Monoscale Ridgelets for the Representation of Images with Edges

Emmanuel J. Candès

Department of Statistics
Stanford University
Stanford, California 94305–4065

In a previous paper [1], the author introduced a new system for representing multivariate functions, namely, the *ridgelets*. In a following article [3], ridgelets were shown to be optimal for representing functions that are smooth away from hyperplanes, e.g. in two dimensions ridgelets provide optimally sparse representations of smooth functions that are discontinuous along lines, i.e. straight edges. This is unlike Fourier or wavelet methods.

This paper applies the localization principle for constructing local ridgelet frames, the *monoscale ridgelets.* This simple refinement allows efficient representations of smooth images with smooth edges. A model of such images is introduced and the present work shows that naive thresholding of the monoscale ridgelet expansion gives optimal nonlinear approximation bounds. The method is constructive and computationally attractive. Potential applications of these results include image compression and statistical estimation.

Key Words and Phrases. Edges. Partitioning. Local Fourier Transform (JPEG). Ridge functions. Ridgelets. Radon transform. Nonlinear approximation. Thresholding of ridgelet coefficients.

Acknowledgments. This research was supported by National Science Foundation grant DMS 95–05151 and by AFOSR MURI 95–P49620–96–1–0028.

1 Introduction

Candès (1998) developed a new set of ideas, *ridgelet analysis*, and showed how they can be applied to solve important problems such as constructing neural networks or approximating and estimating multivariate functions by linear combinations of ridge functions. This paper argues that ridgelets offer a breakthrough for approximating certain kinds of images with edges.

1.1 The problem of edges

Consider a simple model of images with edges: Let $\Gamma(s, C)$ be the set of curves defined by $\Gamma(s, C) = \{\gamma : [0,1] \to [1/10, 9/10]^2, \ \ \|\gamma\|_{C^3} \le C\}$. (Here, C^3 denotes the curves that are three times differentiable with bounded third derivative.) In addition, we will require that the curves do not loop around: that is, for $\delta > 0$, there must exist $\eta > 0$ such that

$$\|\gamma(t) - \gamma(t')\| \le \delta \Longrightarrow |t - t'| \le \eta(\delta), \quad \eta \to 0 \text{ as } \delta \to 0. \tag{1.1}$$

For a given $\gamma \in \Gamma(s, C)$ satisfying the 'no loop' condition (1.1), we define the set of images with singularity γ as

$$\mathcal{F}_\gamma(r, A) = \{f \in [0,1]^2 \setminus \gamma[0,1], \ \ \|f\|_{C^r} \le A\}.$$

Finally, we define our model to be

$$\mathcal{F}_\Gamma(r, A) = \bigcup_{\gamma \in \Gamma(s,C)} \mathcal{F}_\gamma(r, A). \tag{1.2}$$

Roughly speaking, (1.2) is meant to represent smooth images with discontinuities along arbitrary smooth edges γ.

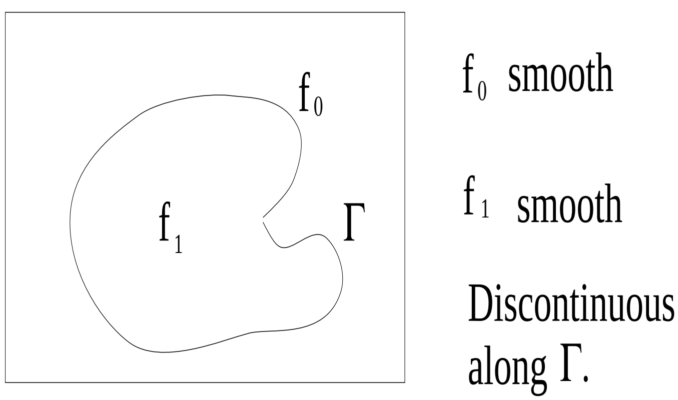

Figure 1: Image Model

Representation. We are interested in sparse representations of images from our model and in finding, ultimately, good m-term approximations.

Traditional methods based on wavelets, Fourier series, local cosine transforms (JPEG), or splines fail at efficiently representing images from our model. The edge limits the approximation error. To detail this claim, let us examine a very simple example: let f be the indicator function of the disk of radius 1/2, $f(x_1, x_2) = 1_{\{\sqrt{x_1^2 + x_2^2} \le 1/2\}}$. In some sense, this is a very simple object and we would like to find sparse representations of objects of this kind. Suppose for instance that one wishes to represent this object in a Fourier basis adapted to the unit square, say. Then the number of coefficients larger than $1/m$ (in absolute value) is bounded below by $c\,m^{4/3}$. This immediately translates into lower bounds for nonlinear approximations. Letting f_m^F be the best m-term partial reconstruction of f, the L_2-squared error of such an approximation obeys

$$\|f - f_m^F\|_{L_2}^2 \geq c\,m^{-1/2}. \tag{1.3}$$

The situation is quite different if one chooses to represent this same object in a wavelet basis. The number of wavelet coefficients larger than $1/m$ is bounded below by $c\,m$. The squared error of approximation of the best m-tem approximation f_m^W now obeys

$$\|f - f_m^W\|_{L_2}^2 \geq c\,m^{-1}. \tag{1.4}$$

Both bounds (1.3) and (1.4) are sharp in the sense that the partial reconstructions f_m^F and f_m^W converge at the rates $m^{-1/2}$ and m^{-1} respectively. In this sense, one might say that wavelets represent a significant improvement over sinusoids.

The previous example illustrates a general fact: let f be an object $f \in \mathcal{F}_\Gamma(r, A)$ taken from our model, $1 \leq r$; the methods mentioned above would give at the very best a L_2-squared error of approximation of order m^{-1}. This rate is not optimal and the point of this paper is to derive a very simple scheme that improves on these bounds.

Estimation. Suppose now that from our model one wishes to recover an image from noisy data. Although we do not attempt to go that far in this paper, it is interesting to observe that the inability of the aforementioned methods to approximate smooth images with edges efficiently has a 'statistical corollary': these methods do not remove noise from images with edges efficiently. For instance, a thresholding estimator in a nice wavelet basis does not provide good estimation bounds. A more quantitative discussion is provided in section 5.

1.2 Ridgelets

In [1], the author introduced a new discretization of the frequency plane that led to the construction of ridgelet frames. We will briefly explain the ridgelet construction in two dimensions. Suppose that we have a univariate function ψ satisfying an oscillatory condition, namely,

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

A ridgelet is a function of the form

$$\frac{1}{a^{1/2}} \psi\left(\frac{x\cos\theta + y\sin\theta - b}{a}\right). \tag{1.5}$$

By definition, a ridgelet is a ridge function: that is, a function of a linear combination of the variables, therefore, constant along lines. A ridgelet has a scale a, an orientation θ, and a location parameter b. In two dimensions, ridgelets are concentrated around lines: roughly speaking, the ridgelet (1.5) is supported near the line $\{x, \, |u \cdot x - b| \leq a\}$.

To cut a long story short, one can find a discretization such that the collection

$$\{\psi_{j,\ell,k}(x) = 2^{j/2}\psi(2^j(x_1\cos(\theta_{j,\ell}) + x_2\sin(\theta_{j,\ell}) - k)\,\}_{(j \geq j_0, \ell, k)} \tag{1.6}$$

is a frame for the unit square; the frame property says that for any f supported in the square with finite L^2 norm, there exist two constants $A, B > 0$ such that

$$A \, \|f\|_{L_2}^2 \le \sum_{j,\ell,k} |\langle \psi_{j,\ell,k}, f \rangle|^2 \le B \, \|f\|_{L_2}^2. \tag{1.7}$$

Then the discrete collection of ridgelets is complete in $L_2[0,1]^2$ and any function f can be reconstructed from the knowledge of its coefficients $(\langle \psi_{j,\ell,k}, f \rangle)$. In addition, the frame property (1.7) implies the existence of a dual set $(\widetilde{\psi_{j,\ell,k}})$ together with the classical decomposition

$$f = \sum_{j,\ell,k} \langle f, \widetilde{\psi_{j,\ell,k}} \rangle \psi_{j,\ell,k} = \sum_{j,\ell,k} \langle f, \psi_{j,\ell,k} \rangle \widetilde{\psi_{j,\ell,k}}, \tag{1.8}$$

with equality holding in an L_2 sense.

Ridgelets are directional and, here, the interesting aspect is the discretization of the angular variable,

$$\theta_{j,\ell} = 2\pi\theta_0 \ell 2^{-j} :$$

that is, the sampling step is inversely proportional to the scale. A detailed exposition on the ridgelet construction may be found in [1].

We would like to emphasize that there is an important connection between ridgelet analysis and wavelet analysis of the Radon transform. Let $\mathcal{R}$ be the Radon transform of f [7]

$$\mathcal{R}f(\theta, t) = \int f(x_1, x_2) \, \delta(x_1 \cos\theta + x_2 \sin\theta - t) \, dx_1 dx_2.$$

Then a ridgelet coefficient may be viewed as a kind of wavelet coefficient of the Radon transform; i.e.,

$$\langle f, \psi_{j,\ell,k} \rangle = \langle \mathcal{R}f(\theta_{j,\ell}, \cdot), \psi_{j,k} \rangle,$$

where $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$. This fact will be used implicitly throughout the remainder of the paper.

We will use the compact notation ψ_α $(\alpha \in A)$ and, therefore, we will keep in mind that the index α runs through an enumeration of the triples (j, ℓ, k). It will then be handy to use the notation A_j to refer to the subset of indices for which the scale is equal to j.

1.3 Ridgelets and singularities

Previous work [3] showed that 2-dimensional ridgelets were optimal for representing functions that are smooth away from lines (and in higher dimensions, for representing functions that are smooth away from hyperplanes). For instance, suppose that g is a C^r function that may be discontinuous along an arbitrary line; then selecting those terms corresponding to the m largest coefficients in the ridgelet series gives a degree of approximation obeying

$$\|f - f_m^R\|_{L_2}^2 \le C \, m^{-r}; \tag{1.9}$$

that is, the approximation error is as if there were no discontinuity. The result is true for any degree of smoothness r, and a remarkable feature of the result is that we do not need to know anything about the location of the singularity; the partial reconstruction is simply based on naive thresholding. In comparison, wavelets are unable to attain similar rates; wavelets give a degree of approximation only scaling like m^{-1} (1.4) regardless of the degree of smoothness $r \geq 1$.

Unfortunately, because of their straight and elongated structure, ridgelets are not able efficiently to represent smooth functions with curved singularities. Indeed, if f is an object from our class $\mathcal{F}$, the rate scales like $\max(m^{-1}, m^{-r})$. In other words, ridgelets are asymptotically equally as good as wavelets for this class of objects. The method of localization, however, adapts to this apparent difficulty; one can improve on our rate simply by localizing the ridgelets.

1.4 Localized ridgelets

This section presents the localization idea applied to the ridgelets. Let us consider, for instance, the uniform partitioning of the unit square with squares of side-length δ: this is illustrated on Figure 2. Suppose now that one wishes to reconstruct a piece of f that is supported on a square Q which intersects with the edge Γ. At a superficial level, this seems attractive as the edge has very low curvature relative to the scale and, hence, is fairly straight and should be easy to reconstruct. A more sophisticated argument developed in section 2 shows that truncating the ridgelet expansion provides very good approximations as long as the number of terms retained is less than the inverse of the curvature. The strategy that we develop exploits this fact and is essentially equivalent to constructing local ridgelet approximations on each square and piecing them up.

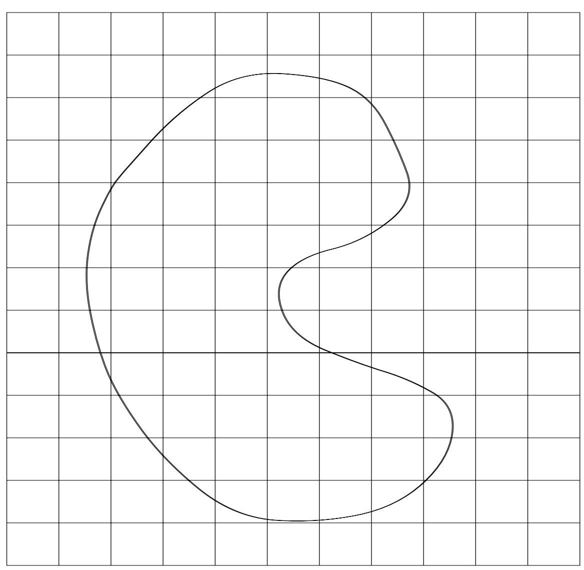

Figure 2: Uniform Partitioning

We now present the details of the construction and our main result. Suppose, we are given a ridgelet frame $(\psi_\alpha)_{\alpha \in A}$ of $L_2[0,1]^2$ as in [2] with frame bounds A and B, say. First, observe that if Q is a square of side-length $|Q|^{1/2}$ and lower-left corner x_Q, the collection $\left(|Q|^{-1/2}\psi(|Q|^{-1/2}(\cdot - x_Q))\right)$ is a frame for $L_2(Q)$. For a fixed square Q, we will abuse notations and let $(\psi_{Q,\alpha})_{\alpha \in A}$ denote a ridgelet frame for $2Q$ (this convention will hold throughout the remainder of the paper).

For an integer $s \geq 0$ and integers k_1, k_2, we let Q be the dyadic square defined by $Q = [k_1/2^s, (k_1+1)/2^s) \times [k_2/2^s, (k_2+1)/2^s)$. The collection of all dyadic squares at scale s will be denoted by $\mathcal{Q}_s$. The idea is to smoothly localize the function f we wish to represent near each of the dyadic squares of $\mathcal{Q}_s$. We choose an orthonormal partition of unity w_Q; that is, a collection of windows such that w_Q^2 is a partition of unity

$$\sum_{Q \in \mathcal{Q}_s} w_Q^2 = 1. \tag{1.10}$$

The following details a way of making up such an orthonormal partition: take a C^∞ univariate window ν supported in $[-1/4, 5/4]$ such that $\nu(t) = 1$ on $[0,1]$; define $v_Q = \nu(2^s x_1 - k_1)\,\nu(2^s x_2 - k_2)$; and renormalize the windows v_Q with

$$w_Q = v_Q / (\sum_{Q \in \mathcal{Q}_s} v_Q^2)^{1/2}.$$

It is then clear that the w_Q's obey the desired condition.

With this setup, we have the following definition-proposition:

Definition 1.1 *The collection $(w_Q\, \psi_{Q,\alpha})_{Q \in \mathcal{Q}_s, \alpha \in A}$ is a frame for $L_2(\mathbb{R}^2)$ and we will refer to it as a* **monoscale ridgelet frame with base-scale** *s. Moreover, if $(\psi_\alpha)_{\alpha \in A}$ is a tight frame ($A = B$ in (1.7)), then $(w_Q\, \psi_{Q,\alpha})_{Q \in \mathcal{Q}_s, \alpha \in A}$ is also a tight frame.*

We will briefly justify the claim. Since $\psi_{Q,\alpha}$ is a frame for $L_2(2Q)$ and supp $w_Q \subset 2Q$, it follows that there exist two constants (independent of Q) such that

$$A \|f\, w_Q\|_{L_2}^2 \leq \sum_{\alpha \in A} |\langle f, w_Q\, \psi_{Q,\alpha} \rangle|^2 \leq B \|f\, w_Q\|_{L_2}^2.$$

Summing this inequality over squares and using (1.10) give the frame property. The second part of the claim is trivial.

It will simplify notations to index the monoscale ridgelets via

$$\{\psi_\mu \equiv w_Q\, \psi_{Q,\alpha},\; Q \in \mathcal{Q}_s,\; \alpha \in A\}$$

where the set of all indices μ will be denoted by M_s; the subscript s reminding us of the value of the base scale.

As we will now see, local ridgelet frames provide good approximations to our model of images with edges.

1.5 Achievements

Approximation. There is a reasonable way to get good approximations using the simple modification of the ridgelet construction presented in the previous section. Indeed, suppose that one wishes to construct an m-term approximation to an element f taken from our class $\mathcal{F}(r, \Gamma)$. It suffices to expand the object in a localized ridgelet frame ψ_μ with base scale s (as discussed in the previous section) where s is chosen such that $2^{2s} \sim m/2$, and from the exact series

$$f = \sum_{\mu \in M_s} \langle f, \psi_\mu \rangle \widetilde{\psi_\mu},$$

extract the m-term partial reconstruction f_m^{MR} corresponding to the m largest coefficients $a_\mu = \langle f, \psi_\mu \rangle$. In section 4, it will be shown that the degree of approximation is given by

$$\|f - f_m^{MR}\|_{L_2}^2 \le C \max(m^{-r}, m^{-3/2}). \tag{1.11}$$

There are two main comments about this result:

1. Our result (1.11) shows that for $1 \le r \le 3/2$ (and $s \ge r$), there are approximations of L_2-squared error of order $O(m^{-r})$. This rate is comparable to the rate of approximation of C^r functions without edges. That is to say, thresholding the expansion of the object in the local ridgelet dictionary gives approximation bounds as if there were no singularities.

2. If $3/2 < r < 2$, our result only guarantees a squared error of order $m^{-3/2}$ whereas it can be shown that m-term approximations with an L_2-squared error of order $O(m^{-r})$ exist. Constructive, stable and very simple. Nevertheless, the proposed method represents a substantial gain over classical methods: a gain of order m^{-1} over trigonometric series, of order $m^{-1/2}$ over wavelets, etc.

As previously suggested, the argument relies on the fact that ridgelet expansions are very efficient at representing objects that are singular along curves with low curvature: that is, the decay of the ridgelet coefficient sequence is optimal up to a point that depends on the curvature of the singularity. In order to give a mathematical content to our statement, suppose that we have a C^2 function supported on the unit disk but that may be discontinuous along a curve of constant curvature $1/R$. Then, the m-term ridgelet approximation has the following property:

$$\|f - f_m\|_{L_2}^2 \le C\, m^{-2} \qquad m \le R.$$

This result is the object of section 2. In section 4, we will use this result to derive the approximation bound (1.11). We will comment on possible extensions and future work in section 5.

2 Representation of slowly curving singularities

In this section, we will suppose that we are given a ridgelet frame as in Candès (1998, Chapter 3), with φ and ψ being R times differentiable and ψ having vanishing moments through order D; $\min(R, D) \ge 2$.

The sequence of ridgelet coefficients of a given function f will be denoted by a: $a_\alpha = \langle f, \psi_\alpha \rangle$ and we let $|a|_{(n)}$ be the nth largest entry in the sequence $(|a_\alpha|)$.

The main result of this section is as follows:

Theorem 2.1 *Let w be a C^2 function supported on the unit disk $D \equiv \{x_1^2 + x_2^2 \le 1\}$ and let f be the function defined by*

$$f(x_1, x_2) = w(x_1, x_2) 1_{\{(x_1 - R)^2 + x_2^2 \le R^2\}}.$$

We have

$$\sum_{n > R} |a|_{(n)}^2 \le C\, R^{-2} \sup_{|m|=2} \|D^m w\|_\infty^2. \tag{2.1}$$

where the constant C does not depend on f.

As we shall see in Section 3, a consequence of Theorem 2.1 is that the partial reconstruction

$$f_n = \sum_\alpha a_\alpha \, 1_{\{|a_\alpha| > |a|_{(n)}\}} \tilde{\psi}_\alpha$$

obtained by naive thresholding obeys

$$\|f - f_n\|_{L_2}^2 \le C\, n^{-2}, \quad n \le R.$$

Remark 1. Theorem 2.1 is in some sense optimal as it cannot be fundamentally improved since the rate R^{-2} corresponds to m-term approximation error of C^2 functions. (The result would perhaps continue to hold under slightly weaker conditions such as $w \in W_2^2$, for instance; such refinements, however, are not the point of this paper.)

Remark 2. It will simplify the analysis to take ψ to be compactly supported. It is not a loss of generality as the theorem continues to hold if one also assumes sufficient decay on ψ together with its derivatives.

2.1 The Radon transform

The analysis turns out to be slightly simpler if one chooses to work with a translated version of f rather than with f itself. Let g be defined by

$$g(x_1, x_2) = f(x_1 - R, x_2);$$

it is then straightforward that it suffices to prove the desired property for g (the argument being simply to transport the estimates that one obtains for the wavelet coefficient of $Rg(t, \theta)$ to those of $Rf(t, \theta)$ using the trivial relationship $Rf(t, \theta) = Rg(t + R\cos\theta, \theta)$).

The Radon transform of g may be calculated as follows:

$$\begin{aligned} \mathcal{R}g(t,\theta) &= \int w(x_1 - R, x_2) 1_{\{x_1^2 + x_2^2 \le R^2\}} \delta(x_1 \cos\theta + x_2 \sin\theta - t)\, dx_1 dx_2 \\ &= \int w(t_1 \cos\theta - t_2 \sin\theta - R, t_1 \sin\theta + t_2 \cos\theta) 1_{\{t_1^2 + t_2^2 \le R^2\}}\, dt_2 \\ &= \int_{-\sqrt{R^2 - t_1^2}}^{\sqrt{R^2 - t_1^2}} w(t_1 \cos\theta - t_2 \sin\theta - R, t_1 \sin\theta + t_2 \cos\theta)\, dt_2, \end{aligned}$$

where the last equation holds if, of course, t_1^2 is chosen to be less than R^2.

Now, set

$$G_\theta(t_1, t_2) = \int_{-\infty}^{t_2} w(t_1 \cos\theta - u \sin\theta - R, t_1 \sin\theta + u \cos\theta)\, du;$$

with these notations, we have

$$\mathcal{R}g(t,\theta) = \begin{cases} G_\theta(t, \sqrt{R^2 - t^2}) - G_\theta(t, -\sqrt{R^2 - t^2}) & t^2 \le R^2 \\ 0 & t^2 > R^2 \end{cases}.$$

G_θ is, of course, two times differentiable; in particular, we have

$$\partial_2 G_\theta(t_1, t_2) = w(t_1 \cos\theta - t_2 \sin\theta - R, t_1 \sin\theta + t_2 \cos\theta).$$

Hence, the derivative $\partial_2 G_\theta(t_1, t_2)$ vanishes whenever $(t_1 \cos\theta - t_2 \sin\theta - R)^2 + (t_1 \sin\theta + t_2 \cos\theta)^2 \geq 1$. In addition, straightforward calculations show that

$$\sup_\theta \sup_{|m|=2} \|D^m G_\theta(\cdot, \cdot)\|_\infty \leq C \sup_{|m|=2} \|D^m w\|_\infty \tag{2.2}$$

for some universal constant C (we used the notation $D^m = \partial_1^{m_1} \partial_2^{m_2}$ for any double $m = (m_1, m_2)$).

We will now study the behavior of the Radon transform $\mathcal{R}g(t, \theta)$; we will limit our study to $\theta \in [-\pi/2, \pi/2]$ because of the relationship $\mathcal{R}g(t, \theta) = \mathcal{R}g(-t, (\theta + \pi))$. We start by making a couple of observations.

- The Radon transform $\mathcal{R}g(\theta, \cdot)$ is obviously twice differentiable at any point $t \neq R$.
- The support constraint on w implies that the function G_θ is supported on a strip $|t_1 - R\cos\theta| \leq 1$, which means that $\mathcal{R}g(t, \theta) = 0$ whenever $|t - R\cos\theta| \geq 1$.
- The Radon transform $\mathcal{R}g(\theta, \cdot)$ may not be differentiable at the point $t = R$. There are two ranges of angular values for which $\mathcal{R}g(t, \theta)$ exhibit a very different behavior. Let Θ_R be defined by

$$\Theta_R = \{\theta, \, |\sin\theta| \leq 2/R\}. \tag{2.3}$$

For θ in the complement of Θ_R, the Radon transfrom $\mathcal{R}g(\theta, \cdot)$ is identically zero in a neighborhood of the point $t = R$, and therefore $\mathcal{R}g(\theta, \cdot)$ is at least twice differentiable over the real line whereas if $\theta \in \Theta_R$, $\mathcal{R}g(\theta, \cdot)$ may have a singularity of degree 1/2 at the point $t = R$. Indeed, for a fixed $\theta \in \Theta_R$, the Radon transform is possibly nonzero only if the support of g and the line of integration $\mathcal{L}_{t,\theta}$ have a nonempty intersection. The support of g is included in the disk $\{(x_1, x_2), \, x_1^2 + x_2^2 \leq R^2\}$ and in the disk $\{(x_1, x_2), \, (x_1 - R)^2 + x_2^2 \leq 1\}$. Using polar coordinates, for a point $x = (x_1, x_2) = (\rho\cos\phi, \rho\sin\phi) \in \mathcal{L}_{t,\theta}$ to be in the support of g, we must have

$$\rho < R, \text{ and}$$

$$(\rho\cos\phi - R)^2 + \rho^2 \sin^2\phi \leq 1.$$

In particular this implies that ϕ must obey

$$|\sin\phi| \leq R^{-1}.$$

Then, for a given value of ϕ obeying the previous condition, ρ must now obey

$$R\cos\phi - \sqrt{1 - R^2\sin^2\phi} \leq \rho \leq \max(R, R\cos\phi + \sqrt{1 - R^2\sin^2\phi}).$$

Finally, since $(\rho\cos\phi, \rho\sin\phi) \in \mathcal{L}_{t,\theta}$, we have

$$\rho\cos(\theta - \phi) = t$$

and, therefore, the Radon transform of g is possibly nonvanishing only in the interval

$$\left(R\cos\phi - \sqrt{1 - R^2\sin^2\phi}\right)\cos(\theta - \phi) \leq \rho \leq \max\left(R, R\cos\phi + \sqrt{1 - R^2\sin^2\phi}\right)\cos(\theta - \phi).$$

Of course, if θ is such that $|\sin\theta| > R^{-1}$, then $\cos(\theta - \phi) < 1$ ($|\sin\phi| < R^{-1}$) and hence the Radon transform is identically zero in a neighborhood of $t = R$ which is our claim.

The terminology 'in or near the wavefront' will refer to the set of angles Θ_R while the terminology 'away from the wavefront' will refer to the set $\{\theta,\, \theta \notin \Theta_R\}$.

We prove the following lemma:

Lemma 2.2 *Suppose that $\theta \in [-\pi/2, \pi/2]$. For $t \neq R$, the Radon transform $\mathcal{R}g(\theta, \cdot)$ is twice differentiable and the second derivative can be decomposed as follows:*

$$\frac{d^2}{dt^2}\mathcal{R}g(\theta,t) = F_1(\theta,t) + F_2(\theta,t), \tag{2.4}$$

where

$$F_1(\theta,t) = \cos^2\theta\, T_{1,1} + 2\cos\theta\sin\theta\, T_{1,2} + \sin^2\theta\, T_{2,2}, \tag{2.5}$$

with

$$T_{i,j} = \mathcal{R}\{(\partial_i\partial_j w)\, 1_R\}, \quad i,j = 1,2. \tag{2.6}$$

The 'singular' term $F_2(\theta,t)$ has the following properties:

1. *for $|\sin\theta| \leq 2/R$, there exists a constant C such that*

$$|F_2(\theta,t)| \leq C \max\left(R^{1/2}\,|R-t|^{-3/2}, R\,|R-t|^{-1}\right) \sup_{|m|=2} \|D^m w\|_\infty \qquad t \neq R;$$

2. *for $|\sin\theta| > 2/R$, $F_2(\theta,t)$ is differentiable and*

$$|\frac{d}{dt}F_2(\theta,t)| \leq C\,|\sin\theta|^{-3} \sup_{|m|=2} \|D^m w\|_\infty.$$

In both equations the constant C does not depend on g. Moreover, the support of $F_2(\theta,\cdot)$ is at most of length $2(R^{-1} + |\sin\theta|)$.

Proof of Lemma. For $t \neq R$, the first derivative of the Radon transform $\mathcal{R}g(\theta,\cdot)$ is is given by

$$\begin{aligned} \frac{d}{dt}\mathcal{R}g(\theta,t) &= \partial_1 G_\theta(t, \sqrt{R^2-t^2}) - \partial_1 G_\theta(t, -\sqrt{R^2-t^2}) \\ &\quad - \frac{t}{\sqrt{R^2-t^2}}\left(\partial_2 G_\theta(t,\sqrt{R^2-t^2}) + \partial_2 G_\theta(t,-\sqrt{R^2-t^2})\right), \end{aligned}$$

while the second is equal to

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{R}g(\theta,t) &= \partial_1^2 G_\theta(t,\sqrt{R^2-t^2}) - \partial_1^2 G_\theta(t,-\sqrt{R^2-t^2}) \\ &- \frac{2t}{\sqrt{R^2-t^2}}\left(\partial_1\partial_2 G_\theta(t,\sqrt{R^2-t^2}) + \partial_1\partial_2 G_\theta(t,-\sqrt{R^2-t^2})\right) \\ &- \frac{R^2}{(R^2-t^2)^{3/2}}\left(\partial_2 G_\theta(t,\sqrt{R^2-t^2}) + \partial_2 G_\theta(t,-\sqrt{R^2-t^2})\right) \\ &+ \frac{t^2}{(R^2-t^2)}\left(\partial_2^2 G_\theta(t,\sqrt{R^2-t^2}) - \partial_2^2 G_\theta(t,-\sqrt{R^2-t^2})\right). \end{aligned}$$

We will rewrite the previous decomposition as (2.4)

$$\frac{d^2}{dt^2}\mathcal{R}g(\theta,t) = F_1(\theta,t) + F_2(\theta,t),$$

where

$$F_1(\theta,t) = \partial_1^2 G_\theta(t,\sqrt{R^2-t^2}) - \partial_1^2 G_\theta(t,-\sqrt{R^2-t^2}), \quad F_2(\theta,t) = \frac{d^2}{dt^2}\mathcal{R}g(\theta,t) - F_1(\theta,t).$$

It is now easy to check that F_1 enjoys the decomposition (2.6). This decomposition states that $F_1(\theta,t)$ may more or less be rewritten as a sum of Radon transforms (with weights depending on the angular variable θ), a fact that will prove to be especially useful.

For $t \neq R$, thanks to (2.2) we have

$$\begin{aligned}\left|\frac{2t}{\sqrt{R^2-t^2}}\Big(\partial_1\partial_2 G_\theta(t,\sqrt{R^2-t^2}) + \partial_1\partial_2 G_\theta(t,-\sqrt{R^2-t^2})\Big)\right|& \\ &= C\,\frac{1}{\sqrt{R-|t|}}\,\frac{|t|}{\sqrt{R+|t|}}\sup_{|m|=1}\|D^m w\|_\infty \\ &\le C\,\frac{R^{1/2}}{\sqrt{R-|t|}}\sup_{|m|=1}\|D^m w\|_\infty.\end{aligned}$$

The other terms are handled in the same way, namely,

$$\begin{aligned}\left|\frac{R^2}{(R^2-t^2)^{3/2}}\Big(\partial_2 G_\theta(t,\sqrt{R^2-t^2}) + \partial_2 G_\theta(t,-\sqrt{R^2-t^2})\Big)\right|& \\ &\le C\,\frac{1}{(R-|t|)^{3/2}}\,\frac{R^2}{(R+|t|)^{3/2}}\sup_{|m|=2}\|w\|_\infty \\ &\le C\,\frac{R^{1/2}}{(R-|t|)^{3/2}}\sup_{|m|=2}\|w\|_\infty,\end{aligned}$$

while

$$\begin{aligned}\left|\frac{t^2}{(R^2-t^2)}\Big(\partial_2^2 G_\theta(t,\sqrt{R^2-t^2}) - \partial_2^2 G_\theta(t,-\sqrt{R^2-t^2})\Big)\right|& \\ &\le C\,\frac{1}{(R-|t|)}\,\frac{t^2}{(R+|t|)}\sup_{|m|=1}\|D^m w\|_\infty \\ &\le C\,\frac{R}{(R-|t|)}\sup_{|m|=1}\|D^m w\|_\infty,\end{aligned}$$

Now for $\theta \in [-\pi/2,\pi/2]$ and $|\sin\theta| \le 2/R$, the support of $F_2(\theta,\cdot) \subset [R\cos\theta - 1, R] \subset [R-2,R]$ which allows to drop the absolute value of t and gives (**??**).

Suppose now that $\theta \in [-\pi/2,\pi/2]$ is fixed so that $|\sin\theta| \ge 2R^{-1}$. It then not difficult to check that $\partial_2 G(t,\sqrt{R^2-t^2}) = w(t\cos\theta - \sqrt{R^2-t^2}\sin\theta, t\sin\theta + \sqrt{R^2-t^2}\cos\theta)$ is identically zero. For $F_2(\cdot,\theta)$ to be nonzero, the point $(t\cos\theta + \sqrt{R^2-t^2}\sin\theta, t\sin\theta - \sqrt{R^2-t^2}\cos\theta)$ must be in the support of w. i.e.

$$(t\cos\theta + \sqrt{R^2-t^2}\sin\theta - R)^2 + (t\sin\theta - \sqrt{R^2-t^2}\cos\theta)^2 \le 1.$$

The analysis is parallel to the one leading to (**??**). Indeed, letting $t = R\cos(\theta - \phi)$, the previous inequality becomes

$$(R\cos\phi - R)^2 + \rho^2 \sin^2\phi \le 1.$$

which implies that

$$R \inf_{\phi, |\sin\phi| \le R^{-1}} \cos(\theta - \phi) \le t \le R \sup_{\phi, |\sin\phi| \le R^{-1}} \cos(\theta - \phi).$$

Then using the identity

$$\cos(\theta - \phi) - \cos\theta = -2\cos\theta \sin^2\frac{\phi}{2} + \sin\theta\sin\phi,$$

we obtain that for any ϕ s.t. $|\sin\phi| \le R^{-1}$ we have

$$-2\cos\theta R^{-2} - R^{-1}|\sin\theta| \le \cos(\theta - \phi) - \cos\theta \le R^{-1}|\sin\theta|.$$

Then of course, this implies that $F_2(t,\theta)$ to be nonzero, t must obey

$$-2R^{-1} - |\sin\theta| \le t - R\cos\theta \le |\sin\theta|,$$

which shows that the support of $F_2(\cdot,\theta)$ is at most of length $2(R^{-1} + |\sin\theta|)$.

Finally, since $F_2(\theta,\cdot)$ is identically zero in a neighborhood of $t = R$, $F_2(\theta,\cdot)$ is differentiable (because w is twice differentiable) and an analysis similar to the one

This finishes the proof of the lemma. ∎

2.2 Ridgelet analysis

To clarify notations, $\psi_{j,k}$ will be the dyadic wavelet defined as $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ and $\psi_{j,\ell,k}$ the ridgelet $\psi_{j,\ell,k}(x_1,x_2) = \psi_{j,k}(x_1\cos\theta_{j,\ell} + x_2\sin\theta_{j,\ell})$. Recall that we have chosen ψ to be compactly supported with at least 3 vanishing moments. Therefore, if we let the function ρ be defined as

$$\rho(t) = \int_{-\infty}^{t} \psi(u)(t-u)\,du,$$

then ρ is compactly supported, has at least one vanishing moment ($\int t^k \rho(t)dt = 0$, $k = 0,1$) and obviously satisfies

$$\frac{d^2}{dt^2}\rho(t) = \psi(t).$$

The function ρ is *frameable* according to the definition given in [1]. As a consequence, the collection $\rho_{j,\ell,k}(x_1,x_2) = \rho_{j,k}(x_1\cos\theta_{j,\ell} + x_2\sin\theta_{j,\ell})$ is a Bessel sequence (and may very well be a ridgelet frame), i.e. for any f in $L_2[0,1]^2$, we have

$$\sum_{j,\ell,k} |\langle f, \rho_{j,\ell,k}\rangle|^2 \le B\,\|f\|_2^2.$$

We will split the ridgelet index set A and let A' be the subset of indices corresponding to angular values in or near the wavefront defined as follows:

$$A' \equiv \{(j,\ell,k), \theta_{j,\ell} \in \Theta_R\}.$$

In the sequel, the index α will enumerate the triples (j, ℓ, k) so that

$$a_\alpha = \langle g, \psi_{j,\ell,k} \rangle.$$

Let $a' = (a_\alpha)_{\alpha \in A'}$ be the subsequence corresponding to the angles in the wavefront, and similarly $a'' = (a_\alpha)_{\alpha \in A \setminus A'}$, the one corresponding to the angles away from the wavefront. It is trivial that the theorem will be proved if one shows

$$\sum_{n>R} |a'|^2_{(n)} \le C_1 \, R^{-2} \quad \text{and} \quad \sum_{n>R} |a''|^2_{(n)} \le C_2 \, R^{-2}. \tag{2.7}$$

We will now establish the above inequalities for the angles in and away from the wavefront separately.

2.3 Away from the wavefront

Whenever $\theta_{j,\ell} \in \Theta_R$, $\mathcal{R}g(\theta_{j,\ell}, \cdot)$ is twice differentiable and integrating by parts gives

$$\langle g, \psi_{j,\ell,k} \rangle = \langle \mathcal{R}g(\theta_{j,\ell}, .), \psi_{j,k} \rangle = 2^{-2j} \langle \frac{d^2}{dt^2} \mathcal{R}g(\theta_{j,\ell}, .), \rho_{j,k} \rangle.$$

In this case, one can view a ridgelet coefficient as the sum of two terms (recall (2.4)):

$$\langle g, \psi_{j,\ell,k} \rangle = a^{(0)}_{j,\ell,k} + a^{(1)}_{j,\ell,k}$$

where

$$a^{(0)}_{j,\ell,k} = 2^{-2j} \langle F_1(\theta_{j,\ell}, .), \rho_{j,k} \rangle \quad \text{and} \quad a^{(1)}_{j,\ell,k} = 2^{-2j} \langle F_2(\theta_{j,\ell}, .), \rho_{j,k} \rangle.$$

The decomposition (2.5)-(2.6) gives that $a^{(0)}_{j,\ell,k}$ is the sum of three terms

$$a^{(0)}_{j,\ell,k} = 2^{-2j} \left(\cos^2 \theta_{j,\ell} \, b^{(1,1)}_{j,\ell,k} + 2 \sin \theta_{j,\ell} \cos \theta_{j,\ell} \, b^{(1,2)}_{j,\ell,k} + \sin^2 \theta_{j,\ell} \, b^{(2,2)}_{j,\ell,k} \right),$$

where

$$b^{(i,i')}_{j,\ell,k} = \langle \mathcal{R}\{(\partial_i \partial_{i'} w) \, 1_R\}(\theta_{j,\ell}, \cdot), \rho_{j,\ell} \rangle = \langle (\partial_i \partial_{i'} w) \, 1_R, \rho_{j,\ell,k} \rangle, \quad i, i' = 1, 2.$$

Since the sequence $\rho_{j,\ell,k}$ is a Bessel sequence, there exists a constant C such that for any $i, i' = 1, 2$

$$\sum_{j,\ell,k} |b^{(i,i')}_{j,\ell,k}|^2 \le C \, \|(\partial_i \partial_{i'} w) \, 1_R\|_2^2.$$

Hence,

$$\sum_{j,\ell,k} 2^{4j} |a^{(0)}_{j,\ell,k}|^2 \le C \sum_{i<i'} \|(\partial_i \partial_{i'} w) \, 1_R\|_2^2 \le C \sup_{|m|=2} \|D^m w\|_\infty^2. \tag{2.8}$$

Then it easily follows that for any $J > 0$, we have

$$\sum_{j>J} \sum_{\ell,k} |a^{(0)}_{j,\ell,k}|^2 \le C \, 2^{-4J} \sup_{|m|=2} \|D^m w\|_\infty^2.$$

At any given scale j, the number of nonzero coefficients $a^{(0)}_{j,\ell,k}$ is bounded by $C \, 2^{2j}$, where C is a constant only depending on the ridgelet frame $\psi_{j,\ell,k}$. This comes from the following two facts:

- the support of $\rho_{j,\ell,k}$ is included in $\{x,\, |\cos\theta_{j,\ell}x_1 + \sin\theta_{j,\ell}x_2 - k2^{-j}| \le A2^{-j}\}$. Since the support of g is included in $\{x,\, |\cos\theta_{j,\ell}x_1 + \sin\theta_{j,\ell}x_2 - R\cos\theta_{j,\ell}| \le 1\}$, the number of locations k (for a given value of $\theta_{j,\ell}$) for which the support of $\rho_{j,\ell,k}$ has a non empty intersection with that of g is bounded by $C\,2^j$;
- the scale j counts a number of angular values $\theta_{j,\ell}$ bounded by $C\,2^j$.

Let m_J be the number of coefficients α with scale $j \le J$ that may a priori be nonzero. Of course, $m_J \le C\,2^{2J}$ and

$$\sum_{n>m_J} |a^{(0)}|^2_{(n)} \le \sum_{j>J}\sum_{\ell,k} |a^{(0)}_{j,\ell,k}|^2 \le C\,m_J^{-2} \sup_{|m|=2} \|D^m w\|^2_\infty.$$

This implies that for any $m > 0$ we have

$$\sum_{n>m} |a^{(0)}|^2_{(n)} \le C\,m^{-2} \sup_{|m|=2} \|D^m w\|^2_\infty.$$

The analysis of the coefficient sequence $a^{(1)}$ is very similar to the one exposed in Candès (1998)[pages 84–85]. Recall that

$$a^{(1)}_{j,\ell,k} = 2^{-2j}\langle F_2(\theta_{j,\ell},.), \rho_{j,k}\rangle.$$

For $\theta_{j,\ell} \in \Theta_R$ we argued that $F_2(\theta_{j,\ell},.)$ is differentiable (Lemma 2.2). The function $\rho_{j,k}$ has one vanishing moment and a classical argument [12] then shows that

$$\begin{aligned} |\langle F_2(\theta_{j,\ell},\cdot), \rho_{j,k}\rangle| &\le C\,2^{-3j/2} \sup_{t\,\in\, supp\,\rho_{j,k}} |\frac{d}{dt}F_2(\theta_{j,\ell},t)| \\ &\le C\,2^{-3j/2}\,|\sin\theta_{j,\ell}|^{-3} \sup_{|m|=2} \|D^m w\|_\infty. \end{aligned} \qquad (2.9)$$

(The proof of the above inequality consists in integrating by parts and we omit it.) We then have

$$|a^{(1)}_{j,\ell,k}| \le C\,2^{-j(3+1/2)}|\sin\theta_{j,\ell}|^{-3} \sup_{|m|=2} \|D^m w\|_\infty. \qquad (2.10)$$

To control the decay the coefficient sequence $a^{(1)}$, we will use the following property proved in [8], for example: let (b_n) be an arbitrary sequence (b_n), then

$$\left(\sum_{n>m} |b|^2_{(n)}\right)^{1/2} \le C\,m^{-r}\|b\|_{\ell_p}, \quad r = 1/p - 1/2. \qquad (2.11)$$

Hence, it will suffice for our purpose to show that the sequence $a^{(1)}$ is bounded in $\ell_{2/3}$. The length of the support of $F_2(\theta,\cdot)$ is bounded by $2(R^{-1} + |\sin\theta|)$ (see Lemma 2.2) and hence the set of possibly nonvanishing coefficients $\{k,\, a^{(1)}_{j,\ell,k} \ne 0\}$ has a cardinality bounded by $C\,2^j|\sin\theta|$ (recall $|\sin\theta| \ge 2/R$) for some constant C depending on the length of the support of ρ and hence only depending on that ψ. Then summing inequality (2.10) over the location parameter k gives

$$\sum_k |a^{(1)}_{j,\ell,k}|^p \le C\,2^{-jp(3+1/2)}|\sin\theta_{j,\ell}|^{-3p}2^j|\sin\theta_{j,\ell}| \sup_{|m|=2} \|D^m w\|^p_\infty.$$

For $\theta \in [-\pi/2, \pi/2]$, we have

$$2|\theta|/\pi \le |\sin\theta| \le |\theta|,$$

(of course, a similar statement is available in the range $\theta \in [\pi/2, 3\pi/2]$) and therefore for any $\theta_{j,\ell} \in [-\pi/2, \pi/2]$, one has

$$|\sin\theta_{j,\ell}|^{1-3p} \le C\,|\theta_{j,\ell}| = C\,(\theta_0\,\ell\,2^{-j})^{1-3p}.$$

The following bound is now deduced from the previous inequality: for any p s.t. $1-3p > -1$ (i.e. $p < 2/3$)

$$\sum_{\ell:|\sin\theta_{j,\ell}|>2\,R^{-1}} |\sin\theta_{j,\ell}|^{1-3p} \le C\,2^j.$$

Thus, at a given scale j we have

$$\sum_{\ell:|\sin\theta_{j,\ell}|>2R^{-1}} \sum_k |a^{(1)}_{j,\ell,k}|^p \le C\,2^{-jp(3+1/2)}2^j2^j \sup_{|m|=2}\|D^m w\|^p_\infty.,$$

provided that $p < 2/3$. It is useful to recall that the constant C in the above display only depends on parameters associated with the ridgelet frame.

Finally, the boundedness of the ℓ_p norm follows from the previous estimate since

$$\begin{aligned}\sum_{j\ge 0}\sum_{\ell:|\sin\theta_{j,\ell}|>2R^{-1}} \sum_k |a^{(1)}_{j,\ell,k}|^p &\le C\sum_j 2^{-jp(3+1/2)}2^{2j} \sup_{|m|=2}\|D^m w\|^p_\infty\\ &\le C\sup_{|m|=2}\|D^m w\|^p_\infty,\end{aligned}$$

as soon as $p(3+1/2) > 2$, i.e. $p > 4/7$. In particular for $p = 2/3$, the sequence $(a^{(1)}_{j,\ell,k})$ (restricted to $A \setminus A'$) satisfies

$$\|a^{(1)}\|_{\ell_{2/3}} \le C\sup_{|m|=2}\|D^m w\|,$$

which is what needed to be proved (2.11).

The proof for the angles away from the wavefront is now completed as of course one has

$$\sum_{n>R} |a^{(0)} + a^{(1)}|^2_{(n)} \le C\,R^{-2}\sup_{|m|=2}\|D^m w\|_\infty,$$

which proves the first half of (2.7).

2.4 In or near the wavefront

As a corollary of Lemma 2.2, there exists a constant C such that for any $\theta_{j,\ell}$ in Θ_R we have

$$|\langle F_2(\theta_{j,\ell},\cdot), \rho_{j,k}\rangle| \le C\max\left(R^{1/2}2^j(1+|k|)^{-3/2}, R2^{j/2}(1+|k|)^{-1}\right), \qquad (2.12)$$

provided that $\rho_{j,k}$ is identically zero in a neighborhood of $t = R$. Inequality (2.12) simply follows from

$$|\langle F_2(\theta_{j,\ell},\cdot), \rho_{j,k}\rangle| \le \sup_{t\,\in\, supp\,\rho_{j,k}} |F_2(\theta_{j,\ell}, t)|\,\|\rho_{j,k}\|_{L_1}.$$

As a consequence, the identity

$$a_{j,\ell,k} = \langle \mathcal{R}g(\theta_{j,\ell},\cdot), \psi_{j,k}\rangle = \langle \frac{d^2}{dt^2}\mathcal{R}g)(\theta_{j,\ell},\cdot), \rho_{j,k}\rangle = a^{(0)}_{j,\ell,k} + a^{(1)}_{j,\ell,k}$$

which holds whenever $t = R$ is not in the support of $\rho_{j,k}$ implies that

$$a_{j,\ell,k} \le C \max\left(R^{1/2}\, 2^{-j}(1+|k|)^{-3/2}, R\, 2^{-3j/2}(1+|k|)^{-1}\right), \quad \theta_{j,\ell} \in \Theta_R. \tag{2.13}$$

On the other hand, if $t = R$ is in the support of $\psi_{j,k}$ we have

$$|a_{j,\ell,k}| \le \sup_{t \in supp\, \psi_{j,k}} |\mathcal{R}g(\theta_{j,\ell}, t)|\, \|\psi_{j,k}\|_{L_1} \le C\, R^{1/2}\, 2^{-j/2} 2^{-j/2} = C\, R^{1/2}\, 2^{-j},$$

and therefore (2.13) continues to hold whenever $t = R \in \text{supp}\ \psi_{j,k}$.

Without loss of generality, assume that $R = 2^s$ and let A'_s be the subset of A' defined as follows: $(j, \ell, k) \in A'_s$ if and only if

1. $j \le s$ and $\theta_{j,\ell} \in \Theta_R$, or
2. $s < j \le 2s$, $\theta_{j,\ell} \in \Theta_R$ and $|k| \le K_j$.

For a given scale $j \le 2s$ and value of the angular variable $\theta_{j,\ell} \in \Theta_R$, $(j, \theta_{j,\ell}, k)$ is in A'_s for all values of the location parameter k if $j \le s$, whereas if $s < j \le 2s$, A'_s retains only the coefficients corresponding to $|k| \le K_j$. We will set

$$K_j = \epsilon_j 2^{-j} 2^{2s},$$

where ϵ_j might be chosen as follows:

$$\epsilon_j = \max\left(\frac{1}{1+(s-j)^2}, \frac{1}{1+(2s-j)^2}\right).$$

We observe that

$$\sum_{s<j\le 2s} \epsilon_j \le C,$$

for some constant C not depending on s.

Now, at a fixed scale j, the cardinality of the subset $\{\ell,\, \theta_{j,\ell} \in \Theta_R\}$ does not exceed $C\, \max(1, 2^{j-s})$, where the constant C only depends on the ridgelet frame. Altogether, the cardinality of $|A'_s|$ is bounded as follows:

$$|A'_s| \sim \sum_{j\le s} C\, 2^j + \sum_{s<j\le 2s} C\, 2^{j-s} K_j \le C\, 2^s + C\, 2^s \sum_{s<j\le 2s} \epsilon_j \le C\, 2^s.$$

Looking at the contributions of the terms that are not in A'_s, we have

$$\begin{aligned} \sum_{(j,\theta_{j,\ell},k)\in A'\setminus A'_s} |a_{j,\ell,k}|^2 &= \sum_{s<j\le 2s} \sum_{\ell:\theta_{j,\ell}\in\Theta_R} \sum_{k:|k|>K_j} |a_{j,\ell,k}|^2 + \sum_{j>2s} \sum_{\ell:\theta_{j,\ell}\in\Theta_R} \sum_k |a_{j,\ell,k}|^2 \\ &\equiv B_1 + B_2. \end{aligned}$$

The bound (**??**) gives

$$\sum_{k:|k|>K_j} |a_{j,\ell,k}|^2 \le C\, \max\left(2^s\, 2^{-2j}\, K_j^{-2},\, 2^{2s}\, 2^{-3j}\, K_j^{-1}\right).$$

Now the bound on the cardinality of the subset $\{\ell,\, \theta_{j,\ell} \in \Theta_R\}$ gives

$$\sum_{\ell:\theta_{j,\ell}\in\Theta_R} \sum_{k:|k|>K_j} |a_{j,\ell,k}|^2 \le C \,\max\left(2^{-j}\, K_j^{-2},\, 2^s\, 2^{-2j}\, K_j^{-1}\right).$$

Hence, using the definition of $K_j \equiv \epsilon_j 2^{2s-j}$ we obtain

$$\begin{aligned} B_1 &\le C \sum_{s<j\le 2s} \max\left(2^{-4s} 2^j \epsilon_j^2, 2^{-s} 2^{-j} \epsilon_j\right) \\ &\le C \sum_{s<j\le 2s} 2^j 2^{-4s} 2^j (1+(2s-j)^2)^2 + 2^{-s} 2^{-j}(1+(2s-j)^2) \\ &\le C\, 2^{2s} 2^{-4s} + 2^{-s} 2^{-s} = C\, 2^{-2s}. \end{aligned}$$

For the sake of completeness, the previous display made use of both

$$\sum_{j\le 2s} (1+(2s-j)^2)^2 2^j \sim 2^{2s}$$

(that is, the last term of this series dominates the sum of the others) and

$$\sum_{j>s} 2^{-j}(1+(j-s)^2) \sim 2^{-s}$$

(i.e., the first term now dominates the others). As far as the second term is concerned, we simply have

$$\begin{aligned} B_2 &\le C \sum_{j>2s} \sum_{\ell:\theta_{j,\ell}\in\Theta_R} 2^s\, 2^{-2j} + 2^{2s}\, 2^{-3j} \\ &\le C \sum_{j>2s} 2^{-j} + 2^s 2^{-2j} \le C\, 2^{-2s}. \end{aligned}$$

To summarize,

$$\sum_{i\in A\setminus A'_s} |a_\alpha|^2 \le C\, 2^{-2s}, \tag{2.14}$$

which proves the second half of (2.7). The proof of our theorem is complete. ∎

There is nothing special about the location and orientation of our singularity. In fact, repeating the exact same steps of our proof would give:

Corollary 2.3 *Let f be supported in the unit square, say, and be twice differentiable (with bounded second derivative) away from a singular curve Γ with constant curvature $1/R$. Then, if a_α is the ridgelet coefficient sequence of f, we have*

$$\sum_{n>R} |a|_{(n)}^2 \le C \sup_{|m|=2} \|D^m f\|^2_{L_\infty([0,1]^2\setminus\Gamma)}\, R^{-2}.$$

3 Main Result

In this section, we introduce a slightly different definition of monoscale ridgelets. This will ease the statement of our main approximation theorem which follows.

We add an "extra layer of coarse scale coefficients" to eliminate various artifacts. Consider a standard multiresolution analysis that is adapted to the unit square [7] so that the set of translates $\{2^s \varphi(2^s \cdot -k)\}$, $k = (k_1, k_2)$, $k_i = 0, 1, \dots, 2^s - 1$ is orthonormal (underlying there is, of course, a wavelet orthobasis of $L_2[0,1]^2$ $\{2^j \psi(2^j \cdot -k)\}$, $k = (k_1, k_2)$, $k_i = 0, 1, \dots, 2^j - 1$). Let P_s be the orthogonal projector onto V_s, the span of the φ_k's; i.e.,

$$P_s f \equiv \sum_k \langle f, \varphi_k \rangle \varphi_k \equiv \sum_k \beta_k \varphi_k, \quad \beta_k = \langle f, \varphi_k \rangle. \tag{3.1}$$

The following Pythagorean relationship holds

$$\|f\|_2^2 = \|P_s f\|_2^2 + \|(I - P_s) f\|_2^2. \tag{3.2}$$

Finally, define the coefficients

$$a_\mu = \langle (I - P_s) f, w_Q \psi_{Q,\alpha} \rangle \quad \mu = (Q, \alpha), \quad Q \in \mathcal{Q}_s, \alpha \in A. \tag{3.3}$$

Definition 3.1 *The monoscale ridgelet transform with base scale s is the mapping from functions $f \in L_2(\mathbb{R}^2)$ to the amalgamation of coefficients (β_k) and (a_μ).*

It will be more convenient to use a single notation to index the set of curvelet coefficients; the notation M_s' will stand for the union of M_s and $k \in \mathbf{Z}^2$ so that whenever $\mu \in M_s' \setminus M_s$, we let $a_\mu = \beta_k$.

The sum of squares of the coarse-scale-coefficients obey

$$\sum_k |\beta_k|^2 = \|P_s f\|^2$$

while that of the coefficients a_μ verifies

$$\sum_{\mu \in M_s} |a_\mu|^2 \sim \|(I - P_s) f\|^2.$$

It then follows that we have a quasi-Parseval relation, namely

$$\|f\|_2^2 \sim \sum_{\mu \in M_s'} |a_\mu|^2,$$

thanks to the Pythagorean relationship (3.2). Hence, the collection

$$\psi_\mu = \phi_{k_1,k_2} \; \mu \in M_s' \setminus M_s, \quad \psi_\mu = (I - P_s)(w_Q \, \psi_{Q,\alpha}), \; \mu \in M_s$$

is a frame of $L_2(\mathbb{R}^2)$. Dual elements $(\tilde{\psi}_\mu)_{\mu \in M_s'}$ exist with the property

$$f = \sum_{\mu \in M_s'} \langle f, \psi_\mu \rangle \tilde{\psi}_\mu. \tag{3.4}$$

From this exact series, one can extract finite approximations and we will let f_m^{MR} be the m-term approximation corresponding to the m-largest coefficients.

With these notations we have the following approximation theorem for elements of our model of images:

Theorem 3.2 *Let the base scale s be the largest integer such that $2^{2s} \leq m/2$. Then for any element f taken from our model, we have*

$$\|f - f_m^{MR}\|_2^2 \leq C \max(m^{-\alpha}, m^{-3/2}) \|f\|^2_{C^\alpha([0,1]^d \setminus \Gamma)},$$

where the constant C is, of course, independent of f.

Proof of Theorem. It is of course sufficient to prove the result for m of the form $m = 2^{2s+1}$, $s = 0, 1, \ldots$. We claim – and that is, indeed, our main result – that

$$\sum_{n>m} |a|^2_{(n)} \leq C \max(m^{-\alpha}, m^{-3/2}), \tag{3.5}$$

where a is the monoscale ridgelet coefficient sequence of f. The theorem follows immediately from (3.5) since the quasi-Parseval relationship that is available for frames transforms this inequality into an approximation bound. We give a proof of this fact. Let F be the frame analysis operator $Ff = (\langle f, \psi_\mu \rangle)$ and $\tilde{F}$ be the synthesis operator $\tilde{F} b_\mu = \sum b_\mu \tilde{\psi}_\mu$; then $\tilde{F} F = Id$ and $F \tilde{F}$ is the orthogonal projector onto the range of F and has, therefore, a norm (as an operator from ℓ_2 onto itself) bounded by 1. Recall the frame property

$$A \|f\|_2^2 \leq \sum |\langle f, \psi_\mu \rangle|^2 = \|Ff\|^2_{\ell_2} \leq B \|f\|_2^2.$$

Now, let b be an arbitrary sequence and define $\tilde{f} = \tilde{F} b = \sum b_\mu \tilde{\psi}_\mu$. Then, $\|\tilde{f}\|_2^2 \leq A^{-1} \|b\|^2_{\ell_2}$ where A is the lower frame bound that appear on the right-hand side of the previous display. This simply follows from

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

The theorem is proved provided that one establishes (3.5); this is the object of the next section.

4 Proof of the Main Result

Throughout this section, $Q \in \mathcal{Q}_s$ will denote a dyadic square of side-length 2^{-s}. Next, we will work with a unit-speed parameterization of the edge Γ: that is, $\Gamma = \{\gamma(t), 0 \leq t \leq \text{length}(\Gamma)\}$ with

$$\|\gamma'(t)\| = 1,$$

where the notation γ' stands the first derivative of γ. Similarly, γ'' will denote the second derivative of γ.

In order to prove our claim (3.5), it will be sufficient to establish that letting

$$a_\alpha = \langle (I - P_s) f, w_Q \psi_{Q,\alpha} \rangle,$$

we have

$$\sum_{n > 2^{2s}} |a|^2_{(n)} \leq C \max(2^{-2s\alpha}, 2^{-3s}).$$

We will distinguish two types of squares: the ones that intersect or are 'close' to the singular curve Γ

$$\mathcal{Q}_s' = \{Q \in \mathcal{Q}_s, \ d(Q, \Gamma) \leq 2^{-s+2}\}, \tag{4.1}$$

and the others $Q \in \mathcal{Q}_s$ such that $d(Q, \Gamma) > 2^{-s+2}$.

4.1 The squares away from the edge

Let Q be a square such that $Q \in \mathcal{Q}_s \setminus \mathcal{Q}'_s$; a rather classical argument shows that

$$\|(I - P_s)f\, w_Q\|_2^2 \leq C\, 2^{-2s(1+\alpha)}\, |f|^2_{\dot{C}^\alpha}, \tag{4.2}$$

for some constant C not depending on f. Indeed,

$$f(x) = \sum_k \beta_k \varphi_k(x) + \sum_{j \geq s} \sum_k \alpha_{j,k} \psi_{j,k}(x).$$

For any $x \in \mathbb{R}^2$, there are only a finite number of terms per scale j that are potentially nonzero in the above series, namely the terms corresponding to those indices (j,k) for which $x \in \text{supp } \psi_{j,k}$. Since the support of $\psi_{j,k}$ has a radius at most equal to $C \cdot 2^{-j}$ $(j \geq s)$ and $d(Q, \Gamma) \geq 2^{-s+2}$, whenever $Q \in \mathcal{Q}_s \setminus \mathcal{Q}'_s$ we have that

$$\text{supp } \psi_{j,k} \cap Q \neq \emptyset \;\Rightarrow\; \text{supp } \psi_{j,k} \cap \Gamma = \emptyset.$$

Let $\psi_{j,k}$ be such that $\text{supp } \psi_{j,k} \cap \Gamma = \emptyset$, then we have

$$|\alpha_{j,k}| \leq C\, 2^{-j(1+\alpha)}\, |f|_{\dot{C}^\alpha}, \tag{4.3}$$

since the function f is C^α over the support of $\psi_{j,k}$, see [12] for details.

At each scale j, we have

$$\#\Lambda_j^Q \equiv \#\{k,\ \text{supp } \psi_{j,k} \cap Q \neq \emptyset\} \leq C\, 2^{2j-2s}$$

and, therefore, since the corresponding coefficients are bounded by (4.3) we have

$$\begin{aligned}
\|(I - P_s)f\, w_Q\|_2^2 &\leq \sum_{j \geq s} \sum_{k \in \Lambda_j^Q} |\alpha_{j,k}|^2 \\
&\leq C \sum_{j \geq s} 2^{2j-2s} 2^{-2j(1+\alpha)}\, |f|^2_{\dot{C}^\alpha} \\
&\leq C\, 2^{-2s(1+\alpha)}\, |f|^2_{\dot{C}^\alpha}.
\end{aligned}$$

Now inequality (4.2) implies that

$$\sum_{Q \in \mathcal{Q}_s \setminus \mathcal{Q}'_s} \sum_\alpha |a_{Q,\alpha}|^2 \leq C \sum_{Q \in \mathcal{Q}_s \setminus \mathcal{Q}'_s} \|(I - P_s)f\, w_Q\|_2^2 \leq C\, 2^{-2s\alpha} |f|^2_{\dot{C}^\alpha}. \tag{4.4}$$

4.2 The squares near the edge

We start this section with a lemma that gives an upper bound on the number of squares $Q \in \mathcal{Q}'_s$ (4.1).

Lemma 4.1 *Suppose that we have a curve γ with finite arc-length and let Γ be the graph of this curve. Then,*

$$\#\{Q,\, Q \cap \Gamma \neq \emptyset\} \leq 2\, \text{length}\,(\Gamma)\, 2^s + 8.$$

Proof of Lemma. Without loss of generality, we will work with a unit-speed parameterization of the curve γ. Let $t_0 = 0 < t_1 = 2\delta < t_2 = 4\delta < \ldots < t_{n-1} = 2(n-1)\delta < t_n =$ length (Γ), where $\delta = 2^{-s}$ and n is the largest integer such that $2(n-1)\delta <$ length (Γ). We claim that

$$\bigcup_{Q: Q\cap\Gamma\neq\emptyset} Q \subset \bigcup_i Q_{\gamma(t_i)}(2\delta), \tag{4.5}$$

where $Q_{\gamma(t_i)}(2\delta) \equiv \{x, \, \|x-\gamma(t_i)\|_\infty \leq \delta\}$. By construction, we first observe that

$$x \in \bigcup_{Q: Q\cap\Gamma\neq\emptyset} Q \Longrightarrow \inf_{u\in\Gamma} \|x-u\|_\infty \leq \delta.$$

Next, we note that

$$\inf_{u\in\Gamma} \|x-u\|_\infty \leq \delta \Longrightarrow \exists i, \; \|x-\gamma(t_i)\|_\infty \leq 2\delta.$$

The reason is that if u is a point achieving the minimum with $u = \gamma(t)$, we have $t_{i-1} \leq t \leq t_i$ for some index i, $1 \leq i \leq n$; now,

$$\|x-\gamma(t_i)\|_\infty \leq \|x-\gamma(t)\|_\infty + \|\gamma(t)-\gamma(t_i)\|_\infty \leq \delta + \|\gamma(t)-\gamma(t_i)\|_2 \leq \delta + |t-t_i|$$

and similarly for the index $i-1$. The claim follows since $\inf_{j\in\{i-1,i\}} |t-t_j| \leq |t_{i-1}-t_i|/2 \leq \delta$. We finish the proof using the simple observation that the volume of the left-hand side of (4.5) is less or equal to the volume of the right-hand side of the same display: that is,

$$\#\{Q, \, Q\cap\Gamma\neq\emptyset\}\delta^2 \leq (n+1)\,(2\delta)^2.$$

Now $n \leq 1 +$ length $(\gamma)/(2\delta)$ which implies that

$$\#\{Q, \, Q\cap\Gamma\neq\emptyset\} \leq 2 \text{ length}\,(\gamma)/\delta^{-1} + 8,$$

which is what we sought to establish. ∎

There is a corollary to this lemma.

Corollary 4.2 *For* $\ell \leq n$

$$\#\{Q \in \mathcal{Q}_s, \; d_\infty(Q,\Gamma) \leq 2^{s+\ell+1} \neq \emptyset\} \leq 2^{\ell+1}\text{length }(\Gamma)\,2^s + 2^{\ell+4}.$$

The proof of this corollary is parallel to that of Lemma (4.1) and we omit it.

Lemma 4.3 *Suppose that supp* $w_Q \cap \Gamma \neq \emptyset$. *Then, one can find two polynomials* π_0 *and* π_1, *each at most of degree 1, and a circle of center* x_Q *and radius* r_Q *such that if* g *is set to be* $g(x) = \pi_0(x)\,1_{\{\|x-x_Q\|\leq r_Q\}} + \pi_1(x)\,1_{\{\|x-x_Q\|>r_Q\}}$

$$\|(f-g)\,w_Q\|_2^2 \leq C\,\max(2^{-2s(1+\alpha)}, 2^{-4s})$$

for some constant C.

Proof of Lemma. Let t_0 be such that $\gamma(t_0) \in \text{supp } w_Q$. We first observe that

$$|\gamma(t) - \gamma(t_0) - \gamma'(t_0)(t - t_0)| \le \frac{1}{2} \sup_s |\gamma''(s)| \, (t - t_0)^2 \le \frac{1}{2}|t - t_0|,$$

provided that $\sup_s |\gamma''(s)| \, |t - t_0| \le 1$. Now, this means that $|\gamma(t) - \gamma(t_0)| \ge 1/2|t - t_0|$ as long as $\sup_s |\gamma''(s)| \, |t - t_0| \le 1$. This observation then implies that for $|t - t_0| \ge 4\sqrt{2}\, 2^{-s}$, we have $|\gamma(t) - \gamma(t_0)| \ge 2\sqrt{2}\, 2^{-s}$, i.e. $\gamma(t) \notin 2Q \supset \text{supp } w_Q$. Our no-loop assumption (1.1) then guarantees that $\{t, \, \gamma(t) \in \text{supp } w_Q\} \subset [t_0 - 4\sqrt{2}\, 2^{-s}, t_0 + 4\sqrt{2}\, 2^{-s}]$.

Now suppose that the absolute value of the angle between $\gamma'(t_0)$ and the horizontal line is less than $\pi/4$, say. Then, the continuity of γ' implies that one can locally parameterize the curve like $(x_1, x_2 = h(x_1))$ and will let L be the subset of $2Q$ which is above the curve, i.e. $L \equiv 2Q \cap \{(x_1, x_2), \, x_2 \ge h(x_1)\}$; similarly, R will denote the region of $2Q$ which is below the curve. (If the absolute value of the angle is greater than $\pi/4$, we simply rotate the picture.) Now, going to back our model definition, the restriction of f to $2Q$ may be written as $f = f_0 \, 1_L + f_1 \, 1_R$.

We finish the proof with two elementary remarks. First, we clearly have two polynomials π_i, $i \in \{0, 1\}$, each at most of degree one, such that

$$\sup_{x \in 2Q} |f_i(x) - \pi_i(x)| \le C \, 2^{-s\alpha},$$

and we will write f_π the function $f_\pi = \pi_0 \, 1_L + \pi_1 \, 1_R$. It follows from the previous inequality that

$$\|(f - f_\pi) \, w_Q\|_2^2 \le C \, 2^{-2s(1+\alpha)}.$$

Now, put $\rho_Q = \gamma''(t_0)$: that is, the curvature of γ at the point $\gamma(t_0) \in Q$. Further, let $p(t)$ be the equation of the tangent circle (so that $p(t_0) = \gamma(t_0)$ and $p'(t_0) = \gamma'(t_0)$). It is then clear that

$$\begin{aligned} \int |p(t) - \gamma(t)| 1_{\{\gamma(t) \in \text{ supp } w_Q\}} \, dt &\le C \, \|\gamma\|_{\dot{C}^3} \int |t - t_0|^3 1_{\{\gamma(t) \in \text{ supp } w_Q\}} \, dt \\ &\le C \, \|\gamma\|_{\dot{C}^3} \, 2^{-4s}. \end{aligned}$$

In short, we have $\|f_\pi - g\|_2^2 \le C \, \|f\|_\infty^2 \, \|\gamma\|_{\dot{C}^3} \, 2^{-4s}$. The corollary now follows from the trivial inequality

$$\|f - g\|_2^2 \le 2 \left(\|f - f_\pi\|_2^2 + \|f_\pi - g\|_2^2 \right). \quad \blacksquare$$

The lemma has a useful corollary:

Corollary 4.4 *Suppose that $Q \in \mathcal{Q}_s'$ and let w_Q be a nice window supported in Q. Let a_Q be the coefficient sequence defined by $a_{Q,\alpha} = \langle (I - P_s) f, w_Q \psi_{Q,\alpha} \rangle$. Then,*

$$\sum_{n > (2^{-s} \rho_Q)^{-1}} |a_Q|_{(n)}^2 \le C \, \delta^2 \, (2^{-s} \rho_Q)^2. \tag{4.6}$$

Proof of Corollary. Suppose, that $\text{supp } w_Q \cap \Gamma \neq \emptyset$, then Lemma 4.3 implies that for any $x \in 2Q$ one may write $f(x) = g(x) + (f - g)(x)$ where g is of the following form

$$g(x) = \pi_0(x) \, 1_{\{\|x - x_Q\| \le 1/\rho_Q\}} + \pi_1(x) \, 1_{\{\|x - x_Q\| > 1/\rho_Q\}}.$$

We then use the above formula to extend g to the entire plane $\mathbb{R}^2$ and we slightly abuse notations in calling this extension g. We have $(I - P_s)f = (I - P_s)g + (I - P_s)(f - g)$, and it follows from our lemma that

$$\|((I - P_s)(f - g))\, w_Q\|_2^2 \le C\, 2^{-2s(1+\alpha)}. \tag{4.7}$$

We rewrite $(I - P_s)g$ as

$$(I - P_s)g \equiv g - h$$

and one easily verifies that the function h is twice differentiable and obey

$$\|h\|_\infty \le C \quad \text{and} \quad \|h\|_{\dot{C}^2} \le C\, 2^{2s}.$$

Define now the rescaling operator $T_Q g$ by

$$T_Q g = g(2^s x_1 - k_1, 2^s x_2 - k_2),$$

and rescale the function $(g-h)w_Q$ to the unit square using the inverse operator T_Q^{-1}. Then,

$$(T_Q^{-1}(g - h))\, w_Q = T_Q^{-1} g\, w - T_Q^{-1} h\, w,$$

where on one hand $T_Q^{-1}h$ is a C^2 function such that

$$\|T_Q^{-1}h\|_{C^2} \le C,$$

and on the other, *away from an edge of constant curvature* $2^{-s}\rho_Q$, we have

$$\|T_Q^{-1}g\|_{C^2} \le C.$$

Therefore, $T_Q^{-1}(g)\, w - T_Q^{-1}(h)\, w$ verifies the conditions of Theorem 2.1. With these notations we have

$$b_{Q,\alpha} \equiv \langle (I - P_s)g, w_Q \psi_{Q,\alpha}\rangle = 2^{-s}\langle (T_Q^{-1}(g - h))\, w, \psi_\alpha\rangle,$$

and applying the conclusion of Theorem 2.1 gives that the coefficient sequence b_Q (Q is fixed) obeys the following property:

$$\sum_{n > (2^{-s}\rho)^{-1}} |b_Q|_{(n)}^2 \le C\, 2^{-2s}(2^{-s}\rho)^2.$$

The previous inequality together with (4.7) prove the lemma. (The reasoning is simpler in the case supp $w_Q \cap \Gamma = \emptyset$; in this situation, one may take g to be polynomial, without having to split.) ∎

4.3 Piecing up

Finally, we collect the results obtained in the previous two sections to derive our main result. One one hand, the estimate (4.4) obtained for the squares away from the edge gives

$$\sum_{Q \in \mathcal{Q}_s \setminus \mathcal{Q}_s'} \sum_\alpha |a_{Q,\alpha}|^2 \le C\, 2^{-2s\alpha}.$$

On the other hand, the cardinality of $\mathcal{Q}'_s$ is bounded by (Lemma 4.1)

$$\#\mathcal{Q}'_s \leq 2^{\ell+1} \,\mathrm{length}\,(\Gamma)\, 2^s + 2^{\ell+4}$$

and, thus, summing the inequality (4.6) over the squares $Q \in \mathcal{Q}'_s$ gives

$$\sum_{n > 2^{2s}} |a|^2_{(n)} \leq C\, 2^{-3s}.$$

Hence, the last two inequalities give

$$\sum_{n > 2^{2s}} |\alpha|^2_{(n)} \leq C\, \max(2^{-2s\alpha}, 2^{-3s}),$$

which is the desired conclusion.

5 Discussion

5.1 Orthonormal ridgelets

In [9], Donoho introduced a new orthonormal basis whose elements he called 'orthonormal ridgelets.' We will not detail why these elements relate to ridgelets. We quote from [4]: " 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 $L^2(\mathbb{R})$ (Lemarié & Meyer, 1986), and let $(w^0_{i_0,\ell}(\theta),\ \ell = 0, \ldots, 2^{i_0} - 1;\ w^1_{i,\ell}(\theta),\ i \geq i_0,\ \ell = 0, \ldots, 2^i - 1)$ be an orthonormal basis for $L^2[0, 2\pi)$ made of periodized Lemarié scaling functions $w^0_{i_0,\ell}$ at level i_0 and periodized Meyer wavelets $w^1_{i,\ell}$ 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

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

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

Now, one could also localize the orthonormal ridgelets and – following our Definition 1 – could define a tight frame $\rho_{Q,\lambda}$. The same argument as the one developed in this paper would prove that for any f taken from our image model $\mathcal{F}$, the local orthonormal ridgelet sequence $\alpha_{Q,\lambda}$ satisfies the same property as the standard local ridgelet sequence: namely, from the series

$$f = \sum_{Q \in \mathcal{Q}_s} \sum_{\lambda \in \Lambda} \langle f, \rho_{Q,\lambda} \rangle \rho_{Q,\lambda},$$

extract the m-term series f_m corresponding to the m largest coefficients $\alpha_{Q,\lambda} = \langle f, \rho_{Q,\lambda} \rangle$. Then,

$$\|f - f_m\|_2^2 \leq C\, \max(m^{-\alpha}, m^{-3/2}).$$

The proof is very similar to the one we have derived in this paper and is not included here.

5.2 Refinements

The condition of our main approximation theorem says that the edge must be three times differentiable with a bounded third derivative. However, this is not a necessary condition. The author is confident that further detailed studies will allow proof of versions of Theorem (1.11) with milder assumptions: it should be sufficient to require the edge to be continuous and piecewise twice differentiable with a uniformly bounded second derivative.

In addition, further extensions of our analysis would include cases where one has X-crossings and T-junctions (i.e. edges are allowed to cross). Preliminary work shows that ridgelets adapt well to X-crossings; the 'no-loop' condition work is not needed. We hope to report on these extensions in a later paper.

5.3 Computational issues

A fast algorithm has been developed to code up the ridgelet transform: the details of the algorithm are not yet published. At the present stage, the algorithm works in the case of the dimension $n = 2$. The algorithm takes data on a cartesian grid and computes "pseudo-ridgelet" coefficients. Here, the discrete transform is not orthonormal but provides a frame of redundancy factor 2 and is numerically tight. Finally this discrete transform has low complexity since it runs in $O(n^2 \log(n))$ flops for an $n \times n$ image. Of course, there is an associated inverse transform that reconstructs an image from the data of its "pseudo-ridgelet" coefficients; its order of complexity is of the same order as the one of the forward transform. The major part of the work described in this paragraph has been done by David Donoho.

It follows that the local ridgelet transform can obviously be computed in $O(n^2 \log(n))$ flops and, hence, the method we have described in this paper appears to be very attractive from a practical point of view.

5.4 Statistical estimation

Consider now the problem of removing noise from image data. Suppose we observe noisy measurements

$$y_{i,j} = \tilde{f}(i,j) + \sigma z_{i,j},$$

where $z_{i,j} \overset{i.i.d.}{\sim} N(0,1)$ is a Gaussian noise term. We wish to recover f with small per-pixel mean-squared error $MSE(\hat{f}, f) = En^{-2} \sum_{i,j} (\hat{f}(i,j) - \tilde{f}(i,j))^2$.

Classical methods such as wavelet and Fourier methods do not efficiently remove noise from images with edges – and thus from our model. For example, for a nice wavelet thresholding estimate $\hat{f}^{wave}$, we have in the range $\alpha \geq 1$

$$\sup_{\mathcal{F}_\Gamma(\alpha,A)} MSE(\hat{f}^{wave}, f) \geq Cn^{-1}, \quad n \to \infty.$$

The same speed limit applies to other methods. However, I suspect that there is a way to invoke the fast algorithm so that when applied to noisy empirical data from our model it would achieve near-minimax behavior for recovering the underlying noiseless object at least in the range $1 \leq \alpha \leq \gamma \leq 3/2$. The procedure I have in mind is adaptive (with respect to the unknown smoothness parameters) and I expect to get estimates as if there were no singularities: that is, estimates having – up to log-like factors – the following asymptotics:

$$\sup_{\mathcal{F}_\Gamma(\alpha,A)} MSE(\hat{f}^{ridge}, f) \sim n^{-2\alpha/(2\alpha+2)}, \quad n \to \infty.$$

These rates are minimax.

Acknowledgements

I am especially grateful to David Donoho for many fruitful discussions. This work was partially supported by NSF DMS-95-05151, AFOSR MURI-F49620-96-1-0028, and a fellowship from the D.R.E.T. (French Authority).

References

[1] E. J. Candès, Harmonic analysis of neural networks, *Appl. Comput. Harmon. Anal.*, **6**, 197-218, 1999.

[2] E. J. Candès, Ridgelets: theory and applications, Ph. D. thesis, Department of Statistics, Stanford University, 1998.

[3] E. J. Candès, On the Representation of Mutilated Sobolev Functions. Stanford Technical Report, Department of Statistics, Stanford University, 1999. Submitted for publication.

[4] E.J. Candès, and D.L. Donoho, Ridgelets: the Key to Higher-dimensional Intermittency?, to appear in *Phil. Trans. R. Soc. Lond. A.* (1999).

[5] A. Cohen, I. Daubechies, and P. Vial, Wavelets on the interval and fast wavelet transforms, *Appl. Comput. Harmon. Anal.* **1**, 54–81, 1993.

[6] I. Daubechies, "Ten Lectures on Wavelets," Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[7] S. R. Deans, "The Radon transform and some of its applications," John Wiley & Sons, 1983.

[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, Orthonormal ridgelets and linear singularities, Report no. 1998-19, Department of Statistics, Stanford University.

[10] M. Frazier, B. Jawerth, and G. Weiss, "Littlewood Theory and the Study of Function Spaces," NSF-CBMS Regional Conf. Ser. in Mathematics, Vol 79, American Math. Soc., Providence, RI, 1991.

[11] P.G. Lemarié, and Y. Meyer, Ondelettes et bases Hilbertiennes. *Rev. Mat. Iberoamericana* **2**, 1-18, 1986.

[12] Y. Meyer, "Wavelets and Operators," Cambridge University Press, 1992.

[13] N. Murata, An integral representation of functions using three-layered networks and their approximation bounds, *Neural Networks* **9** (1996), 947-956.

[14] R.M. Young, "An Introduction to Nonharmonic Fourier Series," Academic Press, New York, 1980.