactly the same and because $S_2 \subseteq S_1 \cup S_2$, we have that, for some constant $\tilde{C}$, depending on $\tau$,  
\[
\|\Psi_{jk_1} \cdot \Psi_{jk_2}\|_{L^2(S^2)} \leq \frac{\tilde{C}}{(1 + B'd(\xi_{jk_1}, \xi_{jk_2}))^n},
\]
yielding the desired conclusion. \hfill \square

**Remark 5.2.** Assuming that $d(\xi_{jk_1}, \xi_{jk_2}) > \delta$ uniformly for all $j$, we have immediately
\[
|E_{jk_1}^r \hat{\beta}_{jk_2}^r| \leq \kappa_{\tau, \xi_2} \times B^{-j'},
\]
where the constant $\kappa_{\tau, \xi_2}$ only depend on $\tau, \xi_2$.

**Remark 5.3.** The previous lemma provides a tight bound, of some independent interest, on the high frequency behavior of covariances among wavelet coefficients for Poisson random fields. For Gaussian isotropic random fields, analogous results were provided by [3], in the case of standard needlets (bounded support), and by [27,29–32], in the “Mexican” case where support may be unbounded in multipole space. It should be noted how asymptotic uncorrelation holds in much greater generality for Poisson random fields than for Gaussian field: indeed in the latter case a regular variation condition had to be imposed on the tail behavior of the angular power spectrum, and in the Mexican case this condition had to be strengthened imposing an upper bound on the decay of the spectrum itself. The reason for such discrepancy is easily understood: for Poisson random fields, non overlapping regions are independent, whence (heuristically) localization in pixel space is sufficient to ensure asymptotic uncorrelation; on the contrary, in the Gaussian isotropic case different regions of the field are correlated at any angular distance, and asymptotic uncorrelation for the coefficients requires a much more delicate cancellation argument.

### 5.2. Fixed dimension

Fix $d \geq 2$ and $j \geq 1$, consider a fixed number of sampling points $\{\xi_{jk_1}, \ldots, \xi_{jk_d}\}$, and define the associated $d$-dimensional vector
\[
\hat{\beta}_j^{(R_k)} := (\hat{\beta}_{jk_1}^{(R_k)}, \ldots, \hat{\beta}_{jk_d}^{(R_k)}),
\]
whose covariance matrix will be denoted by $\Gamma_j$ (note that, by construction, $\Gamma_j(i, i) = 1$ for every $i = 1, \ldots, d$). Our aim is to apply the rough bound (2.15) in order to estimate the distance between the law of $\hat{\beta}_j^{(R_k)}$ and the law of a random Gaussian vector $Z \sim N_d(0, I_d)$, where $C = I_d$ stands for the identity $d \times d$ matrix. Using Lemma 5.1, one has the following basic estimates:
\[
\|C^{-1}\|_{op} = \|C\|_{op} = 1,
\]
\[
\|C - \Gamma_j\|_{HS} \leq \sup_{k_1 \neq k_2 = 1, \ldots, d} \frac{1}{\sigma_{jk_1}\sigma_{jk_2}} \left(1 + B'd(\xi_{jk_1}, \xi_{jk_2})\right)^n \leq \frac{d}{\xi_j q_2^2} \times \frac{1}{(1 + B') \inf_{k_1 \neq k_2 = 1, \ldots, d} d(\xi_{jk_1}, \xi_{jk_2})} = \mathcal{A}(t). \tag{5.29}
\]
Applying (2.15) yields therefore that
\[
d_2(\beta_j^{(R_k)}, Z) \leq \mathcal{A}(t) + d^2 \frac{\sqrt{2\pi}}{8} \sum_{k=1}^{d} R_k \int_{S^2} |h_{jk}^{(R_k)}(x)|^3 f(x) dx
\]
\[
= \mathcal{A}(t) + d^2 \frac{\sqrt{2\pi}}{8} \frac{\xi_j R_k}{\sqrt{\xi_j q_2^2}} \sum_{k=1}^{d} \int_{S^2} |\psi_{jk}(x)|^3 \sigma_{jk}^3 dx
\]
\[
\leq \mathcal{A}(t) + \frac{d^3 \xi_j}{\sqrt{\xi_j q_2^2}} \frac{\sqrt{2\pi}}{8} \|\psi_k\|_{L^3(S^2)}^3
\]
\[
\leq \mathcal{A}(t) + \frac{(q_j)^3 d^3 \xi_j}{\sqrt{\xi_j q_2^2}} \frac{\sqrt{2\pi}}{8} B^j,
\]
where we used (3.21) and (3.24) to yield $\sigma_{jk}^3 \geq \xi_j^{3/2} q_2^3$. We write this result as a separate statement.
Proposition 5.4. Under the above notation and assumptions,
\[ d_2 \left( \tilde{\beta}^{(R_i)}_{j(t)}, Z \right) \leq \frac{d_2 c_{t, j} B^{-j}}{\zeta_2 q_2^2 \left( 1 + \inf_{k_1 \neq k_2=1, \ldots, d} d \left( \xi_{j(t)k_1}, \xi_{j(t)k_2} \right) \right)^{\tau}} + \frac{(q_3^3 d_2^{3/2} \sqrt{2\pi})}{8 R_{t}^{3/2} q_2^2}. \]

Because \( \tau \) can be chosen arbitrarily large, it is immediately seen that the leading term in the \( d_2 \) distance is decaying with the same rate as in the univariate case, e.g. \( B^{j}/\sqrt{R_t} \). Assuming however that \( d = d_1 \), i.e. the case where the number of coefficients is itself growing with \( t \), the previous bound may become too large to be applicable. We shall hence try to establish a tighter bound, as detailed in the next section.

5.3. Growing dimension

In this section we allow for a growing number of coefficients to be evaluated simultaneously, and investigate the bounds that can be obtained under these circumstances. More precisely, we are now focusing on
\[ \tilde{\beta}^{(R_i)}_{j(t)} := \tilde{\beta}^{(R_i)}_{j(t)k_1}, \ldots, \tilde{\beta}^{(R_i)}_{j(t)k_q}, \]
where \( d_t \to \infty \), as \( t \to \infty \). Throughout the sequel, we shall assume that the points at which these coefficients are evaluated satisfy the condition:
\[ \inf_{k_1 \neq k_2=1, \ldots, d_t} d \left( \xi_{j(t)k_1}, \xi_{j(t)k_2} \right) \approx \frac{1}{\sqrt{d_t}}. \] (5.30)

Condition (5.30) is rather minimal; in fact, the cubature points for a standard needlet/wavelet construction can be taken to form a maximal \( (d_t)^{-1/2} \)-net (see [4,12,36,43] for more details and discussion). The following result is the main achievement of the paper.

Theorem 5.5. Let the previous assumptions and notation prevail. Then for all \( \tau = 2, 3, \ldots \), there exist positive constants \( c \) and \( c' \) (depending on \( \tau, \zeta_1, \zeta_2 \) but not from \( t, j(t), d(t) \)) such that we have
\[ d_2 \left( \tilde{\beta}^{(R_i)}_{j(t)}, Z \right) \leq \frac{c d_t}{\left( 1 + B(t) \inf_{k_1 \neq k_2=1, \ldots, d_t} d \left( \xi_{j(t)k_1}, \xi_{j(t)k_2} \right) \right)^{\tau}} + \frac{\sqrt{2\pi} c' d_t B(t)}{8 R_{t}^{3/2} q_2^2}. \] (5.31)

Proof. In view of (2.14) and (5.29), we just have to prove that the quantity
\[ \frac{\sqrt{2\pi} R_{t}^{3/2} q_2^2}{8} \sum_{k_1 \neq k_2} \int_{\mathbb{R}^2} \left| \psi_{j(t)k_1} (z) \right| \left| \psi_{j(t)k_2} (z) \right| \left| \psi_{j(t)k_3} (z) \right| f(z) dz \]
is smaller than the second summand on the RHS of (5.31). Now note that
\[ \sum_{k_1 \neq k_2} \int_{\mathbb{R}^2} \left| \psi_{j(t)k_1} (z) \right| \left| \psi_{j(t)k_2} (z) \right| \left| \psi_{j(t)k_3} (z) \right| dz \leq \sum_{k} \int_{B(\xi_{j(t)k}, B^{-j(t)})} \left| \sum_{k} \psi_{j(t)k} (z) \right|^3 dz, \]
where, for any \( z \in B(\xi_{j(t)k}, B^{-j(t)}) \)
\[ \sum_{k} \left| \psi_{j(t)k} (z) \right| \leq \sum_{k} \frac{C_{t} B(t)}{\left( 1 + B(t) d(\xi_{j(t)k}, z) \right)^{\tau}} \leq C_{t} B(t) + \sum_{k : \xi_{j(t)k} \not\in B(\xi_{j(t)k}, B^{-j(t)})} \frac{C_{t} B(t)}{\left( 1 + B(t) [d(\xi_{j(t)k}, \xi_{j(t)}) - d(z, \xi_{j(t)})] \right)^{\tau}} \leq C_{t} B(t) + \sum_{k : \xi_{j(t)k} \not\in B(\xi_{j(t)k}, B^{-j(t)})} \frac{C_{t} B(t)}{B(t) d(\xi_{j(t)k}, \xi_{j(t)})} \cdot \]
Now for \( \xi_{j(t)k} \not\in B(\xi_{j(t)k}, B^{-j(t)}), x \in B(\xi_{j(t)k}, B^{-j(t)}) \), we have by triangle inequality
\[ d(\xi_{j(t)k}, \xi_{j(t)}) + d(\xi_{j(t)k}, x) \geq d(\xi_{j(t)k}, x), \]
and because
\[ d(\xi_j(t)k, \xi_j(t)\lambda) \geq d(\xi_j(t)k, x), \quad \text{and} \quad 2d(\xi_j(t)k, \xi_j(t)\lambda) \geq d(\xi_j(t)\lambda, x), \]
we obtain
\[
\sum_{k: \xi_j(t)k \notin B(\xi_j(t)k, B^{-j(t)})} \left\{ \frac{C}{\mu B^{(j)}} \left[ B^{(j)} d(\xi_j(t)k, \xi_j(t)\lambda) \right] \right\}^\ell \\
= \sum_{k: \xi_j(t)k \notin B(\xi_j(t)k, B^{-j(t)})} \frac{1}{\mu \text{meas}(B(\xi_j(t)k, B^{-j(t)}))} \int_{B(\xi_j(t)k, B^{-j(t)})} \frac{k \lambda B^{(j)}}{\left[ B^{(j)} d(\xi_j(t)k, \xi_j(t)\lambda) \right]^{\ell}} dx \\
\leq \sum_{k: \xi_j(t)k \notin B(\xi_j(t)k, B^{-j(t)})} \frac{1}{\mu \text{meas}(B(\xi_j(t)k, B^{-j(t)}))} \int_{B(\xi_j(t)k, B^{-j(t)})} \frac{k \lambda B^{(j)}}{\left[ B^{(j)} d(\xi_j(t)\lambda, x) \right]^{\ell}} dx \leq k' B^{(j)},
\]
arguing as in [3, Lemma 6]. Hence
\[
\sum_{k} d_{k} |\psi_{j(t)k}(z)| \leq k'' B^{(j)}, \tag{5.32}
\]
uniformly over \( z \in S^2 \), which immediately provides the bound.

\[
\sum_{k} \int_{B(\xi_j(t)k, B^{-j(t)})} \left\{ \sum_{k} d_{k} |\psi_{j(t)k}(z)| \right\}^3 dz \leq (k'' B^{(j)})^3 \sum_{k} \int_{B(\xi_j(t)k, B^{-j(t)})} dz = (k'' B^{(j)})^3.
\]
Finally, to establish the sharper constraint
\[
\int_{S^2} \left\{ \sum_{k} d_{k} |\psi_{j(t)k}(z)| \right\}^3 dz \leq \tilde{c}_1 d, B^{(j)},
\]
it is sufficient to note that, exploiting (5.32)

\[
\sum_{k_1} \int_{S^2} |\psi_{j(t)k_1}(z)| \sum_{k_2} |\psi_{j(t)k_2}(z)| \sum_{k_3} |\psi_{j(t)k_3}(z)| \ dz \\
\leq k^2 B^{(j)} \sum_{k_1} \int_{S^2} |\psi_{j(t)k_1}(z)| \ dz = k^2 B^{(j)} \| \psi_{j(t)} \|_{L^1(S^2)} \\
\leq d_{t} \kappa^2 B^{(j)} B^{-j(t)} = d_{t} \kappa^2 B^{(j)},
\]
where we have used again \( \| \psi_{j(t)} \|_{L^p(S^2)} \leq O(B^{(j)}(1 + \frac{1}{j})^p) = O(B^{(j)(p-2)}) \), for \( p = 1 \). Thus (5.31) is established. \( \square \)

For definiteness, we shall also impose tighter conditions on the rate of growth of \( d_{t}, B^{(j)} \) with respect to \( R_{t} \), so that we can obtain a much more explicit bound, as follows:

**Corollary 5.6.** Let the previous assumptions and notation prevail, and assume moreover that there exists \( \alpha, \beta \) such that, as \( t \to \infty \)

\[
B^{(j)} \approx R_{t}^{\alpha}, \quad 0 < \alpha < 1, \quad d_{t} \approx R_{t}^{\beta}, \quad 0 < \beta < \alpha.
\]

There exists a constant \( \kappa \) (depending on \( \xi_1, \xi_2 \), but not on \( j, d_{j}, B \)) such that

\[
d_{2} \left( \tilde{\beta}_{j(t)}(R_{t}), \bar{z} \right) \leq \kappa \frac{d_{t} B^{(j)}}{R_{t}^{\beta}}, \tag{5.33}
\]
for all vectors \( \tilde{\beta}_{j_{1}}(R_{t}), \ldots, \tilde{\beta}_{j_{d}}(R_{t}) \), such that (5.30) holds.
Proof. It suffices to note that
\[
\frac{d_{t}k^{(t)}_r}{\sqrt{\tau}} \left( 1 + \frac{B(t)}{B^{(t)}} \inf_{k_1 \neq k_2} d(\xi_{j_1}, \xi_{j_2}) \right)^{\tau} = O(B^{-\tau(t)}d_t^{1+\tau/2})
\]

\[
= O \left( \frac{d_tB^{(t)}}{\sqrt{R_t}} \left( \frac{R_t d_t^2}{B^{(t+1)2j(t)}} \right)^{1/2} \right)
\]

and
\[
\frac{R_t d_t^2}{B^{(t+1)2j(t)}} = \frac{R_t^{1+\beta}}{R_t^{(t+1)w_t}} = R_t^{-\alpha + \tau(\beta - \alpha) + 1} = o(1), \quad \text{for } \tau > 1 - \alpha \frac{\alpha - \beta}{\alpha - \beta}. \quad \Box
\]

Remark 5.7. From (5.33), it follows that for \( R_t \approx 10^{12} \) we can establish asymptotic joint Gaussianity for all sequences of coefficients \( (\widetilde{p}_{j(t)k_1(t)}, \ldots, \widetilde{p}_{j(t)k_d(t)}) \) of dimensions such that
\[
\frac{d_tB^{(t)}}{\sqrt{R_t}} = o(1),
\]

i.e., we can take \( d_t \approx o(\sqrt{R_t}/B^{(t)}) \approx o(10^{6}/B^{(t)}) \), so that even at multipoles in the order of \( B^{(t)} = O(10^{3}) \) we might take around \( 10^{3} \) coefficients with the multivariate Gaussian approximation still holding. These arrays would not be sufficient for the map reconstruction at this scale, but would indeed provide a basis for joint multiple testing procedures as those described earlier.

Remark 5.8. Assume that \( d_t \) scales as \( B^{3(t)} \); loosely speaking, this corresponds to the situation where one focuses on the whole set of coefficients corresponding to scale \( j \), so that exact reconstruction for bandlimited functions with \( l = O(B^3) \) is feasible. Under this requirement, however, the “covariance” term \( \Lambda(t) \), i.e. the first element on the right-hand side of (5.31), is no longer asymptotically negligible and the approximation with Gaussian independent variables cannot be expected to hold. The approximation may however be implemented in terms of a Gaussian vector with dependent components. For the second term, convergence to zero when \( d_t \approx B^{2(t)} \) requires \( B^{3(t)} = o(\sqrt{R_t}) \). In terms of astrophysical applications, for \( R_t \approx 10^{12} \) this implies that one can focus on scales until \( 180^\circ/B' \approx 180^\circ/10^{2} \approx 2^\circ \); this is close to the resolution level considered for ground-based Cosmic Rays experiments such as ARGO-YBJ [17]. Of course, this value is much lower than the factor \( B' = o(\sqrt{R_t}) = o(10^{6}) \) required for the Gaussian approximation to hold in the one-dimensional case (e.g., on a univariate sequence of coefficients, for instance corresponding to a single location on the sphere).

Remark 5.9. As mentioned in the introduction, in this paper we decided to focus on a specific framework (spherical Poisson fields), which we believe of interest from the theoretical and the applied point of view. It is readily verified, however, how our results continue to hold with trivial modifications in a much greater span of circumstances, indeed in some cases with simpler proofs. Assume for instance we observe a sample of i.i.d. random variables \( \{X_t\} \), with probability density function \( f(\cdot) \) which is bounded and has support in \( [a, b] \subset \mathbb{R} \). Consider the kernel estimates
\[
\hat{f}_n(x_{nk}) := \frac{1}{nB^{-j}} \sum_{t=1}^{n} K \left( \frac{X_t - x_{nk}}{B^{-j}} \right), \quad (5.34)
\]

where \( K(\cdot) \) denotes a compactly supported and bounded kernel satisfying standard regularity conditions, and for each \( j \) the evaluation points \( (x_{n0}, \ldots, x_{nb^j}) \) form a \( B^{-j} \)-net; for instance
\[
a = x_{n0} < x_{n1} \ldots < x_{nb^j} = b, \quad x_{nk} = a + k \frac{b - a}{B^j}, \quad k = 0, 1, \ldots, B^j.
\]

As argued earlier, conditionally on \( N_t([a, b]) = n \), (5.34) has the same distribution as
\[
\hat{f}_n(x_{nk}) := \frac{1}{N_t([a, b])B^{-j}} \int_{a}^{b} K \left( \frac{u - x_{nk}}{B^{-j}} \right) dN_t(u),
\]

where \( N_t \) is a Poisson measure governed by \( R_t \times \int_{a} f(x)dx \) for all \( A \subset [a, b] \). Considering that \( \frac{N_t}{R_t} \rightarrow a.s. 1 \), a bound analogous to (5.33) can be established with little efforts for the vector \( \hat{f}_n(x_{n0}) := \{\hat{f}_n(x_{n1}), \ldots, \hat{f}_n(x_{nb^j})\} \). We leave this and related developments for further research.

Acknowledgment

The second author’s research was supported by the ERC Grant Pascal, n. 277742.