

# Ridgelets and the Representation of Mutilated Sobolev Functions

EMMANUEL J. CANDÈS

Department of Statistics  
Stanford University  
Stanford, California 94305–4065

We show that ridgelets, a system introduced in [4], are optimal to represent smooth multivariate functions that may exhibit linear singularities. For instance, let  $\{u \cdot x - b > 0\}$  be an arbitrary hyperplane and consider the singular function  $f(x) = 1_{\{u \cdot x - b > 0\}}g(x)$ , where  $g$  is compactly supported with finite Sobolev  $L_2$  norm  $\|g\|_{H^s}$ ,  $s > 0$ . The ridgelet coefficient sequence of such an object is as sparse as if  $f$  were without singularity, allowing optimal partial reconstructions. For instance, the  $n$ -term approximation obtained by keeping the terms corresponding to the  $n$  largest coefficients in the ridgelet series achieves a rate of approximation of order  $n^{-s/d}$ ; the presence of the singularity does not spoil the quality of the ridgelet approximation. This is unlike all systems currently in use and especially Fourier or wavelet representations.

**Key Words and Phrases.** Sobolev spaces, Fourier transform, singularities, ridgelets, orthonormal ridgelets, nonlinear approximation, sparsity. AMS subject classifications: 41A46, 42B99.

**Acknowledgments.** I am especially grateful to David Donoho for many fruitful discussions. I would also like to thank one referee for some very helpful comments on the original version of the manuscript. This research was supported by National Science Foundation grant DMS 98–72890 (KDI) and grant DMS 95–05151 and by AFOSR MURI 95–P49620–96–1–0028.

Some of the results were briefly described at the Royal Society meeting “Wavelets: a key to intermittent information?” held in London, February 1999.

# 1 Introduction

## 1.1 Ideal representations of Sobolev classes

It is well known that trigonometric series and wavelets are well adapted to represent functions taken from  $L_2$  Sobolev classes [1]. For a nonnegative integer  $s$ , the  $L_2$  Sobolev norm is

$$\|f\|_{H^s}^2 = \|f\|_2^2 + \|f^{(s)}\|_2^2$$

where  $f^{(s)}$  is the  $s$ -th derivative of  $f$ ; and, more generally, the norm of  $f$  is defined by means of the Fourier transform; let  $\mathcal{F}$  be the classical Fourier transform,

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) = \int f(x)e^{-ix \cdot \xi} dx; \tag{1.1}$$

then,

$$\|f\|_{H^s}^2 = \int |\hat{f}(\xi)|^2 (1 + |\xi|^{2s}) d\xi$$

when  $s > 0$  is not necessarily an integer. (Of course, when  $s$  is an integer, the two definitions are equivalent thanks to the Plancherel formula, see [13], for example.)

Both wavelet and Fourier bases provide unconditional bases for these Sobolev spaces  $H^s$  defined on the torus, say. Abstractly, a basis  $(\phi_i)_{i \in \mathcal{I}}$  is an unconditional basis for a functional class  $\mathcal{F}$  if shrinking the coefficients preserves the norm of the object: i.e., if we let

$$\theta_i(f) = \langle f, \phi_i \rangle$$

and consider

$$\tilde{f} = \sum_i \theta'_i \phi_i, \quad |\theta'_i| \leq |\theta_i|,$$

then

$$\|\tilde{f}\|_{\mathcal{F}} \leq C \|f\|_{\mathcal{F}}.$$

We quote Donoho [8], “An orthogonal basis of  $L_2$  which is also an unconditional basis of a functional space  $\mathcal{F}$  is an optimal basis for compressing, estimating, and recovering functions in  $\mathcal{F}$ .”

For instance, suppose that  $f$  is a function defined on the circle  $\mathbf{T}$  with bounded Sobolev norm and let  $f_n$  be the  $n$ -term trigonometric nonlinear approximation of  $f$  obtained by keeping the terms corresponding to the  $n$  largest coefficients in the expansion. Then,

$$\|f - f_n\|_2 \leq C n^{-s} \|f\|_{H^s(T)}.$$

The same is true for nice periodic wavelets and essentially, no orthogonal basis would give a better rate of approximation: that is, for any orthobasis  $(\phi_i)_{i \in \mathcal{I}}$ , let  $Q_n(f)$  be the best  $n$ -term approximation in that basis

$$Q_n(f) = \arg \min \|f - g\|_2, \quad g = \sum_{m=1}^n \lambda_m \phi_{i_m};$$

then, letting  $\mathcal{F}$  be the Sobolev ball  $\mathcal{F} = \{f, \|f\|_{H^s(T)} \leq 1\}$ , there is a lower bound on the error of approximation

$$\sup_{f \in \mathcal{F}} \|f - Q_n(f)\|_2 \geq C n^{-s}.$$

Another instance of this property is that in any orthobasis  $(\phi_i)_{i \in \mathcal{I}}$  the number of terms greater than  $1/n$  is greater than  $c \cdot n^{2/(2s+1)}$ . In both Fourier and wavelet bases,  $n^{2/(2s+1)}$  is the order of the number of coefficients that exceed  $1/n$  and in this sense, we may say that these bases are the most “economical” for representing elements from  $H^s(T)$ .

## 1.2 Singularities: the one-dimensional case

However, these nice properties are very fragile. For instance, it is well known that trigonometric series provide poor reconstructions of discontinuous functions. On the interval  $[0, 1]$ , let  $f$  be the periodic function defined by  $f(t) = t - H(t - t_0)$  where  $H(t)$  is the step function  $1_{\{t > 0\}}$ . The best  $L_2$   $n$ -term approximation of  $f$  by trigonometric series gives only an  $L_2$  error of order  $O(n^{-1/2})$ . This is a general fact: if  $g$  is a nice function taken from the Sobolev class  $H^s$  (with support contained in  $(0, 1)$ ), then the rate of approximation of  $H(t - b)g(t)$  is no better than  $O(n^{-1/2})$ . The discontinuity spoils the representation, and we need a lot of different terms to reconstruct the discontinuity with good accuracy. (This phenomenon is well known from engineers and is often referred to as the Gibbs phenomenon or ringing effect.)

One of the reasons why wavelets are so attractive is that they are the best bases for representing objects composed with singularities (see the discussion of Mallat’s heuristics in [8]). As an example, our simple discontinuous object  $H(\cdot - b)g(\cdot)$  has a rate of approximation in a nice wavelet basis of order  $O(n^{-s})$ . Whereas the singularity had a dramatic effect on the sparsity of Fourier coefficients, it does not affect the sparsity of wavelet coefficients as the number of wavelet coefficients exceeding  $1/n$  is still of order  $n^{2/(2s+1)}$ . The singularity does not spoil the wavelet representation. This miracle may explain the spread of wavelet methods in data compression, statistical estimation, inverse problems, etc., as in practical applications, the signals that are to be recovered exhibit these kinds of discontinuities (see the survey paper [11]).

### 1.3 Singularities: the higher-dimensional case

Under a certain viewpoint, however, the picture changes dramatically when the dimension is greater than one. On  $[0, 1]^d$ , suppose now that we want to represent the simple object

$$f(x) = H(u \cdot x - t_0)g(x), \quad g \in H^s \text{ and } \text{supp } g \subset [0, 1]^d. \quad (1.2)$$

The object  $f$  is singular on the hyperplane  $u \cdot x = t_0$  ( $u$  is a unit vector) but may be very smooth elsewhere. Then, the number of wavelet coefficients exceeding  $1/n$  is greater than  $n^{2(1-1/d)}$  yielding  $L_2$  rates of approximation only of order  $O(n^{-\frac{1}{2(d-1)}})$ . This lower bound holds even when  $g$  is as nice as we want, i.e.,  $g \in C^\infty$ . Translated into the framework of image compression, it says that both wavelet bases and Fourier bases are severely inefficient at representing edges in images. Wavelets can deal with point-like phenomena, but cannot deal with line-like phenomena in dimension 2, plane-like phenomena in dimension 3, etc.

In harmonic analysis, there has recently been much interest in finding new dictionaries and ways of representing functions by linear combinations of elements of those. Examples include wavelets, wavelet-packets, Gabor functions, brushlets, etc. The purpose of this paper is to show that ridgelets, a system introduced by [4], are as efficient for representing objects with discontinuities like (1.2) as wavelets are for representing discontinuous functions in one dimension.

### 1.4 Achievements and overview

The ridgelet construction will briefly be reviewed in section 2. In a nutshell, a ridgelet is a ridge function of the form

$$\psi_{a,u,b}(x) = \frac{1}{a^{1/2}} \psi\left(\frac{u \cdot x - b}{a}\right), \quad a > 0, u \in S^{d-1}, b \in \mathbb{R}, \quad (1.3)$$

where  $\psi$  is univariate and oscillatory. The fundamental result is that there is a discrete family  $(\psi_{a_n, u_n, b_n})$  which is a frame for  $L_2$  spaces of compactly supported functions. (We will simply refer to this family as  $\psi_n$ .) The frame property says that for any element  $f \in L_2[0, 1]^d$  there exist two constants  $A, B > 0$  with the property

$$A \|f\|^2 \leq \sum_n |\langle f, \psi_n \rangle|^2 \leq B \|f\|^2.$$

A consequence of the previous display is the existence of a dual set of ridgelets  $(\widetilde{\psi}_n)$  (the dual frame) and of the decomposition

$$f = \sum_n \langle f, \widetilde{\psi}_n \rangle \psi_n = \sum_n \langle f, \psi_n \rangle, \widetilde{\psi}_n \quad (1.4)$$

with equality holding in an  $L_2$  sense.

To measure the sparsity of a sequence  $(\theta_n)$ , we will use the weak- $\ell_p$  or Marcinkiewicz quasi-norm, defined as follows: let  $|\theta|_{(n)}$  be the  $n$ th largest entry in the sequence  $(|\theta_n|)$ ; we set

$$|\theta|_{w\ell_p} = \sup_{n>0} n^{1/p} |\theta|_{(n)}. \quad (1.5)$$

Equipped with a nice ridgelet frame, the key result of our paper (section 4) is the following: let us consider a template  $f$  like in (1.2) and let  $\alpha$  ( $\alpha_n = \langle f, \psi_n \rangle$ ) denote the ridgelet coefficient sequence of  $f$ . Then, the sequence  $\alpha$  is sparse as if  $f$  were not singular in the sense that

$$\|\alpha\|_{w\ell_p} \leq C \|g\|_{H^s}, \quad \text{with } 1/p = s/d + 1/2, \quad (1.6)$$

where the constant  $C$  does not depend on  $f$ ; or equivalently, the number of ridgelet coefficients exceeding  $1/n$  is bounded by  $C n^p \|g\|_{H^s}$ . (Throughout the paper, the letter  $C$  will denote a positive constant whose value may differ at different occurrences, even within a single formula.) There might be some ambiguity about the notation  $\|g\|_{H^s}$  since  $g$  is not uniquely determined by  $f$ . In this paper, we will implicitly take the norm  $\|g\|_{H^s}$  as being the minimum norm of all those elements in  $H^s$  whose restriction to  $\{u \cdot x > t_0\}$  coincide with  $f$ ; i.e.,

$$\|g\|_{H^s} := \inf\{\|h\|_{H^s}, f(x) = H(u \cdot x - t_0)h(x), \text{supp } h \subset [0, 1]^d\}.$$

There is a direct consequence of this result. Consider the  $n$ -term  $f_n$  ridgelet approximation obtained by extracting from the exact series (1.4) the terms corresponding to the  $n$  largest coefficients. Then,

$$\|f - f_n\| \leq C n^{-s/d} \|g\|_{H^s}, \quad (1.7)$$

where, again, the constant  $C$  is independent of  $f$ . *The presence of the singularity does not ruin the sparsity of the ridgelet series.* This is unlike wavelet or Fourier analysis. Hence, we have a very concrete, constructive and stable procedure – namely, the thresholding of ridgelet coefficients – to obtain near-optimal nonlinear approximations. The author is not aware of any other system with similar features.

In dimension 2, Donoho introduced an orthonormal basis, closely related to the ridgelet system, that he calls “orthonormal ridgelets.” Section 5 will show that both results (1.6) and (1.7) continue to hold with orthonormal ridgelets in place of ‘pure’ ridgelets.

## 1.5 Methodology

The method that is used to prove (1.6) and (1.7) involves the study of the Fourier transform along rays going through the origin (section 3). Before we proceed further,  $(r, \theta)$  will index the standard polar coordinates system and throughout the paper we will abuse notation in writing  $f(r, \theta)$  instead

of  $(f \circ \mathcal{C})(r, \theta)$  where  $\mathcal{C}$  is the change of coordinates from polar to cartesian. In two dimensions, let us now consider the singular function  $f$  defined by

$$f(x_1, x_2) = 1_{\{x_1 > 0\}} g(x_1, x_2),$$

with  $g$  in  $H^s$ ,  $s \in \mathbb{N}$  and  $\text{supp } g \subset [0, 1]^d$ . The argument relies on a bound that is available on the integral over the ‘polar’ segment  $\{(r, \theta), 2^j \leq r \leq 2^{j+1}\}$  of the squared modulus of the Fourier transform. Indeed, there exists a constant  $C$  not depending on  $f$  such that

$$\int_{2^j \leq r \leq 2^{j+1}} |\hat{f}(r, \theta)|^2 dr \leq C \epsilon_j^2(\theta) 2^{-j} 2^{-2js} \|g\|_{H^s}^2 + C 2^{-j} \min(1, 2^{-2js} |\sin \theta|^{-2s}) \|g\|_{H^s}^2, \quad (1.8)$$

with  $\sum_j \int_0^{2\pi} \epsilon_j^2(\theta) d\theta \leq 1$ . A  $d$ -dimensional version of (1.8) will be given in section 3.

The singularity  $1_{\{x_1 > 0\}}$  causes the Fourier transform to decay very slowly in the critical directions  $\theta = 0, \pi$  (this set of directions is sometimes referred to as the wavefront). Indeed, for  $\theta = 0$ , say,  $|\hat{f}(r, \theta)| \sim r^{-1}$  and, therefore, for this critical value of  $\theta$ ,  $\int_{2^j \leq r \leq 2^{j+1}} |\hat{f}(r, \theta)|^2 dr \sim 2^{-j}$  which is the content of (1.8). However, this effect is really local and our estimate (1.8) pictures the decay of the Fourier transform as  $\theta$  moves away from the singular rays. The result is nonasymptotic since it describes the situation at a finite distance  $2^j$  ( $j \geq 0$ ) of the origin. For instance, in dimension 2 the order of magnitude of the modulus of the Fourier transform at a point with polar coordinates  $(2^j, \theta)$  is  $2^{-j(s+1)} |\sin \theta|^{-s}$ . It is interesting to observe that the smoothness of the object governs the size of the Fourier transform as  $\theta$  approaches  $0, \pi$ . Although this phenomenon may not have been extensively studied in the literature, it perhaps corresponds to some new kind of microlocal analysis and we believe that this is of independent interest.

The localization of the Fourier transform near the wavefront is the key property driving our main results (1.6) and (1.7). Extensions and limitations of these results will be discussed in section 6.

## 2 Ridgelets

In this section,  $\hat{g}$  will denote the Fourier transform of  $g$ . In  $d$  dimensions, the ridgelet construction starts with a univariate function  $\psi$  satisfying an oscillatory condition, namely,

$$\int |\hat{\psi}(\xi)|^2 / |\xi|^d d\xi < \infty. \quad (2.1)$$

A ridgelet is a function of the form

$$\frac{1}{a^{1/2}} \psi\left(\frac{u \cdot x - b}{a}\right), \quad (2.2)$$

where  $a$  and  $b$  are scalar parameters and  $u$  is a vector of unit length. In the sequel, we will suppose that  $\psi$  is normalized so that  $\int |\hat{\psi}(\xi)|^2 |\xi|^{-d} d\xi = 1$ . Of course, a ridgelet is a ridge function whose

profile displays an oscillatory behavior (like a wavelet). A ridgelet has a scale  $a$ , an orientation  $u$ , and a location parameter  $b$ . Ridgelets are concentrated around hyperplanes: roughly speaking the ridgelet (2.2) is supported near the strip  $\{x, |u \cdot x - b| \leq a\}$ .

Remarkably, one can represent any function as a superposition of these ridgelets. Define the ridgelet coefficients

$$\mathcal{R}_f(a, u, b) = \int f(x) a^{-1/2} \psi\left(\frac{u \cdot x - b}{a}\right) dx; \quad (2.3)$$

then, for any  $f \in L_1 \cap L_2(\mathbb{R}^d)$ , we have

$$f(x) = (2\pi)^{-(d-1)} \int \mathcal{R}_f(a, u, b) a^{-1/2} \psi\left(\frac{u \cdot x - b}{a}\right) d\mu(a, u, b), \quad (2.4)$$

where  $d\mu(a, u, b) = da/a^{d+1} du db$  ( $du$  being the uniform measure on the sphere). Furthermore, this formula is stable as one has a Parseval relation

$$\|f\|_2^2 = (2\pi)^{-(d-1)} \int |\mathcal{R}_f(a, u, b)|^2 d\mu(a, u, b). \quad (2.5)$$

Similar to the continuous transform, there is a discrete transform. Consider the following discrete collection of ridgelets

$$\{\psi_{j,\ell,k}(x) = 2^{j/2} \psi(2^j u_{j,\ell} \cdot x - kb_0), j \geq j_0, u_{j,\ell} \in \Sigma_j, k \in \mathbb{Z}\}. \quad (2.6)$$

The scale  $a$  and location parameter  $b$  are discretized dyadically, as in the theory of wavelets. However, unlike wavelets, ridgelets are directional and, here, the interesting aspect is the discretization of the directional variable  $u$ . This variable is sampled at increasing resolution, so that at scale  $j$  the discretized set  $\Sigma_j$  is a net of nearly equispaced points at a distance of order  $2^{-j}$ . A detailed exposition on the ridgelet construction is given in [4]. In two dimensions, for instance, a ridgelet is of the following form

$$\{2^{j/2} \psi(2^j (x_1 \cos \theta_{j,\ell} + x_2 \sin \theta_{j,\ell} - 2\pi k 2^{-j}))\}_{(j \geq j_0, \ell, k)},$$

where the directional parameter  $\theta_{j,\ell}$  is sampled with increasing angular resolution at increasingly fine scales, something like the following:

$$\theta_{j,\ell} = 2\pi \ell 2^{-j}.$$

The key result [4] is that the discrete collection  $(\psi_{j,\ell,k})$  is a frame for square integrable functions supported on the the unit cube. There exists two constants  $A$  and  $B$  such that for any  $f \in L_2([0, 1]^d)$ , we have

$$A \|f\|_{L_2}^2 \leq \sum_{j,\ell,k} |\langle f, \psi_{j,\ell,k} \rangle|^2 \leq B \|f\|_{L_2}^2. \quad (2.7)$$

The previous equation says that the datum of the ridgelet transform at the points  $(a, u, b) = (2^j, u_{j,\ell}, k2^{-j})$  –with the parameter range as in (2.6)– suffices to reconstruct the function perfectly. In this sense, this is analogous to the Shannon sampling theorem for the reconstruction of bandlimited functions. Indeed, standard arguments show that there exists a dual collection  $(\tilde{\psi}_{j,\ell,k})$  with the property

$$f = \sum_{j,\ell,k} \langle f, \tilde{\psi}_{j,\ell,k} \rangle \psi_{j,\ell,k} = \sum_{j,\ell,k} \langle f, \psi_{j,\ell,k} \rangle \tilde{\psi}_{j,\ell,k}, \quad (2.8)$$

where the notation  $\langle \cdot, \cdot \rangle$  stands here and throughout the remainder of this paper for the usual inner product of  $L_2$ :  $\langle f, g \rangle = \int f(x)g(x)dx$ .

At times, we will use the compact notation  $\psi_\nu$  ( $\nu \in \mathcal{N}$ ) for our ridgelet frames and, therefore, we will keep in mind that the index runs  $\nu$  through an enumeration of the triples  $(j, \ell, k)$ .

### 3 Localization of the Fourier transform

The purpose of this section is to quantify the size of the Fourier transform of an object  $f$ , where  $f$  is given by

$$f(x) = H(x_1) g(x)$$

where  $g$  is compactly supported and with finite Sobolev norm (recall  $H(t) = 1_{\{t>0\}}$ ).

To formulate our statement in  $d$  dimensions, we need to introduce the spherical coordinates defined by  $x_1 = r \cos \theta_1$ ,  $x_2 = r \sin \theta_1 \cos \theta_2$ ,  $\dots$ ,  $x_d = r \sin \theta_1 \sin \theta_2 \dots \sin \theta_{d-1}$ ,  $0 \leq \theta_1, \dots, \theta_{d-2} \leq \pi$ ,  $0 \leq \theta_{d-1} < 2\pi$ . In what follows, we will simply refer to  $(\theta_2, \dots, \theta_{d-1})$  as  $\varphi$ , and  $d\varphi$  will denote the element of the surface area on  $S^{d-2}$ , i.e.  $d\varphi = \sin \theta_2^{d-3} \dots \sin \theta_{d-2} d\theta_2 \dots d\theta_{d-1}$ . With these notations, the uniform measure  $du$  on the sphere may thus be rewritten as  $du = (\sin \theta_1)^{d-2} d\theta_1 d\varphi$ . From now on, we will often refer to a unit vector  $u$  by means of its polar coordinates  $(\theta, \varphi)$ ,  $\theta \in [0, \pi]$ ,  $\varphi \in S^{d-2}$ .

We now state our  $d$ -dimensional localization result about the modulus of the Fourier transform.

**Theorem 3.1** *Let  $f$  be given by  $f(x) = H(x_1)g(x)$  with  $g$  in  $H^s$ ,  $s = 0, 1, 2, \dots$ , and  $\text{supp } g \subset [-1, 1]^d$ , and put  $\sigma = s + (d - 2)/2$ . Then, there exists a universal constant  $C$  such that for any  $j \geq 0$ ,*

$$\int_{2^j \leq r \leq 2^{j+1}} \int |\hat{f}(r, \theta, \varphi)|^2 dr d\varphi \leq C \epsilon_j^2(\theta) 2^{-j} 2^{-2j\sigma} \|g\|_{H^s}^2 + C 2^{-j} \min(1, 2^{-2j\sigma} |\sin \theta|^{-2\sigma}) \|g\|_{H^s}^2, \quad (3.1)$$

where  $\sum_j |S^{d-2}| \int \epsilon_j^2(\theta) (\sin \theta)^{d-2} d\theta \leq 1$ .



As we emphasized earlier, the Fourier transform decays very slowly in the directions  $\theta = 0, \pi$  because of the singularity  $H$ . However, (3.1) is not a statement about the decay of  $\hat{f}$  along the singular rays  $\theta = 0, \pi$ , rather it is about the decay of the Fourier transform as  $\theta$  moves away from the critical directions  $\theta = 0, \pi$ . Roughly speaking, the order of magnitude of the modulus of the Fourier transform at a point with polar coordinates  $(2^j, \theta)$  is  $2^{-j(\sigma+1)}|\sin \theta|^{-\sigma}$  with  $\sigma = s + (d-2)/2$ .

*Remark.* The inequality involves a regular term (the first term of the right-hand side of (3.1)) as if one were simply analyzing an object from  $H^s$  and a singular term (the second one) essentially due to the discontinuity across the hyperplane  $x_1 = 0$ .

*Proof of Theorem 3.1.* We will prove the result by induction. The result is true for  $s = 0$  since letting  $I_j(\theta)$  be the left-hand side of (3.1)

$$I_j(\theta) \equiv \int_{2^j \leq r \leq 2^{j+1}} \int |\hat{f}(r, \theta, \varphi)|^2 dr d\varphi,$$

we have, by definition,

$$\begin{aligned} \sum_{j \geq 0} 2^{j(d-1)} \int I_j(\theta) (\sin \theta)^{d-2} d\theta &= \sum_{j \geq 0} 2^{j(d-1)} \int_{2^j}^{2^{j+1}} \int |\hat{f}(r, \theta, \varphi)|^2 dr d\theta d\varphi \\ &\leq \sum_{j \geq 0} \int_{2^j \leq |\xi| \leq 2^{j+1}} |\hat{f}(\xi)|^2 d\xi \leq \|f\|_{L_2}^2 \leq \|g\|_{L_2}^2. \end{aligned}$$

Assume now that the result holds until  $n - 1$  ( $n \in \mathbb{N}$ ), and take  $g \in H^n$ . For any tempered distribution in  $\mathbb{R}^d S$ , we have the well-known relationship

$$\mathcal{F}\{\partial_\ell S\} = i\xi_\ell \hat{S},$$

where in the previous display  $i^2 = -1$ , and  $\partial_\ell$  is the partial derivative with respect to the  $\ell$ th coordinate. We will simply apply this formula to the tempered distribution  $f = Hg$ . First, for any  $1 \leq \ell \leq d$ , we have

$$\partial_\ell f = H \partial_\ell g + g \partial_\ell H. \quad (3.2)$$

We observe that the second term,  $g \partial_\ell H$ , is nonzero only if  $\ell = 1$  in which case it is a distribution supported on  $x_1 = 0$ , namely,  $g \delta_{\{x_1=0\}}$ . Let  $h$  be the restriction of  $g$  on  $x_1 = 0$ . By the trace theorem [15] we know that  $h$  is in  $H^{n-1/2}(\mathbb{R}^{d-1})$  and, more precisely,

$$\|h\|_{H^{n-1/2}} \leq C \|g\|_{H^n}.$$

Let us now choose  $u = \xi/|\xi|$  and let  $\xi = (\xi_1, \xi')$  so that  $\xi' = \pi(\xi)$ , where  $\pi$  is the orthogonal projection onto  $\xi_1 = 0$ . For this particular choice of  $u$ , we have

$$i|\xi| \hat{f}(\xi) = u \cdot \mathcal{F}\{\nabla f\}(\xi) = u \cdot \mathcal{F}\{H \nabla g\}(\xi) + \xi_1/|\xi| \hat{h}(\pi(\xi)) \quad (3.3)$$

since the Fourier transform of  $g \delta_{\{x_1=0\}}$  is given by  $\hat{h}(\pi(\xi)) = (\hat{h} \circ \pi)(\xi)$ . The first term of the right-hand side of (3.3) is effortlessly going through the induction step. Indeed, we have

$$|u \cdot \mathcal{F}\{H \nabla g\}|^2(\xi) \leq \sum_{i=1}^d |\mathcal{F}\{H \partial_{\ell} g\}|^2(\xi);$$

it is clear that for any  $\ell$ ,  $\partial_{\ell} g \in H^{n-1}$  and therefore the induction hypothesis implies that

$$\int_{2^j \leq r \leq 2^{j+1}} \int |u \cdot \mathcal{F}\{H \nabla g\}|^2(r, \theta, \varphi) dr d\varphi \leq C 2^{-j} \epsilon_j^2(\theta) 2^{-2j(\sigma-1)} + C 2^{-j} \min(1, 2^{-2j(\sigma-1)}) |\sin \theta|^{-2(\sigma-1)}. \quad (3.4)$$

We split the analysis of the second term of the right-hand side of (3.3) into two separate cases: namely,  $\sin \theta \geq 2^{-j}$  and  $\sin \theta < 2^{-j}$ . In the former case, we have

$$\begin{aligned} \int_{2^j}^{2^{j+1}} \int |(\hat{h} \circ \pi)(r, \theta, \varphi)|^2 dr d\varphi &= \int_{2^j}^{2^{j+1}} \int |\hat{h}(r \sin \theta, \varphi)|^2 dr d\varphi \\ &= |\sin \theta|^{-1} \int_{2^j |\sin \theta|}^{2^{j+1} |\sin \theta|} \int |\hat{h}(\rho, \varphi)|^2 d\rho d\varphi \\ &\leq |\sin \theta|^{-1} |2^j \sin \theta|^{-(d-2)} \int_{2^j |\sin \theta| \leq |\xi'| \leq 2^{j+1} |\sin \theta|} |\hat{h}(\xi')|^2 d\xi'. \end{aligned}$$

The degree of smoothness of  $h$  ( $h \in H^{n-1/2}$ ) now allows us to bound the right-hand side of the previous display; i.e.,

$$\sum_{j=-\infty}^{\infty} |2^j \sin \theta|^{2(n-1/2)} \int_{2^j |\sin \theta| \leq |\xi'| \leq 2^{j+1} |\sin \theta|} |\hat{h}(\xi')|^2 d\xi' \sim \|h\|_{\dot{H}^{n-1/2}}^2 \leq C \|g\|_{H^n}^2,$$

which implies

$$\int_{2^j |\sin \theta| \leq |\xi'| \leq 2^{j+1} |\sin \theta|} |\hat{h}(\xi')|^2 d\xi' \leq C \eta_j^2(\theta) |2^j \sin \theta|^{-2(n-1/2)} \|g\|_{H^n}^2$$

with  $\sum_j \eta_j^2(\theta) \leq 1$ .

To summarize, we have

$$\int_{2^j \leq r \leq 2^{j+1}} \int |(\hat{h} \circ \pi)(r, \theta, \varphi)|^2 dr d\varphi \leq C 2^{-2j(\sigma-1/2)} |\sin \theta|^{-2\sigma} \|g\|_{H^s}^2 \quad (3.5)$$

in any dimension  $d \geq 2$ .

To finish the proof, we simply recall (3.3) which gives the inequality

$$|\hat{f}(\xi)|^2 = 2|\xi|^{-2} \left( |u \cdot \mathcal{F}\{H \nabla g\}|^2(\xi) + |\hat{h}(\pi(\xi))|^2 \right).$$

The polar integral of each term of the right-hand side of this inequality is bounded via (3.4) and (3.5), respectively, yielding the desired conclusion. The case  $\sin \theta \geq 2^{-j}$  is now fully proved.

We finally treat the case  $\sin \theta < 2^{-j}$ . On one hand  $h$  is bounded in  $H^{n-1/2}$  and therefore in  $L_2$ , since  $n \geq 1$ . On the other  $h$  is compactly supported and hence

$$\sup_{|\xi'| \leq 1} |\hat{h}(\xi')| \leq \|h\|_{L_1} \leq C \|h\|_{L_2} \leq C \|g\|_{H^n}.$$

In this case, we simply write

$$\begin{aligned} \int_{2^j \leq r \leq 2^{j+1}} \int |\hat{h}(r \sin \theta, \varphi)|^2 dr d\varphi &\leq 2^j |S^{d-2}| \sup_{2^j |\sin \theta| \leq |\xi'| \leq 2^{j+1} |\sin \theta|} |\hat{h}(\xi')|^2 \\ &\leq C 2^j \|g\|_{H^n}^2, \end{aligned}$$

and the result for  $\sin \theta < 2^{-j}$  now follows from (3.3). The proof of the theorem is complete.  $\blacksquare$

## 4 Main result

In this section, we will suppose that we are given a ridgelet frame satisfying the following mild assumptions:

1.  $\psi$  is  $R$  times differentiable and has vanishing moments through order  $D$ ;  $\min(R, D) \geq s + (d-1)/2$ .
2.  $\psi$  is of rapid decay, namely, for any  $\gamma > 0$  and  $0 \leq r \leq R$ , one can find a constant  $C$  such that

$$|\psi^{(r)}(t)| \leq C \cdot (1 + |t|)^{-\gamma}.$$

The sequence of ridgelet coefficients of a given function  $f$  will be denoted by  $\alpha$ :  $\alpha_{j,\ell,k} = \langle f, \psi_{j,\ell,k} \rangle$ .

We state our main result.

**Theorem 4.1** *Let  $g \in H^s$ ,  $s > 0$ , with  $\text{supp } g \subset [-1, 1]^d$  and put  $f(x) = H(u \cdot x - b)g(x)$  where  $H$  is the step function  $H(t) = 1_{\{t>0\}}$ . Then, the ridgelet coefficient sequence  $\alpha$  of  $f$  satisfies*

$$\|\alpha\|_{w\ell_{p^*}} \leq C \|g\|_{H^s}, \quad \text{with } 1/p^* = s/d + 1/2,$$

where  $d$  is the dimension of the space.

*Preliminary remark.* For any  $(j, \ell, k)$ , we have the following basic inequality:

$$|\alpha_{j,\ell,k}| \leq 2^{j/2} (1 + |k|)^{-\gamma} \|f\|_2, \quad |k| \geq 2^{j+1}$$

because of the rapid decay of  $\psi$ . Indeed, we have

$$|\psi_{j,\ell,k}(x)| \leq C (1 + 2^j |u_{j,\ell} \cdot x - k 2^{-j}|)^{-\gamma},$$

and, therefore, is not hard to check that for  $|k| \geq 2^{j+1}$

$$\sup_{[-1,1]^d} |\psi_{j,\ell,k}(x)| \leq C 2^{j/2} (1 + |k|)^{-\gamma}.$$

Our claim is then a simple consequence of this last inequality. Thus, if  $\psi$  has a sufficient decay, then the subsequence  $\{(\alpha_{j,\ell,k}), k \geq 2^{j+1}\}$  is in  $\ell_p$ , for any  $p > 0$ ; hence it is enough to restrict our attention to the set  $|k| \leq 2^{j+1}$ .

In order to prove the theorem, we will need a result which is a corollary of Theorem 3.1.

**Corollary 4.2** *Under the assumptions of Theorem 3.1, the ridgelet coefficient sequence  $\alpha$  of  $f$  may be decomposed as*

$$\alpha_{j,\ell,k} = a_{j,\ell,k} + b_{j,\ell,k},$$

where the sequences  $a$  and  $b$  enjoy the following properties:

1. the sequence  $a$  verifies

$$\sum_{\ell,k} |a_{j,\ell,k}|^2 \leq C \epsilon_j^2 2^{-2js} \|g\|_{H^s}^2 \quad (4.1)$$

with  $\sum_j \epsilon_j^2 \leq 1$  and,

2. the sequence  $b$  is localized both in angle and in location.

(a) LOCALIZATION IN ANGLE. For  $1 \leq m < j$ , let  $\Lambda_{j,m}$  be the set of indices such that

$$\Lambda_{j,m} := \{\ell, 2^{-m} \leq |\sin \theta_{j,\ell}| \leq 2^{-m+1}\} \quad (4.2)$$

(for  $m = j$ , we will take  $\Lambda_{j,m}$  to be  $\{\ell, |\sin \theta_{j,\ell}| \leq 2^{-(j-1)}\}$ ); then,

$$\sum_{\ell \in \Lambda_{j,m}} \sum_k |b_{j,\ell,k}|^2 \leq C 2^{-j} 2^{-(j-m)(2s-1)} \|g\|_{H^s}^2. \quad (4.3)$$

(b) LOCALIZATION IN RIDGE LOCATION. For any  $n > 0$ , there is a constant  $C$  (independent of  $f$ ) such that

$$|b_{j,\ell,k}| \leq C 2^{j/2} (1 + ||k| - |2^j \sin \theta_{j,\ell}||)^{-n} \|g\|_{H^s}. \quad (4.4)$$

Not surprisingly, this decomposition involves a regular and a singular contribution as well.

*Proof of Corollary.* Again, we prove the result by induction. For any compactly supported element of  $L_2$ , we have

$$\sum_j \sum_{\ell, k} |\alpha_{j, \ell, k}|^2 \leq C \|f\|_{L_2}^2 \leq C \|g\|_{L_2}^2,$$

which proves the claim in this case since one can simply take  $b \equiv 0$ .

Suppose now that the claim is true up to  $s - 1 \in \mathbb{N}$  and take  $g$  in  $H^s$ . Recall that the ridgelet  $\psi_{j, \ell, k}$  is given by  $2^{j/2} \psi(2^j u_{j, \ell} \cdot x - k)$ . The starting point is to express the ridgelet coefficient  $\alpha_{j, \ell, k}$  as a line integral in the Fourier domain [4]

$$\alpha_{j, \ell, k} = \int_{\mathbb{R}} \hat{f}(\lambda, u_{j, \ell}) 2^{-j/2} \hat{\psi}(2^{-j} \lambda) e^{-ik2^{-j} \lambda} d\lambda. \quad (4.5)$$

where  $\hat{f}(\lambda, u) = \hat{f}(\lambda u_1, \dots, \lambda u_d)$ . In the previous equation, the range of  $\lambda$  is the real line and not only the positive axis (polar coordinates). However, we can convert  $(\lambda, u)$  to classical polar coordinates  $(r, \theta, \varphi)$  via the obvious relationship  $(\lambda, u) = (-\lambda, -u)$ . The decomposition (3.3) then suggests rewriting  $\alpha_{j, \ell, k}$  as

$$\alpha_{j, \ell, k} = a_{j, \ell, k}^{(0)} + b_{j, \ell, k}^{(0)},$$

where

$$a_{j, \ell, k}^{(0)} = 2^{-j} u_{j, \ell} \cdot \int_{\mathbb{R}} \mathcal{F}\{H \nabla g\}(\lambda, u_{j, \ell}) 2^{-j/2} \frac{\hat{\psi}(2^{-j} \lambda)}{2^{-j} \lambda} e^{-ik2^{-j} \lambda} d\lambda$$

and

$$b_{j, \ell, k}^{(0)} = 2^{-j} \cos \theta_{j, \ell} \int_{\mathbb{R}} \hat{h}(\lambda \sin \theta_{j, \ell}, \varphi_{j, \ell}) \frac{\hat{\psi}(2^{-j} \lambda)}{2^{-j} \lambda} e^{-ik2^{-j} \lambda} d\lambda.$$

Let  $\Psi$  be the primitive of  $\psi$  defined by  $\Psi(x) = \int_{-\infty}^x \psi(t) dt$ . Then,  $\Psi$  satisfies the conditions listed at the beginning of the section (with the obvious modification  $\min(R, D) \geq s - 1 + (d - 1)/2$ ) and  $\hat{\Psi}(\lambda) = -i\hat{\psi}(\lambda)/\lambda$ . Therefore, we may apply the induction hypothesis to the sequence  $a$  and obtain

$$a_{j, \ell, k}^{(0)} = 2^{-j} a_{j, \ell, k}^{(1)} + 2^{-j} b_{j, \ell, k}^{(1)},$$

where  $a^{(1)}$  and  $b^{(1)}$ , respectively, satisfy properties (4.1) and (4.3)–(4.4) with  $(s - 1)$  in place of  $s$ . Now, define the sequences  $a$  and  $b$  by

$$a_{j, \ell, k} = 2^{-j} a_{j, \ell, k}^{(1)}$$

and

$$b_{j, \ell, k} = 2^{-j} b_{j, \ell, k}^{(1)} + b_{j, \ell, k}^{(0)}.$$

It is clear that  $a_{j,\ell,k}$  and  $2^{-j}b_{j,\ell,k}^{(1)}$  satisfy conditions (4.1) and (4.3)–(4.4), respectively. Thus we only need to check that the sequence  $b^{(0)}$  verifies (4.3) and (4.4). In the original domain,  $b_{j,\ell,k}^{(0)}$  is given by

$$b_{j,\ell,k}^{(0)} = \langle g \delta_{\{x_1=0\}}, \Psi_{j,\ell,k} \rangle.$$

On the support of  $g \delta_{\{x_1=0\}}$ , it is easy to see that  $\Psi_{j,\ell,k}$  is bounded by  $C 2^{j/2} (1 + ||k| - |2^j \sin \theta_{j,\ell}||)^{-n}$ . Therefore, with the notations of section 3, we have

$$\begin{aligned} |b_{j,\ell,k}^{(0)}| &\leq \|h\|_{L_1} \sup_{x \in \text{supp } g \delta_{\{x_1=0\}}} |\Psi_{j,\ell,k}(x)| \leq C 2^{j/2} (1 + ||k| - |2^j \sin \theta_{j,\ell}||)^{-n} \|h\|_{L_2} \\ &\leq C 2^{j/2} (1 + ||k| - |2^j \sin \theta_{j,\ell}||)^{-n} \|g\|_{H^{1/2}} \end{aligned}$$

which is bounded since  $g \in H^s$ ,  $s \geq 1$ . This finishes the verification of (4.4). It remains to check (4.3).

*Sampling results.* In a separate paper, we have established the following sampling results: let  $\alpha_{j,\ell,k}$  be the ridgelet coefficients of a compactly supported distribution  $S$ ; first,

$$\sum_k |\alpha_{j,\ell,k}|^2 \leq C \int_{\mathbb{R}} |\hat{S}(\lambda, u_{j,\ell})|^2 |\hat{\psi}(2^{-j}\lambda)|^2 (1 + |2^{-j}\lambda|^2) d\lambda; \quad (4.6)$$

second, we recall that at scale  $j$ , the set of discrete angular variables  $\{u_{j,\ell}, \ell \in \Lambda_j\}$  consists of points approximately uniformly distributed on the sphere; for any subset  $\Lambda'_j$  of  $\Lambda_j$ , we have

$$\sum_{\ell \in \Lambda'_j} \sum_k |\alpha_{j,\ell,k}|^2 \leq C 2^{j(d-1)} \int_{\mathbb{R}} |\hat{\psi}(2^{-j}\lambda)|^2 (1 + |2^{-j}\lambda|^{2d}) d\lambda \int_{\Sigma'_j} \sum_{|\alpha| \leq d-1} |D^\alpha \hat{S}(\lambda, u)|^2 du, \quad (4.7)$$

where  $\Sigma'_j$  is the set of points on the sphere defined by

$$\Sigma'_j \equiv \{u \in S^{d-1}, \inf_{\ell \in \Lambda'_j} \|u - u_{j,\ell}\|_2 \leq 2^{-j}\}.$$

Here  $\alpha$  is a multi-index  $\alpha = (\alpha_1, \dots, \alpha_d)$  and  $D^\alpha$  stands for the classical partial derivative with respect to the cartesian coordinate system  $D^\alpha S = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}$ . Thus, (4.7) is a kind of uniform sampling inequality. In a nutshell, (4.7) holds because the points  $\{u_{j,\ell}, \ell \in \Lambda_j\}$  are quasi uniformly distributed on the sphere (at a distance of order  $2^{-j}$ ); that is, for any point  $u \in S^{d-1}$ ,

$$\#\{\ell, \|u_{j,\ell} - u\|_2 \leq \delta\} \leq C 2^{j(d-1)} \delta^{d-1}.$$

We apply this result to the distribution  $S = g \delta_{\{x_1=0\}}$ ; that is, to the restriction of  $f$  to the hyperplane  $\{x_1 = 0\}$  (see section 3 for details). The Fourier transform of  $S$  is the function  $\hat{S} = \hat{h} \circ \pi$  that we introduced in section 3. With  $\Lambda_{j,m}$ ,  $0 \leq m < j$ , as in (4.2), we have

$$\{u \in S^{d-1}, \inf_{\ell \in \Lambda_{j,m}} \|u - u_{j,\ell}\|_2 \leq 2^{-j}\} \subset \{u \in S^{d-1}, 2^{-m} - 2^{-j} \leq \sin \theta \leq 2^{-m+1} + 2^{-j}\}$$

and we omit the proof of this simple inclusion. Therefore, in this context (4.7) gives

$$\sum_{\ell \in \Lambda_{j,m}} \sum_k |b_{j,\ell,k}^{(0)}|^2 \leq C 2^{j(d-1)} \int_{2^{-m-2^{-j}} \leq \sin \theta \leq 2^{-m+1+2^{-j}}} I(\theta) (\sin \theta)^{d-2} d\theta, \quad (4.8)$$

where  $I(\theta)$  is given by:

$$\int_{\mathbf{S}^{d-2}} \int_{\mathbb{R}} \sum_{|\alpha| \leq d-1} |D^\alpha \hat{S}(\lambda, \theta, \varphi)|^2 |\hat{\psi}(2^{-j}\lambda)|^2 (1 + |2^{-j}\lambda|^{2d}) d\lambda d\varphi.$$

Now, if  $\psi$  has  $r$  vanishing moments and is of regularity  $r$ , we have

$$\sup_{2^\ell \leq |\lambda| \leq 2^{\ell+1}} |\hat{\psi}(2^{-j}\lambda)| \leq C 2^{-|j-\ell|r}. \quad (4.9)$$

It is then easy to check that

$$I(\theta) \leq C 2^{-j} 2^{-2j\sigma} |\sin \theta|^{-2\sigma} \|g\|_{H^s}^2. \quad (4.10)$$

To see why this is true, we simply write

$$I(\theta) \leq \sum_{\ell} \sup_{2^\ell \leq |\lambda| \leq 2^{\ell+1}} |\hat{\psi}(2^{-j}\lambda)|^2 (1 + |2^{-j}\lambda|^{2d}) I_\ell(\theta),$$

where

$$I_\ell(\theta) = \int_{2^\ell \leq |\lambda| \leq 2^{\ell+1}} \int_{|\alpha| \leq d-1} |D^\alpha \hat{S}(\lambda, \theta, \varphi)|^2 d\lambda d\varphi.$$

In the proof of Theorem 3.1 (3.5), we obtained

$$\int_{2^\ell \leq |\lambda| \leq 2^{\ell+1}} \int |\hat{S}(\lambda, \theta, \varphi)|^2 d\lambda d\varphi \leq C 2^\ell 2^{-2\ell\sigma} |\sin \theta|^{-2\sigma} \|g\|_{H^s}^2. \quad (4.11)$$

Now,  $D^\alpha \hat{S}$  is the Fourier transform of the distribution  $(-i)^{|\alpha|} x^\alpha S$ , which is the restriction of  $(-i)^{|\alpha|} x^\alpha g$  to the hyperplane  $\{x_1 = 0\}$ . Because  $g$  is compactly supported, we have that

$$\|x^\alpha g\|_{H^s} \leq C \|g\|_{H^s}$$

since the multiplication by a  $C_0^\infty$  function is a bounded operation from  $H^s$  onto itself. Therefore, inequality applies (4.11)  $D^\alpha \hat{S}$  and we have the upper bound

$$I_\ell(\theta) \leq C 2^\ell 2^{-2\ell\sigma} |\sin \theta|^{-2\sigma} \|g\|_{H^s}^2.$$

Inequality (4.10) comes from the previous inequality together with the size estimate (4.9).

Combining (4.10) and (4.8) finally gives (recall  $2\sigma = 2s + d - 2$ )

$$\sum_{\ell \in \Lambda_{j,m}} \sum_k |b_{j,\ell,k}^{(0)}|^2 \leq C 2^{-2js} \|g\|_{H^s}^2 \int_{2^{-m-2^{-j}} \leq \sin \theta \leq 2^{-m+1+2^{-j}}} |\sin \theta|^{-2s} d\theta,$$

which, in turn, gives the desired conclusion

$$\sum_{\ell \in \Lambda_{j,m}} \sum_k |b_{j,\ell,k}^{(0)}|^2 \leq C 2^{-m} 2^{-2(j-m)s} \|g\|_{H^s}^2.$$

The corollary is established.  $\blacksquare$

*Proof of Theorem 4.1.* Let  $s$  be a positive integer. Following on Corollary 4.2, to prove that  $\alpha$  is in  $w_{\ell_{p^*}}$  ( $1/p^* = s/d + 1/2$ ), it is sufficient to prove that both  $a$  and  $b$  are in  $w_{\ell_{p^*}}$ . The membership of  $a$  to  $w_{\ell_{p^*}}$  follows from well-known arguments and is straightforward.

The  $w_{\ell_{p^*}}$  boundedness of the sequence  $(b_{j,\ell,k})$  will be deduced from Corollary 4.2. We identify two subsequences corresponding, respectively, to the indices  $|k| > 2^{j+1} |\sin \theta_{j,\ell}|$  and  $|k| \leq 2^{j+1} |\sin \theta_{j,\ell}|$ ; the interesting contribution concerns the latter subsequence. We prove that

1. the subsequence  $\{b_{j,\ell,k}, |k| \leq 2^{j+1} |\sin \theta_{j,\ell}|\}$  has a finite  $w_{\ell_{p^*}}$  norm, and
2. the  $\ell_p$  norm of the subsequence  $\{b_{j,\ell,k}, |k| > 2^{j+1} |\sin \theta_{j,\ell}|\}$  is bounded for any  $p > 0$ .

We prove the first assertion. Letting  $N(\epsilon)$  be the cardinality of those elements whose absolute value exceeds  $\epsilon$ , namely,

$$N(\epsilon) = \# \{ (j, \ell, k), |k| \leq 2^{j+1} |\sin \theta_{j,\ell}|, s.t. |b_{j,\ell,k}| \geq \epsilon \},$$

we want to show that

$$\sup_{\epsilon > 0} \epsilon N^{1/p^*}(\epsilon) \leq C \|g\|_{H^s}.$$

since the left-hand side is an equivalent definition of the weak- $\ell_{p^*}$  norm (1.5).

Put

$$N_{j,m}(\epsilon) = \# \{ (\ell, k), \ell \in \Lambda_{j,m}, |k| \leq 2^{j+1} |\sin \theta_{j,\ell}|, s.t. |b_{j,\ell,k}| \geq \epsilon \}.$$

Corollary 4.2 posits the existence of a constant  $K$  such that  $|b_{j,\ell,k}|^2 \leq K 2^{-j} \|g\|_{H^s}^2$  (4.3) and therefore, it is clear that  $N_{j,m}(\epsilon) = 0$  if  $2^j \geq K \epsilon^{-2} \|g\|_{H^s}^2$ . In what follows, we will let  $\eta$  be defined by  $\eta = \epsilon / \|g\|_{H^s}$ . Regardless of the condition  $|b_{j,\ell,k}| \geq \epsilon$ , the cardinality of the index set  $\{(\ell, k), \ell \in \Lambda_{j,m}, |k| \leq 2^{j+1} |\sin \theta_{j,\ell}|\}$  is bounded by  $C 2^{d(j-m)}$ . Further, the bound on the  $\ell_2$  norm of the  $b_{j,\ell,k}$ 's (Corollary 4.2) gives

$$N_{j,m}(\epsilon) \leq C \min(2^{(j-m)d}, \eta^{-2} 2^{-j} 2^{(j-m)(1-2s)})$$

whenever  $2^j \leq K \eta^{-2}$ .

Let  $N_j(\epsilon)$  be the number of coefficients whose absolute values exceed  $\epsilon$ , i.e.,

$$N_j(\epsilon) = \# \{ (\ell, k), |k| \leq 2^{j+1} |\sin \theta_{j,\ell}|, |b_{j,\ell,k}| \geq \epsilon \}.$$



Then, a simple calculation gives

$$\begin{aligned} N_j(\epsilon) = \sum_m N_{j,m}(\epsilon) &\leq C \sum_m \min(2^{(j-m)d}, \eta^{-2} 2^{-j} 2^{(j-m)(1-2s)}) \\ &\leq C \min(2^{jd}, \eta^{-2d/\alpha} 2^{-jd/\alpha}), \end{aligned}$$

where  $\alpha = d + 2s - 1$ . To summarize, we have

$$N_j(\epsilon) \leq C \begin{cases} 0 & 2^j \geq K \eta^{-2} \\ \eta^{-2d/\alpha} 2^{-jd/\alpha} & \eta^{-2/(1+\alpha)} \leq 2^j \leq K \eta^{-2} \\ 2^{jd} & 2^j \leq \eta^{-2/(1+\alpha)} \end{cases}.$$

Summing over the scales yields

$$\begin{aligned} N(\epsilon) = \sum_{j=0}^{\infty} N_j(\epsilon) &\leq C \sum_{j: 2^j \leq \eta^{-2/(1+\alpha)}} 2^{jd} + C \sum_{j: \eta^{-2/(1+\alpha)} \leq 2^j \leq K \eta^{-2}} \eta^{-2d/\alpha} 2^{-jd/\alpha} \\ &\leq C \eta^{-2d/(1+\alpha)} = C \eta^{-p^*} = C \epsilon^{-p^*} \|g\|_{H^s}^{p^*}, \end{aligned}$$

with  $1/p^* = s/d + 1/2$ . This finishes the proof of the first assertion.

We now turn to the second assertion. It clearly follows from (4.4) that for any  $q > 0$  we have

$$\sum_{k: |k| > 2^{j+1} |\sin \theta_{j,\ell}|} |b_{j,\ell,k}|^q \leq C 2^{jq/2} (2^j |\sin \theta_{j,\ell}|)^{1-nq} \|g\|_{H^s}^q,$$

since  $n$  may be chosen arbitrarily large and, in particular, greater than  $1/q$ . Summing over the  $\ell$ 's,  $\ell \in \Lambda_{j,m}$  gives

$$\sum_{\ell \in \Lambda_{j,m}} \sum_{k: |k| > 2^{j+1} |\sin \theta_{j,\ell}|} |b_{j,\ell,k}|^q \leq C 2^{jq/2} 2^{(1-nq)(j-m)} 2^{(j-m)(d-1)} \|g\|_{H^s}^q.$$

Now, we must keep in mind that we have available a bound on the  $\ell_2$  norm (4.3); i.e.,

$$\sum_{\ell \in \Lambda_{j,m}} \sum_{k: |k| > 2^{j+1} |\sin \theta_{j,\ell}|} |b_{j,\ell,k}|^2 \leq C 2^{-j} 2^{-(j-m)(2s-1)} \|g\|_{H^s}^2.$$

The interpolation inequality will yield the  $\ell_p$  boundedness. Recall that for any sequence  $a_n$  we have

$$\|a\|_{\ell_p} \leq \|a\|_{\ell_q}^\theta \|a\|_{\ell_2}^{1-\theta}, \quad 1/p = \theta/q + (1-\theta)/2. \quad (4.12)$$

This interpolation inequality applied to our subsequence gives

$$\left( \sum_{\ell \in \Lambda_{j,m}} \sum_{k: |k| > 2^{j+1} |\sin \theta_{j,\ell}|} |b_{j,\ell,k}|^p \right)^{1/p} \leq C \left[ 2^{j/2} 2^{-(j-m)(n-d/q)} \right]^\theta \left[ 2^{-j/2} 2^{-(j-m)(s-1/2)} \right]^{1-\theta} \|g\|_{H^s}.$$

In the previous inequality, the value of  $n$  may be chosen arbitrarily large and, hence, summing up the previous inequalities results in the upper bound

$$\sum_{\ell} \sum_{k: |k| > 2^{j+1} |\sin \theta_{j,\ell}|} |b_{j,\ell,k}|^p \leq C 2^{-jp(1/2-\theta)} \|g\|_{H^s}^p. \quad (4.13)$$

This establishes the boundedness in  $\ell_p$  for any  $p > 0$ . Indeed for  $p > 0$ , choose  $q$  small enough so that  $\theta < 1/2$  (4.12) – i.e.,  $1/q > 2/p + 1/2$  – and apply (4.13). The theorem is proved for  $s = 1, 2, \dots$

Interpolation theory allows us to extend the result to the half line  $s > 0$ . Indeed, let  $T$  be the operator

$$T : g \mapsto (\alpha_\nu)$$

that maps  $g$  into the ridgelet coefficient sequence  $(\alpha_\nu)$  of  $f$ ,  $f(x) = H(u \cdot x - b)g(x)$ , with  $u$  and  $b$  fixed. We abuse notations –as it is understood that we are concerned with elements supported on the unit cube– and let  $H^s$  be the Banach space defined by

$$H^s := \{g, g \in H^s \text{ and } \text{supp } g \subset [0, 1]^d\}$$

equipped with the norm  $\|\cdot\|_{H^s}$ . We proved that for any  $n \geq 1$ ,  $\|T\|$  is a bounded operator from  $H^n$  to  $w\ell_p$ ,  $1/p = n/d + 1/2$ . In addition,  $T$  is bounded from  $L_2$  to  $\ell_2$  (where again we understand  $L_2([0, 1]^d)$ ). On one hand, it is well-known that  $(L_2, H^n)$  is an interpolation couple [2] and that for any  $n > 0$  and any  $0 < \theta < 1$ , we have

$$(L_2, H^n)_{\theta, 2} = H^{n\theta},$$

see [14], for example. On the other, letting  $\ell_2$  be the space of real valued sequences

$$\ell_2 = \{a, \sum_{n \geq 1} |a_n|^2 < \infty\},$$

and similarly for  $w\ell_p$ ,  $p > 0$ , we have

$$(\ell_2, w\ell_p)_{\theta, 2} = \ell_{p^*, 2}, \quad 1/p^* = (1 - \theta)/2 + \theta/p.$$

Here,  $\ell_{p, 2}$ ,  $p > 0$  is the Lorentz space of real sequences

$$\left( \sum_{n \geq 1} |a|_{(n)}^2 n^{2/p-1} \right)^{1/2} < \infty,$$

where we recall that  $|a|_{(n)}$  is the  $n$ th largest entry in the sequence  $(|a_n|)$ . The interpolation theorem [2] gives that

$$T : H^{n\theta} \rightarrow \ell_{p^*, 2}$$

is bounded and further

$$\|T\|_{H^{n\theta} \rightarrow \ell_{p^*,2}} \leq C \|T\|_{L_2 \rightarrow \ell_2}^{1-\theta} \|T\|_{H^n \rightarrow w\ell_p}^\theta.$$

Hence, for any  $s > 0$ , pick  $n > s$  and put  $\theta = s/n$ . We have

$$\frac{1}{p^*} = \frac{1}{2} \left(1 - \frac{s}{n}\right) + \frac{s}{n} \left(\frac{n}{d} + \frac{1}{2}\right) = \frac{s}{d} + \frac{1}{2},$$

and, therefore, our analysis gives that  $T$  is bounded from  $H^s$  to  $\ell_{p^*,2}$ . This completes the proof of our theorem since for any sequence  $a$  and any  $p > 0$ , we have

$$\|a\|_{\ell_{p,2}} \leq \|w\ell_p\|. \quad \blacksquare$$

*Remark:* We proved a slightly stronger result than that announced in our theorem since for any  $s \geq 0$  the ridgelet coefficient sequence obeys

$$\|\alpha\|_{\ell_{p,2}} \leq C \|g\|_{H^s}, \quad 1/p = s/d + 1/2.$$

#### 4.1 Finite approximations

We now exploit Theorem 4.1 to derive nonlinear approximation bounds. The compact notation  $(\psi_\nu)_{\nu \in \mathcal{N}}$  introduced in section 2 will be used to denote the frame elements.

Suppose that  $f$  is of the form

$$f(x) = g_0(x) + H(u \cdot x - b)g_1(x), \quad (4.14)$$

where

$$\|g_i\|_{H^s} \leq C, \quad i = 0, 1.$$

From the exact series

$$f = \sum_{\nu \in \mathcal{N}} \alpha_\nu \tilde{\psi}_\nu,$$

extract the  $n$ -term approximation  $f_n$  obtained by keeping the  $n$  terms corresponding to the  $n$  largest coefficients. Then, we have the following result:

**Corollary 4.3** *With the previous assumptions, there exists a constant  $C$  (not depending on  $f$ ) such that*

$$\|f - f_n\|_2 \leq C n^{-s/d} \sup_{i=0,1} \|g_i\|_{H^s(\mathbb{R}^d)}. \quad (4.15)$$

As we will see below, the convergence rate of  $n$ -term ridgelet approximations is, in some sense, optimal.

Theorem 4.1 gives that the coefficients  $(\alpha_\nu)$  of  $f$  are bounded in  $w\ell_{p^*}$ . Letting  $|\alpha|_{(n)}$  be the  $n$ th largest entry in  $\alpha$  (in absolute values), we have

$$f - f_n = \sum_{\nu} \alpha_\nu 1_{\{|\alpha_\nu| \geq |\alpha|_{(n)}\}} \tilde{\psi}_\nu.$$

The lemma stated below then gives the desired conclusion, namely

$$\|f - f_n\|_2^2 \leq A^{-1} \sum_{m>n} |\alpha|_{(m)}^2 \leq A^{-1} C n^{-2s/d} \|\alpha\|_{w\ell_{p^*}}^2,$$

where  $A$  is the constant appearing on the left-hand side of (2.7).

**Lemma 4.4** *Let  $(a_\nu)_{\nu \in \mathcal{N}}$  be a sequence in  $\ell_2$  and let*

$$\tilde{f} = \sum_{\nu \in \mathcal{N}} a_\nu \tilde{\psi}_\nu.$$

*Then,*

$$\|\tilde{f}\|_2^2 \leq A^{-1} \|a\|_{\ell_2}^2.$$

*Proof of Lemma.* We let  $\tilde{F}$  be the synthesis operator defined by  $\tilde{F}a = \sum a_\nu \tilde{\psi}_\nu$  and  $F$  be the analysis operator  $Ff = (\langle f, \psi_\nu \rangle)_{\nu \in \mathcal{N}}$ . The property (2.7) gives

$$\|\tilde{f}\|^2 = \|\tilde{F}a\|^2 \leq A^{-1} \|F\tilde{F}a\|_{\ell_2}^2.$$

Now, it is easy to see that  $F\tilde{F}$  is the orthogonal projector onto the range of  $F$  and has, therefore, a norm (as an operator from  $\ell_2$  onto itself) bounded by 1. Consequently, we have

$$\|\tilde{f}\|^2 \leq A^{-1} \|F\tilde{F}a\|_{\ell_2}^2 \leq A^{-1} \|a\|_{\ell_2}^2,$$

which is what needed to be shown.  $\blacksquare$

## 4.2 Optimality

In this section, we detail the sense in which Corollary 4.3 is optimal. Consider a class of templates of the form (4.14): i.e., let  $\mathcal{F}(C)$  be the class defined by

$$\mathcal{F}(C) = \{f, f \text{ satisfies (4.14), } \|g_i\|_{H^s} \leq C, \text{ and } \text{supp } g_i \subset [0, 1]^d, i = 0, 1\}. \quad (4.16)$$

In the above definition, the singular hyperplane is not fixed; two elements from  $\mathcal{F}(C)$  may be singular along two different hyperplanes.

The class  $\mathcal{F}(C)$  contains, of course, the Sobolev ball  $H^s(C) = \{f, \|f\|_{H^s} \leq C, \text{ and } \text{supp } f \subset [0, 1]^d\}$ . In any orthobasis  $(\phi)_{i \in \mathcal{I}}$ , there is a lower bound on the convergence of the best  $n$ -term approximation  $Q_n(f)$  in that basis,

$$\sup_{f \in H^s(C)} \|f - Q_n(f)\|_2 \geq C n^{-s/2}.$$

As a consequence, no orthobasis exists that provides better rates than those obtained in Corollary 4.3. There is even a broader notion of optimality based on information theoretic concepts such as the Kolmogorov  $\epsilon$ -entropy or the Minimum Description Length (MDL) paradigm.

Let  $\mathcal{F}$  be a compact set of functions in  $L^2([0, 1]^d)$ . The Kolmogorov  $\epsilon$ -entropy  $N(\epsilon, \mathcal{F})$  of the class  $\mathcal{F}$  is the minimum number of bits that is required to specify any element  $f$  from  $\mathcal{F}$  within an accuracy of  $\epsilon$ . In other words, let  $\ell$  be a fixed counting number and let  $E_\ell : \mathcal{F} \rightarrow \{0, 1\}^\ell$  be a functional which assigns a bit string of length  $\ell$  to each  $f \in \mathcal{F}$ . Let  $D_\ell : \{0, 1\}^\ell \rightarrow L_2[0, 1]^d$  be a mapping which assigns to each bit string of length  $\ell$  a function. The coder-decoder pair  $(E_\ell, D_\ell)$  will be said to achieve a distortion  $\leq \epsilon$  over  $\mathcal{F}$  if

$$\sup_{f \in \mathcal{F}} \|D_\ell(E_\ell(f)) - f\| \leq \epsilon.$$

The Kolmogorov  $\epsilon$ -entropy (minimax description length) may then be defined as

$$L^*(\epsilon, \mathcal{F}) = \min\{\ell : \exists(E_\ell, D_\ell) \text{ achieving distortion } \leq \epsilon \text{ over } \mathcal{F}\}.$$

The minimum number of bits needed to reconstruct any  $f$  taken from our class of templates  $\mathcal{F}(C)$  (4.16) satisfies

$$N(\epsilon, \mathcal{F}(C)) \geq N(\epsilon, H^s) \geq C \epsilon^{2/s}.$$

A strategy identical to that developed in [9][Theorem 2], however, gives a simple way to exploit the sparsity of the ridgelet sequence to construct a coder-decoder pair of length  $O(\log(\epsilon^{-1})\epsilon^{2/s})$  that achieves a distortion of  $\epsilon$ . The construction is based on simple uniform quantization of the ridgelet coefficients  $\alpha_i$ , followed by simple run length coding. Hence, we have available a very concrete way of obtaining near-optimal (possibly within log-like factors) compression rates.

## 5 Orthonormal ridgelets

In dimension 2, Donoho [10] introduced a new orthonormal basis whose elements he called ‘orthonormal ridgelets.’ We will not detail why these elements relate to ridgelets. We quote from [7]: “Such a system can be defined as follows: let  $(\psi_{j,k}(t) : j \in \mathbb{Z}, k \in \mathbb{Z})$  be an orthonormal basis of Meyer wavelets for  $L^2(\mathbb{R})$  [12], and let  $(w_{i_0, \ell}^0(\theta), \ell = 0, \dots, 2^{i_0} - 1; w_{i, \ell}^1(\theta), i \geq i_0, \ell = 0, \dots, 2^i - 1)$

be an orthonormal basis for  $L^2[0, 2\pi)$  made of periodized Lemarié scaling functions  $w_{i_0, \ell}^0$  at level  $i_0$  and periodized Meyer wavelets  $w_{i, \ell}^1$  at levels  $i \geq i_0$ . (We suppose a particular normalization of these functions). Let  $\hat{\psi}_{j, k}(\omega)$  denote the Fourier transform of  $\psi_{j, k}(t)$ , and define ridgelets  $\rho_\lambda(x)$ ,  $\lambda = (j, k; i, \ell, \varepsilon)$  as functions of  $x \in \mathbb{R}^2$  using the frequency-domain definition

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}} (\hat{\psi}_{j, k}(|\xi|) w_{i, \ell}^\varepsilon(\theta) + \hat{\psi}_{j, k}(-|\xi|) w_{i, \ell}^\varepsilon(\theta + \pi)) / 2. \quad (5.1)$$

Here the indices run as follows:  $j, k \in \mathbb{Z}$ ,  $\ell = 0, \dots, 2^{i-1} - 1$ ;  $i \geq i_0$ ,  $i \geq j$ . Notice the restrictions on the range of  $\ell$  and on  $i$ . Let  $\Lambda$  denote the set of all such indices  $\lambda$ . It turns out that  $(\rho_\lambda)_{\lambda \in \Lambda}$  is a complete orthonormal system for  $L^2(\mathbb{R}^2)$ .

There is a close connection between ‘pure’ and orthonormal ridgelets. Pure ridgelets are supported on lines in the Fourier domain: that is, the frequency representation of a pure ridgelet is given by (provided that the profile  $\psi$  is real valued)

$$\hat{\psi}_{j, \ell, k}(\xi) = (\hat{\psi}_{j, k}(|\xi|) \delta(\theta - 2\pi 2^{-j} \ell) + \hat{\psi}_{j, k}(-|\xi|) \delta(\theta + \pi - 2\pi 2^{-j} \ell)) / 2 \quad (5.2)$$

using a formulation emphasizing the resemblance with (5.1). In the ridgelet construction, the angular variable  $\theta$  is uniformly sampled at each scale; the sampling step being inversely proportional to the scale. In contrast, the sampling idea is replaced by the wavelet transform for orthonormal ridgelets. This is the reason why orthonormal ridgelets can perfectly reconstruct objects from  $L^2(\mathbb{R}^2)$  without support constraints. It is interesting to note that the restriction on the range, namely,  $i \geq j$  in the definition (5.1), gives angular scaling functions at scales inversely proportional to the sampling steps of pure ridgelets.

**Theorem 5.1** *Let  $g \in H^s(\mathbb{R}^2)$ ,  $s > 0$ , with compact support and put  $f(x) = H(u \cdot x - b) g(x)$ . Then the orthonormal ridgelet coefficient sequence  $\alpha$  of  $f$  obeys*

$$\|\alpha\|_{w\ell_p} \leq C \|g\|_{H^s}, \quad \text{with } 1/p = s/2 + 1/2,$$

*for some constant  $C$  not depending on  $f$ . It then follows that the truncated  $n$ -term partial reconstruction  $f_n$  achieves the error bound*

$$\|f - f_n\|_2 \leq C n^{-s/2} \|g\|_{H^s}.$$

The proof is an application of Theorem 3.1 and consists of minor modifications to the proof of Theorem 4.1. In the following, we outline the essential steps, thus avoiding worthless repetition.

First, observe that for  $\varepsilon = 0$  ( $i = j$ ) and any choice of  $\gamma > 0$ , the localization of the angular scaling

function gives the upper bound

$$\begin{aligned}
|\langle f, \rho_\lambda \rangle| &= \left| \int \hat{f}(\lambda, \theta) |\lambda|^{1/2} (\hat{\psi}_{j,k}(|\lambda|) w_{j,\ell}^{\varepsilon=0}(\theta) + \hat{\psi}_{j,k}(-|\lambda|) w_{j,\ell}^{\varepsilon=0}(\theta + \pi)) / 2 \, d\lambda d\theta \right| \\
&\leq C 2^j \int (1 + 2^j |\theta - 2\pi \ell 2^{-j}|)^{-\gamma} \, d\theta \left| \int \hat{f}(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(|\lambda|) \, d\lambda \right| \\
&\quad + C 2^j \int (1 + 2^j |\theta + \pi - 2\pi \ell 2^{-j}|)^{-\gamma} \, d\theta \left| \int \hat{f}(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(-|\lambda|) \, d\lambda \right|, \quad (5.3)
\end{aligned}$$

where  $\gamma > 0$  may be chosen arbitrarily large. (The previous inequality used the fact  $|w_{j,\ell}^{\varepsilon=0}(\theta)| \leq C 2^{j/2} (1 + 2^j |\theta - 2\pi \ell 2^{-j}|)^{-\gamma}$ .) The point of this paper has been precisely to bound quantities like  $\left| \int \hat{f}(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(|\lambda|) \, d\lambda \right|$ . For instance, let  $I_{j,\ell} = \{\theta, |\theta - 2\pi 2^{-j} \ell| \leq 2^{-j}\}$  and set

$$\beta_{j,\ell,k} = 2^j \int_{I_{j,\ell}} \left| \int \hat{f}(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(|\lambda|) \, d\lambda \right|.$$

Then, we proved that (dimension 2)

$$\|\beta\|_{w\ell_p} \leq C \|g\|_{H^s}, \quad 1/p = s/2 + 1/2.$$

Compare with (4.5) and Theorem 4.1. Hence, a reasoning similar to the one developed for Theorem 4.1 gives

$$\|\alpha^{\varepsilon=0}\|_{w\ell_p} \leq C \|g\|_{H^s}, \quad 1/p = s/2 + 1/2. \quad (5.4)$$

The point is that the contributions associated with the orthonormal ridgelets corresponding to parameter values  $i > j$  become negligible as  $i$  goes to infinity. This is due to the compactness of the support of  $f$ . Indeed, standard wavelet calculations give

$$\begin{aligned}
|\langle f, \rho_\lambda \rangle| &= \left| \int \hat{f}(\lambda, \theta) |\lambda|^{1/2} (\hat{\psi}_{j,k}(|\lambda|) w_{i,\ell}^{\varepsilon=1}(\theta) + \hat{\psi}_{j,k}(-|\lambda|) w_{i,\ell}^{\varepsilon=1}(\theta + \pi)) / 2 \, d\lambda d\theta \right| \\
&\leq C 2^{-in} 2^{i/2} 2^{j/2} \int (1 + 2^i |\theta - 2\pi \ell 2^{-i}|)^{-\gamma} \, d\theta \left| \int (\partial_\theta^n \hat{f})(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(|\lambda|) \, d\lambda \right| \\
&\quad + C 2^{-in} 2^{i/2} 2^{j/2} \int (1 + 2^i |\theta + \pi - 2\pi \ell 2^{-i}|)^{-\gamma} \, d\theta \left| \int (\partial_\theta^n \hat{f})(\lambda, \theta) |2^{-j} \lambda|^{1/2} \hat{\psi}_{j,k}(-|\lambda|) \, d\lambda \right|.
\end{aligned}$$

The proof of the previous inequality follows from integration by parts together with the vanishing moment properties and the localization of the wavelets  $w_{i,\ell}^{\varepsilon=1}(\theta)$ . (We used the trivial bound on the size of the angular wavelets  $w_{i,\ell}^{\varepsilon=1}$ ; i.e.,  $|w_{i,\ell}^{\varepsilon=1}(\theta)| \leq C 2^{i/2} (1 + 2^i |\theta + \pi - 2\pi \ell 2^{-i}|)^{-\gamma}$ .)

Observe now that

$$\partial_\theta \hat{f}(\lambda, \theta) = \lambda (-\sin \theta (\partial_1 \hat{f})(\lambda, \theta) + \cos \theta (\partial_2 \hat{f})(\lambda, \theta)),$$

and this formula may be iterated to obtain derivatives with respect to the angular variable  $\theta$  of higher orders.

We may then substitute polar derivatives with respect to  $\theta$  by cartesian derivatives and obtain

$$\begin{aligned} |\langle f, \rho_\lambda \rangle| &\leq C 2^j 2^{-(i-j)(n-1/2)} \int (1+2^i |\theta-2\pi \ell 2^{-i}|)^{-\gamma} d\theta \sum_{|\alpha| \leq n} \left| \int (D^\alpha \hat{f})(\lambda, \theta) |2^{-j} \lambda|^{|\alpha|+1/2} \hat{\psi}_{j,k}(|\lambda|) d\lambda \right| \\ &+ C 2^j 2^{-(i-j)(n-1/2)} \int (1+2^i |\theta+\pi-2\pi \ell 2^{-i}|)^{-\gamma} d\theta \sum_{|\alpha| \leq n} \left| \int (D^\alpha \hat{f})(\lambda, \theta) |2^{-j} \lambda|^{|\alpha|+1/2} \hat{\psi}_{j,k}(-|\lambda|) d\lambda \right|. \end{aligned}$$

We already argued in the proof of Corollary 4.2 that, because of the compactness of the support of the distribution  $f$ , the estimates we obtained for  $\hat{f}$  are valid for the derivatives  $D^\alpha \hat{f}$ . Hence, we essentially have the same bound as in (5.3) but for an exponentially decaying factor  $2^{-(i-j)(n-1/2)}$  where  $n$  might be chosen as large as we want. It is then not too difficult to check that the sequence  $\alpha^{\varepsilon=1}$  satisfies

$$\|\alpha^{\varepsilon=1}\|_{w\ell_p} \leq C \|g\|_{H^s}, \quad 1/p = s/2 + 1/2.$$

The  $w\ell_p$  boundedness of the sequence  $\alpha$  naturally follows from this last display and (5.4).  $\blacksquare$

## 6 Discussion

Unlike any known system, ridgelets allow optimal partial reconstructions of  $L_2$  Sobolev functions with linear singularities. These good approximations are, moreover, simply obtained by thresholding the exact ridgelet series (1.4).

### 6.1 Ridgelets and functional classes

As we pointed out in the introduction, wavelets are optimal to represent smooth functions with point-singularities. From a functional viewpoint, we may say that wavelets provide unconditional bases for the Besov spaces and the Triebel spaces [13] and, therefore, provide near-optimal approximations to elements taken from functional balls of such spaces. A natural question would be: what are the functional spaces that are naturally associated with ridgelets? The analysis that we presented already suggests an answer. It is certainly possible to build new functional spaces whose typical elements resemble our mutilated Sobolev objects. In this direction, we might be tempted to consider, for instance, convex combinations of objects like (1.2); let

$$\mathcal{S}_H = \left\{ f = \sum_i a_i f_i, \sum_i |a_i| \leq 1 \right\},$$

where the  $f_i$ 's are our templates; i.e., functions of the form

$$f_i(x) = H(u_i \cdot x - b_i) g_i(x), \quad \|g_i\|_{H^s} \leq 1, \quad \text{supp } g \subset [0, 1]^d.$$



Our functional class  $\mathcal{S}_H$  would then be meant to represent objects composed of singularities across hyperplanes: typical elements of this class are smooth away and discontinuous across these same hyperplanes. There may be an arbitrary number of singularities which may be located in all orientations and positions. In the author’s unpublished thesis [3], it is then proved that ridgelets provide near-optimal representations of objects of this kind, as expected.

This is, indeed, part of a larger picture. A new notion of smoothness may be introduced leading to new functional classes that are naturally associated with ridgelets. This new notion of smoothness is nonclassical; it is discussed in [3] and briefly exposed in [7]. Full details will be provided in a separate paper.

## 6.2 Curved singularities

We would like to emphasize that this paper only considered linear singularities. Ridgelets are not able efficiently to represent smooth functions with curved singularities. For instance, in dimension  $d$ , consider the indicator function of the unit ball

$$f(x) = 1_{\{|x| \leq 1\}},$$

and let  $\alpha$  denote the ridgelet coefficient sequence of  $f$ . Then, [3] shows that

$$\#\{n, \text{ s.t. } |\alpha_n| \geq 1/n\} \geq C n^{2(1-1/d)}, \tag{6.1}$$

yielding partial reconstructions converging only at the rate  $n^{-\frac{1}{2(d-1)}}$ . We quote from [7]: “Unfortunately, the task that ridgelets must face is somewhat more difficult than the task which wavelets must face, since zero-dimensional singularities are inherently simpler objects than higher-dimensional singularities. In effect, zero-dimensional singularities are all the same – points – while a one-dimensional singularity – lying along a 1-dimensional set – can be curved or straight.” It is remarkable, however, that both wavelet and ridgelets, two fundamentally different systems achieve the same degree of sparsity.

The method of localization enables us to obtain sharper approximation bounds on objects with curved singularities. The localization idea is rather straightforward and has been for instance previously deployed in the time frequency literature. We outline this idea in dimension 2: first, partition the unit square into small squares, and smoothly localize the function into smooth pieces supported on or near those squares; then take the ridgelet transform on each piece. This is the basis of the so called *monoscale ridgelet transform* [5]. Again, partial reconstructions simply obtained by keeping the largest coefficients are shown to provide good approximation bounds (of higher order than wavelet or ridgelet approximations).

Further, [6] developed a new approach, namely, the *curvelet transform* that combines ideas from ridgelet analysis and wavelet analysis. In two dimensions, the curvelet transform provides optimal

representations of smooth functions with twice differentiable singularities, a fact whose roots are grounded on the results presented in this paper.

## References

- [1] R. A. Adams. *Sobolev spaces*. Academic Press, New York, 1975.
- [2] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*, volume 223 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin-New York, 1976.
- [3] E. J. Candes. *Ridgelets: theory and applications*. PhD thesis, Department of Statistics, Stanford University, 1998.
- [4] E. J. Candes. Harmonic analysis of neural networks. *Applied and Computational Harmonic Analysis*, 6:197–218, 1999.
- [5] E. J. Candes. *Monoscale ridgelets for the representation of images with edges*. Technical report, Department of Statistics, Stanford University, 1999. Submitted for publication.
- [6] E. J. Candès and D. L. Donoho. *Curvelets*. Manuscript. <http://www-stat.stanford.edu/~donoho/Reports/1998/curvelets.zip>, 1999.
- [7] E. J. Candes and D. L. Donoho. Ridgelets: the Key to Higher-dimensional Intermittency? *Phil. Trans. R. Soc. Lond. A.*, 357:2495–2509, 1999.
- [8] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1:100–115, 1993.
- [9] D. L. Donoho. Unconditional bases and bit-level compression. *Applied and Computational Harmonic Analysis*, 3:388–392, 1996.
- [10] D. L. Donoho. *Orthonormal ridgelets and linear singularities*. Technical report, Department of Statistics, Stanford University, 1998. Submitted for publication.
- [11] D. L. Donoho, M Vetterli, R. A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory*, 44:2435–2476, 1998.
- [12] P. G. Lemarié and Y. Meyer. Ondelettes et bases Hilbertiennes. *Rev. Mat. Iberoamericana*, 2:1–18, 1986.
- [13] Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992.

- [14] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.
- [15] H. Triebel. *Theory of Function Spaces. II.*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992.