Journal of Multivariate Analysis 115 (2013) 138–156

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

The dictionary approach for spherical deconvolution

Thanh Mai Pham Ngoc [a,*], Vincent Rivoirard [b]

[a] *Laboratoire de Mathématique, UMR CNRS 8628, Université Paris Sud, 91405 Orsay Cedex, France*

[b] *CEREMADE UMR CNRS 7534, Université Paris Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 PARIS Cedex 16, France*

ARTICLE INFO

Article history:
Received 19 August 2011
Available online 10 September 2012

AMS subject classifications:
62G07
62G05
62G20

Keywords:
Density deconvolution
Dictionary
Lasso estimate
Oracle inequalities
Calibration
Sparsity
Second generation wavelets

ABSTRACT

We consider the problem of estimating a density of probability from indirect data in the spherical convolution model. We aim at building an estimate of the unknown density as a linear combination of functions of an overcomplete dictionary. The procedure is devised through a well-calibrated ℓ_1-penalized criterion. The spherical deconvolution setting has been barely studied so far, and the two main approaches to this problem, namely the SVD and the hard thresholding ones considered only one basis at a time. The dictionary approach allows to combine various bases and thus enhances estimates sparsity. We provide an oracle inequality under global coherence assumptions. Moreover, the calibrated procedure that we put forward gives quite satisfying results in the numerical study when compared with other procedures.

1. Introduction

We consider the spherical deconvolution problem. We observe:

$$Z_i = \varepsilon_i X_i, \quad i = 1, \ldots, N \tag{1}$$

where the ε_i are i.i.d. random variables of $\mathbb{SO}(3)$ the rotation group in $\mathbb{R}^3$ and the X_i's are i.i.d. random variables of $\mathbb{S}^2$, the unit sphere of $\mathbb{R}^3$. We suppose that X_i and ε_i are independent. We also assume that the distributions of Z_i and X_i are absolutely continuous with respect to the uniform measure on $\mathbb{S}^2$ and we set f_Z and f the densities of Z_i and X_i respectively. The distribution of ε_i is absolutely continuous with respect to the Haar measure on $\mathbb{SO}(3)$ and we will denote it f_ε. In this paper, we consider that f_ε is known.

Then we have

$$f_Z = f_\varepsilon * f,$$

where $*$ denotes the convolution product which is defined below in (10).

The aim of the present paper is to recover the unknown density f from the noisy observations Z_i thanks to a well-calibrated ℓ_1-penalized least squares criterion. Roughly speaking, each genuine observation X_i is contaminated by a small random rotation. Although the problem of deconvolution has been extensively addressed in the case of the real line, it has been barely the case on the sphere. The spherical geometry has its own characteristics and includes more complex analytical tools.

[*] Corresponding author.
E-mail addresses: thanh.pham_ngoc@math.u-psud.fr (T.M. Pham Ngoc), Vincent.Rivoirard@dauphine.fr (V. Rivoirard).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.08.011

The model of spherical convolution, as expressed in (1), has applications in medical imaging and in astrophysics. In medical imaging, people are interested in estimating a fiber orientation density from high angular resolution diffusion MRI data (MRI stands for Magnetic Resonance Imaging) see the work of Tournier et al. [28]. In their paper, the spherical convolution (1) models the situation where the density of the MRI data is viewed as a convolution of a response function and the density of interest. In astrophysics, the so-called UHECR (Ultra High Energy Cosmic Rays) are at the core of astrophysics concerns. In order to understand the mechanisms of the UHECR, a crucial challenge is the estimation of the density probability of the incidence directions with which the UHECR arrive on earth. The convolution model takes into account a natural noise which corrupts the genuine observations X_i.

The first authors who actually solved this problem were Healy, Hendriks and Kim in their pioneering work, see [14]. They introduced an orthogonal series method on the Fourier basis of $\mathbb{L}_2(\mathbb{S}^2)$ namely the spherical harmonics and assessed its theoretical performances by presenting convergence rates for Sobolev type regularities. Moreover, the spherical harmonics constitute the SVD (Singular Value Decomposition) basis in the spherical deconvolution setting and hence allow to invert the convolution operator f_ε in a stable way. Subsequently, Kim and Koo [17] proved that those rates of convergence were optimal and refined those results by enhancing sharp minimaxity under a super-smooth condition on the error distribution, see [18]. The SVD procedure is of course appealing for its simplicity and its ability to invert quickly the operator f_ε but has poor local performances. Indeed, the spherical harmonics which are spread all over the sphere might be a drawback if one is interested in highlighting some local features of the density of interest. It is the case whenever one concentrates on the infinity norm or on adapting to inhomogeneous smoothness. To circumvent these problems, Kerkyacharian et al. [16] considered a thresholding procedure on needlets. The needlets due to Narcowich et al. [24] is a tight frame constructed on the spherical harmonics. They enjoy very good localization properties. This procedure turned out to be profitable both in theory with consideration of $\mathbb{L}^p$ loss with $1 \leq p \leq \infty$ and in practice.

Nonetheless as one may have noticed, each approach mentioned above leans on only one basis, the spherical harmonics or the needlet one. Consequently, instead of sticking to only one basis, it may be relevant to consider an overcomplete dictionary. Moreover, K can be larger than the number of observations N contrary to thresholding techniques where $K \leq N$.

With this aim in view, we would like to build an estimate of f as a linear combination of functions of a dictionary $(\varphi_1, \ldots, \varphi_K)$ with $\varphi_k \in \mathbb{L}_2(\mathbb{S}^2)$. Denote by f_λ the linear combination

$$f_\lambda(x) = \sum_{k=1}^{K} \lambda_k \varphi_k(x), \quad x \in \mathbb{S}^2,\ \lambda = (\lambda_1, \ldots, \lambda_K) \in \mathbb{C}^K. \tag{2}$$

By considering an overcomplete dictionary which cardinality K can be larger than the sample size N, we tacitly believe that the estimates of f are sparse, namely that very few coordinates of $\hat{\lambda}$ is non zero. To our knowledge, the dictionary approach has not been used to face the spherical convolution model (1) or its analogous on the real line expressed as $Y = X + \varepsilon$.

A question immediately comes into sight. Because we precisely treat a convolution problem which provides a relevant setup where observations may come from one source observed through some noise, is there some hope to keep a sparse structure of the estimates of λ? In other words, is the dictionary approach capable to retrieve sparsity despite the action of the convolution operator? The answer seems to be affirmative at least in the present numerical study that we conducted with a ℓ_1-penalized criterion.

Indeed, we suggest a data-driven choice of $\hat{\lambda}$ that will be obtained with a well-calibrated ℓ_1-penalized criterion. The so-called popular Lasso first introduced by Tibshirani [27] has been widely used since then in the statistical literature. In a fairly general Gaussian framework we may cite the recent work of Massart and Meynet [22], in the linear model regression, see [11,13,23,27] and for nonparametric regression with general fixed or random design, see [4,8,6,5]. The Lasso performances were also studied in the density estimation framework by Bunea et al. [7,9], van de Geer [29] and Bertin et al. [3]. In addition, many efforts have also been provided to prove model selection consistency of the Lasso, see [4,20,23,31–33].

ℓ_1-penalty methods have been also investigated to solve inverse problems. We may cite among others the work of Loubes (see [19]) who tackled the classical inverse regression model with independent errors by minimizing an empirical contrast built upon the SVD basis of the operator with an ℓ_1 penalty term. But we stress that it was by no means a dictionary approach. In addition, we point out that the model of interest in [19] and papers devoted to inverse problems in general are much more linked to regression problem whereas our convolution model (1) is to be more connected to a density estimation problem.

In this paper, we aim at showing that the dictionary approach conducted with the Lasso minimization algorithm can be used successfully to face the spherical deconvolution problem in theory and more especially in practice. Indeed, in the simulation study, we compare it with the hard thresholding procedure on needlets of Kerkyacharian et al. [16], showing that the dictionary approach does pretty well both in graphics reconstructions and in terms of quadratic and sup-norm losses. The Lasso estimates actually enhance the sparsity of the representation of the signal. Moreover, the choice of tuning parameters turns out to be easy to calibrate and match what the theory states contrary to thresholding techniques where theorems are too conservative about the allowed values of tuning parameters.

Here is the outline of the present paper. In Section 2 we give some basic tools of Fourier analysis on $\mathbb{L}_2(\mathbb{SO}(3))$ and $\mathbb{L}_2(\mathbb{S}^2)$ and the construction of the Lasso estimates. In Section 3 we obtain oracle inequalities under mild assumptions on the dictionaries. In Section 4 we present our simulation results. Appendix is devoted to the proofs of our results.

2. Lasso-type estimates of the density *f*

2.1. Preliminaries about harmonic analysis on $\mathbb{S}^2$ and $\mathbb{SO}(3)$

Let us begin with some notations and some elements of harmonic analysis on $\mathbb{S}^2$ and $\mathbb{SO}(3)$ which will be useful throughout the paper.

For two functions g and h we denote $\langle g, h\rangle$ the $\mathbb{L}_2$-hermitian product between g and h:

$$\langle g, h\rangle = \int_{x\in\mathbb{S}^2} g(x)\overline{h(x)}dx,$$

and $\|\cdot\|_2$ is the associated norm.

$|\cdot|$ will denote the modulus, $\Re$ and $\Im$ the real and the imaginary part of a complex number respectively.

For any vector $\lambda \in \mathbb{C}^K$ and any set of indices J, we denote $\lambda_J \in \mathbb{C}^K$ the vector which has the same coordinates as λ on J and 0 elsewhere. We set for any $1 \le q < \infty$,

$$\|\lambda\|_{\ell_q} = \left(\sum_{k=1}^{K} |\lambda_k|^q\right)^{\frac{1}{q}}.$$

We shall now recall some elements of harmonic analysis on $\mathbb{SO}(3)$ and $\mathbb{S}^2$. We shall refer the reader to Healy et al. [14] and Kim and Koo [17] for more precisions. Consider the functions, known as the rotational harmonics,

$$D^l_{mn}(\phi,\theta,\psi) = e^{-i(m\phi+n\psi)}P^l_{mn}(\cos\theta), \quad (m,n)\in \mathcal{I}_l^2,\ l=0,1,\dots \tag{3}$$

where $\theta \in [0,\pi)$, $\phi \in [0,2\pi)$, $\psi \in [0,2\pi)$ and $\mathcal{I}_l = [-l,-l+1,\dots,l-1,l]$. The generalized Legendre associated functions P^l_{mn} are fully described in [30]. The $\sqrt{2l+1}D^l_{mn}$ form a complete orthonormal basis of $\mathbb{L}_2(\mathbb{SO}(3))$ with respect to the Haar probability measure.

For $f \in \mathbb{L}_2(\mathbb{SO}(3))$, we define the rotational Fourier transform on $\mathbb{SO}(3)$ by

$$(f^{*l})_{mn} = \int_{\mathbb{SO}(3)} f(g)D^l_{mn}(g)dg. \tag{4}$$

Then (f^{*l}) is a matrix of size $(2l+1)\times(2l+1)$ which entrance is given by the element $(f^{*l})_{mn}$ with $m\in\mathcal{I}_l$ and $n\in\mathcal{I}_l$. The rotational inversion can be obtained by

$$\begin{aligned} f(g) &= \sum_l \sum_{-l\le m,\, n\le l} (f^{*l})_{mn}\overline{D^l_{mn}(g)} \\ &= \sum_l \sum_{-l\le m,\, n\le l} (f^{*l})_{mn}D^l_{mn}(g^{-1}), \end{aligned} \tag{5}$$

Eq. (5) is to be understood in $\mathbb{L}_2$-sense although with additional smoothness conditions, it can hold pointwise.

A parallel spherical Fourier analysis is available on $\mathbb{S}^2$. Any point on $\mathbb{S}^2$ can be represented by

$$\omega = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)^t,$$

with, $\phi\in[0,2\pi)$, $\theta\in[0,\pi)$. We also define the functions:

$$Y^l_m(\omega) = Y^l_m(\theta,\phi) = (-1)^m\sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P^l_m(\cos\theta)e^{im\varphi}, \quad m\in\mathcal{I}_l,\ l=0,1,\dots, \tag{6}$$

with $\phi\in[0,2\pi)$, $\theta\in[0,\pi)$ and $P^l_m(\cos\theta)$ are the associated Legendre functions.

The functions Y^l_m obey

$$Y^l_{-m}(\theta,\phi) = (-1)^m\overline{Y^l_m(\theta,\phi)}. \tag{7}$$

The set $\{Y^l_m,\ m\in\mathcal{I}_l,\ l=0,\ 1,\dots\}$ is forming an orthonormal basis of $\mathbb{L}_2(\mathbb{S}^2)$, generally referred to as the spherical harmonic basis.

Again, as above for $f\in\mathbb{L}_2(\mathbb{S}^2)$, we define the spherical Fourier transform on $\mathbb{S}^2$ by

$$(f^{*l})_m = \int_{\mathbb{S}^2} f(x)\overline{Y^l_m(x)}dx, \tag{8}$$

where dx is the uniform probability measure on the sphere $\mathbb{S}^2$.

Then (f^{*l}) is a vector of size $2l+1$ which entrance is given by the element $(f^{*l})_m$ with $m \in \mathcal{I}_l$. The spherical inversion can be obtained by

$$f(x) = \sum_l \sum_{m \in \mathcal{I}_l} (f^{*l})_m Y^l_m(x). \tag{9}$$

The bases detailed above are important because they realize a singular value decomposition of the convolution operator created by our model. In effect, we define for $f_\varepsilon \in \mathbb{L}_2(\mathbb{SO}(3))$, $f \in \mathbb{L}_2(\mathbb{S}^2)$ the convolution by the following formula:

$$f_\varepsilon * f(x) = \int_{\mathbb{SO}(3)} f_\varepsilon(u) f(u^{-1}x)du, \tag{10}$$

and we have for all $m \in \mathcal{I}_l$, $l = 0, 1, \dots,$

$$(f_\varepsilon * f)^{*l}_m = \sum_{n \in \mathcal{I}_l} (f^{*l}_\varepsilon)_{mn} (f^{*l})_n. \tag{11}$$

2.2. The Lasso estimator of the density f

In the sequel, the estimate of f will be a linear combination of functions of the dictionary $\Upsilon = (\varphi_k)_{k=1,\dots,K}$. For any $\lambda \in \mathbb{C}^K$ we set:

$$f_\lambda = \sum_{k=1}^K \lambda_k \varphi_k, \quad \lambda = (\lambda_k)_{k=1\cdots K}.$$

We assume that for any k, $\|\varphi_k\|_2 = 1$. For the moment we do not need more assumptions on the dictionary. It will not be the case for oracle inequalities of Section 3.1. We set for any k,

$$\beta_k = \int_{\mathbb{S}^2} \varphi_k(x) f(x) dx.$$

If we denote for any function φ_k, $(\varphi_k^{*l})_m$ the (l, m)-Fourier coefficient of φ_k:

$$(\varphi_k^{*l})_m = \langle \varphi_k, Y^l_m \rangle = \int_{\mathbb{S}^2} \varphi_k(x) \overline{Y^l_m}(x) dx,$$

the Parseval equality yields

$$\beta_k = \sum_{l=0}^\infty \sum_{m \in \mathcal{I}_l} (\varphi_k^*)^l_m (f^*)^l_m.$$

The SVD method leads to the following unbiased estimate of $(f^*)^l_m$ (see [14,16,17]).

$$(\hat{f}^{*l})_m = \frac{1}{N} \sum_{i=1}^N \sum_{n \in \mathcal{I}_l} (f^{*l}_\varepsilon)^{-1}_{mn} \overline{Y^l_n}(Z_i),$$

where $(f^{*l}_\varepsilon)^{-1}_{mn}$ denotes the inverse of the rotational Fourier transform of f_ε. Precisely, one first considers the matrix f^{*l}_ε of size $(2l+1) \times (2l+1)$ which element at line m and column n is given by the Fourier transform $(f^{*l}_\varepsilon)_{mn}$, then one inverts this matrix and takes the element at line m and column n given by $(f^{*l}_\varepsilon)^{-1}_{mn}$. Consequently,

$$\hat{\beta}_k = \frac{1}{N} \sum_{i=1}^N \sum_{l=0}^\infty \sum_{m \in \mathcal{I}_l} \sum_{n \in \mathcal{I}_l} (\varphi_k^{*l})_m (f^{*l}_\varepsilon)^{-1}_{mn} \overline{Y^l_n}(Z_i),$$

is an unbiased estimate of β_k. In particular, if we set for any $x \in \mathbb{S}^2$,

$$\phi_k(x) = \sum_{l=0}^\infty \sum_{m \in \mathcal{I}_l} \sum_{n \in \mathcal{I}_l} (\varphi_k^{*l})_m (f^{*l}_\varepsilon)^{-1}_{mn} \overline{Y^l_n}(x),$$

then

$$\hat{\beta}_k = \frac{1}{N} \sum_{i=1}^N \phi_k(Z_i).$$

We now introduce the *Lasso estimator* $\hat{f}^L$ of the density f.

Definition 1. The Lasso estimate is $\hat{f}^L = \max\left\{\Re(f_{\hat{\lambda}^L}), 0\right\}$ where $\hat{\lambda}^L$ is the solution of the following minimization problem

$$\hat{\lambda}^L = \underset{\lambda \in \mathbb{C}^K}{\operatorname{argmin}} \left\{ C(\lambda) + 2\sum_{k=1}^{K} \left(\eta_{1,k}|\Re(\lambda_k)| + \eta_{2,k}|\Im(\lambda_k)|\right) \right\}, \tag{12}$$

where

$$C(\lambda) = \|f_\lambda\|_2^2 - 2\Re\left(\sum_{k=1}^{K} \lambda_k \hat{\beta}_k\right)$$

and $(\eta_{1,k})_{k\in\{1,\dots,K\}}$ and $(\eta_{2,k})_{k\in\{1,\dots,K\}}$ are two sequences of positive real numbers chosen in (21) and (22) subsequently.

The next proposition shows that $\hat{\lambda}^L$ is obtained by minimizing a ℓ_1-penalized empirical contrast.

Proposition 1. *For any $\lambda \in \mathbb{C}^K$,*

$$\mathbb{E}[C(\lambda)] = \|f_\lambda - f\|_2^2 - \|f\|_2^2,$$

which yields

$$\underset{\lambda \in \mathbb{C}^K}{\operatorname{argmin}}\, \mathbb{E}[C(\lambda)] = \underset{\lambda \in \mathbb{C}^K}{\operatorname{argmin}} \|f_\lambda - f\|_2^2.$$

Let G the Gram matrix associated to the dictionary Υ given for any $1 \le k, k' \le K$ by

$$G_{kk'} = \int_{\mathbb{S}^2} \varphi_k(x)\overline{\varphi_{k'}(x)}dx. \tag{13}$$

We have for any k and k', $G_{kk'} = \overline{G_{k'k}}$ which entails that the matrix G is hermitian. Now, the key point for establishing oracle properties of $\hat{f}^L$ is the following result.

Proposition 2. *A necessary condition for λ to be a solution of (12) is*

$$|\Re((G\overline{\lambda})_k - \hat{\beta}_k)| \le \eta_{1,k} \quad \textit{and} \quad |\Im((G\overline{\lambda})_k - \hat{\beta}_k)| \le \eta_{2,k} \quad \forall k \in \{1,\dots,K\}. \tag{14}$$

In particular, for any k,

$$\left|(G\overline{\lambda})_k - \hat{\beta}_k\right| \le |\eta_k|$$

where

$$\eta_k = \eta_{1,k} + i\eta_{2,k}. \tag{15}$$

Now we shall give some comments about the condition (14). To this purpose, let us denote $\Pi_\Upsilon(f)$ the projection of f on the linear space spanned by the functions of the dictionary Υ. There exists $\lambda(f) \in \mathbb{C}^k$ such that

$$\Pi_\Upsilon(f) = \sum_{k=1}^{K} \lambda(f)_k \varphi_k. \tag{16}$$

If, as expected, our wealthy dictionary Υ provides a sparse linear combination of the functions of Υ that approximates f accurately, we can hope that the well calibrated Lasso procedure does a good job for estimating f. Indeed, for fixed k, we have:

$$(G\overline{\lambda(f)})_k = \int \varphi_k(x) \sum_{k'=1}^{K} \overline{\lambda(f)_{k'}\varphi_{k'}(x)}dx = \int \varphi_k(x)\overline{\Pi_\Upsilon(f)(x)}dx = \int \varphi_k(x)f(x)dx = \beta_k = \mathbb{E}(\hat{\beta}_k).$$

Hence inequalities (14) express a control of the fluctuations of $\hat{\beta}_k$ around its expectation. Consequently, it is natural to find parameters $\eta_{1,k}$ and $\eta_{1,k}$ satisfying (14) which expressions involve the unbiased variance $\sigma_{1,k}^2$ and $\sigma_{2,k}^2$ of respectively $\Re(\hat{\beta}_k)$ and $\Im(\hat{\beta}_k)$.

Precisely, we have the following result providing the values of the Lasso parameters $(\eta_{1,k})_k$ and $(\eta_{2,k})_k$.

Theorem 1. *We set*

$$\hat{\sigma}_{1,k}^2 = \frac{1}{2N(N-1)} \sum_{i \ne j} (\Re(\phi_k(Z_i)) - \Re(\phi_k(Z_j)))^2 \tag{17}$$

and

$$\hat{\sigma}^2_{2,k} = \frac{1}{2N(N-1)} \sum_{i \neq j} (\Im(\phi_k(Z_i)) - \Im(\phi_k(Z_j)))^2 \tag{18}$$

the unbiased estimates of

$$\sigma^2_{1,k} = \mathrm{Var}(\Re(\phi_k(Z_1))) \quad \textit{and} \quad \sigma^2_{2,k} = \mathrm{Var}(\Im(\phi_k(Z_1))).$$

Then we introduce

$$\tilde{\sigma}^2_{1,k} = \hat{\sigma}^2_{1,k} + 2\|\Re(\phi_k)\|_\infty \sqrt{\frac{2\hat{\sigma}^2_{1,k}\gamma \log K}{N}} + \frac{8\|\Re(\phi_k)\|^2_\infty \gamma \log K}{N}, \tag{19}$$

$$\tilde{\sigma}^2_{2,k} = \hat{\sigma}^2_{2,k} + 2\|\Im(\phi_k)\|_\infty \sqrt{\frac{2\hat{\sigma}^2_{2,k}\gamma \log K}{N}} + \frac{8\|\Im(\phi_k)\|^2_\infty \gamma \log K}{N} \tag{20}$$

$$\eta_{1,k} = \sqrt{\frac{2\tilde{\sigma}^2_{1,k}\gamma \log K}{N}} + \frac{2\|\Re(\phi_k)\|_\infty \gamma \log K}{3N} \tag{21}$$

and

$$\eta_{2,k} = \sqrt{\frac{2\tilde{\sigma}^2_{2,k}\gamma \log K}{N}} + \frac{2\|\Im(\phi_k)\|_\infty \gamma \log K}{3N}. \tag{22}$$

Let us assume that K satisfies

$$N \leq K \leq \exp(N^\delta)$$

for $\delta < 1$. Let $\gamma > 1$. Then, for any $\varepsilon > 0$, there exists a constant $C_1(\varepsilon, \delta, \gamma)$ depending on ε, δ and γ such that if Ω is the random set such that for any $k \in \{1, \dots, K\}$

$$|\Re(\beta_k - \hat{\beta}_k)| \leq \eta_{1,k} \quad \textit{and} \quad |\Im(\beta_k - \hat{\beta}_k)| \leq \eta_{2,k},$$

then

$$\mathbb{P}\left(\Omega^c\right) \leq C_1(\varepsilon, \delta, \gamma) K^{1-\frac{\gamma}{1+\varepsilon}}.$$

Let us make some comments about the parameter γ. It is a tuning parameter which appears in the statistical procedure through $\eta_{1,k}$ and $\eta_{2,k}$, see (21) and (22), (it is not the case for ε). The sharp concentration inequality of Theorem 1 is established under the assumption $\gamma > 1$ which ensures that probabilities converge to 0.

In nonparametric statistics it is a well-known fact that theoretical results rarely yield optimal choices for tuning parameters from a practical point of view. Judisky and Lambert–Lacroix (in Remark 5 in [15]) underlined this gap between theory and practice. They highlighted that values for tuning parameters given by theory are extremely conservative and proposed ad hoc choices of thresholds. In our paper we aim at fulfilling this gap by proposing data-driven choices of thresholds based on sharp concentration inequalities for i.i.d. variables.

Furthermore as conditions on tuning parameters provided by theory proved to be most of the time too conservative and that convenient tuning parameters are smaller in practice, for practical considerations, a natural way for choosing γ consists in saturating the inequality $\gamma > 1$ by taking γ greater than 1 but very close to 1.

3. Oracle and minimax properties satisfied by lasso-type estimates

3.1. Oracle inequalities under coherence assumptions for general dictionaries

In the sequel, we establish oracle inequalities under classical assumptions on the dictionary. We first introduce the minimal "restricted" eigenvalue of the Gram matrix G: for $1 \leq l \leq K$, we denote

$$\xi_{\min}(l) = \min_{|J| \leq l} \min_{\substack{\lambda \in \mathbb{C}^K \\ \lambda_J \neq 0}} \frac{\|f_{\lambda_J}\|_2^2}{\|\lambda_J\|^2_{\ell_2}}.$$

Since the functions of the dictionary satisfy $\|\varphi_k\|_2 = 1$ for any k, we have $\xi_{\min}(l) \in [0, 1]$ for any l. When the dictionary constitutes an orthonormal system, we have $\xi_{\min}(l) = 1$ for any $1 \leq l \leq K$. By contrast, if two functions of the dictionary are proportional, then $\xi_{\min}(l) = 0$ for any $2 \le l \le K$. So, assuming that $\xi_{\min}(l)$ is close to 1 means that every set of columns of G with cardinality less than l behaves like an orthonormal system. We also consider the restricted correlations: for $1 \le l, l' \le K$, we denote

$$\theta_{l,l'} = \max_{\substack{|J|\le l\\ |J'|\le l'\\ J\cap J'=\emptyset}} \max_{\substack{\lambda,\lambda'\in\mathbb{C}^K\\ \lambda_J\neq 0,\lambda'_{J'}\neq 0}} \frac{\langle f_{\lambda_J}, f_{\lambda'_{J'}}\rangle}{\|\lambda_J\|_{\ell_2}\|\lambda'_{J'}\|_{\ell_2}}.$$

Small values of $\theta_{l,l'}$ mean that two disjoint sets of columns of G with cardinality less than l and l' span nearly orthogonal spaces. We shall use the following assumption:

Assumption 1. For s an integer such that $1 \le s \le K/2$ and c_0 a positive real number, we have

$$\xi_{\min}(2s) > c_0\theta_{s,2s}.$$

Oracle inequalities for the Dantzig selector were established under Assumption 1 with $c_0 = 1$ in the parametric linear model by Candès and Tao in [10]. It was also considered by Bickel et al. [4] for non-parametric regression and for the Lasso estimate for larger values of c_0.

For any $J \subset \{1, \dots, K\}$, let us set $J^C = \{1, \dots, K\} \setminus J$ and define λ_J the vector which has the same coordinates as λ on J and zero coordinates on J^C.

We have now the following theorem:

Theorem 2. *Let us assume that Assumption 1 is true for some positive integer s and with $c_0 = 1$. On Ω, the random set introduced in Theorem 1, we have for any $\alpha > 0$*

$$\|\hat{f}^L - f\|_2^2 \le \inf_{\substack{\lambda\in\mathbb{C}^K\\ \|\hat{\lambda}\|_{\ell_1}\le\|\lambda\|_{\ell_1}}} \inf_{\substack{J\subset\{1,\dots,K\}\\ |J|=s}} \left\{ \|f_\lambda - f\|_2^2 + \alpha\left(1+\frac{2\mu_s}{\kappa_s}\right)^2 \frac{\|\lambda_{J^C}\|_{\ell_1}^2}{s} + 16s\left(\frac{1}{\alpha}+\frac{1}{\kappa_s^2}\right)\|\eta\|_{\ell_\infty}^2 \right\} \tag{23}$$

with, using (15),

$$\|\eta\|_{\ell_\infty} = \max_{k\in\{1,\dots,K\}} |\eta_k|,$$

and κ_s and μ_s are defined as follows:

$$\mu_s = \frac{\theta_{s,2s}}{\sqrt{\xi_{\min}(2s)}}, \quad \kappa_s = \sqrt{\xi_{\min}(2s)} - \mu_s.$$

Let us give an interpretation of the right hand side of inequality (23). The value of the infimum depends on three terms. The first two terms are approximation terms that naturally appear since our procedure is based on minimization of an ℓ_1-penalized $\mathbb{L}_2$-criterion and the third one can be viewed as a variance term.

Concerning the behaviour of this variance term $\|\eta\|_{\ell_\infty}^2$ our result can be closely connected to the one recently obtained by Dalalyan and Salmon [12] (see Theorem 1, Remark 4 and Section 3 devoted to ill-posed inverse problems and group weighting in [12]). Indeed, their paper deals with non-parametric regression model with heteroscedastic Gaussian noise which is known to well describe ill-posed inverse problems and they obtain the same behavior for their remaining term.

4. Numerical results

In this section, we present some numerical experiments which make a comparison in practice between the Lasso procedure described in this paper and the thresholding algorithm on needlets of Kerkyacharian et al. [16]. We aim at reconstructing a density defined on $\mathbb{S}^2$ from noisy data corrupted by small random rotations.

The target density (given in Fig. 3 on the left) presents one principal sharp mode and minor fluctuations at its basis which means that the observations are mainly concentrated in one direction and otherwise a little bit spread all over the sphere. When dealing with directional data, the most common and popular density function is given by the well-known von Mises Fisher distribution on $\mathbb{S}^2$. It has the following form:

$$f(x) = C(\sigma)\exp(\sigma\mu^T x).$$

The von Mises Fisher density is characterized by a mean direction μ and a concentration parameter σ. The greater the value of the concentration parameter, the higher the concentration of the distribution around the mean direction. This density is unimodal for $\sigma > 0$. Our test function resembles much a von Mises-Fisher distribution while being a little bit more intricate. Furthermore, if one looks at what have been done for numerical experiments in the context of directional for density estimation (see the work of Baldi et al. [2]), or for spherical deconvolution with hard thresholding techniques (see [16]), one would realize that densities close to ours have been considered.

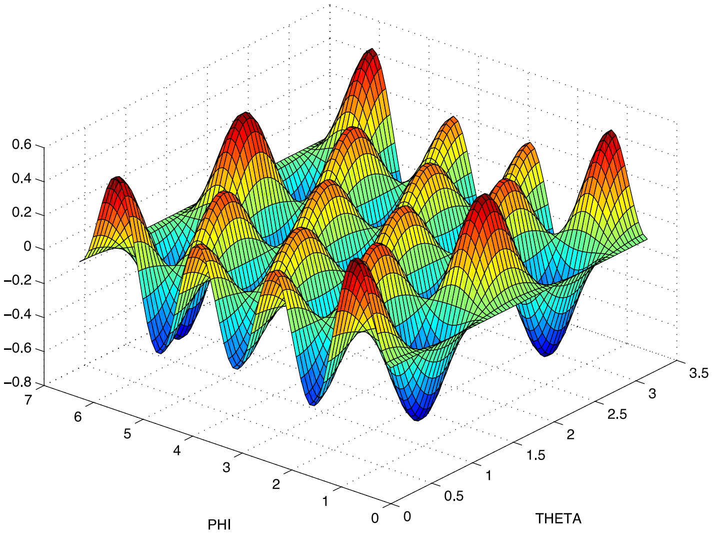

Fig. 1. A spherical harmonics of degree 8.

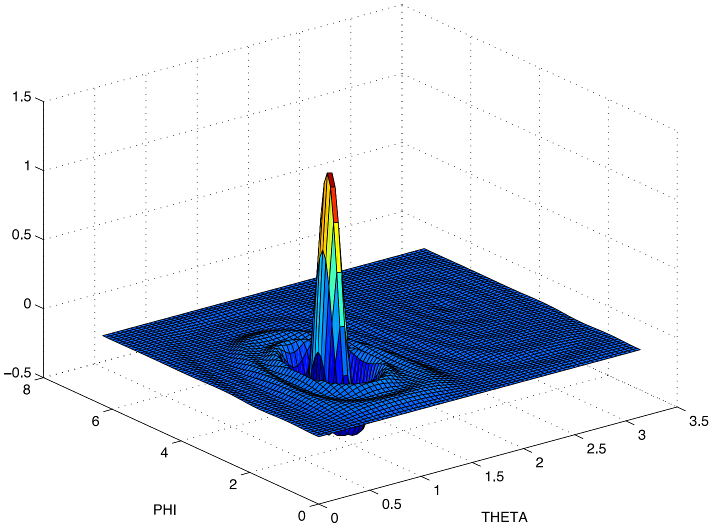

Fig. 2. A needlet at resolution 3.

For each method, we consider a data set of 800 observations generated from our target density. Then we contaminate each data with some noise which consists in a rotation about the $0z$ axis by a random angle. For each observation, the random angle is of course different and follows a uniform law on a certain interval $U[0, \alpha]$, $\alpha > 0$. The larger the interval of the uniform law, the larger the amount of noise.

So far, two approximations basis have been put forward to solve theoretically the problem of density estimation in the spherical deconvolution setting: spherical harmonics and needlets. On the one hand, spherical harmonics can be seen as the Fourier basis of $\mathbb{L}_2(\mathbb{S}^2)$ (see Fig. 1). On the other hand, roughly speaking, needlets can be seen as "spherical wavelets". Their construction is rather intricate and is based on two main steps, a Littlewood Paley decomposition and a discretization step (for further information see [24]). The needlets form a semi-orthogonal family in the sense that every two needlets which are from at least two levels apart are orthogonal. Hence they are close to an orthonormal basis. They enjoy excellent localization properties as shown in Fig. 2. We specify that in our numerical experiments, we had to compute them on our own.

Here is the expression of a needlet for a resolution level j and centered at a quadrature point $\xi_{j\eta}$:

$$\psi_{j\eta}(x) = \sqrt{\lambda_{j\eta}} \sum_{l=2^{j-1}}^{2^{j+1}} b\left(\frac{l}{2^j}\right) \sum_{m=-l}^{l} Y_m^l(\xi_{j\eta})\overline{Y_m^l(x)},$$

where $\lambda_{j\eta}$ is a quadrature weight, b is a cut-off function and Y_m^l denotes the spherical harmonics. In our experiments, to compute the quadrature points, we used the spherical pixelization HEALPix software package. HEALPix provides an approximate quadrature of the sphere with a number of data points of order 12.2^{2J} and a number of quadrature weights of order $\frac{1}{12.2^{2J}}$. This approximation is considered as reliable enough and commonly used in astrophysics.

Our dictionary mixes spherical harmonics and needlets and thus ensures a sufficiently incoherent design in the same spirit of Theorem 2. The dictionary is composed of 81 spherical harmonics and 1020 needlets, hence the cardinality of the dictionary is $K = 1101$. The number of spherical harmonics used corresponds to a maximal degree $L = 8$ and the needlets to a maximal resolution $J = 3$.

More precisely, let us describe our statistical procedure scheme.

1. Compute the $\hat{\beta}_k$ for all $k = 1 \cdots K$.
2. Compute $\hat{\sigma}^2_{1,k}$, $\hat{\sigma}^2_{2,k}$, $\tilde{\sigma}^2_{1,k}$ and $\tilde{\sigma}^2_{2,k}$ given by (17)–(20).
3. Compute the $\eta_{1,k}$ and $\eta_{2,k}$ defined in (21) and (22) with $\gamma = 1.01$.
4. Compute the coefficients $\hat{\lambda}^L$ by the Lasso minimization described in (12).
5. Select the support $\hat{J}^L$ of the estimate $\hat{\lambda}^L$.$\hat{J}^L$ defines a subset of the dictionary on which the density is regressed:
$$(\hat{\lambda}^{L'})_{\hat{J}^L} = G^{-1}_{\hat{J}^L}(\hat{\beta}_k)_{\hat{J}^L},$$
where $G_{\hat{J}^L}$ is the submatrix of the Gram matrix G corresponding to the subset $\hat{J}^L$. The values of $\hat{\lambda}^{L'}$ outside $\hat{J}^L$ are set to 0.
6. Compute the final estimate $\hat{f}^{L'} = \Re(f_{\hat{\lambda}^{L'}}) = \Re(\sum_{k=1}^K \hat{\lambda}_k^{L'}\varphi_k)$.

We shall give some remarks about this scheme. The Lasso minimization of step 4 is solved using "the homotopy method" proposed by Asif and Romberg [1]. This procedure has been fully described by Osborne et al. [25]. The Gram matrix has been pre-computed, the scalar products between the dictionary functions being computed thanks to the spherical quadrature formula, see [24]. Step 5 is a least squares step as advocated in [10] which is intended to decrease the bias introduced by the Lasso. The tuning paramater γ in the expression of $\eta_{1,k}$ and $\eta_{2,k}$ is set to 1.01.

Let us now describe briefly the needlet thresholding algorithm, all details for this procedure can be found in [16]. A needlet is denoted $\psi_{j\eta}$, $\hat{\beta}_{j\eta}$ is the estimate of the scalar product between f and $\psi_{j\eta}$, j is the resolution level and η is the quadrature point around which the corresponding needlet is almost exponentially localized. Each η belongs to a quadrature set $\mathcal{Z}_j$ provides by HEALPix and which cardinality is equal to 12.2^{2j}. We have set $J = 3$. The estimator of f is given by:

$$\hat{f}^T = \Re\left(\sum_{j=0}^{J}\sum_{\eta\in\mathcal{Z}_j}\hat{\beta}_{j\eta}\mathbf{1}\{|\hat{\beta}_{j\eta}| \geq \kappa t_N|\sigma_j|\}\psi_{j\eta}\right),$$

with

$$t_N = \sqrt{\frac{\log N}{N}},$$
$$\sigma_j^2 = A\sum_{l=2^{j-1}}^{2^{j+1}}\sum_{n\in\mathcal{I}_l}\left|\sum_{m\in\mathcal{I}_l}\psi^{*l}_{j\eta,m}(f_\varepsilon^{*l})^{-1}_{mn}\right|^2,$$

with $A \geq \|f_Z\|_\infty$. The quantity σ_j^2 constitutes an upper bound for the variance of the estimated coefficients $\hat{\beta}_{j\eta}$. Here, we have decided to estimate directly the variance of $\hat{\beta}_{j\eta}$ and plug it in the expression of $\hat{f}^T$, like in [16].

As for the tuning parameter κ, we set it to $\kappa = 1$ which gives the best results in terms of $\mathbb{L}_2$ loss.

Let us give some comments to highlight the choice of tuning parameters in both methods. This latter value of $\kappa = 1$ was not easy to find and relies on the data at stake and the type of noise, in other words it is an ad hoc choice. Moreover, $\kappa = 1$ does not correspond to what the theory says. Theorems in [16] states that for the $\mathbb{L}_2$ loss for instance, κ should be taken greater than $\frac{16}{\sqrt{3\pi}\|f\|_\infty}$ which is equal to 17 with our target density, not to mention that for real data $\|f\|_\infty$ is unknown. This ad hoc choice of κ constitutes a real drawback especially when simulations are conducted in dimension 2. On the other hand, for the Lasso, once one has set $\gamma = 1.01$ which is the smallest value allowed by theoretical arguments, the Lasso offers a full-calibrated procedure and gives good results.

Now, we present the graphics reconstructions, $\mathbb{L}_2$ and $\mathbb{L}_\infty$ losses. The estimated losses are computed over 15 runs. The sup-norm loss is computed on an almost uniform grid of $\mathbb{S}^2$ of 192 points provided by the software HEALPix.

Concerning the graphics reconstructions, we present the target density, the Lasso estimate and the needlet thresholding one for various amounts of noise.

$\mathbb{L}_2$ loss

Estimator	$\phi = \frac{\pi}{8}$	$\phi = \frac{3\pi}{16}$
Thresholding estimator	0.0014	0.0015
Calibrated lasso	0.0008	0.0027

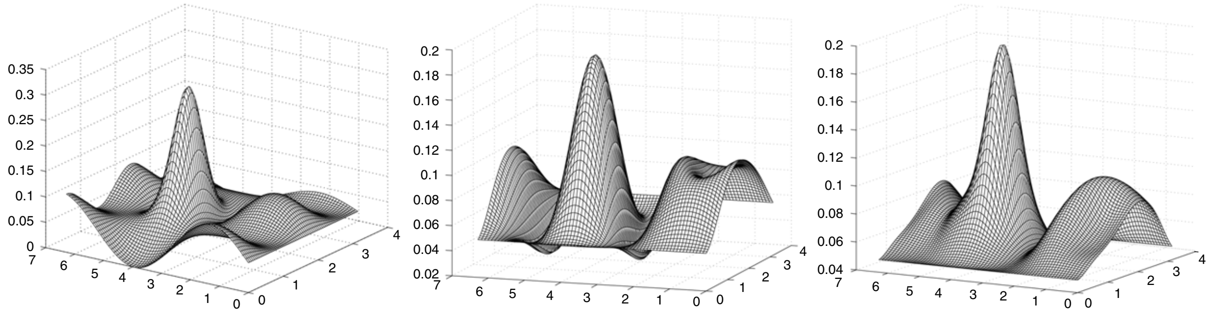

Fig. 3. The exact density, the Lasso, the needlet thresholding, $\phi \sim U[0, \frac{\pi}{16}]$.

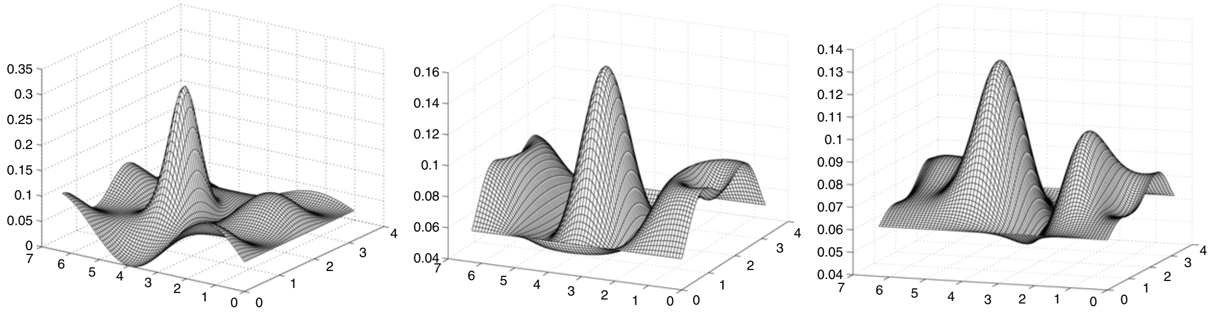

Fig. 4. The exact density, the Lasso, the needlet thresholding, $\phi \sim U[0, \frac{3\pi}{16}]$.

$\mathbb{L}_\infty$ loss

Estimator	$\phi = \frac{\pi}{8}$	$\phi = \frac{3\pi}{16}$
Thresholding estimator	0.1801	0.2056
Calibrated lasso	0.1226	0.1968

Analyzing the graphic reconstructions in Figs. 3 and 4, it appears that for the first case of noise with $\phi \sim U[0, \frac{\pi}{8}]$, the Lasso selects only four coefficients whereas the needlet thresholding procedure keeps 31 coefficients. Consequently, the Lasso enhances much better the sparsity of the signal. As we increase the noise, with $\phi \sim U[0, \frac{3\pi}{16}]$, the Lasso keeps only one coefficient, whereas the needlet thresholding algorithm keeps 27 ones. Once again, the Lasso highlights the very sparse feature of the signal. Of course, for both methods, the intensity of the peaks decreases as the random rotations scatter the observations. We precise that on the graphics, the left extremity of the estimated density is the continuation of the right one because of the spherical symmetry.

At closer inspection, both methods manage to recover the principal mode even if the small fluctuations for the Lasso are slightly flattened on the top. Although the graphic reconstructions seem a bit better for the thresholding method, both procedures are capable to localize the main peak which constitutes the most important fact in directional problems. That said, as the Lasso only keeps very few coefficients, it is pretty normal that the reconstructions are visually a bit more distorted but on the other hand we stress that we gain sparsity and clearer interpretation of our signal.

Considering the $\mathbb{L}^2$ loss and the sup-norm loss, the Lasso performs better in three of four cases and always better for the sup-norm which is a nice result.

Acknowledgment

The research of Thanh Mai Pham Ngoc and Vincent Rivoirard is partly supported by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

Appendix

A.1. Proof of Proposition 1

Straightforward computations establish Proposition 1. We have:

$$\begin{aligned}
\mathbb{E}[C(\lambda)] &= \int_{\mathbb{S}^2}\left|\sum_{k=1}^K \lambda_k\varphi_k(x)\right|^2 dx - \sum_{k=1}^K \lambda_k\beta_k - \overline{\sum_{k=1}^K \lambda_k\beta_k}\\
&= \int_{\mathbb{S}^2}\left|\sum_{k=1}^K \lambda_k\varphi_k(x)\right|^2 dx - \sum_{k=1}^K \lambda_k\int_{\mathbb{S}^2} f(x)\varphi_k(x)dx - \overline{\sum_{k=1}^K \lambda_k\int_{\mathbb{S}^2} f(x)\varphi_k(x)dx}\\
&= \int_{\mathbb{S}^2}\left|\sum_{k=1}^K \lambda_k\varphi_k(x)\right|^2 dx - \sum_{k=1}^K \lambda_k\int_{\mathbb{S}^2} f(x)\varphi_k(x)dx - \sum_{k=1}^K \overline{\lambda_k}\int_{\mathbb{S}^2} f(x)\overline{\varphi_k(x)}dx\\
&= \int_{\mathbb{S}^2}\left|\sum_{k=1}^K \lambda_k\varphi_k(x)\right|^2 dx - \int_{\mathbb{S}^2} f(x)\left(\sum_{k=1}^K \lambda_k\varphi_k(x) + \overline{\lambda_k\varphi_k(x)}\right)dx.
\end{aligned}$$

But,

$$\begin{aligned}
\left\|\sum_{k=1}^K \lambda_k\varphi_k - f\right\|_2^2 &= \int_{\mathbb{S}^2}\left(\sum_{k=1}^K \lambda_k\varphi_k(x) - f(x)\right)\left(\sum_{k=1}^K \overline{\lambda_k\varphi_k(x)} - f(x)\right)dx\\
&= \int_{\mathbb{S}^2}\left(\left|\sum_{k=1}^K \lambda_k\varphi_k(x)\right|^2 + f^2(x) - f(x)\sum_{k=1}^K \overline{\lambda_k\varphi_k(x)} - f(x)\sum_{k=1}^K \lambda_k\varphi_k(x)\right)dx\\
&= \mathbb{E}[C(\lambda)] + \|f\|_2^2,
\end{aligned}$$

which proves the result.

A.2. Proof of Proposition 2

We set for any k,

$$\begin{aligned}
\lambda_k^{(1)} &= \Re(\lambda_k), \qquad \lambda_k^{(2)} = \Im(\lambda_k),\\
\varphi_k^{(1)} &= \Re(\varphi_k), \qquad \varphi_k^{(2)} = \Im(\varphi_k)
\end{aligned}$$

and

$$\hat\beta_k^{(1)} = \Re(\hat\beta_k), \qquad \hat\beta_k^{(2)} = \Im(\hat\beta_k).$$

We now show that for any k,

$$\frac{\partial C(\lambda)}{\partial\lambda_k^{(1)}} = 2\Re((G\bar\lambda)_k - \hat\beta_k), \tag{24}$$

and

$$\frac{\partial C(\lambda)}{\partial\lambda_k^{(2)}} = 2\Im(\hat\beta_k - (G\bar\lambda)_k). \tag{25}$$

We have:

$$\begin{aligned}
\|f_\lambda\|_2^2 = \int &\left[\sum_{k=1}^K \lambda_k^{(1)}\varphi_k^{(1)}(x) - \lambda_k^{(2)}\varphi_k^{(2)}(x) + i(\lambda_k^{(2)}\varphi_k^{(1)}(x) + \lambda_k^{(1)}\varphi_k^{(2)}(x))\right]\\
&\times\left[\sum_{k=1}^K \lambda_k^{(1)}\varphi_k^{(1)}(x) - \lambda_k^{(2)}\varphi_k^{(2)}(x) - i(\lambda_k^{(2)}\varphi_k^{(1)}(x) + \lambda_k^{(1)}\varphi_k^{(2)}(x))\right]dx.
\end{aligned}$$

Let us compute partial derivatives:

$$\begin{aligned}
\frac{\partial \|f_\lambda\|_2^2}{\lambda_1^{(1)}} &= \int (\varphi_1^{(1)}(x) + i\varphi_1^{(2)}(x))\overline{f_\lambda(x)} + f_\lambda(x)(\varphi_1^{(1)}(x) - i\varphi_1^{(2)}(x))dx \\
&= \int 2\Re\left(\varphi_1(x)\overline{\sum_{k=1}^K \lambda_k\varphi_k(x)}\right)dx \\
&= 2\Re\left(\sum_{k=1}^K \overline{\lambda_k}\int \varphi_1(x)\overline{\varphi_k}(x)dx\right) \\
&= 2\Re\left(\sum_{k=1}^K \overline{\lambda_k}G_{1k}\right) = 2\Re((G\overline{\lambda})_1).
\end{aligned}$$

Besides, we have

$$\begin{aligned}
\frac{\partial \|f_\lambda\|_2^2}{\lambda_1^{(2)}} &= \int \Big((-\varphi_1^{(2)}(x) + i\varphi_1^{(1)}(x))\overline{f_\lambda(x)} + f_\lambda(x)(-\varphi_1^{(2)}(x) - i\varphi_1^{(1)}(x))\Big)dx \\
&= \int \Big(i\varphi_1(x)\overline{f_\lambda(x)} + f_\lambda(x)(-i\overline{\varphi_1}(x))\Big)dx \\
&= 2\Re\int\left(i\varphi_1(x)\overline{\sum_{k=1}^K \lambda_k\varphi_k(x)}\right)dx \\
&= 2\Re\left(i\sum_{k=1}^K \overline{\lambda_k}\int \varphi_1(x)\overline{\varphi_k(x)}dx\right) \\
&= 2\Re(i(G\overline{\lambda})_1) = -2\Im((G\overline{\lambda})_1).
\end{aligned}$$

Finally, if we set $A = 2\Re(\sum_{k=1}^K \lambda_k\hat{\beta}_k)$,

$$\begin{aligned}
A &= 2\Re\left(\sum_{k=1}^K (\lambda_k^{(1)} + i\lambda_k^{(2)})(\hat{\beta}_k^{(1)} + i\hat{\beta}_k^{(2)})\right) \\
&= 2\sum_{k=1}^K (\lambda_k^{(1)}\hat{\beta}_k^{(1)} - \lambda_k^{(2)}\hat{\beta}_k^{(2)})
\end{aligned}$$

hence we have

$$\frac{\partial A}{\lambda_1^{(1)}} = 2\Re(\hat{\beta}_1) \qquad \frac{\partial A}{\lambda_1^{(2)}} = -2\Im(\hat{\beta}_1),$$

which completes the proofs of (24) and (25). KKT first-order conditions end the proof of Proposition 2.

A.3. Proof of Theorem 1

We only make the proof for the real part. Identical arguments hold for the imaginary part.

First of all, let us establish that $\hat{\sigma}_{1,k}^2$ is an unbiased estimator of the variance $\sigma_{1,k}^2$. We have

$$\hat{\sigma}_{1,k}^2 = \frac{1}{2N(N-1)}\sum_{i\neq j}(\Re^2(\phi_k(Z_i)) + \Re^2(\phi_k(Z_j)) - 2\Re(\phi_k(Z_i))\Re(\phi_k(Z_j))).$$

So,

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}_{1,k}^2) &= \frac{N(N-1)}{N(N-1)}\mathbb{E}(\Re^2(\phi_k(Z_1))) - \frac{1}{N(N-1)}\sum_{i\neq j}\mathbb{E}(\Re(\phi_k(Z_i)))\mathbb{E}(\Re(\phi_k(Z_j))) \\
&= \mathbb{E}(\Re^2(\phi_k(Z_1))) - (\mathbb{E}(\Re(\phi_k(Z_1))))^2 \\
&= \sigma_{1,k}^2.
\end{aligned}$$

Now, we set

$$W_i = \Re\left(\frac{1}{N}(\phi_k(Z_i) - \beta_k)\right).$$

As

$$W_i = \Re\left(\frac{1}{N}\left(\phi_k(Z_i) - \int_{\mathbb{S}^2}\phi_k(x)f_Z(x)dx\right)\right)$$

then the W_i satisfy almost surely

$$|W_i| \le \frac{2\|\Re(\phi_k)\|_\infty}{N}.$$

Then, we apply Bernstein's Inequality (see [21, on pp. 24 and 26)]) with the variables W_i and $-W_i$: for any $u > 0$,

$$\mathbb{P}\left(|\Re(\hat{\beta}_k - \beta_k)| \ge \sqrt{\frac{2\sigma_{1,k}^2 u}{N}} + \frac{2u\|\Re(\phi_k)\|_\infty}{3N}\right) \le 2e^{-u}. \tag{26}$$

Now, let us decompose $\hat{\sigma}_{1,k}^2$ in two terms:

$$\begin{aligned}\hat{\sigma}_{1,k}^2 &= \frac{1}{2N(N-1)}\sum_{i\neq j}(\Re(\phi_k(Z_i) - \phi_k(Z_j)))^2 \\ &= \frac{1}{2N}\sum_{i=1}^N(\Re(\phi_k(Z_i) - \beta_k))^2 + \frac{1}{2N}\sum_{j=1}^N(\Re(\phi_k(Z_j) - \beta_k))^2 \\ &\quad - \frac{2}{N(N-1)}\sum_{i=2}^N\sum_{j=1}^{i-1}(\Re(\phi_k(X_i) - \beta_k))(\Re(\phi_k(Z_j) - \beta_k)) \\ &= s_N - \frac{2}{N(N-1)}u_N\end{aligned}$$

with

$$s_N = \frac{1}{N}\sum_{i=1}^N(\Re(\phi_k(Z_i) - \beta_k))^2 \quad \text{and} \quad u_N = \sum_{i=2}^N\sum_{j=1}^{i-1}(\Re(\phi_k(Z_i) - \beta_k))(\Re(\phi_k(Z_j) - \beta_k)). \tag{27}$$

Let us first focus on s_N that is the main term of $\hat{\sigma}_{1,k}^2$ by applying again Bernstein's Inequality with

$$Y_i = \frac{\sigma_{1,k}^2 - (\Re(\phi_k(Z_i) - \beta_k))^2}{N}$$

which satisfies

$$Y_i \le \frac{\sigma_{1,k}^2}{N}.$$

One has that for any $u > 0$

$$\mathbb{P}\left(\sigma_{1,k}^2 \ge s_N + \sqrt{2v_k u} + \frac{\sigma_{1,k}^2 u}{3N}\right) \le e^{-u}$$

with

$$v_k = \frac{1}{N}\mathbb{E}\left(\left[\sigma_{1,k}^2 - (\Re(\phi_k(Z_i) - \beta_k))^2\right]^2\right).$$

But we have

$$\begin{aligned}v_k &= \frac{1}{N}\left(\sigma_{1,k}^4 + \mathbb{E}\left[\Re(\phi_k(Z_i) - \beta_k)^4\right] - 2\sigma_{1,k}^2\mathbb{E}\left[\Re(\phi_k(Z_i) - \beta_k)^2\right]\right) \\ &= \frac{1}{N}\left(\mathbb{E}\left[\Re(\phi_k(Z_i) - \beta_k)^4\right] - \sigma_{1,k}^4\right) \\ &\le \frac{\sigma_{1,k}^2}{N}\left(\|\Re(\phi_k)\|_\infty + |\Re(\beta_k)|\right)^2 \\ &\le \frac{4\sigma_{1,k}^2}{N}\|\Re(\phi_k)\|_\infty^2.\end{aligned}$$

Finally, with for any $u > 0$

$$S(u) = 2\sqrt{2}\sigma_{1,k}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}} + \frac{\sigma_{1,k}^2 u}{3N},$$

we have

$$\mathbb{P}(\sigma_{1,k}^2 \geq s_N + S(u)) \leq e^{-u}. \tag{28}$$

The term u_N is a degenerate U-statistics that satisfies for any $u > 0$

$$\mathbb{P}(|u_N| \geq U(u)) \leq 6e^{-u}, \tag{29}$$

with for any $u > 0$

$$U(u) = \frac{4}{3}Au^2 + \left(4\sqrt{2} + \frac{2}{3}\right)Bu^{\frac{3}{2}} + \left(2D + \frac{2}{3}F\right)u + 2\sqrt{2}C\sqrt{u},$$

where A, B, C, D and F are constants not depending on u that satisfy

$$\begin{aligned} &A \leq 4\|\Re(\phi_k)\|_\infty^2,\\ &B \leq 2\sqrt{N-1}\|\Re(\phi_k)\|_\infty^2,\\ &C \leq \sqrt{\frac{N(N-1)}{2}}\sigma_{1,k}^2,\\ &D \leq \sqrt{\frac{N(N-1)}{2}}\sigma_{1,k}^2, \end{aligned}$$

and

$$F \leq 2\sqrt{2}\|\Re(\phi_k)\|_\infty^2\sqrt{(N-1)\log(2N)}$$

(see [26]). Then, we have for any $u > 0$,

$$\begin{aligned} \frac{2}{N(N-1)}U(u) \leq\; & \frac{32}{3}\frac{\|\Re(\phi_k)\|_\infty^2}{N(N-1)}u^2 + \left(16\sqrt{2} + \frac{8}{3}\right)\frac{\|\Re(\phi_k)\|_\infty^2}{N\sqrt{N-1}}u^{\frac{3}{2}}\\ & + \left(2\sqrt{2}\frac{\sigma_{1,k}^2}{\sqrt{N(N-1)}} + \frac{8\sqrt{2}}{3}\frac{\sqrt{\log(2N)}\|\Re(\phi_k)\|_\infty^2}{N\sqrt{N-1}}\right)u + \frac{4\sigma_{1,k}^2}{\sqrt{N(N-1)}}\sqrt{u}. \end{aligned}$$

Now, we take u that satisfies

$$u = o(N) \tag{30}$$

and

$$\sqrt{\log(2N)} \leq \sqrt{2u}. \tag{31}$$

Therefore, for any $\varepsilon_1 > 0$, we have for N large enough,

$$\frac{2}{N(N-1)}U(u) \leq \varepsilon_1\sigma_{1,k}^2 + \left(16\sqrt{2} + 8\right)\frac{\|\Re(\phi_k)\|_\infty^2}{N\sqrt{N-1}}u^{\frac{3}{2}} + \frac{32}{3}\frac{\|\Re(\phi_k)\|_\infty^2}{N(N-1)}u^2.$$

So, for N large enough,

$$\frac{2}{N(N-1)}U(u) \leq \varepsilon_1\sigma_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}, \tag{32}$$

where $C_1 = 16\sqrt{2} + 19$. Using Inequalities (28) and (29), we obtain

$$\begin{aligned} \mathbb{P}\left(\sigma_{1,k}^2 \geq \hat{\sigma}_{1,k}^2 + S(u) + \frac{2}{N(N-1)}U(u)\right) &= \mathbb{P}\left(\sigma_{1,k}^2 \geq s_N - \frac{2}{N(N-1)}u_N + S(u) + \frac{2}{N(N-1)}U(u)\right)\\ &\leq \mathbb{P}\left(\sigma_{1,k}^2 \geq s_N + S(u)\right) + \mathbb{P}\left(u_N \geq U(u)\right)\\ &\leq 7e^{-u}. \end{aligned}$$

Now, using (32), for any $0 < \varepsilon_2 < 1$, we have for N large enough,

$$\begin{aligned}\hat{\sigma}_{1,k}^2 + S(u) + \frac{2}{N(N-1)}U(u) &= \hat{\sigma}_{1,k}^2 + 2\sqrt{2}\sigma_{1,k}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}} + \frac{\sigma_{1,k}^2 u}{3N} + \frac{2}{N(N-1)}U(u)\\ &\le \hat{\sigma}_{1,k}^2 + 2\sqrt{2}\sigma_{1,k}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}} + \frac{\sigma_{1,k}^2 u}{3N} + \varepsilon_1\sigma_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}\\ &\le \hat{\sigma}_{1,k}^2 + 2\sqrt{2}\sigma_{1,k}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}} + \varepsilon_2\sigma_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}.\end{aligned}$$

Therefore,

$$\mathbb{P}\left((1-\varepsilon_2)\sigma_{1,k}^2 \ge \hat{\sigma}_{1,k}^2 + 2\sqrt{2}\sigma_{1,k}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}} + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}\right) \le 7e^{-u}. \tag{33}$$

Now, let us set

$$a = 1-\varepsilon_2, \qquad b = \sqrt{2}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}}, \qquad c = \hat{\sigma}_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}$$

and consider the polynomial

$$P(x) = ax^2 - 2bx - c,$$

with roots $\frac{b\pm\sqrt{b^2+ac}}{a}$. So, we have

$$\begin{aligned}P(\sigma_{1,k}) \ge 0 &\iff \sigma_{1,k} \ge \frac{b+\sqrt{b^2+ac}}{a}\\ &\iff \sigma_{1,k}^2 \ge \frac{c}{a} + \frac{2b^2}{a^2} + \frac{2b\sqrt{b^2+ac}}{a^2}.\end{aligned}$$

It yields

$$\mathbb{P}\left(\sigma_{1,k}^2 \ge \frac{c}{a} + \frac{2b^2}{a^2} + \frac{2b\sqrt{b^2+ac}}{a^2}\right) \le 7e^{-u},$$

so,

$$\mathbb{P}\left(\sigma_{1,k}^2 \ge \frac{c}{a} + \frac{4b^2}{a^2} + \frac{2b\sqrt{c}}{a\sqrt{a}}\right) \le 7e^{-u},$$

which means that for any $0 < \varepsilon_3 < 1$, we have for N large enough,

$$\begin{aligned}\mathbb{P}\Bigg(\sigma_{1,k}^2 \ge (1+\varepsilon_3)\Bigg(&\hat{\sigma}_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}} + 8\|\Re(\phi_k)\|_\infty^2\frac{u}{N}\\ &+ 2\sqrt{2}\|\Re(\phi_k)\|_\infty\sqrt{\frac{u}{N}}\sqrt{\hat{\sigma}_{1,k}^2 + C_1\|\Re(\phi_k)\|_\infty^2\left(\frac{u}{N}\right)^{\frac{3}{2}}}\Bigg)\Bigg) \le 7e^{-u}.\end{aligned}$$

Finally, we can claim that for any $0 < \varepsilon_4 < 1$, we have for N large enough,

$$\mathbb{P}\left(\sigma_{1,k}^2 \ge (1+\varepsilon_4)\left(\hat{\sigma}_{1,k}^2 + 8\|\Re(\phi_k)\|_\infty^2\frac{u}{N} + 2\|\Re(\phi_k)\|_\infty\sqrt{2\hat{\sigma}_{1,k}^2\frac{u}{N}}\right)\right) \le 7e^{-u}.$$

Now, we take $u = \gamma\log K$ with $\gamma > 1$. Since $N \le K \le \exp(N^\delta)$ with $\delta < 1$ then (30) and (31) are satisfied. The previous concentration inequality means that

$$\mathbb{P}\left(\sigma_{1,k}^2 \ge (1+\varepsilon_4)\tilde{\sigma}_{1,k}^2\right) \le 7K^{-\gamma}.$$

Now, using (26), we have for N large enough,

$$\begin{aligned}\mathbb{P}\left(|\Re(\beta_k-\hat{\beta}_k)| \ge \eta_{1,k}\right) &= \mathbb{P}\left(|\Re(\beta_k-\hat{\beta}_k)| \ge \sqrt{\frac{2\tilde{\sigma}_{1,k}^2\gamma\log K}{N}} + \frac{2\|\Re(\phi_k)\|_\infty\gamma\log K}{3N}, \sigma_{1,k}^2 < (1+\varepsilon_4)\tilde{\sigma}_{1,k}^2\right)\\ &\quad + \mathbb{P}\left(|\Re(\beta_k-\hat{\beta}_k)| \ge \eta_{1,k}, \sigma_{1,k}^2 \ge (1+\varepsilon_4)\tilde{\sigma}_{1,k}^2\right)\end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{P}\left(|\Re(\beta_k-\hat{\beta}_k)| \geq \sqrt{\frac{2\sigma_{1,k}^2\gamma(1+\varepsilon_4)^{-1}\log K}{N}} + \frac{2\|\Re(\phi_k)\|_\infty\gamma(1+\varepsilon_4)^{-1}\log K}{3N}\right)\\
&\quad + \mathbb{P}\left(\sigma_{1,k}^2 \geq (1+\varepsilon_4)\tilde{\sigma}_{1,k}^2\right)\\
&\leq 2K^{-\gamma(1+\varepsilon_4)^{-1}} + 7K^{-\gamma}.
\end{aligned}$$

Then, the first part of Theorem 1 is proved: for any $\varepsilon > 0$,

$$\mathbb{P}\left(|\Re(\beta_k-\hat{\beta}_k)| \geq \eta_{1,k}\right) \leq C_1(\varepsilon,\delta,\gamma)K^{-\frac{\gamma}{1+\varepsilon}},$$

where $C_1(\varepsilon,\delta,\gamma)$ is a constant that depends on ε, δ and γ.

As explained in the beginning of the proof, a similar result holds for $|\Im(\beta_k-\hat{\beta}_k)|$, so Theorem 1 is true.

A.4. Proof of Theorem 2

We first state the following lemma.

Lemma 1. *Let $J_0 \subset \{1,\dots,K\}$ with cardinality $|J_0| = s$ and $\Delta \in \mathbb{C}^K$. We have:*

$$\|f_\Delta\|_2 \geq \sqrt{\xi_{\min}(2s)}\|\Delta_{J_0}\|_{\ell_2} - \frac{\mu_s}{\sqrt{s}}\|\Delta_{J_0^c}\|_{\ell_1},$$

with

$$\mu_s = \frac{\theta_{s,2s}}{\sqrt{\xi_{\min}(2s)}}.$$

Proof. We denote by J_1 the subset of $\{1,\dots,K\}$ corresponding to the s largest coordinates of Δ (in modulus) outside J_0 and we set $J_{01} = J_0 \cup J_1$. We denote by $P_{J_{01}}$ the projector on the linear space spanned by $(\varphi_k)_{k\in J_{01}}$. For $k > 1$, we denote by J_k the indices corresponding to the coordinates of Δ outside J_0 whose absolute values are between the $((k-1)\times s+1)$-th and the $(k\times s)$-th largest ones (in absolute value). Note that this definition is consistent with the definition of J_1. Using this notation, we have

$$\begin{aligned}
\|P_{J_{01}}f_\Delta\|_2 &\geq \|P_{J_{01}}f_{\Delta_{J_{01}}}\|_2 - \left\|\sum_{k\geq 2}P_{J_{01}}f_{\Delta_{J_k}}\right\|_2\\
&\geq \|f_{\Delta_{J_{01}}}\|_2 - \sum_{k\geq 2}\|P_{J_{01}}f_{\Delta_{J_k}}\|_2.
\end{aligned}$$

Since J_{01} has $2s$ elements, we have

$$\|f_{\Delta_{J_{01}}}\|_2 \geq \sqrt{\xi_{\min}(2s)}\|\Delta_{J_{01}}\|_{\ell_2}.$$

Note that $P_{J_{01}}f_{\Delta_{J_k}} = f_{C_{J_{01}}}$ for some vector $C \in \mathbb{C}^K$. Since,

$$\langle P_{J_{01}}f_{\Delta_{J_k}} - f_{\Delta_{J_k}}, P_{J_{01}}f_{\Delta_{J_k}}\rangle = 0,$$

one obtains that

$$\|P_{J_{01}}f_{\Delta_{J_k}}\|_2^2 = \langle f_{\Delta_{J_k}}, f_{C_{J_{01}}}\rangle$$

and thus

$$\begin{aligned}
\|P_{J_{01}}f_{\Delta_{J_k}}\|_2^2 &\leq \theta_{s,2s}\|\Delta_{J_k}\|_{\ell_2}\|C_{J_{01}}\|_{\ell_2} \leq \theta_{s,2s}\|\Delta_{J_k}\|_{\ell_2}\frac{\|f_{C_{J_{01}}}\|_2}{\sqrt{\xi_{\min}(2s)}}\\
&\leq \frac{\theta_{s,2s}}{\sqrt{\xi_{\min}(2s)}}\|\Delta_{J_k}\|_{\ell_2}\|P_{J_{01}}f_{\Delta_{J_k}}\|_2.
\end{aligned}$$

This implies that

$$\|P_{J_{01}}f_{\Delta_{J_k}}\|_2 \leq \frac{\theta_{s,2s}}{\sqrt{\xi_{\min}(2s)}}\|\Delta_{J_k}\|_{\ell_2} = \mu_s\|\Delta_{J_k}\|_{\ell_2}.$$

Now using that $\|\Delta_{J_{k+1}}\|_{\ell_2} \le \|\Delta_{J_k}\|_{\ell_1}/\sqrt{s}$, we obtain

$$\sum_{k\ge 2} \|P_{J_{01}} f_{\Delta_{J_k}}\|_2 \le \frac{\mu_s}{\sqrt{s}} \|\Delta_{J_0^c}\|_{\ell_1}$$

and

$$\|P_{J_{01}} f_\Delta\|_2 \ge \sqrt{\xi_{\min}(2s)}\|\Delta_{J_{01}}\|_{\ell_2} - \frac{\mu_s}{\sqrt{s}}\|\Delta_{J_0^c}\|_{\ell_1},$$

which finally leads to

$$\|f_\Delta\|_2 \ge \sqrt{\xi_{\min}(2s)}\|\Delta_{J_0}\|_{\ell_2} - \frac{\mu_s}{\sqrt{s}}\|\Delta_{J_0^c}\|_{\ell_1}. \quad \square$$

Now, let $\lambda \in \mathbb{C}^K$ and $J \subset \{1, \dots, K\}$ such that $|J| = s$. We set $\Delta = \lambda - \hat{\lambda}$ where $\hat{\lambda}$ stands for $\hat{\lambda}^L$. The ℓ_1-norm of Δ satisfies the inequality stated in the following lemma.

Lemma 2. *Using assumptions of Theorem 2, we have:*

$$\|\Delta\|_{\ell_1} \le \frac{2\sqrt{|J|}}{\kappa_s}\|f_\Delta\|_2 + 2\|\lambda_{J^c}\|_{\ell_1}\left(1 + \frac{2\mu_s}{\kappa_s}\right).$$

Proof. Since

$$\|\hat{\lambda}\|_{\ell_1} \le \|\lambda\|_{\ell_1},$$

we have

$$\|\Delta_J - \lambda_J\|_{\ell_1} + \|\Delta_{J^c} - \lambda_{J^c}\|_{\ell_1} \le \|\lambda_J\|_{\ell_1} + \|\lambda_{J^c}\|_{\ell_1},$$

and thus

$$\|\lambda_J\|_{\ell_1} - \|\Delta_J\|_{\ell_1} + \|\Delta_{J^c}\|_{\ell_1} - \|\lambda_{J^c}\|_{\ell_1} \le \|\lambda_J\|_{\ell_1} + \|\lambda_{J^c}\|_{\ell_1}.$$

So, we have

$$\|\Delta_{J^c}\|_{\ell_1} - \|\Delta_J\|_{\ell_1} \le 2\|\lambda_{J^c}\|_{\ell_1}. \tag{34}$$

Using Lemma 1 with $J_0 = J$, we obtain that

$$\|f_\Delta\|_2 \ge \sqrt{\xi_{\min}(2s)}\|\Delta_J\|_{\ell_2} - \frac{\mu_s}{\sqrt{|J|}}(\|\Delta_J\|_{\ell_1} + 2\|\lambda_{J^c}\|_{\ell_1}).$$

Using $\|\Delta_J\|_{\ell_1} \le \sqrt{|J|}\|\Delta_J\|_{\ell_2}$, we deduce that

$$\begin{aligned}\|f_\Delta\|_2 &\ge \left(\sqrt{\xi_{\min}(2s)} - \mu_s\right)\|\Delta_J\|_{\ell_2} - \frac{2\mu_s}{\sqrt{|J|}}\|\lambda_{J^c}\|_{\ell_1}\\ &\ge \kappa_s\|\Delta_J\|_{\ell_2} - \frac{2\mu_s}{\sqrt{|J|}}\|\lambda_{J^c}\|_{\ell_1},\end{aligned}$$

and thus

$$\|\Delta_J\|_{\ell_2} \le \frac{1}{\kappa_s}\|f_\Delta\|_2 + 2\frac{\mu_s}{\sqrt{|J|}\kappa_s}\|\lambda_{J^c}\|_{\ell_1}.$$

By using again (34), we deduce then

$$\begin{aligned}\|\Delta\|_{\ell_1} &\le 2\|\Delta_J\|_{\ell_1} + 2\|\lambda_{J^c}\|_{\ell_1}\\ &\le 2\sqrt{|J|}\|\Delta_J\|_{\ell_2} + 2\|\lambda_{J^c}\|_{\ell_1}\\ &\le \frac{2\sqrt{|J|}}{\kappa_s}\|f_\Delta\|_2 + 2\|\lambda_{J^c}\|_{\ell_1}\left(1 + \frac{2\mu_s}{\kappa_s}\right),\end{aligned}$$

which ends the proof of the lemma. $\square$

Now, let us focus on the proof of Theorem 2. We have:

$$\begin{aligned}\|f_\lambda - f\|_2^2 &= \int (f_\lambda(x) - f(x))\left(\overline{f_\lambda(x)} - f(x)\right) dx\\ &= \|f_\lambda - f_{\hat\lambda}\|_2^2 + \|f - f_{\hat\lambda}\|_2^2 + 2\Re\left[\int \left(f_\lambda(x) - f_{\hat\lambda}(x)\right)\left(\overline{f_{\hat\lambda}(x)} - f(x)\right) dx\right]\\ &= \|f_\Delta\|_2^2 + \|f - f_{\hat\lambda}\|_2^2 + 2\Re\left[\int \sum_{k=1}^{K} \Delta_k \varphi_k(x) \times \left(\sum_{k'=1}^{K} \overline{\hat\lambda_{k'}\varphi_{k'}(x)} - f(x)\right) dx\right].\end{aligned}$$

So, using Proposition 2, on Ω,

$$\begin{aligned}\|f_{\hat\lambda} - f\|_2^2 &= \|f_\lambda - f\|_2^2 - \|f_\Delta\|_2^2 - 2\Re\left[\sum_{k=1}^{K} \Delta_k \left((G\overline{\hat\lambda})_k - \hat\beta_k + \hat\beta_k - \beta_k\right)\right]\\ &\le \|f_\lambda - f\|_2^2 - \|f_\Delta\|_2^2 + 2\sum_{k=1}^{K} |\Delta_k| \times \left(2\sqrt{\eta_{1,k}^2 + \eta_{2,k}^2}\right)\\ &\le \|f_\lambda - f\|_2^2 - \|f_\Delta\|_2^2 + 4\|\eta\|_{\ell_\infty}\|\Delta\|_{\ell_1}. \end{aligned} \tag{35}$$

Now, we use Lemma 2 to obtain

$$\begin{aligned}4\|\eta\|_{\ell_\infty}\|\Delta\|_{\ell_1} &\le \frac{8\sqrt{|J|}}{\kappa_s}\|\eta\|_{\ell_\infty}\|f_\Delta\|_2 + 8\|\lambda_{J^c}\|_{\ell_1}\left(1 + \frac{2\mu_s}{\kappa_s}\right)\|\eta\|_{\ell_\infty}\\ &\le \frac{16|J|}{\kappa_s^2}\|\eta\|_{\ell_\infty}^2 + \|f_\Delta\|_2^2 + 8\|\lambda_{J^c}\|_{\ell_1}\left(1 + \frac{2\mu_s}{\kappa_s}\right)\|\eta\|_{\ell_\infty},\end{aligned}$$

so, we have for any $\alpha > 0$,

$$4\|\eta\|_{\ell_\infty}\|\Delta\|_{\ell_1} - \|f_\Delta\|_2^2 \le 16|J|\left(\frac{1}{\alpha} + \frac{1}{\kappa_s^2}\right)\|\eta\|_{\ell_\infty}^2 + \alpha\frac{\|\lambda_{J^c}\|_{\ell_1}^2}{|J|}\left(1 + \frac{2\mu_s}{\kappa_s}\right)^2. \tag{36}$$

Since $\|\hat f^L - f\|_2 \le \|f_{\hat\lambda^L} - f\|_2$, (35) and (36) yield the result.

References

[1] M.S. Asif, J. Romberg, Dantzig selector homotopy with dynamic measurements, in: Proceedings of SPIE Computational Imaging VII, 2009.

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

[3] K. Bertin, E. Le Pennec, V. Rivoirard, Adaptive Dantzig density estimation, Annales de l'Institut Henri Poincaré. Probabilités et Statistiques 47 (1) (2011) 43–74.

[4] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, Annals of Statistics 37 (4) (2009) 1705–1732.

[5] F. Bunea, A.B. Tsybakov, M.H. Wegkamp, Aggregation and sparsity via ℓ_1 penalized least squares, in: G. Lugosi, H.U. Simon (Eds.), Proceedings of 19th Annual Conference on Learning Theory, COLT 2006, in: Lecture Notes in Artificial Intelligence, vol. 4005, Springer-Verlag, Berlin, Heidelberg, 2006.

[6] F. Bunea, A.B. Tsybakov, M.H. Wegkamp, Aggregation for Gaussian regression, Annals of Statistics 35 (4) (2007) 1674–1697.

[7] F. Bunea, A.B. Tsybakov, M.H. Wegkamp, Sparse density estimation with l_1 penalties, Lecture Notes in Artificial Intelligence 4539 (2007) 530–543.

[8] F. Bunea, A.B. Tsybakov, M.H. Wegkamp, Sparsity oracle inequalities for the LASSO, Electronic Journal of Statistics 1 (2007) 169–194.

[9] F. Bunea, A.B. Tsybakov, M.H. Wegkamp, Spades and mixture models, Annals of Statistics 38 (4) (2010) 2525–2558.

[10] E.J. Candès, T. Tao, The Dantzig selector: statistical estimation when p is much larger than n, Annals of Statistics 35 (6) (2007) 2313–2351.

[11] D. Chen, D.L. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43 (2001) 129–159.

[12] A. Dalalyan, J. Salmon, Sharp oracle inequalities for aggregation of affine estimators, 2011. Preprint.

[13] D.L. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Transactions on Information Theory 52 (2006) 6–18.

[14] D.M. Healy, H. Hendriks, P.T. Kim, Spherical deconvolution, Journal of Multivariate Analysis 67 (1) (2002) 1–22.

[15] A. Juditsky, S. Lambert-Lacroix, On minimax density estimation on $\mathbb{R}$, Bernoulli 10 (2) (2004) 187–220.

[16] G. Kerkyacharian, T.M. Pham Ngoc, D. Picard, Localized spherical deconvolution, Annals of Statistics 39 (2) (2011) 1042–1068.

[17] Peter T. Kim, J.Y. Koo, Optimal spherical deconvolution, Journal of Multivariate Analysis 80 (1) (2002) 21–42.

[18] P.T. Kim, J.Y. Koo, H.J. Park, Sharp minimaxity and spherical deconvolution for super-smooth error distributions, Journal of Multivariate Analysis 90 (2) (2004) 384–392.

[19] J.M. Loubes, ℓ^1 penalty for ill-posed inverse problems, Communications in Statistics—Theory and Methods 39 (2008) 1399–1411.

[20] K. Lounici, Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators, Electronic Journal of Statistics 2 (2008).

[21] P. Massart, Concentration inequalities and model selection, in: Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, Springer, Berlin, 2007.

[22] P. Massart, C. Meynet, An ℓ_1 oracle inequality for the Lasso, 2010. Preprint.

[23] N. Meinshausen, P. Buhlmann, High dimensional graphs and variable selection with the Lasso, Annals of Statistics 34 (3) (2006) 1436–1462.

[24] F.J. Narcowich, P. Petrushev, J.D. Ward, Localized tight frames on spheres, SIAM Journal on Mathematical Analysis 38 (2) (2006) 574–594.

[25] M.R. Osborne, B. Presnell, B.A. Turlach, A new approach to variable selection in least squares problems, IMA Journal of Numerical Analysis 20 (2000) 389–404.

[26] P. Reynaud-Bouret, V. Rivoirard, C. Tuleau, On the influence of the support of functions for density estimation, Journal of Statistical Planning and Inference 141 (2009) 115–139.

[27] R. Tibshirani, Regression shrinkage and selection via the Lasso, Journal of the Royal Statistical Society, Series B 58 (1996) 267–288.
[28] J.D. Tournier, F. Calamante, D.G. Gadian, A. Connelly, Direct estimation of the fiber orientation density function form diffusion-weihted MRI data using spherical deconvolution, NeuroImage 23 (2004) 1176–1185.
[29] S. van de Geer, High dimensional generalized linear models and the Lasso, Annals of Statistics 36 (2) (2008) 614–645.
[30] N.J. Vilenkin, Fonctions Spéciales et Théorie de la Représentation des Groupes, in: Monographies Universitaires de Mathématiques, vol. 33, Dunod, Paris, 1969.
[31] B. Yu, P. Zhao, On model selection consistency of Lasso estimators, Journal of Machine Learning Research 7 (2006) 2541–2567.
[32] C.H. Zhang, J. Huang, The sparsity and bias of the Lasso selection in high-dimensional linear regression, Annals of Statistics 36 (4) (2008) 1567–1594.
[33] H. Zou, The adaptive Lasso and its oracle properties, Journal of the American Statistical Association 101 (476) (2006) 1418–1429.