



A note on adaptive wavelet estimation in a shifted curves model via block thresholding

Christophe Chesneau, Jalal M. Fadili

► To cite this version:

Christophe Chesneau, Jalal M. Fadili. A note on adaptive wavelet estimation in a shifted curves model via block thresholding. 2009. hal-00421743

HAL Id: hal-00421743

<https://hal.science/hal-00421743>

Preprint submitted on 2 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A note on adaptive wavelet estimation in a shifted curves model via block thresholding

Christophe Chesneau · Jalal Fadili

Received:

Abstract In this paper, the problem of adaptive estimation of a mean pattern in a randomly shifted curve model is considered. Adopting the new point of view of Bigot and Gadat (2008), we develop an adaptive estimator based on wavelet block thresholding. Taking the minimax approach, we prove that it attains near optimal rates of convergence under the quadratic risk over a wide range of Besov balls. In comparison to the procedure of Bigot and Gadat (2008), we gain a logarithmic term in the rates of convergence (for the regular zone).

Keywords Mean pattern estimation · Deconvolution · Adaptive curve estimation · Wavelets · Block thresholding

1 Introduction

1.1 Model and motivation

Suppose that we observe realizations of n noisy and randomly shifted curves $Y_1(t), \dots, Y_n(t)$ defined by the stochastic equation:

$$dY_m(t) = f(t - \tau_m)dt + \epsilon dW_m(t), \quad t \in [0, 1], \quad m \in \{1, \dots, n\}, \quad (1)$$

where f is an unknown function, $\epsilon > 0$ is a fixed constant, $W_1(t), \dots, W_m(t)$ are n non-observed i.i.d. standard Brownian motion and τ_1, \dots, τ_n are i.i.d. random variables. The density function of τ_1 is denoted g . It is supposed to be known. The goal is to estimate f from $Y_1(t), \dots, Y_n(t)$.

C. Chesneau
Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie,
Campus II, Science 3, 14032 Caen, France. E-mail: chesneau@math.unicaen.fr

J. Fadili
GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen France. E-mail:
Jalal.Fadili@greyc.ensicaen.fr

We assume in the sequel that f and g belong to $\mathbb{L}_{\text{per}}^2([0, 1])$, the space of periodic functions of period one that are square-integrable on $[0, 1]$,

$$\mathbb{L}_{\text{per}}^2([0, 1]) = \left\{ h; h \text{ is one-periodic and } \int_0^1 h^2(t) dt < \infty \right\}.$$

Any function $h \in \mathbb{L}_{\text{per}}^2([0, 1])$ can be represented by its Fourier series

$$h(t) = \sum_{\ell \in \mathbb{Z}} \mathcal{F}(h)(\ell) e^{i2\pi\ell t},$$

where the equality is intended in mean-square convergence sense¹, and $\mathcal{F}(h)(\ell)$ denotes the Fourier coefficient given by

$$\mathcal{F}(h)(\ell) = \int_0^1 h(t) e^{-2i\pi\ell t} dt, \quad \ell \in \mathbb{Z},$$

whenever this integral exists. The notation $\overline{\cdot}$ will be used for the complex conjugate.

Model (1) is useful for studying the problem of recovering a mean pattern from a set of similar curves in the presence of random translations and additive noise. It fits the setting of Grenander's theory of random shapes and patterns Grenander (1993). This problem has been the focus of many works in a variety of situations. The interested reader may refer to Bigot and Gadat (2008) and references therein for a comprehensive review. It can also be viewed as a "randomized" version of the standard Gaussian (periodic) deconvolution problem, investigated for instance by Cavalier and Tsybakov (2002), Johnstone et al. (2004) and others.

Before detailing the contribution of this paper, let us briefly describe the approach developed by Bigot and Gadat (2008). It consists in using the properties of the Fourier transform to convert model (1) into a linear inverse problem where g plays the role of a convolution kernel. Starting from this new formulation (see Section 3.2), Bigot and Gadat (2008) constructed an adaptive estimation procedure using wavelet-domain term-by-term hard thresholding. It is closely related to the WaveD procedure of Johnstone et al. (2004) designed for the standard Gaussian deconvolution problem. Taking the minimax point of view, under a standard smoothness assumption on g (to be specified in Section 3), it was shown that their estimator achieves near optimal minimax rates under the quadratic risk over a wide class of Besov balls. More precisely, if \tilde{f}_n denotes the estimate provided by their procedure and $B_{p,r}^s(M)$ the Besov ball of radius M (to be defined in Section 2), then

$$\sup_{f \in B_{p,r}^s(M)} \mathbb{E} \left(\int_0^1 \left(\tilde{f}_n(t) - f(t) \right)^2 dt \right) \leq C (\log n / n)^{2s/(2s+2\delta+1)},$$

¹ This convergence can be sharpened under additional regularity properties, see e.g. assumption (A_g) .

where δ is a regularity parameter coming from the smoothness assumption made on g . Since the optimal rate of convergence for (1) is $n^{-2s/(2s+2\delta+1)}$ (see Bigot and Gadat (2008)), \tilde{f}_n is near optimal; as usual the terminology "near" pertains to the extra logarithmic factor $(\log n)^{2s/(2s+2\delta+1)}$.

1.2 Contribution

In this paper, we adopt the same methodology but consider another type of thresholding: block thresholding. For many nonparametric estimation problems, procedures based on block thresholding achieve better rates of convergence than the term-by-term one (including hard thresholding). See, for instance, Cai (2002); Cavalier and Tsybakov (2002); Cai and Chien (2005); Li and Xiao (2008); Li (2008); Chesneau et al. (2010). This motivates the construction of such a procedure for the estimation of f in (1). We prove that block thresholding indeed exhibits better convergence rates than term-by-term thresholding of Bigot and Gadat (2008). In particular, it gets rid of the logarithmic factor when s , p , r and δ belong to the "regular zone". The proof is based on a general theorem on the minimax performances of wavelet block thresholding procedures established by Chesneau et al. (2010). To apply this result, we need to prove two conditions: a moment condition and a concentration one. The first one is proved in Bigot and Gadat (2008). The second is established in this paper. For the proof, some powerful and sharp concentration inequalities established by Talagrand (1994) and Cirelson, Ibragimov and Sudakov (1976) are used.

1.3 Paper organization

The paper is organized as follows. In Section 2, we briefly describe wavelets and Besov balls. Section 3 clarifies the assumption made on g and reformulates (1) using Fourier analysis. Our wavelet block thresholding procedure is described in Section 4. In Section 5, the main results of the paper are stated. Their proofs are provided in Section 6.

2 Wavelets and Besov balls

2.1 Periodized Meyer Wavelets

We consider an orthonormal wavelet basis generated by dilations and translations of a "father" Meyer-type wavelet ϕ and a "mother" Meyer-type wavelet ψ . The main features of such wavelets are:

- they are bandlimited, i.e. the Fourier transforms of ϕ and ψ have compact supports respectively included in $[-4\pi 3^{-1}, 4\pi 3^{-1}]$ and $[-8\pi 3^{-1}, -2\pi 3^{-1}] \cup [2\pi 3^{-1}, 8\pi 3^{-1}]$.

- for any frequency in $[-2\pi, -\pi] \cup [\pi, 2\pi]$, there exists a constant $c > 0$ such that the magnitude of the Fourier transform of ψ is lower bounded by c .
- the functions (ϕ, ψ) are \mathcal{C}^∞ as their Fourier transforms have a compact support, and ψ has an infinite number of vanishing moments as its Fourier transform vanishes in a neighborhood of the origin:

$$\int_{-\infty}^{\infty} t^u \psi(t) dt = 0, \quad \forall u \in \mathbb{N}.$$

If the Fourier transforms of ϕ and ψ are also in \mathcal{C}^r for a chosen $r \in \mathbb{N}$, then for it can be easily shown that ϕ and ψ decay as

$$|\phi(t)| = O((1 + |t|)^{-r-1}), \quad |\psi(t)| = O((1 + |t|)^{-r-1}),$$

meaning that ϕ and ψ are not very well localized in time. This is why a Meyer wavelet transform is generally implemented in the Fourier domain.

For the purpose of this paper, we use the periodized Meyer wavelet bases on the unit interval. For any $t \in [0, 1]$, any integer j and any $k \in \{0, \dots, 2^j - 1\}$, let

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k)$$

be the elements of the wavelet basis, and

$$\phi_{j,k}^{\text{per}}(t) = \sum_{l \in \mathbb{Z}} \phi_{j,k}(t - l), \quad \psi_{j,k}^{\text{per}}(t) = \sum_{l \in \mathbb{Z}} \psi_{j,k}(t - l),$$

their periodized versions. There exists an integer j_* such that the collection $\{\phi_{j_*,k}^{\text{per}}, k = 0, \dots, 2^{j_*} - 1; \psi_{j,k}^{\text{per}}, j = j_*, \dots, \infty, k = 0, \dots, 2^j - 1\}$ forms an orthonormal basis of $\mathbb{L}_{\text{per}}^2([0, 1])$. In what follows, the superscript "per" will be dropped to lighten the notation.

Let j_c be an integer such that $j_c \geq j_*$. A function $f \in \mathbb{L}_{\text{per}}^2([0, 1])$ can be expanded into a wavelet series as

$$f(t) = \sum_{k=0}^{2^{j_c}-1} \alpha_{j_c,k} \phi_{j_c,k}(t) + \sum_{j=j_c}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k}(t), \quad t \in [0, 1],$$

where

$$\alpha_{j_c,k} = \langle f, \phi_{j_c,k} \rangle = \int_0^1 f(t) \overline{\phi_{j_c,k}}(t) dt, \quad \beta_{j,k} = \langle f, \psi_{j,k} \rangle = \int_0^1 f(t) \overline{\psi_{j,k}}(t) dt,$$

and $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{L}_{\text{per}}^2([0, 1])$. See (Meyer 1992, Vol. 1 Chapter III.11) for a detailed account on periodized orthonormal wavelet bases.

2.2 Besov balls

Let $M \in (0, \infty)$, $s \in (0, \infty)$, $p \in [1, \infty)$ and $r \in [1, \infty)$. Set $\beta_{j_*-1,k} = \alpha_{j_*,k}$. A function f belongs to the Besov balls $B_{p,r}^s(M)$ if and only if there exists a constant $M^* > 0$ such that the associated wavelet coefficients satisfy

$$\left(\sum_{j=j_*-1}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*.$$

For a particular choice of parameters s , p and r , these sets contain the Hölder and Sobolev balls. See Meyer (1992).

3 Preliminaries

3.1 Assumptions on the density g

As a periodic function, the density $g \in \mathbb{L}_{\text{per}}^2([0, 1])$ can be written as the series $g(t) = \sum_{l \in \mathbb{Z}} g_1(t + l)$, and a sufficient condition for g to exist for all $t \in [0, 1]$ is that the template g_1 to have fast decay for the series to converge pointwise.

We suppose that there exist three constants, $c > 0$, $C > 0$ and $\delta > 1$, such that, for any $\ell \in \mathbb{Z}$, the Fourier coefficient of g , i.e. $\gamma_\ell = \mathcal{F}(g)(\ell)$, satisfies

$$c(1 + |\ell|^\delta)^{-1} \leq |\gamma_\ell| \leq C(1 + |\ell|^\delta)^{-1}. \quad (\text{A}_g)$$

This assumption controls the decay of the Fourier coefficients of g , and thus the smoothness of g . It is a standard hypothesis usually adopted in the field of nonparametric estimation for deconvolution problems where the parameter δ quantifies the spectral properties, hence the ill-conditioning, of the convolution operator associated to g . See, for instance, Pensky and Vidakovic (1999); Fan and Koo (2002); Johnstone et al. (2004); Bigot and Gadat (2008).

For example, if $g_1(t) = e^{-|t|}$, then $g(t)$ exists for $t \in [0, 1]$ and satisfies (A_g) . Indeed, for any $\ell \in \mathbb{Z}$, by a change of variable one can show that $\gamma_\ell = 2(1 + 4\pi^2\ell^2)^{-1}$. Hence (A_g) is satisfied with $c = (2\pi^2)^{-1}$, $C = 2$ and $\delta = 2$.

3.2 Reformulation as a linear inverse problem

In order to estimate the unknown wavelet coefficients of f from (1), we recall briefly the methodology of Bigot and Gadat (2008). It consists in casting (1) as a linear inverse (deconvolution) problem. More precisely, for any $\ell \in \mathbb{Z}$ and $m \in \{1, \dots, n\}$, let $c_{m,\ell} = \int_0^1 e^{-2i\pi\ell t} dY_m(t)$, then by simple properties of the Fourier transform

$$c_{m,\ell} = \theta_\ell e^{-2i\pi\ell\tau_m} + \epsilon z_{m,\ell}, \quad (2)$$

where

$$\theta_\ell = \mathcal{F}(f)(\ell), \quad z_{m,\ell} = \int_0^1 e^{-2i\pi\ell t} dW_m(t).$$

Now, let $\hat{c}_\ell = n^{-1} \sum_{m=1}^n c_{m,\ell}$, it follows from (2) that

$$\hat{c}_\ell = \theta_\ell \hat{\gamma}_\ell + n^{-1/2} \epsilon \eta_\ell, \quad (3)$$

where

$$\hat{\gamma}_\ell = n^{-1} \sum_{m=1}^n e^{-2i\pi\ell\tau_m}, \quad \eta_\ell = n^{-1/2} \sum_{m=1}^n z_{m,\ell}.$$

Notice that $\mathbb{E}(\hat{\gamma}_\ell) = \int_0^1 e^{-2i\pi\ell x} g(x) dx = \gamma_\ell$ and η_1, \dots, η_n are i.i.d. $\mathcal{N}(0, 1)$. Embarking from (3), if we set, for any $j \geq j_c$ and $k \in \{0, \dots, 2^j - 1\}$,

$$\hat{\beta}_{j,k} = \sum_{\ell \in \mathbb{Z}} (\hat{c}_\ell / \gamma_\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell), \quad (4)$$

then the Parseval theorem implies

$$\begin{aligned} \mathbb{E}(\hat{\beta}_{j,k}) &= \sum_{\ell \in \mathbb{Z}} (\theta_\ell / \gamma_\ell) \mathbb{E}(\hat{\gamma}_\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell) = \sum_{\ell \in \mathbb{Z}} \theta_\ell \overline{\mathcal{F}(\psi_{j,k})}(\ell) \\ &= \sum_{\ell \in \mathbb{Z}} \mathcal{F}(f)(\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell) = \int_0^1 f(x) \overline{\psi_{j,k}}(x) dx = \beta_{j,k}. \end{aligned}$$

In plain words, $\hat{\beta}_{j,k}$ is an unbiased estimator of $\beta_{j,k}$. It satisfies additional interesting properties. Two of them are the core of our theoretical contribution and are provided in Section 5.

4 Our estimator

We use the notations introduced in Section 3. We assume that (A_g) is fulfilled with a regularity parameter $\delta > 1$, and the upper-bound in $\sup_{\ell \in \mathbb{Z}} |\theta_\ell| \leq C_*$ is a known constant. We now describe the proposed adaptive procedure for estimating f from (3). It combines James-Stein rule (see Stein (1990)) with the wavelet methodology. In the same vein as in Cai (2002), we call it BlockJS.

Let $L = \lfloor \log n \rfloor$ be the block length, $J_1 = \lfloor \log_2 L \rfloor$ is the coarsest decomposition scale, and $J_2 = \lfloor (1/(2\delta + 1)) \log_2(n/\log n) \rfloor$. For any scale $j \in \{J_1, \dots, J_2\}$, let $A_j = \{1, \dots, 2^j L^{-1}\}$ be the set indexing the blocks at scale j . For each block index $K \in A_j$, $B_{j,K} = \{k \in \{0, \dots, 2^j - 1\}; (K-1)L \leq k \leq KL-1\}$ is the set indexing the positions of coefficients within the K th block. The sets A_j and $B_{j,K}$ are chosen such that the blocks partition $\{0, \dots, 2^j - 1\}$ without overlapping, i.e. $\bigcup_{K \in A_j} B_{j,K} = \{0, \dots, 2^j - 1\}$, $B_{j,K} \cap B_{j,K'} = \emptyset$ for any $K \neq K'$ with $K, K' \in A_j$, and $\text{Card}(B_{j,K}) = L$.

As devised by the discussion of Section 3.2, dividing (3) by γ_ℓ (this always makes sense under (A_g)) and taking the wavelet transform yields the observed coefficients sequence

$$\hat{\alpha}_{J_1,k} = \sum_{\ell \in \mathcal{D}_{J_1}} (\hat{c}_\ell / \gamma_\ell) \overline{\mathcal{F}(\phi_{J_1,k})}(\ell), \quad \hat{\beta}_{j,k} = \sum_{\ell \in \mathcal{C}_j} (\hat{c}_\ell / \gamma_\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell), \quad (5)$$

where \mathcal{D}_{J_1} denotes the support of $\mathcal{F}(\phi_{J_1,k})$, and \mathcal{C}_j that of $\mathcal{F}(\psi_{j,k})$.

We define the BlockJS estimator by

$$\hat{f}_n(t) = \sum_{k=0}^{2^{J_1}-1} \hat{\alpha}_{J_1,k} \phi_{J_1,k}(t) + \sum_{j=J_1}^{J_2} \sum_{K \in A_j} \sum_{k \in B_{j,K}} \hat{\beta}_{j,k}^* \psi_{j,k}(t), \quad t \in [0, 1], \quad (6)$$

where

$$\hat{\beta}_{j,k}^* = \hat{\beta}_{j,k} \left(1 - \frac{\lambda(\epsilon \vee C_*)^2 n^{-1} 2^{2\delta j}}{\frac{1}{L} \sum_{k \in B_{j,K}} |\hat{\beta}_{j,k}|^2} \right)_+, \quad (7)$$

with, for any $(a, b) \in \mathbb{R}^2$, $(a)_+ = \max(a, 0)$, $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ and $\lambda > 0$ is a threshold parameter to be discussed later.

The differences between \hat{f}_n and the procedure of Bigot and Gadat (2008) are the threshold (to be discussed below) and the thresholding rule: instead of term-by-term selection of $\hat{\beta}_{j,k}$, we operate by group selection with a suitable length for each block, i.e. $L = \lfloor \log n \rfloor$. This length is optimal for numerous nonparametric problems, and, as we will see in Section 5, it is adequate for (1).

For recent minimax (or oracle) results on BlockJS procedures for other problems, we refer to Cai (2002); Cavalier and Tsybakov (2002); Cai and Chicken (2005); Li and Xiao (2008); Li (2008); Chesneau et al. (2010). Details on the BlockJS for the standard Gaussian white noise model can be found in Tsybakov (2004).

5 Main results

5.1 Minimax theorem

Theorem 1 *Consider the random shift model defined by (1). Suppose that (A_g) is satisfied and the upper-bound in $\sup_{\ell \in \mathbb{Z}} |\theta_\ell| \leq C_*$ is a known constant. Let \hat{f}_n be the estimator defined by (6) with a large enough λ . Then there exists a constant $C > 0$ such that, for any $p \in [1, \infty)$, $r \in [1, \infty)$, $s \in (1/p, \infty)$, and n large enough, we have*

$$\sup_{f \in B_{p,r}^s(M)} \mathbb{E} \left(\int_0^1 \left(\hat{f}_n(t) - f(t) \right)^2 dt \right) \leq C \varphi_n,$$

where

$$\varphi_n = \begin{cases} n^{-2s/(2s+2\delta+1)}, & \text{when } p \geq 2, \\ (\log n/n)^{2s/(2s+2\delta+1)}, & \text{when } p \in [1, 2) \text{ and } s > (1/p - 1/2)(2\delta + 1). \end{cases}$$

With regard to Bigot and Gadat (2008, Theorem 1.1), φ_n is near optimal. There is only an extra logarithmic factor $(\log n)^{2s/(2s+2\delta+1)}$ for the case $p \in [1, 2)$ and $s > (1/p - 1/2)(2\delta + 1)$. In comparison to the procedure of Bigot and Gadat (2008), we achieve a better rate as we gain $(\log n)^{2s/(2s+2\delta+1)}$ for the case $p \geq 2$.

Theorem 1 can be proved via a more general result on the minimax performance of BlockJS. This general result is (Chesneau et al. 2010, Theorem 3.1). To apply this theorem, two conditions on $(\hat{\alpha}_{J_1,k})_{k \in \{0, \dots, 2^{J_1}-1\}}$ and $(\hat{\beta}_{j,k})_{k \in \{0, \dots, 2^j-1\}}$ are required.

- *Moment condition.* There exists a constant $C > 0$ such that, for any $j \in \{J_1, \dots, J_2\}$ and $k \in \{0, \dots, 2^j - 1\}$,

$$\mathbb{E}(|\hat{\alpha}_{J_1,k} - \alpha_{J_1,k}|^2) \leq C 2^{2\delta J_1} n^{-1}, \quad \mathbb{E}(|\hat{\beta}_{j,k} - \beta_{j,k}|^4) \leq C 2^{4\delta j} n^{-2}.$$

This is proved in (Bigot and Gadat 2008, Proposition 3.1).

- *Concentration condition.* This is formalized in Proposition 1 below.

Proposition 1 *Consider the framework of Theorem 1. Then there exists a constant $\lambda > 0$ such that, for any $j \in \{J_1, \dots, J_2\}$, any $K \in A_j$ and n large enough, the estimator $(\hat{\beta}_{j,k}^*)_{k \in B_{j,K}}$ of (7) obeys*

$$\mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |\hat{\beta}_{j,k} - \beta_{j,k}|^2 \right)^{1/2} \geq \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \leq n^{-2}.$$

Consequently, only the proof of Proposition 1 needs to be set to prove Theorem 1. This is done in Section 6.

5.2 On the choice of the threshold

The estimator \hat{f}_n is defined with the threshold

$$t_j = \sqrt{\lambda}(\epsilon \vee C_*) 2^{\delta j} \sqrt{L/n}.$$

Recall that C_* is supposed to be a known constant such that $\sup_{\ell \in \mathbb{Z}} |\theta_\ell| \leq C_*$. Since f , and a fortiori, θ_ℓ , is unknown, this condition is not realistic in practice. To solve this problem, Bigot and Gadat (2008) propose to estimate $|\theta_\ell|$ by $|\hat{\theta}_\ell| = \sqrt{\frac{1}{n} \sum_{m=1}^n |c_{m,\ell}|^2} - \epsilon$ and consider a more sophisticated threshold:

$$t_j^* = 2 \left(\left(\hat{\sigma}_j \sqrt{2 \log n/n} \right) \vee \left(\hat{\delta}_j (\log n/(3n)) \right) \right),$$

where $\hat{\sigma}_j^2 = \epsilon^2 \sum_{\ell \in C_j} |\mathcal{F}(\psi_{j,0})(\ell)/\gamma_\ell|^2$ and $\hat{\delta}_j = \sum_{\ell \in C_j} |\mathcal{F}(\psi_{j,0})(\ell) \hat{\theta}_\ell|/|\gamma_\ell|$.

6 Proofs

In this section, c and C denote positive constants which can take different values for each mathematical term. They are independent of f and n .

6.1 Proof of Proposition 1

We have

$$\widehat{\beta}_{j,k} - \beta_{j,k} = U_{j,k} + V_{j,k},$$

where

$$U_{j,k} = \sum_{\ell \in \mathcal{C}_j} \theta_\ell (\widehat{\gamma}_\ell / \gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})}(\ell), \quad V_{j,k} = \epsilon n^{-1/2} \sum_{\ell \in \mathcal{C}_j} (\eta_\ell / \gamma_\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell).$$

For any $\lambda > 0$, the Minkowski inequality implies

$$\begin{aligned} & \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |\widehat{\beta}_{j,k} - \beta_{j,k}|^2 \right)^{1/2} \geq \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \\ &= \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |U_{j,k} + V_{j,k}|^2 \right)^{1/2} \geq \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \\ &\leq \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |U_{j,k}|^2 \right)^{1/2} + \left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \geq \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \\ &\leq A + B, \end{aligned} \tag{8}$$

where

$$A = \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |U_{j,k}|^2 \right)^{1/2} \geq 2^{-1} \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \tag{9}$$

and

$$B = \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \geq 2^{-1} \lambda(\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right). \tag{10}$$

Let us now investigate the upper bounds for A and B .

The upper bound for A (see (9)). For this, we need the following inequality due to Talagrand (1994).

Lemma 1 (Talagrand (1994)) *Let V_1, \dots, V_n be i.i.d. random variables, $\epsilon_1, \dots, \epsilon_n$ be independent Rademacher variables, also independent of V_1, \dots, V_n , F be a class of functions uniformly bounded by T and $r_n : F \rightarrow \mathbb{R}$ be the operator defined by*

$$r_n(h) = n^{-1} \sum_{i=1}^n h(V_i) - \mathbb{E}(h(V_1)).$$

Suppose that

$$\sup_{h \in F} \mathbb{V}(h(V_1)) \leq v \quad \text{and} \quad \mathbb{E} \left(\sup_{h \in F} \sum_{i=1}^n \epsilon_i h(V_i) \right) \leq nH.$$

Then, there exist two absolute constants $C_1 > 0$ and $C_2 > 0$ such that, for any $t > 0$, we have

$$\mathbb{P} \left(\sup_{h \in F} r_n(h) \geq t + C_2 H \right) \leq \exp(-nC_1 (t^2 v^{-1} \wedge tT^{-1})).$$

In order to apply the Talagrand inequality, consider the unit ball $\Omega = \{a = (a_k) \in \mathbb{C}; \sum_{k \in B_{j,K}} |a_k|^2 \leq 1\}$ and the class F of functions defined by $F = \{h; h(x) = \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} \theta_\ell (e^{-2i\pi\ell x} / \gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})}(\ell), a \in \Omega\}$. By classical results of convex analysis, and more precisely the Legendre-Fenchel conjugate, we have

$$\begin{aligned} & \left(\sum_{k \in B_{j,K}} |U_{j,k}|^2 \right)^{1/2} = \sup_{a \in \Omega} \sum_{k \in B_{j,K}} a_k U_{j,k} \\ &= \sup_{a \in \Omega} \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} \theta_\ell \left(n^{-1} \sum_{m=1}^n e^{-2i\pi\ell\tau_m} / \gamma_\ell - 1 \right) \overline{\mathcal{F}(\psi_{j,k})}(\ell) \\ &= \sup_{h \in F} r_n(h), \end{aligned}$$

where r_n denotes the function defined in Lemma 1. Now, let us evaluate the quantities T , H and v of the Talagrand inequality.

The value of T . Let h be a function in F . Using $|a_k| \leq 1$, by (A_g) , $\sup_{\ell \in \mathcal{C}_j} |e^{-2i\pi\ell x} / \gamma_\ell - 1| \leq \sup_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|) + 1 \leq C2^{\delta j}$, and $\sup_{\ell \in \mathbb{Z}} |\theta_\ell| \leq C_*$, we obtain

$$\begin{aligned} |h(x)| &= \left| \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} \theta_\ell (e^{-2i\pi\ell x} / \gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right| \\ &\leq \sum_{k \in B_{j,K}} |a_k| \sum_{\ell \in \mathcal{C}_j} |e^{-2i\pi\ell x} / \gamma_\ell - 1| |\theta_\ell| |\mathcal{F}(\psi_{j,k})|(\ell) \\ &\leq CC_* 2^{\delta j} \sum_{k \in B_{j,K}} \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})|(\ell). \end{aligned}$$

Since $\sup_{k \in B_{j,K}} \sup_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)| \leq C2^{-j/2}$, $\text{Card}(\mathcal{C}_j) \leq C2^j$ and $\text{Card}(B_{j,K}) \leq \log n$, we have

$$|h(x)| \leq C2^{-j/2} \text{Card}(\mathcal{C}_j) \text{Card}(B_{j,K}) \leq CC_* 2^{\delta j} 2^{j/2} \log n.$$

Hence $T = CC_* 2^{\delta j} 2^{j/2} \log n$.

The value of H . Let $\epsilon_1, \dots, \epsilon_n$ be independent Rademacher variables independent of $\tau = (\tau_1, \dots, \tau_n)$. Since $\sup_{a \in \Omega} \sum_{k \in B_{j,K}} a_k^2 = 1$, and using successively the Cauchy-Schwartz and the Jensen inequalities yield

$$\begin{aligned} & \mathbb{E} \left(\sup_{h \in F} \sum_{m=1}^n \epsilon_m h(\tau_m) \right) \\ &= \mathbb{E} \left(\sup_{a \in \Omega} \sum_{m=1}^n \epsilon_m \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right) \\ &\leq \sup_{a \in \Omega} \left(\sum_{k \in B_{j,K}} a_k^2 \right)^{1/2} \times \\ &\quad \mathbb{E} \left(\left(\sum_{k \in B_{j,K}} \left| \sum_{m=1}^n \epsilon_m \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right)^{1/2} \right) \\ &\leq \left(\sum_{k \in B_{j,K}} \mathbb{E} \left(\left| \sum_{m=1}^n \epsilon_m \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right) \right)^{1/2}. \quad (11) \end{aligned}$$

Since $\epsilon_1, \dots, \epsilon_n$ are independent Rademacher variables, also independent of $\tau = (\tau_1, \dots, \tau_n)$, we have

$$\begin{aligned} & \mathbb{E} \left(\left| \sum_{m=1}^n \epsilon_m \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right) = \\ &= \mathbb{E} \left(\mathbb{E} \left(\left| \sum_{m=1}^n \epsilon_m \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \mid \tau \right) \right) \\ &= \sum_{m=1}^n \mathbb{E} \left(\left| \sum_{\ell \in \mathcal{C}_j} \theta_\ell(e^{-2i\pi\ell\tau_m}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right). \end{aligned}$$

Denote Z_m the random variable $Z_m = \sum_{\ell \in \mathcal{C}_j} \theta_\ell e^{-2i\pi\ell\tau_m} / \gamma_\ell \overline{\mathcal{F}(\psi_{j,k})}(\ell)$. Obviously, for any m , $\mathbb{E}(Z_m) = \sum_{\ell \in \mathcal{C}_j} \theta_\ell \overline{\mathcal{F}(\psi_{j,k})}(\ell)$ and hence

$$\begin{aligned} \mathbb{E} \left(\left| \sum_{\ell \in \mathcal{C}_j} \theta_\ell (e^{-2i\pi\ell\tau_m} / \gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right|^2 \right) &= \mathbb{V}(Z_m) \\ &\leq \mathbb{E}(|Z_m|^2) = \mathbb{E} \left(\left| \sum_{\ell \in \mathcal{C}_j} \theta_\ell e^{-2i\pi\ell\tau_m} / \gamma_\ell \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right|^2 \right). \end{aligned} \quad (12)$$

With this observation, and using the identical distribution of τ_1, \dots, τ_n , we have

$$\begin{aligned} \sum_{m=1}^n \mathbb{V}(Z_m) &= n\mathbb{V}(Z_1) \\ &\leq n\mathbb{E} \left(\left| \sum_{\ell \in \mathcal{C}_j} (\theta_\ell / \gamma_\ell) e^{-2i\pi\ell\tau_1} \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right|^2 \right) \\ &= n \int_0^1 \left| \sum_{\ell \in \mathcal{C}_j} (\theta_\ell / \gamma_\ell) e^{-2i\pi\ell x} \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right|^2 g(x) dx. \end{aligned}$$

By assumption (A_g), for any $t \in [0, 1]$, we have

$$|g(t)| \leq \sum_{\ell \in \mathbb{Z}} |\mathcal{F}(g)(\ell)| \leq C \sum_{\ell \in \mathbb{Z}} (1 + |\ell|)^{-\delta},$$

the sequence is the last right-hand side member is Riemann summable since $\delta > 1$ and thus $|g(t)| < \infty$. Combining this with $\sup_{\ell \in \mathbb{Z}} |\theta_\ell| \leq C_*$, $\sup_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|) \leq C2^{\delta j}$, and the Plancherel formula which implies $\sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)|^2 = \int_0^1 |\psi_{j,k}(t)|^2 dt = 1$, we arrive at

$$\begin{aligned} \sum_{m=1}^n \mathbb{V}(Z_m) &\leq Cn \int_0^1 \left| \sum_{\ell \in \mathcal{C}_j} (\theta_\ell / \gamma_\ell) e^{-2i\pi\ell x} \overline{\mathcal{F}(\psi_{j,k})}(\ell) \right|^2 dx \\ &= Cn \sum_{\ell \in \mathcal{C}_j} (|\theta_\ell|^2 / |\gamma_\ell|^2) |\mathcal{F}(\psi_{j,k})(\ell)|^2 \\ &\leq CC_*^2 n 2^{2\delta j} \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)|^2 \\ &= CC_*^2 n 2^{2\delta j}. \end{aligned} \quad (13)$$

Putting (11) and (13) together, we obtain

$$\mathbb{E} \left(\sup_{h \in F} \sum_{m=1}^n \epsilon_m h(\tau_m) \right) \leq C (C_*^2 n 2^{2\delta j} \text{Card}(B_{j,K}))^{1/2} \leq CC_* 2^{\delta j} (n \log n)^{1/2}.$$

Hence $H = CC_* 2^{\delta j} (\log n/n)^{1/2}$.

The value of v . With the same observation as in (12), we have

$$\begin{aligned} \sup_{h \in F} \mathbb{V}(h(\tau_1)) &\leq \sup_{a \in \Omega} \mathbb{E} \left(\left| \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} \theta_\ell (e^{-2i\pi\ell\tau_1}/\gamma_\ell - 1) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right) \\ &\leq \sup_{a \in \Omega} \mathbb{E} \left(\left| \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} (\theta_\ell/\gamma_\ell) e^{-2i\pi\ell\tau_1} \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 \right) \\ &= \sup_{a \in \Omega} \int_0^1 \left| \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} (\theta_\ell/\gamma_\ell) e^{-2i\pi\ell x} \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 g(x) dx. \end{aligned}$$

Proceeding as in (13), we obtain

$$\begin{aligned} \sup_{h \in F} \mathbb{V}(h(\tau_1)) &\leq C \sup_{a \in \Omega} \int_0^1 \left| \sum_{k \in B_{j,K}} a_k \sum_{\ell \in \mathcal{C}_j} (\theta_\ell/\gamma_\ell) e^{-2i\pi\ell x} \overline{\mathcal{F}(\psi_{j,k})(\ell)} \right|^2 dx \\ &= C \sup_{a \in \Omega} \sum_{\ell \in \mathcal{C}_j} (|\theta_\ell|^2/|\gamma_\ell|^2) \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2 \\ &\leq CC_*^2 2^{2\delta j} \sup_{a \in \Omega} \sum_{\ell \in \mathcal{C}_j} \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2. \end{aligned} \quad (14)$$

By the Plancherel formula, we get

$$\begin{aligned} \sum_{\ell \in \mathcal{C}_j} \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2 &= \sum_{\ell \in \mathcal{C}_j} \left| \mathcal{F} \left(\sum_{k \in B_{j,K}} \bar{a}_k \psi_{j,k} \right) (\ell) \right|^2 \\ &= \int_0^1 \left| \sum_{k \in B_{j,K}} \bar{a}_k \psi_{j,k}(t) \right|^2 dt \\ &= \sum_{k \in B_{j,K}} |a_k|^2 \leq 1. \end{aligned} \quad (15)$$

The last equality follows from the fact that at each scale j , the collection $\{\psi_{j,k}\}_k$ forms an orthonormal basis of the corresponding detail space. It then follows from (14) and (15) that

$$\sup_{h \in F} \mathbb{V}(h(\tau_1)) \leq CC_*^2 2^{2\delta j}.$$

Hence $v = CC_*^2 2^{2\delta j}$.

Combining the obtained values for v , H and T , and taking $t = 4^{-1}\lambda C_* 2^{\delta j} (\log n/n)^{1/2}$, we obtain

$$(t^2 v^{-1} \wedge t T^{-1}) \geq C \left(\lambda^2 (\log n/n) \wedge \lambda (n 2^j \log n)^{-1/2} \right).$$

For any $j \in \{J_1, \dots, J_2\}$ and n large enough, we have

$$n 2^j \log n \leq n 2^{J_2} \log n \leq C n (n/\log n)^{1/(2\delta+1)} \log n \leq \lambda^{-2} (n/\log n)^2.$$

Hence

$$(t^2 v^{-1} \wedge t T^{-1}) \geq C \lambda^2 (\log n/n).$$

Therefore, for λ large enough and $t = 4^{-1}\lambda C_* 2^{\delta j} (\log n/n)^{1/2}$, the Talagrand inequality described in Lemma 1 yields

$$\begin{aligned} A &= \mathbb{P} \left(\left(\sum_{k \in B_{j,\kappa}} |U_{j,k}|^2 \right)^{1/2} \geq 2^{-1} \lambda (\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \\ &\leq \mathbb{P} \left(\left(\sum_{k \in B_{j,\kappa}} |U_{j,k}|^2 \right)^{1/2} \geq 4^{-1} \lambda C_* 2^{\delta j} (\log n/n)^{1/2} + C_2 H \right) \\ &= \mathbb{P} \left(\sup_{h \in F} r_n(h) \geq t + C_2 H \right) \leq \exp(-n C_1 (t^2 v^{-1} \wedge t T^{-1})) \\ &\leq \exp(-n C \lambda^2 (\log n/n)) \leq 2^{-1} n^{-2}. \end{aligned} \tag{16}$$

We obtain the desired upper bound for A .

The upper bound for B (see (10)). We now turn to upper-bounding B . Toward this goal, we need the following inequality due to Cirelson, Ibragimov and Sudakov (1976).

Lemma 2 (Cirelson, Ibragimov and Sudakov (1976)) *Let $(\vartheta_t)_{t \in D}$ be a centered Gaussian process. If $\mathbb{E}(\sup_{t \in D} \vartheta_t) \leq N$ and $\sup_{t \in D} \mathbb{V}(\vartheta_t) \leq V$ then, for any $x > 0$,*

$$\mathbb{P} \left(\sup_{t \in D} \vartheta_t \geq x + N \right) \leq \exp(-x^2/(2V)).$$

The following random variables obey

$$V_{j,k} = \epsilon n^{-1/2} \sum_{\ell \in \mathcal{C}_j} (\eta_\ell / \gamma_\ell) \overline{\mathcal{F}(\psi_{j,k})}(\ell) \sim \mathcal{N}(0, n^{-1} \sigma_{j,k}^2),$$

where

$$\sigma_{j,k}^2 = \epsilon^2 \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell) / \gamma_\ell|^2.$$

Consider again the unit ball Ω as defined above. For any $a \in \Omega$, let $Z(a)$ be the centered Gaussian process defined by

$$Z(a) = \sum_{k \in B_{j,K}} a_k V_{j,k} = \epsilon n^{-1/2} \sum_{\ell \in \mathcal{C}_j} (\eta_\ell / \gamma_\ell) \sum_{k \in B_{j,K}} a_k \overline{\mathcal{F}(\psi_{j,k})}(\ell).$$

Again, by an argument of the Legendre-Fenchel conjugate of the unit Euclidean ball, we have

$$\sup_{a \in \Omega} Z(a) = \left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2}.$$

Let us now apply Lemma 2 to our setting by determining the values of N and V that appear in the Cirelson inequality.

Value of N . Recall that owing to assumption (A_g) , $\sup_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|^2) \leq C 2^{2\delta j}$. In addition, the Plancherel formula implies that

$$\sigma_{j,k}^2 = \epsilon^2 \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)/\gamma_\ell|^2 \leq C \epsilon^2 2^{2\delta j} \sum_{\ell \in \mathcal{C}_j} |\mathcal{F}(\psi_{j,k})(\ell)|^2 = C \epsilon^2 2^{2\delta j}.$$

Combining these facts with the Jensen inequality, we therefore get

$$\begin{aligned} \mathbb{E} \left(\sup_{a \in \Omega} Z(a) \right) &= \mathbb{E} \left(\left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \right) \leq \left(\sum_{k \in B_{j,K}} \mathbb{E} (|V_{j,k}|^2) \right)^{1/2} \\ &\leq C \left(n^{-1} \sum_{k \in B_{j,K}} \sigma_{j,k}^2 \right)^{1/2} \leq C \epsilon 2^{\delta j} n^{-1/2} \text{Card}(B_{j,K})^{1/2} \\ &= C \epsilon 2^{\delta j} (\log n / n)^{1/2}. \end{aligned}$$

Hence $N = C \epsilon 2^{\delta j} (\log n / n)^{1/2}$.

Value of V . Since η_ℓ is a complex white noise, i.e. $\mathbb{E}(\eta_\ell \bar{\eta}_{\ell'}) = 1$ if $\ell = \ell'$ and 0 otherwise, it comes

$$\begin{aligned}
\sup_{a \in \Omega} \mathbb{V}(Z(a)) &= \sup_{a \in \Omega} \mathbb{E} \left(\left| \sum_{k \in B_{j,K}} a_k V_{j,k} \right|^2 \right) \\
&= \sup_{a \in \Omega} \mathbb{E} \left(\sum_{k \in B_{j,K}} \sum_{k' \in B_{j,K}} a_k \bar{a}_{k'} V_{j,k} \bar{V}_{j,k'} \right) \\
&= \epsilon n^{-1} \sup_{a \in \Omega} \sum_{k \in B_{j,K}} \sum_{k' \in B_{j,K}} a_k \bar{a}_{k'} \sum_{\ell \in \mathcal{C}_j} \sum_{\ell' \in \mathcal{C}_j} (1/\gamma_\ell) \overline{\mathcal{F}(\psi_{j,k})(\ell)} (1/\bar{\gamma}_{\ell'}) \mathcal{F}(\psi_{j,k'})(\ell') \mathbb{E}(\eta_\ell \bar{\eta}_{\ell'}) \\
&= \epsilon n^{-1} \sup_{a \in \Omega} \sum_{k \in B_{j,K}} \sum_{k' \in B_{j,K}} a_k \bar{a}_{k'} \sum_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|^2) \overline{\mathcal{F}(\psi_{j,k})(\ell)} \mathcal{F}(\psi_{j,k'})(\ell) \\
&= \epsilon n^{-1} \sup_{a \in \Omega} \sum_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|^2) \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2.
\end{aligned}$$

By assumption (A_g) , we know that $\sup_{\ell \in \mathcal{C}_j} (1/|\gamma_\ell|^2) \leq C 2^{2\delta j}$. Using this and the equality $\sum_{\ell \in \mathcal{C}_j} \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2 = \sum_{k \in B_{j,K}} |a_k|^2$, (see (15)), we obtain

$$\sup_{a \in \Omega} \mathbb{V}(Z(a)) \leq C \epsilon n^{-1} 2^{2\delta j} \sup_{a \in \Omega} \sum_{\ell \in \mathcal{C}_j} \left| \sum_{k \in B_{j,K}} \bar{a}_k \mathcal{F}(\psi_{j,k})(\ell) \right|^2 = C \epsilon n^{-1} 2^{2\delta j}.$$

The last equality is by definition of $a \in \Omega$. Hence $V = C \epsilon n^{-1} 2^{2\delta j}$.

Taking λ large enough and $x = 4^{-1} \lambda \epsilon 2^{\delta j} (\log n/n)^{1/2}$, the inequality of Lemma 2 yields

$$\begin{aligned}
B &= \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \geq 2^{-1} \lambda (\epsilon \vee C_*) 2^{\delta j} (\log n/n)^{1/2} \right) \\
&\leq \mathbb{P} \left(\left(\sum_{k \in B_{j,K}} |V_{j,k}|^2 \right)^{1/2} \geq 4^{-1} \lambda \epsilon 2^{\delta j} (\log n/n)^{1/2} + N \right) \\
&= \mathbb{P} \left(\sup_{a \in \Omega} Z(a) \geq x + N \right) \leq \exp(-x^2/(2V)) \leq \exp(-C \lambda^2 \log n) \\
&\leq 2^{-1} n^{-2}.
\end{aligned} \tag{17}$$

We obtain the desired upper bound for B . Putting (8), (9), (10), (16) and (17) together, the proof of Proposition 1 follows. \square

Acknowledgements This work is supported by ANR grant NatImages, ANR-08-EMER-009.

References

- Bigot, J., Gadat, S. (2008). An inverse problem point of view for adaptive estimation in a shifted curves model. *Technical report*.
- Cai, T., Chicken, E. (2005). Block thresholding for density estimation: Local and global adaptivity. *Journal of Multivariate Analysis*, 95, 76–106.
- Cai, T. (2002). On block thresholding in wavelet regression: Adaptivity, blocksize and threshold level. *Statistica Sinica*, 12, 1241–1273.
- Cavalier, L., Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123, 323–354.
- Chesneau, C., Fadili, J., Starck, J.-L. (2010). Stein block thresholding for image denoising. *Applied and Computational Harmonic Analysis*, to appear.
- Chesneau, C. (2008). Wavelet estimation via block thresholding: A minimax study under the \mathbb{L}^p risk. *Statistica Sinica*, 18, 3, 1007–1024.
- Cirelson, B., Ibragimov, I., Sudakov, V. (1976). *Norm of Gaussian sample functions*, Springer Verlag, Berlin.
- Fan, J., Koo, J.Y. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, 48, 734–747.
- Grenander, U. (1993). *General Pattern Theory: A Mathematical Study of Regular Structures*, Oxford University Press, New York.
- Johnstone, I., Kerkycharian, G., Picard, D., Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society, Serie B*, 66 (3), 547–573.
- Li, L., and Xiao, Y. (2008). On the minimax optimality of block thresholded wavelet estimators with longmemory data. *Journal of Statistical Planning and Inference*, 137, 2850–2869.
- Li, L. (2008). On the block thresholding wavelet estimators with censored data. *Journal of Multivariate Analysis*, 99, 1518–1543.
- Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.
- Pensky, M., Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27, 2033–2053.
- Stein, C. (1990) *Estimation of the mean of a multivariate normal distribution*. *Annals of Statistics*, 10, 1135–1151.
- Talagrand, M. (1994). Sharper bounds for gaussian and empirical processes. *Annals of Probability*, 22, 28–76.
- Tsybakov, A. (2004). *Introduction à l'estimation nonparamétrique*. Springer Verlag, Berlin.
- Walter, G. (1994). *Wavelets and other orthogonal systems in applications*. CRC Press, Boca Raton.
- Zayed, A., Walter, G. (1996). Characterization of analytic functions in term of their wavelet coefficients. *Complex Variables*, 29, 265–276.