Journal of Multivariate Analysis 124 (2014) 398–417

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Application of second generation wavelets to blind spherical deconvolution

T. Vareschi

Université Denis Diderot Paris 7 and CNRS-UMR 7099, 175 rue du Chevaleret 75013 Paris, France

ARTICLE INFO

Article history:
Received 14 January 2013
Available online 4 December 2013

AMS subject classifications:
62G05
62G99
65J20
65J22

Keywords:
Blind deconvolution
Blockwise SVD
Spherical deconvolution
Second generation wavelets
Nonparametric adaptive estimation
Linear inverse problems

ABSTRACT

We address the problem of spherical deconvolution in a non-parametric statistical framework, where both the signal and the operator kernel are subject to measurement errors. After a preliminary treatment of the kernel, we apply a thresholding procedure to the signal in a second generation wavelet basis. Under standard assumptions on the kernel, we study the minimax performances of the resulting algorithm in terms of $\mathbb{L}^p$ losses ($p \geq 1$) on Besov spaces on the sphere. We hereby extend the application of second generation spherical wavelets to the blind spherical deconvolution framework. It is important to stress that the procedure is adaptive with regard to both the target function sparsity and the kernel blurring effect. We end with the study of a concrete example, putting into evidence the improvement of our procedure on the recent blockwise SVD algorithm of Delattre et al. (2012).

1. Introduction

1.1. Statistical framework

Consider the following problem: we aim at recovering a signal $\boldsymbol{f}$ defined on the 2-dimensional sphere $\mathbb{S}^2$. $\boldsymbol{f}$ is not observed directly, but through the action of a blurring process modeled by a linear operator $\boldsymbol{K}$, and further contaminated by an additive Gaussian white noise. This is resumed in the classic white noise model

$$\boldsymbol{Y}_\varepsilon = \boldsymbol{Kf} + \varepsilon\dot{\boldsymbol{W}} \tag{1.1}$$

where $\dot{\boldsymbol{W}}$ is a white noise on $\mathbb{L}^2(\mathbb{S}^2)$ and $\boldsymbol{K} : \mathbb{L}^2(\mathbb{S}^2) \to \mathbb{L}^2(\mathbb{S}^2)$ is a measurable operator. We shall further restrict the shape of $\boldsymbol{K}$ by assuming that it is a convolution operator on $\mathbb{L}^2(\mathbb{S}^2)$, a classic framework [14,19,18] enjoying convenient mathematical properties (see Section 1.2). Namely, we suppose that there exists $\boldsymbol{h} \in \mathbb{L}^2(\mathrm{SO}(3))$ such that

$$\boldsymbol{Kf}(\omega) = \int_{\mathrm{SO}(3)} \boldsymbol{f}(g^{-1}\omega)\boldsymbol{h}(g)dg \tag{1.2}$$

where dg is the Haar measure on SO(3). So to speak, $\boldsymbol{f}$ is averaged on a neighborhood of ω and weighted according to $\boldsymbol{h}(g)$ for each rotation g^{-1} applied to ω. Alternatively, in a density estimation framework, one observes a random n-sample $(\boldsymbol{\theta}_1X_1, \ldots, \boldsymbol{\theta}_nX_n)$ of $Z = \boldsymbol{\theta}X$ with density $\boldsymbol{Kf}$, where $\boldsymbol{\theta}$ is a random element in SO(3) (the group of rotations on $\mathbb{R}^3$) with density $\boldsymbol{h}$, and X has density $\boldsymbol{f} \in \mathbb{L}^2(\mathbb{S}^2)$. Formally we have $\varepsilon \sim n^{-1/2}$, and one can show that (1.2) holds as well [14].

E-mail addresses: thomas.vareschi@univ-paris-diderot.fr, vareschi@math.univ-paris-diderot.fr.

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.11.012

In practice, the blurring operator $\boldsymbol{K}$ is seldom directly observable and rather subject to measurement errors. For example $\boldsymbol{K}$ can be unknown but approximated via preliminary inference, or it can be the result of an unknown perturbation applied to a known operator. Following Efromovich and Koltchinskii [9] and Hoffmann and Reiß [15], we model the error in the operator as an additive Gaussian operator white noise. The observed result is a noisy version $\boldsymbol{K}_\delta$, satisfying

$$\boldsymbol{K}_\delta = \boldsymbol{K} + \delta \dot{\boldsymbol{B}} \tag{1.3}$$

where $\dot{\boldsymbol{B}}$ is a Gaussian operator white noise on $\mathcal{L}\big(\mathbb{L}^2(\mathbb{S}^2)\big)$ the set of linear endomorphisms of $\mathbb{L}^2(\mathbb{S}^2)$, independent from $\dot{\boldsymbol{W}}$. The meaning of models (1.1) and (1.3) is as follows: for $u, v, w, \in \mathbb{L}^2(\mathbb{S}^2)$, observable quantities take the forms

$$\langle \boldsymbol{Kf}, u\rangle + \varepsilon\alpha(u), \qquad \langle \boldsymbol{K}v, w\rangle + \delta\beta(v, w)$$

where $\alpha(u)$ and $\beta(v, w)$ are both Gaussian centered variables with respective variances $\|u\|_2^2$ and $\|v\|_2^2\|w\|_2^2$. Moreover, if $u', v', w' \in \mathbb{L}^2(\mathbb{S}^2)$ are other candidate functions, we have

$$\mathbb{E}[\alpha(u)\alpha(u')] = \langle u, u'\rangle_{\mathbb{L}^2(\mathbb{S}^2)}$$
$$\mathbb{E}[\beta(v, w)\beta(v', w')] = \langle v, v'\rangle_{\mathbb{L}^2(\mathbb{S}^2)}\langle w, w'\rangle_{\mathbb{L}^2(\mathbb{S}^2)}.$$

Many scientific fields call upon simple and efficient tools for the resolution of (1.1). Spherical deconvolution is for example well illustrated by the study of ultra high energy cosmic rays (UHECR), which are high energy radiations hitting the earth from apparently random directions. They could originate from long-lived relic particles from the Big Bang. Alternatively, they could be generated by the acceleration of standard particles, such as protons, in extremely violent astrophysical phenomena. They could also originate from Active Galactic Nuclei (AGN), or from neutron stars surrounded by extremely high magnetic fields. Discriminating among these different hypotheses involves the precise reconstruction of the probability density generating their observations. One could ask for example whether the latter is uniformly distributed among the sphere, or if it is constituted of superimposed localized spikes. In practice however, observations $(X_1, \dots, X_n)$ of such radiations are often subject to various physical perturbations. We model these by a random rotation $\boldsymbol{\theta}$, which is to say we actually observe $(\boldsymbol{\theta}_1 X_1, \dots \boldsymbol{\theta}_n X_n)$, n realizations of the random variable $Z = \boldsymbol{\theta}X$. The difficulty of the problem is characterized by the spreading of $\boldsymbol{h}$, the density of θ, around the neutral element of SO(3): the less localized, the more difficult the estimation of $\boldsymbol{f}$. Moreover, the law of $\boldsymbol{\theta}$ is not known in general, even if some assumptions can restrict its shape. In this case, preliminary inference is necessary, and leads to an estimator $\boldsymbol{K}_\delta$ of $\boldsymbol{K}$ according to (1.3).

Case of a known operator

We shall consider here the case where $\delta = 0$, and expose the path which finally led to the introduction and use of needlets. Spherical harmonics constitute the most natural set of functions to expand $\boldsymbol{f} \in \mathbb{L}^2(\mathbb{S}^2)$. Their frequency localization furthermore makes them ideally suited to spherical deconvolution, as they realize a blockwise SVD of $\boldsymbol{K}$ (as shown in Proposition 1.1), a property which guarantees the stability of its inversion. It prompted Healy et al. [14] to solve the spherical deconvolution problem with their use, hereby reaching optimal $\mathbb{L}^2$ rates of convergence on Sobolev spaces (Kim and Koo [19]). Unfortunately their performances can prove quite poor when the loss is measured by other $\mathbb{L}^p$ norms, $1 \le p \le \infty$, since they lack localization in the spatial domain (see [13]). The recent development of spherical wavelets [27,22] reversed this compromise, the latter being well localized in the spatial domain but very poorly in the frequential one. This makes them useful when a direct estimation of $\boldsymbol{f}$ is involved (see for example Freeden et al. [10] or Freeden et al. [11] for applications to geophysics and atmospheric sciences), but irrelevant in the setting of spherical deconvolution. The solution to this problem was finally brought by Narcowich et al. [24], who introduced a new set of functions, called needlets, which preserve the frequential localization of spherical harmonics and remedy their lack of spatial localization. Thereby, needlets inherit the stability of spherical harmonics in spherical deconvolution. They were subsequently exploited by Kerkyacharian et al. [18], who designed a procedure involving needlets attaining near-minimax rates of convergence for $\mathbb{L}^p$ losses $(1 \le p \le \infty)$ on Besov spaces (which definition is given in Section 2.3). Needlets also found various applications in the case of a direct estimation of $\boldsymbol{f}$, whether in astrophysics [21,13] or brain shape modeling [28].

Case of an unknown operator and Galerkin projection

The main methods in the context of blind deconvolution involve SVD and Galerkin schemes (see [3,4,15] for example). Galerkin projections were for example successfully applied to blind deconvolution on Hilbert spaces [9] or on Besov spaces on $[0, 1]^d$ [15,4]. They are based upon a discretization of (1.1) and (1.3) through the choice of appropriate test functions. Suppose we want to recover a function f from the observation of $g = Kf$. Let $(V_n)_{n\ge 0}$ and $(W_n)_{n\ge 0}$ be two increasing sequences of finite n-dimensional subspaces in $\mathbb{L}^2(\mathbb{S}^2)$, which admit the respective orthogonal bases $\boldsymbol{\varphi} = (\varphi_k)_{k\le n}$ and $\boldsymbol{\psi} = (\psi_k)_{k\le n}$. The Galerkin approximation $f_G \in V_n$ of f is the solution of the equation

$$\langle Kf_G, v\rangle = \langle g, v\rangle, \quad \forall v \in W_n. \tag{1.4}$$

This equation actually amounts to solving a finite dimensional linear system. Indeed, for $\gamma \in V_n$, note γ^n the vector whose components are $(\langle\gamma, \varphi_k\rangle)_{k\le n}$ and K^n the matrix with entries $(\langle K\varphi_k, \psi_k\rangle)_{k,k'\le n}$. Then $f_G \in V_n$ and we have

$$g^n = K^n f_G^n.$$

The presence of noise in the signal and the operator raises additional issues. First, the algorithm must include and articulate two essential steps, namely the inversion of $\boldsymbol{K}$ and the regularization of the data which we will perform through a projection/thresholding scheme. Note that both the signal and the operator $\boldsymbol{K}$ can (and will) be subject to regularization (see [15,6]). The second practical problem concerns the choice of the test functions $\boldsymbol{\varphi}$, $\boldsymbol{\psi}$. This choice should answer the dilemma to find a set which is compatible both with the sparsity of $\boldsymbol{f}$ and with the structure of $\boldsymbol{K}$ (see [15,6]). Spherical harmonics respond optimally to this problem in the case of spherical deconvolution on Sobolev spaces for a $\mathbb{L}^2$ error, since they realize a blockwise SVD of $\boldsymbol{K}$, as shown in Proposition 1.1. More importantly here, they allow a fine treatment of $\boldsymbol{K}_\delta$ thanks to the sparse structure of the original operator $\boldsymbol{K}$ in that basis. This structure was exploited by Delattre et al. [6] in the context of blind spherical deconvolution: by an adequate regularization of $\boldsymbol{K}_\delta$ and $\boldsymbol{Y}_\varepsilon$, the authors exhibited optimal rates of convergence under common assumptions on $\boldsymbol{f}$ and $\boldsymbol{K}$. This procedure is exposed in detail in Section 3.1.2.

1.2. Harmonic analysis on SO(3) and $\mathbb{S}^2$

The present part provides useful mathematical tools in the context of spherical deconvolution. It is a quick overview of harmonic analysis on the spaces $\mathbb{S}^2$ and SO(3) which is mostly inspired by Healy et al. [14], and will end up with the blockwise SVD property.

Let us define the Euler matrices

$$u(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad a(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

where $\varphi \in [0, 2\pi)$, $\theta \in [0, \pi)$.

Every rotation g in SO(3) is the product of 3 elementary rotations:

$$\varepsilon = u(\varphi)a(\theta)u(\psi) \tag{1.5}$$

where $\varphi, \psi \in [0, 2\pi)$, $\theta \in [0, \pi)$ are the Euler angles of g. Let $\ell \in \mathbb{N}$ and $-\ell \le m, n \le \ell$. We also define the rotational harmonics

$$R_{\ell,m,n}(\varphi, \theta, \psi) = e^{-i(m\varphi+n\psi)} P_{\ell,m,n}(\cos(\theta)) \tag{1.6}$$

where $P_{\ell,m,n}$ are the second type Legendre functions (see [29]).

The functions $R_{\ell,m,n}$, $\ell \in \mathbb{N}$, $|m|, |n| \leqslant \ell$ are the eigenfunctions of the Laplace–Beltrami operator on SO(3), associated with the eigenvalues $2\ell + 1$. Therefore, the system $(\sqrt{2\ell+1}R_{\ell,m,n})_{\ell\ge 0, |m|, |n| \le \ell}$ forms a complete orthonormal basis of $\mathbb{L}^2(\mathrm{SO}(3))$.
Let $h \in \mathbb{L}^2(\mathrm{SO}(3))$. For all $\ell \ge 0$, the projection of h on the space of rotational harmonics with degree ℓ is

$$\sum_{m,n=-\ell}^{\ell} \hat{h}_{\ell,m,n} R_{\ell,m,n}$$

where $\hat{h}_{\ell,m,n}$ is the (ℓ, m, n) Fourier coefficient of h, defined by

$$\hat{h}_{\ell,m,n} = \int_{\mathrm{SO}(3)} h(g)\overline{R_{\ell,m,n}(g)}dg \tag{1.7}$$

and dg is the Haar measure on SO(3). An analogous study is available on $\mathbb{S}^2$. Any point $\omega \in \mathbb{S}^2$ is determined by its spherical coordinates (θ, φ):

$$\omega = (\sin(\theta)\cos(\varphi), \sin(\theta)\sin(\varphi), \cos(\theta)) \tag{1.8}$$

where $\theta \in [0, \pi)$ and $\varphi \in [0, 2\pi)$. Let ℓ a positive integer and m, n two integers ranking from $-\ell$ to ℓ. Define the spherical harmonics on $\mathbb{S}^2$ by:

$$Y_{\ell,m}(\theta, \varphi) = (-1)^m \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_{\ell,m}(\cos(\theta))e^{im\varphi} \tag{1.9}$$

where $P_{\ell,m}$ are the Legendre functions (see [29]). The set $(Y_{\ell,m})_{\ell\ge 0, |m|\le\ell}$ constitutes an orthonormal basis of $\mathbb{L}^2(\mathbb{S}^2)$. Note $\mathbb{H}_\ell$ the space of spherical harmonics with degree ℓ and $\boldsymbol{P}_\ell$ the orthogonal projector onto $\mathbb{H}_\ell$. Then for every function $\gamma \in \mathbb{L}^2(\mathbb{S}^2)$,

$$\boldsymbol{P}_\ell \gamma = \sum_{m=-\ell}^{\ell} \hat{\gamma}_{\ell,m} Y_{\ell,m}$$

where $\hat{\gamma}_{\ell,m}$ is the (ℓ, m) Fourier coefficient of f, defined by

$$\hat{\gamma}_{\ell,m} = \int_{\mathbb{S}^2} \gamma(\omega)\overline{Y_{\ell,m}(\omega)}d\omega.$$

The term 'blockwise SVD' finds its roots in the following proposition, which links the Fourier coefficients of $h * \gamma$ to those of h and γ. A proof is present in [14].

Proposition 1.1 (*Blockwise SVD Property*). *Let $h \in \mathbb{L}^2(\mathrm{SO}(3))$ and $\gamma \in \mathbb{L}^2(\mathbb{S}^2)$. The Fourier coefficients of $h * \gamma$ are*

$$\widehat{(h*\gamma)}_{\ell,m} = \sum_{n=-\ell}^{\ell} \hat{h}_{\ell,m,n}\hat{\gamma}_{\ell,n} = \sum_{n=-\ell}^{\ell} \langle h * Y_{\ell,n}, Y_{\ell,m}\rangle\langle\gamma, Y_{\ell,n}\rangle.$$

Consequences on the Galerkin projection

Proposition 1.1 has interesting implications in terms of the Galerkin projection of (1.1) and (1.3) on spherical harmonics. Indeed, note $\boldsymbol{K}_\ell \in M_{2\ell+1}(\mathbb{C})$ the matrix

$$\boldsymbol{K}_\ell = \left(\langle \boldsymbol{K} Y_{\ell,n}, Y_{\ell,m}\rangle\right)_{|m|,|n|\le\ell}$$

and, for $\gamma \in \mathbb{L}^2(\mathbb{S}^2)$, note $\gamma_\ell \in \mathbb{C}^{2\ell+1}$ the vector $\left(\langle\gamma, Y_{\ell,m}\rangle\right)_{|m|\le\ell}$. Proposition 1.1 then translates into

$$\left(\boldsymbol{K}\boldsymbol{f}\right)_\ell = \boldsymbol{K}_\ell \boldsymbol{f}_\ell.$$

Let us hence turn back to the Galerkin scheme (1.4). Take $W_n = V_n$ to be the subspace of $\mathbb{L}^2(\mathbb{S}^2)$ spanned by all spherical harmonics with degree less than n. Proposition 1.1 implies the following:

1. $\boldsymbol{f}_G = \sum_{\ell\le n} \boldsymbol{P}_\ell \boldsymbol{f}$
2. The Galerkin matrix $\boldsymbol{K}^n$ is a sparse block matrix, with blocks $\boldsymbol{K}_\ell$, $\ell \le n$, along its diagonal. This justifies the denomination of blockwise SVD.

In the sequel, if $\gamma \in \mathbb{L}^2(\mathbb{S}^2)$, we will refer indifferently to $\boldsymbol{P}_\ell\gamma$ or γ_ℓ. Similarly, if $\boldsymbol{K}$ is a convolution operator on $\mathbb{L}^2(\mathbb{S}^2)$, we will refer indifferently to $\boldsymbol{P}_\ell\boldsymbol{K}\boldsymbol{P}_\ell$ or $\boldsymbol{K}_\ell$. Due to Parseval's formula, we also have the relations

$$\|\boldsymbol{P}_\ell\,\gamma\|_{\mathbb{L}^2(\mathbb{S}^2)} = \|\gamma_\ell\|_{\ell^2(\mathbb{C}^{2\ell+1})}$$
$$\|\boldsymbol{P}_\ell\boldsymbol{K}\boldsymbol{P}_\ell\|_{\mathbb{L}^2(\mathbb{S}^2)\to\mathbb{L}^2(\mathbb{S}^2)} = \|\boldsymbol{K}_\ell\|_{op}$$

where we have noted $\|.\|_{op}$ the spectral norm of a matrix. Turning back to the original problem and reminding Proposition 1.1, we can reformulate the equivalent problem, obtained by projecting (1.1) and (1.3) onto every space $\mathbb{H}_\ell$:

$$\forall\ell \ge 0, \quad \boldsymbol{Y}_{\varepsilon,\ell} = \boldsymbol{K}_\ell\boldsymbol{f}_\ell + \varepsilon\dot{\boldsymbol{W}}_\ell \tag{1.10}$$

$$\forall\ell \ge 0, \quad \boldsymbol{K}_{\delta,\ell} = \boldsymbol{K}_\ell + \delta\dot{\boldsymbol{B}}_\ell \tag{1.11}$$

where $\dot{\boldsymbol{W}}_\ell$ is a centered Gaussian vector with covariance $\boldsymbol{I}_{2\ell+1}$, and $\dot{\boldsymbol{B}}_\ell$ is a $(2\ell+1)\times(2\ell+1)$ matrix whose entries are i.i.d. $\mathcal{N}(0,1)$ variables.

An alternative point of view

The blockwise SVD property allows to avoid considering the inner products $\langle \boldsymbol{K}_\delta Y_{\ell,m}, Y_{\ell' n}\rangle$ when $\ell \ne \ell'$ and provides hereby $\boldsymbol{K}_\delta$ with a sparse structure consisting in $(2\ell+1)\times(2\ell+1)$ matrices on its diagonal. This is a consequence of the convolutive structure of $\boldsymbol{K}$. Actually, an alternative way to get to (1.11) is to consider that the kernel $\boldsymbol{h}$ of $\boldsymbol{K}$ is directly polluted by an additive Gaussian white noise on $\mathbb{L}^2(\mathrm{SO}(3))$. Namely we would observe $\boldsymbol{h}_\delta = \boldsymbol{h} + \delta\dot{\boldsymbol{b}}$ where $\dot{\boldsymbol{b}}$ is a Gaussian white noise on $\mathbb{L}^2(\mathrm{SO}(3))$. Observations conducted on $\boldsymbol{K}_\delta$ would hence become

$$\langle \boldsymbol{K}_\delta Y_{\ell,m}, Y_{\ell',n}\rangle = \langle \boldsymbol{h}_\delta * Y_{\ell,m}, Y_{\ell',n}\rangle$$

and would be null if $\ell \ne \ell'$ as a consequence of Proposition 1.1.

2. Needlets

2.1. Construction of needlets

Needlets were introduced by Narcowich et al. [24], and used in the framework of density estimation on the sphere by Baldi et al. [1] and Kerkyacharian et al. [18]. As their construction relies on a rearrangement of spherical harmonics, they inherit the useful stability properties of the latter in spherical deconvolution. In addition, whereas the support of a spherical harmonic spreads all over the sphere, each needlet is localized around its center, and decays almost exponentially away from it. This concentration is more and more pronounced as the level of resolution increases, making needlets a handy multi-scale tool on the sphere. Also, they characterize in a way similar to Euclidean wavelets belonging to Besov spaces on the sphere, which will be introduced in Section 2.3.

Needlet framework

The needlet theory relies on the orthogonal decomposition of the space $\mathbb{L}^2(\mathbb{S}^2)$:

$$\mathbb{L}^2(\mathbb{S}^2) = \bigoplus_{\ell\ge 0}^{\perp} \mathbb{H}_\ell$$

where each $\mathbb{H}_\ell$ represents the space of spherical harmonics with degree ℓ. Along with this decomposition naturally come the orthogonal projectors $\boldsymbol{P}_\ell$ on $\mathbb{H}_\ell$ and their associated kernels

$$L_\ell(x,y) = L_\ell\big(\langle x,y\rangle_{\mathbb{R}^{2\ell+1}}\big) = \sum_{|m|\le\ell} Y_{\ell,m}(x)\overline{Y_{\ell,m}(y)}.$$

Since L_ℓ is the kernel of an orthogonal projector, it satisfies the property

$$\int_{\mathbb{S}^2} L_\ell(x,y)L_k(y,z)dy = \delta_{\ell,k}L_\ell(x,z), \quad \text{for all } x,z\in\mathbb{S}^2. \tag{2.1}$$

Paley–Littlewood decomposition

The role of the Paley–Littlewood theory is to generalize some convenient properties, tied to the $\mathbb{L}^2$ framework, to $\mathbb{L}^p$ spaces. This task is performed via a binary filtering of the projectors $\boldsymbol{P}_\ell$.
Let $a \in \mathcal{C}^\infty(\mathbb{R})$ be a symmetric function, compactly supported in $[-1,1]$ and decreasing on $\mathbb{R}^+$, such that for all $x\in\mathbb{R}$, $0\le a(x)\le 1$ and for all $|x|\le 1/2$, $|a(x)|=1$. Define for all $x\in\mathbb{R}$, $b^2(x) = a(x/2)-a(x)$. b^2 is a positive function, supported in $[-2;-1/2]\bigcup[1/2;2]$, satisfying by construction $\sum_{j\ge0} b^2(x2^{-j}) = 1$ for all $|x|\ge 1$. Define also the following projection kernels on $\mathbb{R}^2$

$$\Lambda_j(x,y) = \sum_{\ell\ge0} b^2\left(\frac{\ell}{2^j}\right)L_\ell(x,y) \tag{2.2}$$

$$M_j(x,y) = \sum_{\ell\ge0} b\left(\frac{\ell}{2^j}\right)L_\ell(x,y), \tag{2.3}$$

as well as the associated operators on $\mathbb{L}^2(\mathbb{S}^2)$,

$$B_j : \gamma \mapsto \int_{\mathbb{S}^2}\Lambda_j(x,y)\gamma(y)dy, \qquad A_J : \gamma\mapsto \sum_{j=-1}^{J} B_j\gamma$$

with the convention $B_{-1}\gamma = \boldsymbol{P}_0\gamma$. Note that the sum in (2.2) and (2.3) is finite since $b(\ell 2^{-j}) = 0$ if $\ell\notin L_j$, where we have noted L_j the set of integers ranging from 2^{j-1} to $2^{j+1}-1$. It is straightforward to show that, for all $f\in\mathbb{L}^2(\mathbb{S}^2)$,

$$\|A_J\gamma-\gamma\|_2 \to 0 \quad \text{as } J\to\infty. \tag{2.4}$$

One of the main results in Narcowich et al. [24] states that A_J also mimics the best polynomial approximation of γ with respect to $\|.\|_p$ for all $p\ge1$, as expressed in the following theorem:

Theorem 2.1. *For all $p\in[1,\infty[$, if $f\in\mathbb{L}^p(\mathbb{S}^2)$, then*

$$\|A_J\gamma-\gamma\|_p \to 0 \quad \text{as } J\to\infty,$$

with uniform convergence if $f\in\mathcal{C}^0(\mathbb{S}^2)$.

Space discretization

The second ingredient in the construction of needlets is the polynomial structure of the spaces $\mathbb{H}_\ell$. Indeed, for all $\gamma\in\mathbb{H}_\ell$, $\gamma'\in\mathbb{H}_k$, we have $\gamma\gamma'\in\mathbb{H}_{\ell+k}$. Let us note $\mathcal{P}_\ell = \bigoplus_{k\le\ell}\mathbb{H}_k$. A corresponding quadrature formula on every space $\mathcal{P}_\ell$ is available as well:

Proposition 2.2 (*Quadrature Formula*)**.** *For all $\ell\ge0$, there exists a finite set Z_ℓ of cubature points, associated to the cubature weights $(\lambda_\eta)_{\eta\in Z_\ell}$ such that, for all $\gamma\in\mathcal{P}_\ell$,*

$$\int_{\mathbb{S}^2}\gamma = \sum_{\eta\in Z_\ell}\lambda_\eta\gamma(\eta).$$

Since $b(\ell2^{-j})\ne 0$ only if $2^{j-1}\le\ell<2^{j+1}$ the function $z\mapsto M_j(x,z)$ belongs to $\mathcal{P}_{2^{j+1}-1}$, and $z\mapsto M_j(x,z)M_j(z,y)$ is an element of $\mathcal{P}_{2^{j+2}-2}$. For more convenience, we will note $\mathcal{Z}_j = Z_{2^{j+2}-2}$ the corresponding set of cubature points. It can be shown that the cubature points $\eta\in\mathcal{Z}_j$ and weights $(\lambda_\eta)_{\eta\in\mathcal{Z}_j}$ can be chosen so that the two following conditions are fulfilled:

$$c^{-1}2^{2j}\le \operatorname{card}(\mathcal{Z}_j)\le c2^{2j} \quad\text{and}\quad C^{-1}2^{-2j}\le\lambda_\eta\le C2^{-2j} \tag{2.5}$$

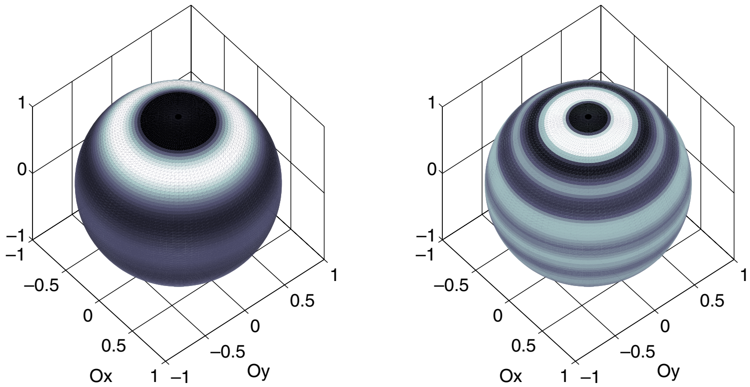


Fig. 1. A Spherical representation of two needlets (level $j = 2, 3$ from left to right) centered around the point (0, 0, 1). The darkened zones correspond to the regions where the needlet is high. The concentration of the needlet around its center is more pronounced as the level j increases.

for some constants $c, C > 0$. For all $j \geq 0$, B_j now satisfies:

$$B_j(\gamma) = \int_{\mathbb{S}^2} \left(\sum_{\eta \in \mathcal{Z}_j} \lambda_\eta M_j(x, \eta) M_j(\eta, y) dz \right) \gamma(y) dy$$
$$= \sum_{\eta \in \mathcal{Z}_j} \sqrt{\lambda_\eta} M_j(x, \eta) \int_{\mathbb{S}^2} \sqrt{\lambda_\eta} M_j(\eta, y) \gamma(y) dy. \tag{2.6}$$

The functions $\boldsymbol{\psi}_{j,\eta} = \sqrt{\lambda_\eta} M_j(., \eta)$ appearing in (2.6) are called needlets. By extension we also define $\boldsymbol{\psi}_{-1,\eta} = \boldsymbol{\psi}_0 = (4\pi)^{-1} \mathbf{1}_{\{\mathbb{S}^2\}}$ the normalized constant on $\mathbb{S}^2$. An immediate consequence of (2.4) is the following: for all $\gamma \in \mathbb{L}^2(\mathbb{S}^2)$,

$$\|\gamma\|_2^2 = \langle \gamma, \boldsymbol{\psi}_0 \rangle^2 + \sum_{j \geq 0} \sum_{\eta \in \mathcal{Z}_j} \langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle^2. \tag{2.7}$$

The next section shows that this property somehow generalizes to other $\mathbb{L}^p$ norms, $1 \leq p \leq \infty$.

2.2. Properties of needlets

By construction, needlets are compactly supported in the frequential domain. A crucial result proved by Narcowich et al. [24] shows that they are furthermore nearly exponentially localized in space:

Theorem 2.3. *Let $j \geq 0$, $\eta \in \mathcal{Z}_j$. For all $M > 0$, there exists $C_M > 0$ such that*

$$\forall x \in \mathbb{S}^2, \ |\boldsymbol{\psi}_{j,\eta}(x)| \leq \frac{C_M 2^j}{(1 + 2^j d(x, \eta))^M} \tag{2.8}$$

where $d(x, y) = \arccos(\langle x, y \rangle)$ is the geodesic distance on the sphere. To illustrate this point, we represented two needlets of level $j = 2, 3$ in Fig. 1. This property is central, since it allows to relate the $\mathbb{L}^p$ norm of the projection $\sum_{\eta \in \mathcal{Z}_j} \langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle \boldsymbol{\psi}_{j,\eta}$ to the discrete ℓ^p norm of the finite sequence $(|\langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle| \|\boldsymbol{\psi}_{j,\eta}\|_p)_{\eta \in \mathcal{Z}_j}$. Indeed, the two following propositions hold [17]:

Proposition 2.4. *For all $1 \leq p \leq \infty$ (with the convention $1/\infty = 0$), there exist $c_p, C_p > 0$ such that*

$$c_p 2^{2j(\frac{1}{2} - \frac{1}{p})} \leq \|\boldsymbol{\psi}_{j,\eta}\|_p \leq C_p 2^{2j(\frac{1}{2} - \frac{1}{p})}. \tag{2.9}$$

Proposition 2.5. *For all $p \geq 1$, there exists a constant C_p such that for all $\gamma \in \mathbb{L}^p(\mathbb{S}^2)$,*

$$\|B_j(\gamma)\|_p \leq C_p \left(\sum_{\eta \in \mathcal{Z}_j} |\langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle|^p \|\boldsymbol{\psi}_{j,\eta}\|_p^p \right)^{1/p}. \tag{2.10}$$

Moreover, if $p = \infty$, there exists $C_\infty > 0$ such that

$$\|B_j(\gamma)\|_\infty \leq C_\infty 2^j \sup_{\eta \in \mathcal{Z}_j} |\langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle|. \tag{2.11}$$

2.3. Besov spaces

Besov spaces on the sphere naturally generalize the usual approximation properties of regular functions while being simply characterized with the help of needlets. A complete description, and the proofs of the results claimed in this part can be found in Narcowich et al. [23] or Kerkyacharian and Picard [17]. Let $\gamma : \mathbb{S}^2 \mapsto \mathbb{R}$ be a measurable function and let $E_{k,\pi}(\gamma)$ $(1 \le \pi \le \infty)$ be the distance of γ to $\mathcal{P}_k$ with respect to $\|.\|_\pi$, that is

$$E_{k,\pi}(\gamma) = \inf_{P \in \mathcal{P}_k} \|\gamma - P\|_\pi .$$

Theorem 2.6. *Let $0 < s < \infty$, $1 \le p \le \infty$ and $0 < r \le \infty$. Let $\gamma \in L^\pi(\mathbb{S}^2)$. The following statements are equivalent and define the Besov space $B^s_{\pi,r}$ on $\mathbb{S}^2$:*

$$\left(\sum_{k \ge 0} k^{rs} E_{k,\pi}(\gamma)^r \frac{1}{k} \right)^{1/r} < \infty \tag{2.12}$$

$$\left(\sum_{j \ge 0} 2^{jrs} E_{2^j,\pi}(\gamma)^r \right)^{1/r} < \infty \tag{2.13}$$

$$\exists\, \xi_j \in \ell^r(\mathbb{N}), \quad \|B_j \gamma\|_\pi = \xi_j 2^{-js} \tag{2.14}$$

$$\exists\, \xi_j \in \ell^r(\mathbb{N}), \quad \left(\sum_{\eta \in \mathcal{Z}_j} |\langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle|^\pi \, \|\boldsymbol{\psi}_{j,\eta}\|_\pi^\pi \right)^{1/\pi} = \xi_j 2^{-js}. \tag{2.15}$$

$B^s_{\pi,q}$ is a Banach space, endowed with the norm

$$\|\gamma\|_{B^s_{\pi,r}} = \left\| \left(2^{j\left(s+2\left(\frac{1}{2}-\frac{1}{\pi}\right)\right)} \left(\sum_{\eta \in \mathcal{Z}_j} |\langle \gamma, \boldsymbol{\psi}_{j,\eta} \rangle|^\pi \right)^{1/\pi} \right)_{j \ge 0} \right\|_{\ell^r(\mathbb{N})} .$$

Besov embeddings

In order to conduct a procedure of estimation on Besov spaces, it is important to understand how they relate to each other as the values of the parameters s, π, r change. This is resumed in the following proposition:

Proposition 2.7 (*Besov Embeddings*)**.** *Let $s > 0$, $1 \le p, \pi, r \le \infty$. We have*

- $B^s_{\pi,r} \subset B^s_{p,r}$ *if* $\pi \ge p$.
- $B^s_{\pi,r} \subset B^{s-2(\frac{1}{\pi}-\frac{1}{p})}_{p,r}$ *if* $\pi < p$ *and* $s - 2(1/\pi - 1/p) > 0$.
- $B^s_{\pi,r} \subset \mathcal{C}^0(\mathbb{S}^2)$ *if* $s > 2/\pi$, *where* $\mathcal{C}^0(\mathbb{S}^2)$ *is the set of continuous functions on* $\mathbb{S}^2$.

3. Estimation procedure

We turn to the presentation of our procedure of Blind Deconvolution using Needlets, which we will denote by **BND**, and derive rates of convergence when the loss is measured in $\mathbb{L}^p$ norm, $1 \le p \le \infty$, on Besov spaces. It is natural to suggest needlets as the test functions to be used in the Galerkin projection (1.4), since they efficiently represent any function $\boldsymbol{f} \in B^s_{\pi,r}$. Unfortunately, the ensuing Galerkin matrix

$$\left(\langle \boldsymbol{K}\boldsymbol{\psi}_{j,\eta}, \boldsymbol{\psi}_{h,\alpha} \rangle \right)_{j \ge 0, \eta \in \mathcal{Z}_j, h \ge 0, \alpha \in \mathcal{Z}_h}$$

has many non-zero entries, due to the fact that the inner product $\langle \boldsymbol{K}\boldsymbol{\psi}_{j,\eta}, \boldsymbol{\psi}_{h,\alpha} \rangle$ is not necessarily null when $|j - h| \le 1$. This is a direct consequence of Proposition 1.1 and the definition of needlets. The functions $Y_{\ell,m}$ constitute a more interesting choice since the inner product $\langle \boldsymbol{Y}_\varepsilon, \boldsymbol{\psi}_{j,\eta} \rangle$ can be expressed in terms of the matrices $\boldsymbol{K}_\ell$. Indeed, Parseval's formula entails

$$\langle \boldsymbol{Y}_\varepsilon, \boldsymbol{\psi}_{j,\eta} \rangle = \sum_{\ell \in L_j} \langle \boldsymbol{K}_\ell \boldsymbol{f}_\ell + \varepsilon \dot{\boldsymbol{W}}_\ell, \boldsymbol{\psi}_{j,\eta,\ell} \rangle.$$

Before entering into the core of our procedure, we need to precise the blurring effect of $\boldsymbol{K}$. This will be realized through the introduction of a constant ν which controls the increase of the norms $\|\boldsymbol{K}_\ell^{-1}\|_{op}$.

Assumption 3.1 (*Degree of Ill-posedness*)**.** There exists $\nu \ge 0$, $Q_1, Q_2 \ge 0$ such that, for all $\ell \in \mathbb{N}$,

$$Q_1\left(\ell^\nu \vee 1\right) \le \|\boldsymbol{K}_\ell^{-1}\|_{op} \le Q_2\left(\ell^\nu \vee 1\right). \tag{3.1}$$

We note $\mathcal{K}_\nu(Q_1, Q_2)$ the set of operators satisfying this assumption, and call ν the degree of ill-posedness (DIP) of $\boldsymbol{K}$.

Assumption 3.1 actually states that even if $\boldsymbol{K}$ is continuous from $\mathbb{L}^2(\mathbb{S}^2)$ to $\mathbb{L}^2(\mathbb{S}^2)$, its inverse is not bounded and hence not computable in a satisfying way. However the weaker continuity of $\boldsymbol{K}^{-1} : \mathcal{W}^{\nu/2} \to \mathcal{W}^{-\nu/2}$ holds, where we have noted $\mathcal{W}^s = B^s_{2,2}$ the Sobolev space with parameter $s > 0$ on $\mathbb{S}^2$.

Let us now give an intuition of the main procedure in this paper. The decomposition of the inner product $\langle \boldsymbol{Kf}, \boldsymbol{\psi}_{j,\eta}\rangle$ onto every space $\mathbb{H}_\ell$, $\ell \geq 0$ via Parseval's formula, coupled with Proposition 1.1 gives

$$\langle \boldsymbol{f}, \boldsymbol{\psi}_{j,\eta}\rangle = \sum_{\ell\in L_j} \langle \boldsymbol{K}_\ell^{-1}(\boldsymbol{Kf})_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle.$$

Hence, a first natural estimator of $\langle \boldsymbol{f}, \boldsymbol{\psi}_{j,\eta}\rangle$ is

$$\widetilde{\boldsymbol{\beta}_{j,\eta}} = \sum_{\ell\in L_j} \langle \boldsymbol{K}_{\delta,\ell}^{-1}\boldsymbol{Y}_{\varepsilon,\ell}, \boldsymbol{\psi}_{j,\eta,\ell}\rangle. \tag{3.2}$$

Remark that the elements $\boldsymbol{\psi}_{j,\eta,\ell}$, $\ell \in L_j$ are easily computable thanks to the identity

$$\langle \boldsymbol{\psi}_{j,\eta}, Y_{\ell,m}\rangle = b\left(\frac{\ell}{2^j}\right)\overline{Y_{\ell,m}(\eta)} \quad \text{for all } \ell \in L_j,\ |m| \leq \ell.$$

However, the presence of noises contaminating both the signal $\boldsymbol{Kf}$ and the operator $\boldsymbol{K}$ requires an additional treatment. Schematically, a regularized version of $\boldsymbol{K}_\delta$ is plugged into (3.2), and the resulting estimator is subsequently thresholded in order to control the variance induced by the two types of noises.

3.1. Main procedure

Suppose that Assumption 3.1 holds, and define J the maximal resolution level such that

$$2^J = \lambda\left\lfloor \left(\varepsilon\sqrt{|\log\varepsilon|}\right)^{-1} \wedge \left(\delta\sqrt{|\log\delta|}\right)^{-2}\right\rfloor \tag{3.3}$$

for a positive parameter λ. Set the operator thresholding level $O_\ell(\delta)$ to

$$O_\ell(\delta) = \kappa\sqrt{2\ell+1}\,\delta\sqrt{|\log\delta|} \tag{3.4}$$

where $\kappa > 0$. For $j \in \mathbb{N}$, let

$$\ell_j = \min\{\ell \in L_j,\ \|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op} \leq O_\ell(\delta)^{-1}\}$$

(with the convention $\min\emptyset = +\infty$), and, for positive constants τ_{sig}, τ_{op}, define the signal thresholding level

$$S_j(\delta,\varepsilon) = \begin{cases} \|\boldsymbol{K}_{\delta,\ell_j}^{-1}\|_{op}\left(\tau_{sig}\varepsilon\sqrt{|\log\varepsilon|} \vee \tau_{op}2^{-j/2}\delta\sqrt{|\log\delta|}\right) & \text{if } \ell_j < \infty \\ +\infty \quad \text{if } \ell_j = +\infty. \end{cases} \tag{3.5}$$

Define now the estimator $\widehat{\boldsymbol{\beta}}_{j,\eta}$ similarly to $\widetilde{\boldsymbol{\beta}_{j,\eta}}$ but where $\boldsymbol{K}_\delta^{-1}$ is replaced by its thresholded version

$$\widehat{\boldsymbol{\beta}}_{j,\eta} = \sum_{\ell=2^{j-1}}^{2^{j+1}} \langle \boldsymbol{K}_{\delta,\ell}^{-1}\mathbf{1}_{\left\{\|\boldsymbol{K}_{\delta,\ell}^{-1}\|\leq O_\ell(\delta)^{-1}\right\}}\boldsymbol{Y}_{\varepsilon,\ell}, \boldsymbol{\psi}_{j,\eta,\ell}\rangle.$$

The final estimator $\widetilde{\boldsymbol{f}}$ of $\boldsymbol{f}$ is

$$\widetilde{\boldsymbol{f}} = \sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j} \widehat{\boldsymbol{\beta}}_{j,\eta}\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>S_j(\delta,\varepsilon)\}}\boldsymbol{\psi}_{j,\eta}.$$

For the sake of brevity, we shall denote this procedure as **BND** (for Blind Deconvolution using Needlets). Before establishing the convergence rates of this algorithm in a minimax framework, let us give some enlightenments about the shape of the thresholding levels. Usual thresholds [15,18] involve an upper bound on the variance of the coefficients $\widehat{\boldsymbol{\beta}}_{j,\eta}$, and thereby the knowledge of the constant ν. The term $\|\boldsymbol{K}_{\delta,\ell_j}^{-1}\|_{op}$ in (3.5) is meant to replace the more often used upper bound $2^{j\nu}$. Indeed, Lemmas 4.1 and 4.2 show that with high probability these two quantities coincide up to a multiplicative constant. This trick endows the procedure with adaptivity with regard to the parameters s, π, r and Q_1, Q_2, ν, and subsequently means that no a priori knowledge on $\boldsymbol{f}$ nor $\boldsymbol{K}$ is required in order to set it up.

Lemma 4.2 however heavily relies on the non negativity of Q_1. As a matter of fact, the rates of convergence derived below fall apart when the latter is null. In that case, a preliminary knowledge of ν is essential, as well as the subsequent following adaptations: the signal level $S_j(\delta,\varepsilon)$ is changed to

$$\widetilde{S_j}(\delta,\varepsilon) = 2^{j\nu}\left(\tau_{sig}\varepsilon\sqrt{|\log\varepsilon|} \vee \tau_{op}2^{-j/2}\delta\sqrt{|\log\delta|}\right)$$

and the new maximal level of resolution $\widetilde{J}$ satisfies

$$2^{\widetilde{J}} = \lambda \lfloor \left(\varepsilon\sqrt{|\log\varepsilon|}\right)^{\frac{-1}{\nu+1}} \wedge \left(\delta\sqrt{|\log\delta|}\right)^{\frac{-1}{\nu+1/2}} \rfloor.$$

The tuning down of the maximal level is not essential, yet it permits to avoid unnecessary calculations. As a matter of fact, computations of the needlet coefficients are quite heavy, due to the absence of a simplifying algorithm (such as the pyramidal algorithm for wavelets, see [20]). Thus, the computation of a single needlet coefficient $\langle\gamma, \boldsymbol{\psi}_{j,\eta,\ell}\rangle$, which number grows exponentially as the resolution level increases, requires the determination of $2.2^j - 1$ inner products $\langle\gamma_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle$.

Let us now turn to the convergence rates of the procedure in a minimax framework, when the loss is measured in $\mathbb{L}^p$ norm ($1 \leq p \leq \infty$) and $\boldsymbol{f}$ belongs to a Besov body.

Theorem 3.2. *Let $\pi \geq 1$, $s > 2/\pi$, $r \geq 1$ and $M > 0$. Let $\nu \geq 0$, let $Q_1 \geq Q_2 > 0$. Then for sufficiently large κ, τ_{sig}, τ_{op}, for all $p \in [1, +\infty[$,*

$$\sup_{\boldsymbol{f}\in B^s_{\pi,r}(M),\, \boldsymbol{K}\in\mathcal{K}_\nu(Q_1,Q_2)} \mathbb{E}\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_p^p \lesssim \mathcal{R}_p(\delta, 2, 1) \vee \mathcal{R}_p(\varepsilon, 2, 2)$$

where $\lesssim$ means inequality up to a multiplicative constant depending only on $p, s, \pi, r, M, \nu, Q_1, Q_2, \lambda, \kappa, \tau_{sig}$ and τ_{op}. The convergence rates $\mathcal{R}_p(x, d, d')$ are defined, for all $x > 0$, $1 \leq p < \infty$ and $d, d' \in \mathbb{N}$ by

$$\mathcal{R}_{x,p}(d, d') = (|\log x|)^{p-1}(x\sqrt{|\log x|})^{p\mu(d,d')}$$

where we noted

$$\mu(d, d') = \begin{cases} \dfrac{s}{s+\nu+\frac{d'}{2}} & \text{if } s > (\nu + d'/2)(p/\pi - 1) \\ & \text{or } s = (\nu + d'/2)(p/\pi - 1) \text{ and } r \leq \pi \\ \dfrac{s - d/\pi + d/p}{s - d/\pi + \nu + d'/2} & \text{if } \dfrac{d}{\pi} < s < (\nu + d'/2)(p/\pi - 1). \end{cases}$$

Theorem 3.3. *Under the same hypotheses as in Theorem 3.2,*

$$\sup_{\boldsymbol{f}\in B^s_{\pi,r}(M),\, \boldsymbol{K}\in\mathcal{K}_\nu(Q_1,Q_2)} \mathbb{E}\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_\infty \lesssim \mathcal{R}_\infty(\delta, 2, 1) \vee \mathcal{R}_\infty(\varepsilon, 2, 2) \tag{3.6}$$

where

$$\mathcal{R}_\infty(x, d, d') = \sqrt{|\log x|}(x\sqrt{|\log x|})^{\mu'(d,d')}$$

and

$$\mu'(d, d') = \frac{s - d/\pi}{s - d/\pi + \nu + d'/2}.$$

The speeds of convergence exhibit an explicit interplay between the noise levels δ and ε, including the possible case where $\delta \gg \varepsilon$. The convergence rates obtained when $\delta = 0$ and $\varepsilon = 0$ are very similar, except that the problem $\varepsilon = 0$ reveals rates corresponding to a problem in dimension 1. This indicates that the denoising of $\boldsymbol{K}_\delta$ results in the same rates as the denoising of $\boldsymbol{Y}_\varepsilon$, but with a dimension parameter given by the size of the blocks appearing in the blockwise SVD (that is to say the integer d' such that $\dim \mathbb{H}_\ell \sim \ell^{d'}$).

If $\delta = 0$, the rates coincide with the results of Kerkyacharian et al. [18] (actually, the algorithms themselves are nearly identical), which can be proved to be optimal in a minimax sense (up to a logarithmic factor, see Willer [30] for a sketch of proof). The optimality when $\delta \gg \varepsilon$ is not yet established, and we will not address it in the present paper. The two regions $s \geq (\nu + d/2)(p/\pi - 1)$ and $s < (\nu + d/2)(p/\pi - 1)$ are classic in non-parametric estimation in inverse problems, and respectively referred to as the regular case and the sparse case.

Although we chose to work in a white noise model for the convenience of calculations, the algorithm and ensuing results should be easily adaptable to the density estimation framework mentioned in Section 1.1, in which one observes direct realizations $(\boldsymbol{\theta}_1 X_1, \ldots, \boldsymbol{\theta}_n X_n)$ of $\boldsymbol{\theta}X$ and a noisy version $\boldsymbol{K}_\delta$ of $\boldsymbol{K}$.

3.1.1. Adaptation to other dimensions and comparison with existing works

A close inspection of the proofs of Theorems 3.2 and 3.3 shows that the presence of a blockwise SVD decomposition, combined with properties of the ensuing needlet frame similar to those in Section 2, ensures the applicability of the scheme with adapted convergence rates.

This includes in particular the corresponding 1-dimensional problem, equivalent to deconvolution in a periodic setting [8]. Here Meyer's periodized wavelets can endorse the role of needlets in the present setting, since they are compactly supported in the frequential domain as well. In the Fourier basis on $[0, 1]$, $\boldsymbol{K}$ is directly diagonalized, which corresponds to a blockwise SVD with dimensionality $d' = 0$. The adapted algorithm henceforth reaches the rates $\mathcal{R}_p(\delta, 1, 0) \vee \mathcal{R}_p(\varepsilon, 1, 1)$, $1 \le p \le \infty$. It outperforms the one developed in Hoffmann and Reiß [15] which corresponds formally to $\mathcal{R}_2(\delta, 1, 1) \vee \mathcal{R}_2(\varepsilon, 1, 1)$. The reason to it is that a Galerkin projection on wavelets is agnostic to the blockwise structure of $\boldsymbol{K}$. Moreover, our procedure widens the possible range of considered $\mathbb{L}^p$ losses.

In image processing, a signal $\boldsymbol{f} \in \mathbb{L}^2([0, 1]^2)$ is observed through its convolution with a function $\boldsymbol{k} \in \mathbb{L}^2([0, 1]^2)$ called the Point Spread Function of the measuring device, which requires to be estimated in first instance (see [25,2]). A careful adaptation of the main results in this paper allows to treat this problem as well (for the definition of needlets on $[0, 1]^2$, see [16]). Another relevant example concerns the operators defined on $\mathbb{S}^d$, $d \ge 1$ via

$$\boldsymbol{K}f(\xi) = \int_{\mathbb{S}^d} \varphi(\langle \xi, \omega \rangle) f(\omega) d\omega$$

where φ is a bounded integrable function on $[-1, 1]$. In this case, as shown by the Funk–Hecke theorem (see [12]), spherical harmonics realize a SVD of $\boldsymbol{K}$. On the other hand, the construction of needlets generalizes naturally to $\mathbb{S}^d$ (see [24]), and the rates derived hence change to $\mathcal{R}_p(\delta, d, 0) \vee \mathcal{R}_p(\varepsilon, d, d)$, $1 \le p \le \infty$.

3.1.2. Comparison with the blockwise SVD algorithm of Delattre et al. [6]

In this section we present the blockwise SVD algorithm **BBD** (for Blind Blockwise Deconvolution) depicted in Delattre et al. [6], and compare it to **BND**. **BBD** also relies on the blockwise SVD property in Proposition 1.1, but tackles the thresholding of the signal and the operator differently. Namely, define the maximal resolution level

$$L \sim \lfloor \left(\varepsilon\sqrt{|\log \varepsilon|}\right)^{-1} \wedge \left(\delta\sqrt{|\log \delta|}\right)^{-2} \rfloor$$

and the signal thresholding level $\mathcal{E}_\ell(\varepsilon) = \tau\sqrt{2\ell + 1}\varepsilon\sqrt{|\log \varepsilon|}$, $\tau > 0$. The estimator $\widetilde{f}$ provided by the **BBD** algorithm is

$$\boldsymbol{f} = \sum_{\ell \le L} \boldsymbol{K}_{\delta,\ell}^{-1} \mathbf{1}_{\left\{\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op} < O_\ell(\delta)^{-1}\right\}} \boldsymbol{Y}_{\varepsilon,\ell} \mathbf{1}_{\left\{\|\boldsymbol{Y}_{\varepsilon,\ell}\| > \mathcal{E}_\ell(\varepsilon)\right\}}. \tag{3.7}$$

It is quickly seen that **BBD** is still adaptive with respect to $\boldsymbol{f}$ and $\boldsymbol{K}$. This estimator is however fundamentally different from ours: first, the signal regularization is directly performed on the observed signal $\boldsymbol{Y}_\varepsilon$ rather than on $\boldsymbol{K}_\delta^{-1}\boldsymbol{Y}_\varepsilon$. Secondly, this regularization step is anterior to the inversion of $\boldsymbol{K}_\delta$, whereas in **BND** the signal is first inverted and then thresholded. These differences imply differences as well in the setting in which **BBD** will perform well. Namely, Delattre et al. [6] require that the two following inequalities hold:

$$\|\boldsymbol{K}_\ell\|_{op} \le R_1 \ell^{-\nu} \quad \text{and} \quad \|\boldsymbol{K}_\ell^{-1}\|_{op} \le R_2 \ell^{\nu} \tag{3.8}$$

for $R_1, R_2 > 0$. Let us note $\mathcal{G}(R_1, R_2)$ the set of operators satisfying (3.8). (3.8) unilaterally entails Assumption 3.1 which means that **BND** applies in the context of **BBD** but the reverse is false. As a matter of fact, (3.8) restricts the scope of application of **BBD** to quasi diagonal operators. The setting of **BND** is much more generic. Finally, due to the shape of the threshold performed on each $\boldsymbol{Y}_{\varepsilon,\ell}$, **BBD** performs well only when the loss in measured in quadratic risk, and when $\boldsymbol{f}$ belongs to a Sobolev space (which corresponds to $\boldsymbol{f} \in B^s_{2,2}$).

The counterpart to such restrictive assumptions is the remarkably fast rates of convergence it attains in the case $\varepsilon = 0$. Indeed, it can be proved that, for $s, M > 0$, $R_1, R_2 > 0$ and $\nu \ge 0$,

$$\sup_{\substack{\boldsymbol{f} \in \mathcal{W}^s(M) \\ \boldsymbol{K} \in \mathcal{G}(R_1,R_2)}} \mathbb{E}\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_2 \lesssim \left(\delta\sqrt{|\log \delta|}\right)^{1 \wedge \frac{2s}{2\nu+1}} \vee \left(\varepsilon\sqrt{|\log \varepsilon|}\right)^{\frac{2s}{2s+2\nu+1}}. \tag{3.9}$$

This clearly outperforms **BND** in the case $\pi = r = 2$, and $p = 2$.

3.2. Practical study

We present the practical numerical performances of **BND** and compare it to the Blind Blockwise Deconvolution algorithm (**BBD**) of Delattre et al. [6]. The sets of cubature points in the simulations that follow have been taken from the web site of R. Womersley http://web.maths.unsw.edu.au/~rsw. We proceed with the following choices of parameters:

Data: the target density $\boldsymbol{f}$ is given by

$$\boldsymbol{f}(\omega) = \exp(-2 * \|\omega - \omega_1\|_{\ell^1(\mathbb{R}^3)})/c$$

with $\omega_1 = (0, 1, 0)$ and $c = 0.6729$. Concerning the operator $\boldsymbol{K}$, we choose it among the class of Rosenthal laws on SO(3). These distributions find their origins in random walks on groups (see [26]). $\boldsymbol{K}$ is said to follow a Rosenthal distribution of parameters $\alpha \in]0; \pi]$ and $\nu > 0$ on SO(3) if, for $\ell \ge 0$, $|m| \le \ell$, we have

$$\boldsymbol{K}_{\ell,m,n} = \left(\frac{\sin((\ell + 1/2)\alpha)}{(2\ell + 1)\sin(\alpha/2)}\right)^{\nu} \mathbf{1}_{\{m=n\}}.$$

A Rosenthal law hence provides a concrete example of operator with DIP $\nu \ge 0$. We will take $\alpha = \pi$ and $\nu = 1$.

Table 1
Choice of κ. Nr_{op} is the average number, computed on the basis of $N = 10$ realizations, of levels $\ell \leq 10$ such that $\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op} \leq O_\ell(\delta)^{-1}$. We have $\delta = 10^{-3}$.

κ	0.3	0.4	0.5	0.6	0.7	0.8
Nr_{op}	10	9	9	8	2	0

Table 2
Choice of τ. For $(\delta_{sig}, \varepsilon_{sig}) = (\varepsilon_{op}, \delta_{op}) = (10^{-4}, 10^{-3})$ and each value of τ, we computed 10 times the described procedure and reported the average number of remaining needlet coefficients at level j.

	τ_{sig}					τ_{op}	
	0.5	0.6	0.7	0.8	0.9	0.1	0.2
$j = 0$	3	0	3	0	0	0	0
$j = 1$	10	6	0	0	0	0	0
$j = 2$	20	9	2	1	0	4	0
$j = 3$	94	22	8	4	0	127	0

Table 3
Average normalized $\mathbb{L}^2$ and $\mathbb{L}^\infty$ loss of **BBD** and **BND**.

δ	ε	$E\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_2$		$E\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_\infty$	
		BBD	**BND**	**BBD**	**BND**
3.10^{-3}	10^{-3}	0.2210	0.1695	0.3867	0.3464
	10^{-4}	0.1013	0.1603	0.2146	0.3374
10^{-3}	10^{-3}	0.2195	0.1242	0.3870	0.2204
	10^{-4}	0.0839	0.0594	0.1931	0.1569
10^{-4}	10^{-3}	0.2194	0.1267	0.3863	0.2257
	10^{-4}	0.0825	0.0584	0.1924	0.1571

Tuning parameters: we set $\lambda = 1$ in (3.3). The concrete choice of adequate thresholding constants κ and τ is a complex issue. Our practical choices will be based on the following remark, inspired from Donoho and Johnstone [7]: in the case of direct estimation on real line, the universal threshold which is both efficient and simple to implement, takes the form $2\sqrt{|\log \varepsilon|}$. A consistent interpretation is to consider that this threshold should kill any pure noise signal. We will adapt this reasoning to the case of interest.

Choice of κ: we use as a benchmark the case where $\boldsymbol{K}_\ell$ is the null matrix of $M_{2\ell+1}(\mathbb{R})$ for $\ell \geq 1$ (this corresponds to the case where the law of θ is uniform over SO(3)). Given δ large enough, the smallest value κ_δ such that

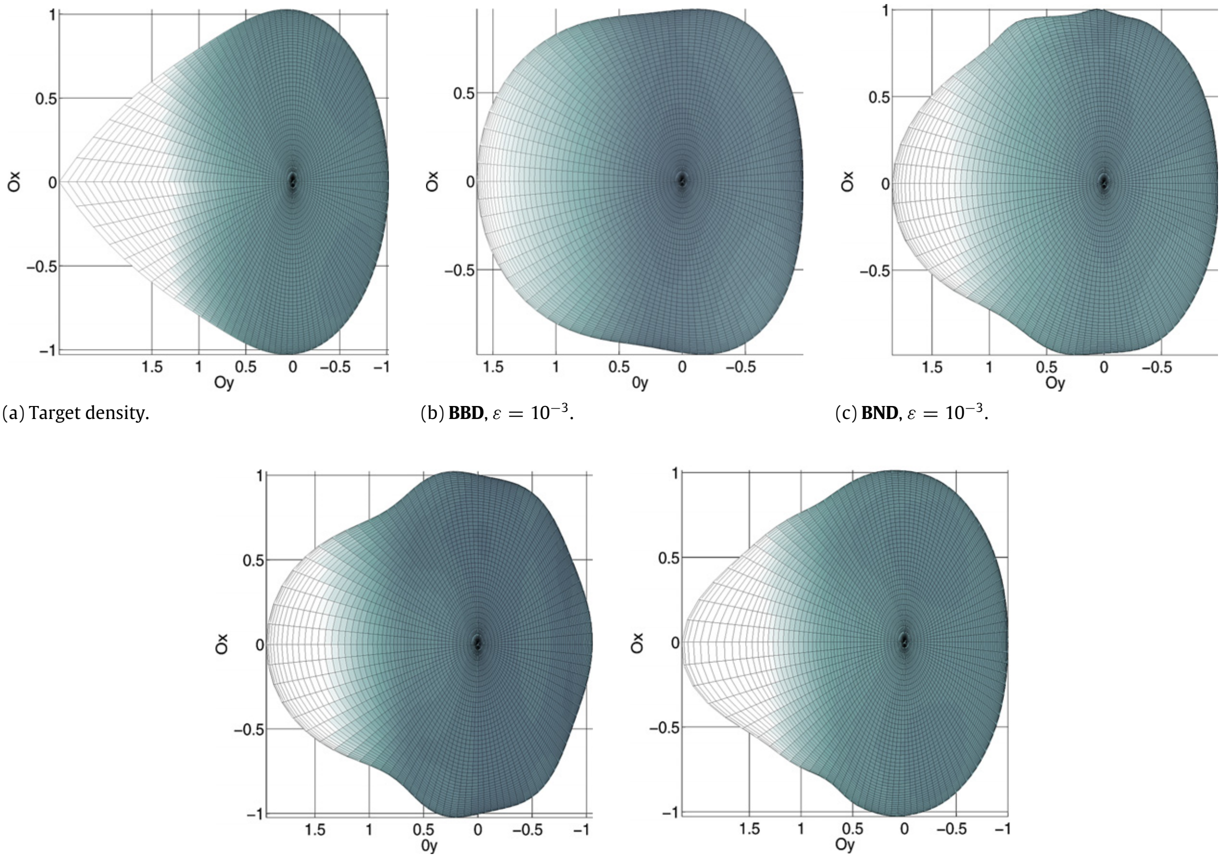
$$\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op} > O_\ell(\delta)^{-1} \quad \text{for all } \ell \leq 10$$

is retained. The results are reported in Table 1 and give $\kappa = 0.8$.

Choice of τ_{sig} *and* τ_{op}: It is clear that the role of τ_{sig} (resp. τ_{op}) is to control the influence of the signal (resp. the operator) error. In order to compute τ_{sig} (resp. τ_{op}), we therefore work with noise levels $\varepsilon_{sig} > \delta_{sig} > 0$ (resp. $\delta_{op} > \varepsilon_{op} > 0$) large enough. We make use of the uniform density $\boldsymbol{u}$ on $\mathbb{S}^2$, satisfying $\boldsymbol{u}_\ell = 0$ for $\ell \geq 1$ as a benchmark. Henceforth $\langle \boldsymbol{u}, \boldsymbol{\psi}_{j,\eta}\rangle = 0$ for $j \geq 0$, $\eta \in \mathcal{Z}_j$, which means that the observations $\langle \boldsymbol{Y}_{\varepsilon_{sig}}, \boldsymbol{\psi}_{j,\eta}\rangle, j \geq 0$ are pure noises. Taking advantage of this remark, we simulate $\boldsymbol{K}_{\delta_{sig}}$ and, integrating the precedently computed value of κ, apply the procedure for increasing values of τ_{sig} (resp. τ_{op}) until all the computed coefficients $\langle \widetilde{\boldsymbol{u}}, \boldsymbol{\psi}_{j,\eta}\rangle$ are killed for $j \leq 3$. The results are reported in Table 2 and give $\tau_{sig} = 0.9$, $\tau_{op} = 0.2$.

We compare the performances of **BBD** (with parameters $\kappa = 0.8$ and $\tau = 1$ taken from [6]) and **BND** for $\delta \in \{3.10^{-3}, 10^{-3}, 10^{-4}\}$, $\varepsilon \in \{10^{-3}, 10^{-4}\}$. The (normalized) mean squared error and the supremum norm error are computed with a Monte Carlo method based on $N = 200$ simulations. Each loss is approximated by its discrete equivalent calculated on a uniform grid of 4096 points on $\mathbb{S}^2$ at each step. Results are reported in Table 3. They clearly illustrate the rates of convergence derived in Theorems 3.2 and 3.3 in that the loss in always higher when the signal noise level ε is predominant. Besides, it also confirms the relationship between the rates in Theorems 3.2, 3.3 and (3.9). Indeed, since $\boldsymbol{K}$ also verifies (3.8), the rates (3.9) are available and Table 3 exposes the outperforming of **BND** over **BBD** in every situation except when the operator noise is highly predominant (corresponding to $(\delta, \varepsilon) = (3.10^{-3}, 10^{-4})$).

For particular realizations of $\boldsymbol{Y}_\varepsilon$ and $\boldsymbol{K}_\delta$, we plot in Fig. 2: the original shape of the density, and the results of the different algorithms in the form of spherical views seen 'from above'. The figures emphasize the better adaptivity of **BND** to the 'spiky' shape of the target density.



(a) Target density. (b) **BBD**, $\varepsilon = 10^{-3}$. (c) **BND**, $\varepsilon = 10^{-3}$.

(d) **BBD**, $\varepsilon = 10^{-4}$. (e) **BND**, $\varepsilon = 10^{-4}$.

Fig. 2. Spherical view from above of the results of the two algorithms with noise level $\delta = 10^{-3}$.

4. Proof of Theorems 3.2 and 3.3

Preliminary lemmas

The establishment of the convergence rates in Theorems 3.2 and 3.3 requires the control of the tails of the variables $|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|$. This will be the subject of Lemma 4.3. Two results upstream of this lemma involve the control of the tails of the variables $\|\dot{\boldsymbol{B}}_\ell\|_{op}$ (Lemma 4.1), as well as a control of $\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op}$ on a particular set (Lemma 4.2). We will not perform the proofs of the two latter results, but will provide references where they are conducted.

Lemma 4.1 (*Davidson and Szarek [5], Theorem 2.4*)**.** *There exist two constants $\beta_0, c_0 \geq 0$ independent from $\ell \in \mathbb{N}$ such that*

$$\forall t \geq \beta_0, \quad \mathbb{P}((2\ell+1)^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{op} > t) \leq \exp(-c_0 t(2\ell+1)^2).$$

A simple corollary is the following upper bound on the moments of $\|\dot{\boldsymbol{B}}_\ell\|_{op}$

$$\mathbb{E}[\|\dot{\boldsymbol{B}}_\ell\|_{op}^p] \lesssim \ell^{p/2}.$$

Lemma 4.2 (*Delattre et al. [6], Proof of Theorem 3.1*)**.** *We introduce further the events $\mathcal{A}_\ell = \{\|(\boldsymbol{K}_{\delta,\ell})^{-1}\|_{op} \leq O_\ell(\delta)^{-1}\}$ and $\mathcal{B}_\ell = \{\|\delta\dot{\boldsymbol{B}}_\ell\|_{op} \leq a_\ell\}$ with $a_\ell = \rho O_\ell(\delta)$ for some $0 < \rho < \frac{1}{2}$. On $\mathcal{A}_\ell \cap \mathcal{B}_\ell$, we have*

$$\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op} \leq \frac{\rho}{1-\rho}\|\boldsymbol{K}_\ell^{-1}\|_{op}$$

$$\textit{and} \quad \|\boldsymbol{K}_\ell^{-1}\|_{op} \leq (1-\rho)^{-1}\|\boldsymbol{K}_{\delta,\ell}^{-1}\|_{op}.$$

Lemma 4.3. *Let $\overline{S}_j(\delta,\varepsilon) = \tau 2^{j\nu}\big(\varepsilon\sqrt{|\log\varepsilon|} \vee 2^{-j/2}\delta\sqrt{|\log\delta|}\big)$ with $\tau = \tau_{siq} \vee \tau_{op}$. In the setting of Theorem 3.2, for all $j \le J$, $\eta \in \mathcal{Z}_j$, for all $p \ge 1$*

$$\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}| > \overline{S}_j(\delta,\varepsilon)) \lesssim \varepsilon^{\kappa^2} \vee \delta^{\kappa^2} \tag{4.1}$$

$$\mathbb{E}[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p] \lesssim (\varepsilon 2^{j\nu})^p \vee (\delta 2^{j(\nu-1/2)})^p \vee |\boldsymbol{\beta}_{j,\eta}|^p \mathbf{1}_{\{j\ge j_0\}} \tag{4.2}$$

$$\mathbb{E}[\sup_{\eta\in\mathcal{Z}_j} |\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p] \lesssim (j+1)^p\Big[(\varepsilon 2^{j\nu})^p \vee (\delta 2^{j(\nu-1/2)})^p\Big] \vee |\boldsymbol{\beta}_{j,\eta}|^p \mathbf{1}_{\{j\ge j_0\}} \tag{4.3}$$

where $2^{j_0} \gtrsim \big(\delta\sqrt{|\log\delta|}\big)^{-\frac{2}{2\nu+1}}$.

Proof (*Proof of* Lemma 4.3)**.** All inequalities can be derived from the study of $\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}| > t)$ in each case. Resorting to the identity

$$\boldsymbol{K}_{\delta,\ell}^{-1}(\boldsymbol{K}_\ell \boldsymbol{f}_\ell + \varepsilon \dot{\boldsymbol{W}}_\ell) - \boldsymbol{f}_\ell = -\delta \boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell + \boldsymbol{K}_{\delta,\ell}^{-1}\varepsilon \dot{\boldsymbol{W}}_\ell \tag{4.4}$$

which holds for every $\ell \in \mathbb{N}$, and using Parseval's formula, we decompose $\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta} = I + II + II$ where

$$I = \sum_{\ell\in L_j} \langle -\delta \boldsymbol{K}_{\delta,\ell}^{-1} \mathbf{1}_{\mathcal{A}_\ell} \dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle$$

$$II = \sum_{\ell\in L_j} \langle \boldsymbol{K}_{\delta,\ell}^{-1} \mathbf{1}_{\mathcal{A}_\ell} \varepsilon \dot{\boldsymbol{W}}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle$$

$$III = -\sum_{\ell\in L_j} \langle \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle \mathbf{1}_{\mathcal{A}_\ell^c}.$$

We now have to study the deviation bounds of I, II, III. Term I can be decomposed as $I = IV + V$ where

$$IV = -\sum_{\ell\in \boldsymbol{L}_j} \langle \delta \boldsymbol{K}_{\delta,\ell}^{-1} \dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle \mathbf{1}_{\mathcal{A}_\ell} \mathbf{1}_{\mathcal{B}_\ell}$$

$$V = -\sum_{\ell\in \boldsymbol{L}_j} \langle \delta \boldsymbol{K}_{\delta,\ell}^{-1} \dot{\boldsymbol{B}}_\ell \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle \mathbf{1}_{\mathcal{A}_\ell} \mathbf{1}_{\mathcal{B}_\ell^c}.$$

In order to treat Term IV, we introduce the operator

$$\mathbf{Q}_j = \sum_{\ell\in L_j} \boldsymbol{K}_{\delta,\ell}^{-1} \mathbf{1}_{\mathcal{A}_\ell} \mathbf{1}_{\mathcal{B}_\ell} \dot{\boldsymbol{B}}_\ell$$

defined for $j \le J$. Since $\boldsymbol{K}$ and $\dot{\boldsymbol{B}}$ are both stable with regard to every space $\mathbb{H}_\ell$, and since $\langle \boldsymbol{\psi}_{j,\eta}, \boldsymbol{\psi}_{h,\alpha}\rangle = 0$ if $|j-h| > 1$, we have

$$IV = -\langle \delta \mathbf{Q}_j f, \boldsymbol{\psi}_{j,\eta}\rangle = -\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} \langle \delta \mathbf{Q}_j \boldsymbol{\psi}_{h,\alpha}, \boldsymbol{\psi}_{j,\eta}\rangle \langle f, \boldsymbol{\psi}_{h,\alpha}\rangle.$$

Henceforth

$$\begin{aligned} |IV| \le{} & \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle \delta \mathbf{Q}_j \boldsymbol{\psi}_{h,\alpha}, \boldsymbol{\psi}_{j,\eta}\rangle|^{\pi'}\right)^{\frac{1}{\pi'}} \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle f, \boldsymbol{\psi}_{h,\alpha}\rangle|^{\pi}\right)^{\frac{1}{\pi}} \mathbf{1}_{\{\pi\le 2\}} \\ & + \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle \delta \mathbf{Q}_j \boldsymbol{\psi}_{h,\alpha}, \boldsymbol{\psi}_{j,\eta}\rangle|^{\pi}\right)^{\frac{1}{\pi}} \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle f, \boldsymbol{\psi}_{h,\alpha}\rangle|^{\pi'}\right)^{\frac{1}{\pi'}} \mathbf{1}_{\{\pi> 2\}} \end{aligned}$$

where we used Hölder's inequality with $\pi^{-1} + (\pi')^{-1} = 1$. Now, if $\pi \le 2$, then $\pi' \ge 2$ and (2.7) together with Proposition 2.4 entail

$$\begin{aligned} \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle \delta \mathbf{Q}_j \boldsymbol{\psi}_{h,\alpha}, \boldsymbol{\psi}_{j,\eta}\rangle|^{\pi'}\right)^{\frac{1}{\pi'}} &\le \left(\sum_{h=j-1}^{j+1} \sum_{\alpha\in\mathcal{Z}_h} |\langle \delta \mathbf{Q}_j \boldsymbol{\psi}_{h,\alpha}, \boldsymbol{\psi}_{j,\eta}\rangle|^{2}\right)^{\frac{1}{2}} \\ &\le \|\delta^T \mathbf{Q}_j \boldsymbol{\psi}_{j,\eta}\| \\ &\lesssim \delta 2^{j(\nu+1/2)}. \end{aligned}$$

Moreover, since $f \in B^s_{\pi,r}$, we have

$$\left(\sum_{h=j-1}^{j+1}\sum_{\alpha\in\mathcal{Z}_h}|\langle f,\boldsymbol{\psi}_{h,\alpha}\rangle|^\pi\right)^{\frac{1}{\pi}} \lesssim 2^{-j(s-\frac{2}{\pi}+1)}.$$

If $\pi > 2$, a similar argument added with the Besov embedding $B^s_{\pi,r} \subset B^{s-2(1/\pi-1/\pi')}_{\pi',r}$ leads to the same bound. Finally,

$$\begin{aligned}\mathbb{P}(|IV| > t) &\le \mathbb{P}\big(\|\delta^T\boldsymbol{Q}_j\|_{op}2^{-j(s-\frac{2}{\pi}+1)} \gtrsim t\big)\\ &\le \mathbb{P}(2^{-j/2}\|\widetilde{\boldsymbol{P}}_{\ell_j}\dot{\boldsymbol{B}}\widetilde{\boldsymbol{P}}_{\ell_j}\|_{op} \gtrsim t\delta^{-1}2^{j(\nu-1/2-(s-2/\pi))})\\ &\le \exp\left(-\frac{c_0t^22^{2j}}{22^{2j(\nu-\frac{1}{2})}}\right)\mathbf{1}_{\left\{t\gtrsim\beta_02^{j(\nu-\frac{1}{2})}\right\}}\end{aligned} \tag{4.5}$$

where we noted $\widetilde{\boldsymbol{P}}_{\ell_j}$ the orthogonal projector onto $\bigoplus_{\ell\in L_j}\mathbb{H}_\ell$ and used Lemmas 4.1 and 4.2 together with the fact that $s > 2/\pi$. Turning to Term V, a direct application of Lemma 4.1 entails

$$\mathbb{P}(\|\delta\dot{\boldsymbol{B}}_\ell\|_{op} > \boldsymbol{a}_\ell) \le \delta^{c_0\rho^2(2\ell+1)^2\kappa^2}. \tag{4.6}$$

So we have

$$\begin{aligned}\mathbb{P}(|V| > t) &\le \mathbb{P}\left(\sum_{\ell\in\boldsymbol{L}_j}\delta\|\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{B}}_\ell\|_{op}\|\boldsymbol{f}_\ell\|\|\boldsymbol{\psi}_{j,\eta,\ell}\|\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell^c} > t\right)\\ &\le \sum_{\ell\in\boldsymbol{L}_j}\mathbb{P}(\|\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{B}}_\ell\|_{op}\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell^c} > t)\\ &\lesssim \sum_{\ell\in\boldsymbol{L}_j}\mathbb{P}(\|\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{B}}_\ell\|_{op}\mathbf{1}_{\mathcal{A}_\ell} > t)^{1/2}\mathbb{P}(\|\delta\dot{\boldsymbol{B}}_\ell\|_{op} > \boldsymbol{a}_\ell)^{1/2}\\ &\lesssim \sum_{\ell\in\boldsymbol{L}_j}\mathbb{P}((2\ell+1)^{-1/2}\|\dot{\boldsymbol{B}}_\ell\|_{op} > t\kappa\log^{1/2}\delta)^{1/2}\delta^{c_0\rho^2(2\ell+1)^2\kappa^2/2}\\ &\lesssim \delta^{c_0\rho^22^{2j}\kappa^2/2}\sum_{\ell\in\boldsymbol{L}_j}\exp(-c_0(2\ell+1)^2t^2\kappa^2\log\delta/2)\\ &\lesssim \delta^{c_0\rho^22^{2j}\kappa^2/2}\exp(-c_02^{2j}t^2\kappa^2\log\delta/2).\end{aligned}$$

Turning to Term II, we decompose in a similar fashion $II = VI + VII$ where

$$VI = \sum_{\ell\in\boldsymbol{L}_j}\langle\varepsilon\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{W}}_\ell,\boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell}$$

$$VII = \sum_{\ell\in\boldsymbol{L}_j}\langle\varepsilon\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{W}}_\ell,\boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell^c}.$$

Conditioning on $(\dot{\boldsymbol{B}}_\ell)_{\ell\in L_j}$ and applying Lemma 4.2, we derive for all $t > 0$,

$$\begin{aligned}\mathbb{P}(|VI| > t) &= \mathbb{P}\left(\left|\sum_{\ell\in\boldsymbol{L}_j}\langle\varepsilon\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{W}}_\ell,\boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell}\right| > t\right)\\ &\le \exp\left(-\frac{t^2}{2\varepsilon^22^{2j\nu}}\right).\end{aligned} \tag{4.7}$$

As for Term VII, employing the Cauchy–Schwarz inequality, (4.6), and conditioning on $(\dot{\boldsymbol{B}}_\ell)_{\ell\in L_j}$, we write

$$\begin{aligned}\mathbb{P}(|VII| > t) &= \mathbb{P}\left(\left|\sum_{\ell\in\boldsymbol{L}_j}\langle\varepsilon\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{W}}_\ell,\boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell^c}\right| > t\right)\\ &\le \sum_{\ell\in\boldsymbol{L}_j}\mathbb{P}(|\langle\varepsilon\boldsymbol{K}_{\delta,\ell}^{-1}\dot{\boldsymbol{W}}_\ell,\boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\mathcal{A}_\ell}\mathbf{1}_{\mathcal{B}_\ell^c}| > t)\end{aligned}$$

$$\le \sum_{\ell\in \boldsymbol{L}_j} \mathbb{P}(|\langle \varepsilon \boldsymbol{K}_{\delta,\ell}^{-1}\mathbf{1}_{\mathcal{A}_\ell}\dot{\boldsymbol{W}}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle| > t)^{1/2}\mathbb{P}(\delta\|\dot{\boldsymbol{B}}_\ell\|_{op} > \boldsymbol{a}_\ell)^{1/2}$$

$$\lesssim \exp\left(-\frac{t^2\delta^2|\log\delta|2^j}{4\varepsilon^2}\right)\delta^{c_0\rho^2 2^{2j}\kappa^2/2}.$$

It remains to treat Term III. We claim that

$$\mathcal{A}_\ell^c \subset \left\{\|\delta\dot{\boldsymbol{B}}_\ell\| \ge O_\ell(\delta)\right\}\bigcup\left\{\|\boldsymbol{K}_\ell^{-1}\|_{op} \ge O_\ell(\delta)^{-1}/2\right\} \tag{4.8}$$

(for a proof, we refer to Delattre et al. [6]). Hence, $III \le VIII + IX$ where

$$VIII = \left|\sum_{\ell\in L_j}\langle \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle\right|\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|\ge O_\ell(\delta)\}}$$

$$IX = \left|\sum_{\ell\in L_j}\langle \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle\right|\mathbf{1}_{\left\{\|\boldsymbol{K}_\ell^{-1}\|_{op}\ge O_\ell(\delta)^{-1}/2\right\}}.$$

As

$$\left\{\|\boldsymbol{K}_\ell^{-1}\|_{op} > O_\ell(\delta)^{-1}/2\right\} \subset \left\{\ell > c(\delta\sqrt{|\log\delta|})^{-\frac{1}{\nu+1/2}}\right\} \tag{4.9}$$

for a constant c depending only on κ and Q_2, we derive that, noting $j_0 = \lfloor c(\delta\sqrt{|\log\delta|})^{-\frac{1}{\nu+1/2}}\rfloor + 1$, we have

$$\mathbb{P}(|IX| > t) \le \mathbf{1}_{\{t<|\boldsymbol{\beta}_{j,\eta}|\}}\mathbf{1}_{\{j\ge j_0\}}. \tag{4.10}$$

Indeed, for all $j < j_0$, for all $\ell \in L_j$, we have $\|\boldsymbol{K}_\ell^{-1}\|_{op} > O_\ell(\delta)^{-1}/2$. Now, a quick application of Lemma 4.1 entails

$$\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\| \ge O_\ell(\delta)\big) \le \delta^{c_0\kappa^2(2\ell+1)^2}.$$

Hence,

$$\begin{aligned}
P(|VIII| > t) &\le \mathbb{P}\left(\left|\sum_{\ell\in L_j}\langle \boldsymbol{f}_\ell, \boldsymbol{\psi}_{j,\eta,\ell}\rangle\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|\ge O_\ell(\delta)\}}\right| > t\right)\\
&\lesssim \sum_{\ell\in L_j} P\big(\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|\ge O_\ell(\delta)\}} > t\big)\\
&\lesssim \sum_{\ell\in L_j}\mathbb{E}[\mathbf{1}_{\{\|\delta\dot{\boldsymbol{B}}_\ell\|\ge O_\ell(\delta)\}}\mathbf{1}_{\{t\le 1\}}]\\
&\lesssim \sum_{\ell\in L_j}\mathbb{P}\big(\|\delta\dot{\boldsymbol{B}}_\ell\| \ge O_\ell(\delta)\big)^{1/2}\mathbf{1}_{\{t\le 1\}}\\
&\lesssim \delta^{c_0\kappa^2 2^{2j}/2}\mathbf{1}_{\{t\le 1\}}.
\end{aligned}$$

This ends the study of the tail of $|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|$. If κ and τ_{sig}, τ_{op} are large enough, the leading terms are given by (4.5), (4.7) and (4.10). (4.1) now results directly from the previous deviation inequalities. (4.2) is an application of the well known formula

$$E[|X|^p] = \int_{u>0} pu^{p-1}\mathbb{P}(|X| > u)du \le p\int_{u>0} u^{p-1}\big(1\wedge \mathbb{P}(|X| > u)\big)du$$

if X is a real random variable. As for inequality (4.3), we have

$$\begin{aligned}
\mathbb{E}[\sup_{\eta\in\mathcal{Z}_j}|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p] &\le \int_{u>0} pu^{p-1}\big(1\wedge \mathbb{P}(\sup_{\eta\in\mathcal{Z}_j}|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}| > u)\big)du\\
&\le p\int_{u>0} u^{p-1}\big(1\wedge 2^{2j}\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}| > u)\big)du.
\end{aligned}$$

Moreover, considering only the terms (4.5), (4.7) and (4.10) as mentioned above, we have

$$2^{2j}\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}| > u) \lesssim e^{-\frac{u^2}{2\varepsilon^2 2^{2j\nu}}+2j\log 2} + e^{-\frac{u^2}{2\delta^2 2^{j(2\nu-1)}}+2j\log 2} + 2^{2j}\mathbf{1}_{\{u\lesssim \delta 2^{j(2\nu-1)}\}} + 2^{2j}\mathbf{1}_{\{u\le|\boldsymbol{\beta}_{j,\eta}|\}}\mathbf{1}_{\{j\ge j_0\}}$$

which entails (4.3). $\square$

4.1. Proof of Theorem 3.2

Proof. We shall only investigate the case where $p > \pi$, since for $p \leq \pi$, we have $B^s_{\pi,r} \subset B^s_{p,r}$. The $\mathbb{L}^p$ loss of the procedure can be decomposed as follows:

$$\mathbb{E}\|\widetilde{\boldsymbol{f}} - \boldsymbol{f}\|_p^p \lesssim \mathbb{E}\left\|\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\langle\widetilde{\boldsymbol{f}} - \boldsymbol{f}, \boldsymbol{\psi}_{j,\eta}\rangle\boldsymbol{\psi}_{j,\eta}\right\|_p^p + \left\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\boldsymbol{\beta}_{j,\eta}\boldsymbol{\psi}_{j,\eta}\right\|_p^p.$$

Since $f \in B^s_{\pi,r}$, we have

$$\left\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\boldsymbol{\beta}_{j,\eta}\boldsymbol{\psi}_{j,\eta}\right\|_p^p \lesssim 2^{-Jp\left(s-2(1/\pi-1/p)\right)}.$$

Now it is proved in Kerkyacharian et al. [18] that for all $\nu, d' \geq 0$, we have

$$\frac{s - 2(1/\pi - 1/p)}{\nu + d'/2} \geq \mu(2, d'). \tag{4.11}$$

Since $2/d' \geq (\nu + d'/2)^{-1}$, the choice of the maximal level J ensures that this term is properly bounded by the desired rates of convergence. Bounding $\mathbb{E}\|\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\langle\widetilde{\boldsymbol{f}} - \boldsymbol{f}, \boldsymbol{\psi}_{j,\eta}\rangle\boldsymbol{\psi}_{j,\eta}\|_p^p$ is more involved. First we apply Hölder's inequality and (2.10) and decompose it as

$$\mathbb{E}\left\|\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\langle\widetilde{\boldsymbol{f}} - \boldsymbol{f}, \boldsymbol{\psi}_{j,\eta}\rangle\boldsymbol{\psi}_{j,\eta}\right\|_p^p \lesssim B + S$$

where

$$B = J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\big[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>S_j(\delta,\varepsilon)\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$
$$S = J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\big[|\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\leq S_j(\delta,\varepsilon)\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p.$$

The first step is to replace $S_j(\delta, \varepsilon)$ in B and S by a quantity explicitly depending on $2^{j\nu}$, namely $\overline{S_j}(\delta, \varepsilon)$. Remark to this end that on the event $\{\ell_j = +\infty\}$, we have $|\widehat{\boldsymbol{\beta}}_{j,\eta}| \leq S_j(\delta, \varepsilon)$ almost surely. We subsequently write

$$\begin{aligned}
B &= J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>S_j(\delta,\varepsilon)\}}\mathbf{1}_{\{\ell_j<+\infty\}}\mathbf{1}_{\mathcal{B}_{\ell_j}}\Big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\quad + J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{\ell_j<+\infty\}}\mathbf{1}_{\mathcal{B}^c_{\ell_j}}\Big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\lesssim J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>\overline{S_j}(\delta,\varepsilon)\}}\Big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\quad + J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\widehat{\boldsymbol{\beta}}_{j,\eta} - \boldsymbol{\beta}_{j,\eta}|^{2p}\Big]^{p/2}\delta^{c_0\rho^2(22^j+1)^2\kappa^2/2}\|\boldsymbol{\psi}_{j,\eta}\|_p^p
\end{aligned}\tag{4.12}$$

where we successively applied Lemma 4.2, (4.6) and the Cauchy–Schwarz inequality. It is clear that (4.12) is negligible for κ large enough. In a similar way,

$$\begin{aligned}
S &= J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\leq S_j(\delta,\varepsilon)\}}\mathbf{1}_{\{\ell_j<+\infty\}}\mathbf{1}_{\mathcal{B}_{\ell_j}}\Big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\quad + J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{P}(\ell_j = +\infty)\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\quad + J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{P}(\mathcal{B}^c_{\ell_j})\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\
&\lesssim J^{p-1}\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\Big[|\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\leq\overline{S_j}(\delta,\varepsilon)\}}\Big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p
\end{aligned}\tag{4.13}$$

$$+J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{P}(\ell_j=+\infty)\|\boldsymbol{\psi}_{j,\eta}\|_p^p \tag{4.14}$$

$$+J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\delta^{c_0\rho^2(22^j+1)^2\kappa^2/2}\|\boldsymbol{\psi}_{j,\eta}\|_p^p \tag{4.15}$$

(4.15) is small enough for κ large enough. Moreover, thanks to (4.8),

$$\big\{\ell_j=+\infty\big\}\subset\mathcal{A}_{2^j}^c\subset\big\{\|\delta\dot{\boldsymbol{B}}_{2^j}\|\ge O_{2^j}(\delta)\big\}\bigcup\big\{\|(\boldsymbol{K}_{2^j})^{-1}\|_{op}\ge O_{2^j}(\delta)^{-1}/2\big\}.$$

Thus, thanks to (4.9) and (4.6), the term (4.14) is negligible as well. We subsequently deduce that

$$\mathbb{E}\Big\|\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\langle\widetilde{\boldsymbol{f}}-\boldsymbol{f},\boldsymbol{\psi}_{j,\eta}\rangle\boldsymbol{\psi}_{j,\eta}\Big\|_p^p\lesssim J^{p-1}\big(Bb+Bs+Sb+Ss\big)$$

with

$$Bb=\sum_{j\le J,\eta\in\mathcal{Z}_j}\mathbb{E}\big[|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$

$$Bs=\sum_{j\le J,\eta\in\mathcal{Z}_j}\mathbb{E}\big[|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\le\overline{S}_j(\delta,\varepsilon)/2\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$

$$Sb=\sum_{j\le J,\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{E}\big[\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\le\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>2\overline{S}_j(\delta,\varepsilon)\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$

$$Ss=\sum_{j\le J,\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{E}\big[\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\le\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\le2\overline{S}_j(\delta,\varepsilon)\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p.$$

We can now treat the terms Bs, Bb, Sb and Ss. Applying (2.10), (4.1) and the Cauchy–Schwarz inequality:

$$\begin{aligned}Bs&\le J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}\big[|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2\}}\big]\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\&\le J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}[|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|^{2p}]^{1/2}\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2)^{1/2}\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\&\lesssim J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\big((\varepsilon2^{j\nu})^p\vee(\delta2^{j(\nu-1/2)})^p\vee|\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{j\ge j_0\}}\big)2^{jp}\big(\varepsilon^{\tau^2}\vee\delta^{\tau^2}\big).\end{aligned}$$

Moreover,

$$\begin{aligned}Sb&\le J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^p\mathbb{P}(|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon))\|\boldsymbol{\psi}_{j,\eta}\|_p^p\\&\lesssim J^{p-1}\big(\varepsilon^{\tau^2}\vee\delta^{\tau^2}\big)\end{aligned}$$

since $\boldsymbol{f}\in B_{p,r}^{s-2(1/\pi-1/p)}$. Hence in both cases the rate of convergence is smaller than what is claimed for sufficiently large τ. Turning to Bb and Ss, we write, for all $z,z'\ge0$,

$$\begin{aligned}Bb&\lesssim J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\mathbb{E}[|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|^p]\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2\}}\|\boldsymbol{\psi}_{j,\eta}\|^p\\&\lesssim\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}\big((\varepsilon2^{j\nu})^p\vee(\delta2^{j(\nu-1/2)})^p\vee|\boldsymbol{\beta}_{j,\eta}|^p\mathbf{1}_{\{j\ge j_0\}}\big)\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2\}}\|\boldsymbol{\psi}_{j,\eta}\|^p\\&\lesssim J^{p-1}\big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{p-z}\sum_{j\le J}2^{j[\nu(p-z)+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^z\\&\quad+J^{p-1}\big(\delta\sqrt{|\log\delta|}\big)^{p-z'}\sum_{j\le J}2^{j[(\nu-1/2)(p-z')+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z'}+J^{p-1}2^{-j_0p\left(s-2(\frac{1}{\pi}-\frac{1}{p})\right)}\end{aligned}$$

and

$$Ss\lesssim J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^z\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\le2\tau2^{j\nu}\varepsilon\sqrt{|\log\varepsilon|}\}}\|\boldsymbol{\psi}_{j,\eta}\|_p^p+J^{p-1}\sum_{j\le J}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^z\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\le2\tau2^{j(\nu-1/2)}\delta\sqrt{|\log\delta|}\}}\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$

$$\lesssim J^{p-1}\big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{p-z}\sum_{j\leq J}2^{j[\nu(p-z)+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z}\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$
$$+J^{p-1}\big(\delta\sqrt{|\log\delta|}\big)^{p-z'}\sum_{j\leq J}2^{j[(\nu-1/2)(p-z')+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z'}\|\boldsymbol{\psi}_{j,\eta}\|_p^p.$$

The term

$$2^{-j_0p\left(s-2(1/\pi-1/p)\right)}\sim(\delta\sqrt{|\log\delta|})^{p\frac{s-2(1/\pi-1/p)}{\nu+1/2}}$$

is readily bounded thanks to (4.11), which leaves us in both cases with the term $R(\varepsilon,\nu,z)+R(\delta,\nu-1/2,z')$ to control, where

$$R(x,y,z)=J^{p-1}\big(x\sqrt{|\log x|}\big)^{p-z}\sum_{j\leq J}2^{j[y(p-z)+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z}\|\boldsymbol{\psi}_{j,\eta}\|_p^p.$$

We only give a brief overview of the treatment of $R(x,y,z)$, a detailed one is present in Kerkyacharian et al. [18]. First, we split it as follows

$$R(x,y,z)=J^{p-1}\Bigg[\big(x\sqrt{|\log x|}\big)^{p-z_1}\sum_{j\leq J_0}2^{j[y(p-z_1)+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z_1}\|\boldsymbol{\psi}_{j,\eta}\|_p^p$$
$$+\big(x\sqrt{|\log x|}\big)^{p-z_2}\sum_{j>J_0}2^{j[y(p-z_2)+p-2]}\sum_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|^{z_2}\|\boldsymbol{\psi}_{j,\eta}\|_p^p\Bigg]$$

where z_1,z_2,J_0 are to be determined. Consider first the case where $s\geq(y+1)(p/\pi-1)$. Note $q=p(y+1)(s+y+1)^{-1}$. Taking $z_2=\pi$, $z_1=\widetilde{q}<q$ and $2^{J_0\frac{p}{q}(y+1)}\sim(x\sqrt{|\log x|})^{-1}$ entails

$$R(x,y,J_0)\lesssim(\log x)^{p-1}(x\sqrt{|\log x|})^{p-q}$$

which is the desired bound. Now consider the case where $s<(y+1)(p/\pi-1)$ and note $q=p(y+1-2/p)(y+1+s-2/\pi)^{-1}$. Take $z_1=\pi$, $z_2=\widetilde{q}>q$ and $2^{J_0\frac{p}{q}(y+1-2/p)}\sim(x\sqrt{|\log x|})^{-1}$. We obtain

$$R(x,y,J_0)\lesssim(\log\varepsilon)^{p-1}(x\sqrt{|\log x|})^{p-q}$$

which ends the proof. □

4.2. *Proof of Theorem 3.3*

Proof. Write similarly

$$\|\widetilde{\boldsymbol{f}}-\boldsymbol{f}\|_\infty\leq\mathbb{E}\Big\|\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\big(\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}\big)\boldsymbol{\psi}_{j,\eta}\Big\|_\infty+\Big\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\boldsymbol{\beta}_{j,\eta}\boldsymbol{\psi}_{j,\eta}\Big\|_\infty.$$

The term $\|\sum_{j>J}\sum_{\eta\in\mathcal{Z}_j}\boldsymbol{\beta}_{j,\eta}\boldsymbol{\psi}_{j,\eta}\|_\infty$ can be handled as in Theorem 3.2. Moreover, (2.11) for $p=\infty$ entails, similarly to the proof of Theorem 3.2,

$$\mathbb{E}\Big\|\sum_{j\leq J}\sum_{\eta\in\mathcal{Z}_j}\big(\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}\big)\boldsymbol{\psi}_{j,\eta}\Big\|_\infty\lesssim Bb+Bs+Sb+Ss$$

with

$$Bb=\sum_{j\leq J}2^j\mathbb{E}\big[\sup_{\eta\in\mathcal{Z}_j}|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)/2\}}\big]$$
$$Bs=\sum_{j\leq J}2^j\mathbb{E}\big[\sup_{\eta\in\mathcal{Z}_j}|\widehat{\boldsymbol{\beta}}_{j,\eta}-\boldsymbol{\beta}_{j,\eta}|\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|>\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\leq\overline{S}_j(\delta,\varepsilon)/2\}}\big]$$
$$Sb=\sum_{j\leq J}2^j\sup_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|\mathbb{E}\big[\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\leq\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|>2\overline{S}_j(\delta,\varepsilon)\}}\big]$$
$$Ss=\sum_{j\leq J}2^j\sup_{\eta\in\mathcal{Z}_j}|\boldsymbol{\beta}_{j,\eta}|\mathbb{E}\big[\mathbf{1}_{\{|\widehat{\boldsymbol{\beta}}_{j,\eta}|\leq\overline{S}_j(\delta,\varepsilon)\}}\mathbf{1}_{\{|\boldsymbol{\beta}_{j,\eta}|\leq2\overline{S}_j(\delta,\varepsilon)\}}\big].$$

Now we have, using inequality (4.3),

$$\begin{aligned}
Bb &\le \sum_{j\le J} 2^j \mathbb{E} \sup_{\eta\in\mathcal{Z}_j} |\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}| \mathbf{1}_{\{|\boldsymbol\beta_{j,\eta}|>\overline{S_j}(\delta,\varepsilon)/2\}} \\
&\le \sum_{j\le J} 2^j \mathbf{1}_{\{\exists\eta\in\mathcal{Z}_j,\,|\boldsymbol\beta_{j,\eta}|\ge\overline{S_j}(\delta,\varepsilon)/2\}} 2^j \mathbb{E} \sup_{\eta\in\mathcal{Z}_j} |\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}| \\
&\lesssim \sum_{j\le J} 2^j \mathbf{1}_{\{\exists\eta\in\mathcal{Z}_j,\,|\boldsymbol\beta_{j,\eta}|\ge\overline{S_j}(\delta,\varepsilon)/2\}} (j+1)\big(\varepsilon 2^{j\nu} \vee \delta 2^{j(\nu-1/2)}\big) \vee |\boldsymbol\beta_{j,\eta}| \mathbf{1}_{\{j\ge j_0\}} \\
&\lesssim 2^{J_1(\nu+1)}(J_1+1)\varepsilon + 2^{I_1(\nu+3/2)}(I_1+1)\delta + \sum_{j\ge j_0} 2^j |\boldsymbol\beta_{j,\eta}|
\end{aligned}$$

where J_1 is chosen so that, for $j \ge J_1$, $|\boldsymbol\beta_{j,\eta}| \le \tau\varepsilon\sqrt{|\log\varepsilon|}2^{j\nu}/2$. We can take for example (see [18]) J_1 verifying, for a certain constant B,

$$2^{J_1} = B\big(\varepsilon\sqrt{|\log\varepsilon|}\big)^{-(s+\nu+1-2/\pi)^{-1}}.$$

Similarly, taking

$$2^{I_1} = C\big(\delta\sqrt{|\log\delta|}\big)^{-(s+\nu+1/2-2/\pi)^{-1}}$$

for a certain constant C implies $|\boldsymbol\beta_{j,\eta}| \le \tau\delta\sqrt{|\log\delta|}2^{j(\nu-1/2)}/2$ for all $j \le I_1$. The term $\sum_{j\ge j_0} 2^j|\boldsymbol\beta_{j,\eta}|$ is easily treated. This finally leads to the rates

$$Bb \lesssim |\log\varepsilon|\varepsilon^{\mu'(2)} \vee |\log\delta|\delta^{\mu'(1)}$$

and

$$\begin{aligned}
Ss &\le \sum_{j\le J} 2^j \sup_{\eta\in\mathcal{Z}_j} |\boldsymbol\beta_{j,\eta}| \mathbf{1}_{\{|\boldsymbol\beta_{j,\eta}|\le 2\overline{S_j}(\delta,\varepsilon)\}} \\
&\lesssim \left[\sum_{j\le J_1} 2^j\varepsilon\sqrt{|\log\varepsilon|}2^{j\nu} + \sum_{j>J_1} 2^j|\boldsymbol\beta_{j,\eta}|\right] \vee \left[\sum_{j\le I_1} 2^j\delta\sqrt{|\log\delta|}2^{j(\nu-1/2)} + \sum_{j>I_1} 2^j|\boldsymbol\beta_{j,\eta}|\right]
\end{aligned}$$

which are of the proper order. Turning to Bs and Sb, we write, using inequalities (4.1) and (4.3)

$$\begin{aligned}
Bs &\le \sum_{j\le J} 2^j \mathbb{E}\big[\sup_{\eta\in\mathcal{Z}_j} |\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}| \mathbf{1}_{\{|\widehat{\boldsymbol\beta}_{j,\eta}-\boldsymbol\beta_{j,\eta}|>\overline{S_j}(\delta,\varepsilon)/2\}}\big] \\
&\le \sum_{j\le J} 2^j \mathbb{E}[\sup_{\eta\in\mathcal{Z}_j} |\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}|^2]^{1/2} \mathbb{P}(\exists\eta\in\mathcal{Z}_j,\ |\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}| > \overline{S_j}(\delta,\varepsilon)/2)^{1/2} \\
&\lesssim \sum_{j\le J} 2^j \Big[(j+1)\big(\varepsilon 2^{j\nu} \vee \delta 2^{j(\nu-1/2)}\big) \vee |\boldsymbol\beta_{j,\eta}|\mathbf{1}_{\{j\ge j_0\}}\Big]\Big[2^{2j}(\varepsilon^{\tau^2} \vee \delta^{\tau^2})\Big]^{1/2}.
\end{aligned}$$

Now apply inequality (4.1) and the fact that $|\boldsymbol\beta_{j,\eta}| \lesssim 2^{-j}$ to derive

$$\begin{aligned}
Sb &\le \sum_{j\le J} 2^j \mathbb{E}\big[\sup_{\eta\in\mathcal{Z}_j} |\boldsymbol\beta_{j,\eta}| \mathbf{1}_{\{|\widehat{\boldsymbol\beta}_{j,\eta}-\boldsymbol\beta_{j,\eta}|>\overline{S_j}(\delta,\varepsilon)\}}\big] \\
&\lesssim \sum_{j\le J} 2^{2j} \mathbb{P}\big(|\widehat{\boldsymbol\beta}_{j,\eta} - \boldsymbol\beta_{j,\eta}| > \overline{S_j}(\delta,\varepsilon)\big) \\
&\lesssim \sum_{j\le J} 2^{2j}(\varepsilon^{\tau^2} \vee \delta^{\tau^2}).
\end{aligned}$$

It is clear that for a well chosen τ these terms are smaller than the announced rates. □

Acknowledgment

The author would like to thank his advisor Dominique Picard for numerous and fruitful discussions and suggestions.

References

[1] P. Baldi, G. Kerkyacharian, D. Marinucci, D. Picard, Adaptive density estimation for directional data using needlets, Ann. Statist. 37 (2009) 3362–3395.
[2] P. Campisi, K. Egiazarian, Blind Image Deconvolution: Theory and Applications, CRC Press, 2007.

[3] L. Cavalier, N.W. Hengartner, Adaptive estimation for inverse problems with noisy operators, Inverse Problems 21 (2005) 1345–1361.
[4] L. Cavalier, M. Raimondo, Wavelet deconvolution with noisy eigenvalues, IEEE Trans. Signal Process. 55 (2007) 2414–2424.
[5] K. Davidson, S. Szarek, Local operator theory, random matrices and Banach spaces, in: Lindenstrauss (Ed.), in: Handbook on the Geometry of Banach spaces, vol. 1, Elsevier Science, 2001, pp. 317–366.
[6] S. Delattre, M. Hoffmann, D. Picard, T. Vareschi, Blockwise SVD with error in the operator and application to blind deconvolution, Electron. J. Stat. 6 (2012) 2274–2308.
[7] D. Donoho, I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[8] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, AOS 24 (2) (1996) 508–539.
[9] S. Efromovich, V. Koltchinskii, On inverse problems with unknown operators, IEEE Trans. Inform. Theory 47 (2001) 2876–2894.
[10] W. Freeden, T. Gervens, M. Schreiner, Constructive Approximation on the Sphere: With Applications to Geomathematics, Clarendon Press, 1998.
[11] W. Freeden, D. Michel, V. Michel, Local multiscale approximations of geostrophic oceanic flow: theoretical background and aspects of scientific computing, Mar. Geod. 28 (2005) 313–329.
[12] H. Groemer, Geometric Applications of Fourier Series and Spherical Harmonics, Cambridge Univ. Press, 1996.
[13] F. Guilloux, G. Faÿ, F. Cardoso, Practical wavelet design on the sphere, Appl. Comput. Harmon. Anal. 26 (2) (2009) 143–160.
[14] D.M. Healy, H. Hendriks, P.T. Kim, Spherical deconvolution, J. Multivariate Anal. 67 (1–22) (1998).
[15] M. Hoffmann, M. Reiß, Nonlinear estimation for linear inverse problems with error in the operator, Ann. Statist. 36 (2008) 310–336.
[16] K. Ivanov, P. Petrushev, Y. Xu, Decomposition of spaces of distributions induced by tensor product bases, J. Funct. Anal. 263 (5) (2012) 1147–1197.
[17] G. Kerkyacharian, D. Picard, New generation wavelets associated with statistical problems, in: 8th Workshop on Stochastic Numerics, Research Institute for Mathematical Sciences, Kyoto University, 2009, pp. 119–146.
[18] G. Kerkyacharian, T.M. Pham Ngoc, D. Picard, Localized spherical deconvolution, Ann. Statist. 39 (2011) 1042–1068.
[19] P.T. Kim, J.Y. Koo, Optimal spherical deconvolution, J. Multivariate Anal. 80 (2002) 21–42.
[20] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[21] D. Marinucci, D. Pietrobon, A. Balbi, P. Baldi, P. Cabella, G. Kerkyacharian, P. Natoli, D. Picard, N. Vittorio, Spherical needlets for CMB data analysis, Mon. Not. R. Astron. Soc. 383 (2) (2008) 143–160.
[22] F. Narcowich, J. Ward, Nonstationary wavelets on the m-sphere for scattered data, Appl. Comput. Harm. Anal. 3 (1996) 324–336.
[23] F. Narcowich, P. Petrushev, J. Ward, Decomposition of Besov and Triebel–Lizorkin spaces on the sphere, J. Funct. Anal. 238 (2) (2006) 530–564.
[24] F. Narcowich, P. Petrushev, J. Ward, Localized tight frames on spheres, J. Funct. Anal. 238 (2) (2006) 530–564.
[25] P. Pankajakshan, L. Blanc-Féraud, B. Zhang, Z. Kam, J.-C. Olivo-Marin, J. Zerubia, Parametric blind deconvolution for confocal laser scanning microscopy (CLSM)-proof of concept, in: Project INRIA Ariana, Research Report 6493, 2008.
[26] J.S. Rosenthal, Random rotations: characters and random walks on SO(N), Ann. Probab. 22 (1) (1994) 398–423.
[27] P. Shröder, W. Sweldens, Spherical wavelets: efficiently representing functions on the sphere, in: Computer Graphics Proceedings, vol. 95, 161–172.
[28] J.D. Tournier, F. Calamante, D.G. Gadian, A. Connelly, Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution, Neuroimage 23 (3) (2004) 1176–1185.
[29] N.J. Vilenkin, Fonctions spéciales et théorie de la représentation des groupes, in: Monographies Universitaires de Mathématiques, vol. 33, Dunod, Paris, 1969.
[30] Thomas Willer, Optimal bounds for inverse problems with Jacobi-type eigenfunctions, Statist. Sinica 19 (2) (2009) 785–800.