New Tight Frames of Curvelets and Optimal Representations of Objects with C^2 Singularities

Emmanuel J. Candès
Applied and Computational Mathematics
California Institute of Technology
Pasadena, California 91125

David L. Donoho
Department of Statistics
Stanford University
Stanford, California, 94305

November 2002

Abstract

This paper introduces new tight frames of curvelets to address the problem of finding optimally sparse representations of objects with discontinuities along C^2 edges. Conceptually, the curvelet transform is a multiscale pyramid with many directions and positions at each length scale, and needle-shaped elements at fine scales. These elements have many useful geometric multiscale features that set them apart from classical multiscale representations such as wavelets. For instance, curvelets obey a parabolic scaling relation which says that at scale 2^{-j}, each element has an envelope which is aligned along a 'ridge' of length $2^{-j/2}$ and width 2^{-j}.

We prove that curvelets provide an essentially optimal representation of typical objects f which are C^2 except for discontinuities along C^2 curves. Such representations are nearly as sparse as if f were not singular and turn out to be far more sparse than the wavelet decomposition of the object. For instance, the n-term partial reconstruction f_n^C obtained by selecting the n largest terms in the curvelet series obeys

$$\|f - f_n^C\|_{L_2}^2 \le C \cdot n^{-2} \cdot (\log n)^3, \quad n \to \infty.$$

This rate of convergence holds uniformly over a class of functions which are C^2 except for discontinuities along C^2 curves and is essentially optimal. In comparison, the squared error of n-term wavelet approximations only converges as n^{-1} as $n \to \infty$, which is considerably worst than the optimal behavior.

Keywords. Curvelets, wavelets, second dyadic decomposition, edges, nonlinear approximation, singularities, thresholding, Radon transform.

Acknowledgments. This research was supported by National Science Foundation grants DMS 98–72890 (KDI) and by an Alfred P. Sloan Fellowship. E. C. thanks the Institute for Pure and Applied Mathematics at UCLA and especially Mark Green and Eilish Hathaway for their warm hospitality. E. C. would also like to acknowledge fruitful conversations with Laurent Demanet.

1 Introduction

1.1 The Problem of Edges

Edges are prominent features of the visual world; from some points of view a visual scene contains little of importance *besides* edges. In fact, neuroscientists have identified edge processing neurons in the earliest and most fundamental stages of the processing pipeline upon which mammalian visual processing is built. Edges are also ubiquitous in synthetic and real digital imagery, and a great deal of technological research aims to find and represent edges. Thus, in medical imaging, detecting and enhancing boundaries between different cavities is of prime importance. Finally, edge-like phenomena exist outside vision, for example in certain physical systems, where 'shock fronts' occur naturally. Research in scientific computing concerns the efficient representation and propagation of such fronts.

This article is motivated by fundamental questions concerning the mathematical representation of objects containing edges: what is the sparsest representation of functions $f(x_1, x_2)$ which contain smooth regions, but also edges? We give a quantitative content to this question, by using a simple mathematical model of images, asking a fundamental approximation-theoretic question about this model, and using harmonic analysis techniques to answer the question. We construct and analyze a new tight frame for representing functions $f(x_1, x_2)$, and establish an essential optimality of this system. Underlying these results is a mathematical insight concerning the central role, for the analysis and synthesis of objects with discontinuities along curves (i.e. 'edges'), played by *parabolic scaling*, in which analysis elements are supported in elongated regions obeying the relation $width \approx length^2$.

The theoretical results established here should be of considerable interest for a wide variety of technological fields. We mention two specific areas where the implications seem most immediate:

- *Image Coding.* Today the most advanced image coders are transform coders; these typically apply a linear transform to the image data, yielding coefficients that are then quantized. Our theoretical results suggest in a mathematically precise model of image encoding that popular classical transforms such as cosine transforms and wavelet transforms can be substantially outperformed by the new class of transforms we describe here. Since those classical transforms underly JPEG and JPEG-2000, this fact may be of substantial interest.

- *Image Reconstruction.* Digitally acquired data which are blurred, noisy, and indirectly measured, are of interest in technological and scientific fields ranging from medical imagery to extragalactic astronomy. Our theoretical results, see the companion paper [8], show that, because images have edges, many of the standard approaches to image restoration and enhancement (e.g. method of regularization, wavelet-vaguelette decomposition) are suboptimal, particularly in very noisy situations; whereas, in our mathematical model of image restoration, the new schemes we introduce are essentially optimal.

In essence, in both areas, parabolic scaling ought to be extremely helpful in better resolving the edgelike components of the images; it gives better accuracy in the vicinity of edges while using many fewer terms in an approximation. When compared to non-parabolic scaling methods like Fourier analysis and wavelets, this can lead to better compression in image coding, and better image restoration in the presence of noise.

1.2 Quantifying the Approximation Performance

To quantify the performance of various representations, we will take the viewpoint of nonlinear approximation. We consider as a model problem the case where typical objects are functions of two variables with discontinuities along edges and which are otherwise smooth. To make things concrete, consider representing in a Fourier basis a binary object f, i.e. the indicator function of a set with a C^2 boundary. Then the number of Fourier coefficients of f exceeding $1/n$ in absolute value grows as rapidly as $c \cdot n^2$ as $n \to \infty$. This 'rapid' rate of growth means that many different terms are needed to obtain good partial reconstructions. Let f_n^F be the best partial reconstruction obtained by selecting the n largest terms in the Fourier series; then the squared error of such an n-term expansion would obey

$$\|f - f_n^F\|_{L_2}^2 \asymp n^{-1/2}, \quad n \to \infty. \tag{1.1}$$

The asymptotics are quite different if we now consider representing f in a nice wavelet basis. For n tending to infinity, the number of coefficients above the threshold $1/n$ now only grows as $c \cdot n$ and the best n-term wavelet approximation would obey

$$\|f - f_n^W\|_{L_2}^2 \asymp n^{-1}, \quad n \to \infty, \tag{1.2}$$

which is significantly better. Nevertheless, this is nowhere close to being optimal.

Despite the fact that wavelets have had a wide impact in image processing, they fail to efficiently represent objects with edges for the simple reason that the wavelet transform does not take advantage of the geometry of the underlying edge curve. Suppose that the edge curve is of length one, say. Then at each fixed scale 2^{-j}, there are about 2^j wavelets which interact with the edge yielding coefficients of size about 2^{-j}. Although this is very crude, this analysis is essentially correct and explains why the wavelet coefficients only decay like $1/n$. In other words, using a wavelet basis we need about 2^j coefficients to reconstruct the frequency content of an edge up to the subband $|\xi| \sim 2^j$. In comparison, this paper will exhibit a construction in which one can achieve a similar feat with only $O(2^{j/2})$ coefficients! The limitation here is that wavelets are non geometrical and do not exploit the regularity of the edge curve. To obtain nearly optimal approximation rates, we need new multiscale ideas and basis functions with a very different geometry.

In fact, one can easily imagine more geometrical means of approximations based on adapted triangulations. Consider a dictionary of indicator functions of triangles with *arbitrary shapes* and locations. Then it is quite clear [19] that for each n, there exists a superposition of n-triangles $f_n^T = \sum_{i=1}^n 1_{T_i}$ with the property

$$\|f - f_n^T\|_{L_2}^2 \asymp n^{-2}, \quad n \to \infty. \tag{1.3}$$

These types of approximation are adaptive, and are of course very different from thresholding ideas in a fixed basis. In this direction, it is important to underline the conceptual problems with the foundations of such results. First, the best or near-best approximation is the solution of a complicated and abstract minimization problem. How to construct such approximations is totally unclear if not intractable. Second, these types of results hardly lead to any kind of realistic implementation in any practical setting. When presented with an array of pixel intensities, it is unclear how to extract an adapted triangulation. One would perhaps need to perform some kind of edge detection which for complicated imagery is already quite problematic. Although this is a very important issue, we choose not to dwell further on this topic and simply refer the reader to the companion paper [7] which contains a comprehensive discussion on this theme.

Despite the lack of the constructive character of such results, we nevertheless find them useful because they provide an objective performance benchmark.

1.3 Optimality

The asymptotic convergence rate (1.3) is actually the correct optimal behavior for approximating general smooth objects having discontinuities along C^2 curves. Consider an image model $\mathcal{E}$ containig binary ('black and white') objects supported in the unit square, for which the curvature of the boundary curve separating 'black' from 'white' is bounded by some constant C.

- No orthogonal bases can yield approximation rates which are better than n^{-2}. We shall not give the argument here and simply say that this follows from information and approximation theoretic arguments which may be found in [19] and [27].
- Even if one considers finite linear combinations of arbitrary dictionaries of waveforms (which do not necessarily build up orthobases or near orthogonal systems), there is no depth-search limited dictionary which can achieve a better rate than n^{-2}, see [19]. By depth-search limited, we suggest that we are allowed sequences of dictionaries whose size grow polynomially in the number of terms to be kept in the approximation, see also [7].
- No pre-existing basis comes even close to the optimal convergence rate. In fact, the wavelet convergence rate (1.2) is the best published nonadaptive result.

These itemized facts raise a fundamental question: "is there is a basis which does nearly this well?" That is, is there a basis or tight frame in which simple thresholding achieves the optimal rate of convergence. This article argues that the answer is "yes."

1.4 New Tight Frames of Curvelets

In this paper, we construct new tight frames of curvelets to address the problem of finding optimally sparse representations of objects with discontinuities along C^2 edges. These tight frames are different from that introduced in [7] and are roughly defined as follows. We let μ be the triple (j, ℓ, k); here, $j = 0, 1, 2, \ldots$ is a scale parameter; $\ell = 0, 1, \ldots 2^j$ is an orientation parameter; and $k = (k_1, k_2)$, $k_1, k_2 \in \mathbb{Z}$ is a translation parameter. Introduce

1. the *parabolic scaling* matrix D_j

$$D_j = \begin{pmatrix} 2^{2j} & 0 \\ 0 & 2^j \end{pmatrix}, \tag{1.4}$$

2. the *rotation angle* $\theta_J = 2\pi \cdot 2^{-j} \cdot \ell$, with J indexing the scale/angle pair $J = (j, \ell)$,
3. and the *translation parameter* $k_\delta = (k_1 \cdot \delta_1, k_2 \cdot \delta_2)$ (see Section 2 for the numerical values of the parameters $\delta_1, \delta_2 > 0$,).

With these notations, we define curvelets as functions of $x \in \mathbb{R}^2$ by

$$\gamma_\mu(x) = 2^{3j/2} \gamma \left(D_j R_{\theta_J} x - k_\delta \right). \tag{1.5}$$

Here the waveform γ is smooth and oscillatory in the horizontal direction and bell-shaped (nonoscillatory) along the vertical direction. (We will see that the waveform actually depends on the scale parameter j but only very weakly). Continuing at this informal level of discussion, it will be useful to think of γ as being roughly of the form $\gamma(x_1, x_2) = \psi(x_1)\cdot\varphi(x_2)$ where ψ is a smooth wavelet and φ a smooth scaling function (in fact γ is not such a direct product).

Hence, curvelet frame elements are obtained by anisotropic dilations, rotations and translations of a collection of unit-scale oscillatory blobs. Some properties are immediate

- The parabolic scaling (1.4) yields an *Anisotropy Scaling Law*: the system is well-localized in space and obeys approximately the relationships

$$\text{length} \approx 2^{-j}, \quad \text{width} \approx 2^{-2j}$$

  and, therefore, the width and length of a curvelet obey the anisotropy scaling relation

$$\text{width} \approx \text{length}^2.$$

- *Directional Sensitivity*: the elements are oriented in the co-direction $\theta_J = \pi \cdot \ell \cdot 2^{-j}$. Identifying the curvelet width 2^{-2j} with the scale, there are 2^j directions at scale 2^{-2j}; that is,

$$\#\ \text{orientations} = 1/\sqrt{\text{scale}}.$$

- *Spatial Localization.* For a given scale and orientations, curvelets are obtained by two dimensional translations; those translations form a Cartesian grid with a spacing proportional to the length in the direction θ_J and width in the normal direction.

- *Oscillatory Nature.* Curvelets elements display oscillatory components across the 'ridge'.

As in wavelet theory, we also have coarse scale elements which are of the form

$$\varphi_{k_1,k_2}(x) = \varphi(x - k_\delta), \quad k_1, k_2 \in \mathbb{Z},$$

i.e. coarse scale curvelets are translates of a waveform $\varphi(x_1, x_2)$ that we shall take to be bandlimited and rapidly decaying.

One of the result of this paper is to show that one can select profiles φ and γ such that the system $(\gamma_\mu)_\mu$ obeys the Parseval relation

$$\sum_\mu |\langle f, \gamma_\mu\rangle|^2 = \|f\|^2_{L_2(\mathbb{R}^2)}, \quad \forall f \in L_2(\mathbb{R}^2). \tag{1.6}$$

This equality says that $(\gamma_\mu)_\mu$ is a tight frame and standard arguments give the reconstruction formula

$$f = \sum_\mu \langle f, \gamma_\mu\rangle \gamma_\mu, \tag{1.7}$$

with equality holding in an L_2-sense. The reconstruction formula says that one can analyze and synthesize any square integrable function as a superposition of curvelet elements in a very concrete way.

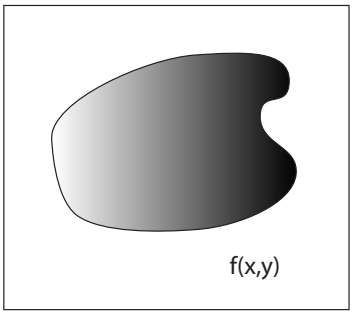

Figure 1: Typical element from our edge model.

1.5 Functions which are C^2 Away from C^2 Edges

We now formally specify a class of objects with discontinuities along edges which is inspired by [18, 21, 19]. It is clear that nothing in the arguments below would depend on the specific assumptions we make here, but the precision allows us to make our arguments uniform over classes of such objects.

We follow [18] and introduce $\textsc{Star}^2(A)$, a class of indicator functions of sets B with C^2 boundaries ∂B. In polar coordinates, we let $\rho(\theta) : [0, 2\pi) \to [0, 1]$ be a radius function and define B by $x \in B$ iff $|x| \le \rho(\theta)$. In particular, the boundary ∂B is given by the curve

$$\beta(\theta) = (\rho(\theta)\cos\theta, \rho(\theta)\sin\theta) \tag{1.8}$$

The class of boundaries of interest to us are defined by

$$\rho \le \rho_0, \quad \|\rho\|_{\dot{C}^2} = \sup|\rho''(\theta)| \le A. \tag{1.9}$$

To fix ideas, take $\rho_0 = 1/10$. We say that a set $B \in \textsc{Star}^2(A)$ if $B \subset [0, 1]^2$ and if B is a translate of a set obeying (1.8) and (1.9).

The geometrical regularity of the members of the class $\textsc{Star}^2(A)$ is useful; it forces very simple interactions of the boundary with dyadic squares at sufficiently fine scales. We use this to guarantee that 'sufficiently fine' has a uniform meaning for every B of interest.

The actual objects of interest to us are functions which are twice continuously differentiable except for discontinuities along edges ∂B of sets in $\textsc{Star}^2(A)$. We define $C_0^2(A)$ to be the collection of twice continuously differentiable functions supported strictly inside $[0, 1]^2$.

Definition 1.1 *Let $\mathcal{E}^2(A)$ denote the collection of functions f on $\mathbb{R}^2$ which are supported in the square $[0, 1]^2$ and obey*

$$f = f_0 + f_1 \cdot 1_B \tag{1.10}$$

where $B \in \textsc{Star}^2(A)$, and each $f_i \in C_0^2(A)$. We speak of $\mathcal{E}^2(A)$ as consisting of **functions which are C^2 away from a C^2 edge.**

Figure 1 gives a graphical indication of a typical element of $\mathcal{E}^2(A)$.

1.6 Sparsity and Nonlinear Approximation

Let f be an object which is C^2 away from a C^2 edge. The main result of this paper is that the curvelet coefficient sequence $(\theta_\mu)_{\mu\in M}$ of f is in some sense, *as sparse as if f were not singular.*

Theorem 1.2 *Let $\mathcal{E}^2(A)$ be the collection (1.10) of objects which are C^2 away from a C^2 curve. Define $|\theta|_{(n)}$ to be the n-th largest entry in the coefficient sequence $(|\theta_\mu|)_{\mu\in M}$ in the curvelet system. Then*

$$\sup_{f\in\mathcal{E}^2(A)} |\theta|_{(n)} \le C\cdot n^{-3/2}\cdot(\log n)^{3/2}. \tag{1.11}$$

There is a natural companion to this theorem. Let f_n^C be the n-term approximation of f obtained by extracting from the curvelet series (1.7) the terms corresponding to the n largest coefficients. The approximation error obeys

$$\|f-f_n^C\|_{L_2}^2 \le \sum_{m>n} |\theta|_{(m)}^2,$$

and, therefore, the rate of decay (1.11) gives the following result.

Theorem 1.3 *Under the assumptions of Theorem 1.2, the n-term approximation f_n^C obtained by simple thresholding in a curvelet frame achieves*

$$\|f-f_n^C\|_{L_2}^2 \le C\cdot n^{-2}\cdot(\log n)^3. \tag{1.12}$$

Simply put, ignoring log-like factors, there is no basis in which coefficients of an object with an arbitrary C^2 singularity would decay faster than in a curvelet frame. Moreover, naive thresholding in a fixed curvelet frame achieves convergence rates rivaling those attainable by adaptive approximation procedures which would attempt to track the discontinuity; a result that the companion paper [7] qualified as quite 'surprising.' We quote from that paper: "In short, in a problem of considerable applied relevance, where one would have thought that adaptive representation was essentially more powerful than fixed nonadaptive representation, it turns out that a new fixed nonadaptive representation is essentially as good as adaptive representation, from the point of view of asymptotic n-term approximation errors."

1.7 Significance

The potential for sparsity is now well-understood for data compression, statistical estimation [15, 22], etc., to the point that the sparsity concept has become a real paradigm in certain research communities.

For instance, consider encoding a function f by the method of *wavelet transform coding.* First, one quantizes its wavelet coefficients $\langle f,\psi_\lambda\rangle$ into integers k_λ using a uniform quantum q:

$$k_\lambda = \mathrm{sgn}(\langle f,\psi_\lambda\rangle)\cdot\lfloor|\langle f,\psi_\lambda\rangle|/q\rfloor.$$

One encodes the position of the nonzero coefficients and the values of the nonzero coefficients as bit strings by standard devices (run-length coding and so forth). Later, an approximate reconstruction of f can be obtained from $\tilde f_q = \sum qk_\lambda\psi_\lambda$. Here we retain the index q to remind us that the quantization stepsize q controls the behavior of the algorithm. This coding method has distortion $\delta(q)=\|f-\tilde f_q\|_{L^2}$; by picking q appropriately, we can arrange

that $\delta(q;f) = \epsilon$ for any desired distortion level $\epsilon > 0$. In return, the number of bits required for a distortion level ϵ is the *description length* $L(\epsilon) = L(\epsilon; f, \text{WAVELETS})$ for wavelet transform coding. Of course for typical functions f, $L(\epsilon) \to \infty$ as $\epsilon \to 0$.

If we encode a function f of the above type—which is smooth away from C^2 edges—by wavelet transform coding, we get that the wavelet description length $L(\epsilon; g, \text{WAVELETS})$ grows as $\epsilon \to 0$ at least as rapidly as $c \cdot \epsilon^{-2}$. Because Fourier series are much denser than wavelet expansions, the Fourier description length is significantly worst; $L(\epsilon; g, \text{FOURIER})$ grows as $\epsilon \to 0$ at least as rapidly as $c \cdot \epsilon^{-4}$. Now a strategy identical to that developed in [17] shows that one can exploit the sparsity (1.11) and develop a curvelet transform coder based on simple ideas such as scalar quantization and run-length coding which ignoring log-like factors, yields a curvelet description length $L(\epsilon; g, \text{CURVELETS})$ growing more slowly than ϵ^{-1}.

In fact, $L(\epsilon; g, \text{CURVELETS})$ is asymptotically nearly optimal in the sense that there is no strategy which can encode elements taken from our edge models with fewer bits than $c \cdot \epsilon^{-1}$ as $\epsilon \to 0$, see [19] for details.

Another implication concerns statistical estimation. Consider the problem of recovering a function $f(x_1, x_2)$ from noisy data. The function f to be recovered is assumed smooth apart from a discontinuity along a C^2 edge. We use the continuum white noise model and observe $y = f + n$ where n is white noise with noise level σ. Then a near corollary of Theorem 1.2 gives that simple strategies based on the shrinkage of curvelet coefficients yielding an estimator $\hat{f}$ achieve—ignoring log-like factors—a Mean Squared Error (MSE) obeying

$$\sup_{f \in \mathcal{F}} E\|\hat{f} - f\|_{L_2}^2 \asymp \sigma^{4/3}, \quad \sigma \to 0.$$

In fact, this is essentially the optimal rate of convergence as the minimax rate scales like $\sigma^{4/3}$. In other words, there are no other estimating procedure which, in an asymptotic sense, give fundamentally better MSEs. In comparison, wavelet shrinkage methods only achieves a MSE which scales like σ as $\sigma \to 0$.

The situation is a little different when one considers adaptive methods which somehow try to estimate the location and size of the discontinuities. Edge detection is a delicate topic and realistic existing methods which are amenable to rigorous optimality results are nearly nonexistent (we are of course aware of [18]). The curvelet shrinkage approach avoids these issues as it does not use edge detectors or any other problematic schemes. The algorithms simply extracts the large curvelet coefficients.

1.8 Relationship with Other Curvelets

A previous article [7] used a radically different machinery to construct tight frames also known under the name of curvelets. For clarity, we shall call the former tight frame curvelets 99. We now briefly review the curvelet 99 transform and explain how it operates on a square integrable object f. This will make explicit the connections and differences with the frames presented in this paper. Before we do so, we need to introduce orthonormal ridgelets. We quote from [6]: "Let $(\psi_{j,k}(t) : j \in \mathbb{Z}, k \in \mathbb{Z})$ be an orthonormal basis of Meyer wavelets for $L^2(\mathbb{R})$ [28], 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

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}}(\hat{\psi_{j,k}}(|\xi|)w^\varepsilon_{i,\ell}(\theta) + \hat{\psi_{j,k}}(-|\xi|)w^\varepsilon_{i,\ell}(\theta+\pi))/2\ . \tag{1.13}$$

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 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 $L^2(\mathbb{R}^2)$."

The curvelet 99 transform makes use of multiscale partitions of unity to localize an object in space. Let Q denote a dyadic square $Q = [k_1/2^s,(k_1+1)/2^s)\times[k_2/2^s,(k_2+1)/2^s)$ and let $\mathcal{Q}$ be the collection of all such dyadic squares. The notation $\mathcal{Q}_s$ will correspond to all dyadic squares of scale s. Let w_Q be a window centered near Q, obtained after dilation and translation of a single w, such that the w_Q^2's, $Q \in \mathcal{Q}_s$, make up a partition of unity. We define multiscale ridgelets by $\{\rho_{Q,\lambda} : s \geq s_0, Q \in \mathcal{Q}_s, \lambda \in \Lambda\}$

$$\rho_{Q,\lambda} = w_Q\, T_Q \rho_\lambda,$$

where

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

The discrete curvelet 99 transform also employs a bank of filters $(P_0 f, \Delta_1 f, \Delta_2 f, \dots)$ with the property that the passband filter Δ_s is concentrated near the frequencies $[2^s; 2^{2s+2}]$ e.g. $\Delta_s(f) = \Psi_{2s} * f$, $\hat{\Psi}_{2s}(\xi) = \hat{\Psi}(2^{-2s}\xi)$. Note that the coronization is nonstandard.

With these preliminaries, the curvelet 99 transform operates as follows.

- *Subband Decomposition.* The object f is filtered into subbands:

$$f \mapsto (P_0 f, \Delta_1 f, \Delta_2 f, \dots).$$

- *Smooth Partitioning.* Each subband is smoothly windowed into "squares" of an appropriate scale:

$$\Delta_s f \mapsto (w_Q \Delta_s f)_{Q\in\mathcal{Q}_s}.$$

- *Renormalization.* Each resulting square is renormalized to unit scale

$$g_Q = (T_Q)^{-1}(w_Q \Delta_s f), \qquad Q \in \mathcal{Q}_s.$$

- *Ridgelet Analysis.* Each square is analyzed in the orthonormal ridgelet system.

$$\alpha_\mu = \langle g_Q, \rho_\lambda \rangle, \qquad \mu = (Q,\lambda).$$

With our notations, we have available a formula for curvelet 99 frame elements, i.e. $\alpha_\mu = \langle f, \gamma_\mu \rangle$ with

$$\gamma_\mu = \Delta_s \rho_{Q,\lambda}, \qquad \mu = (\lambda \in \Lambda, Q \in \mathcal{Q}_s).$$

By linking the filter passband $|\xi| \approx 2^{2s}$ to the scale of spatial localization 2^{-s}, we impose that (1) most curvelets 99 are negligible in norm (most multiscale ridgelets do not survive the bandpass filtering Δ_s); (2) the nonnegligible curvelets 99 obey length $\approx 2^{-s}$ while width $\approx 2^{-2s}$. In short, the system obeys approximately the scaling relationship

$$\text{width} \approx \text{length}^2.$$

Note: it is at this last step that the 2^{2s} coronization scheme comes fully into play. Despite exhibiting novel and interesting properties, the original curvelet construction presents some disadvantaging features we now describe.

First, the construction involves a seven-index structure $\mu = (s, k_1, k_2; j, k; i, \ell, \epsilon)$ whose indices include parameters for *scale* s, *location* $K = (k_1, k_2)$, *ridge scale* j, *ridge location* k, *angular scale* $i \geq max(j, i_0)$, *angular location* ℓ, and a *gender token* ϵ. In addition, we already mentioned that the scaling ratio width $\approx$ length^2 is actually a distortion of the reality. In truth, curvelets 99 assume a wide range of aspect ratios—only their energy decays as the scaling ratio is increasingly less parabolic. The geometry and aspect ratio of orthonormal ridgelets is itself unclear as they are not true ridge functions. This and other facts together with the complicated index structure makes any kind of mathematical and quantitative analysis especially delicate, see for instance the structure of the proof in [8]. For example, when proving results about the sparsity sequence one has to worry about a myriad of coefficients which may sometimes be quite daunting.

In contrast, the new definition exhibits a much simpler structure as it is indexed by only three parameters; namely, scale, orientation (angle) and location—a byproduct being that mathematical analysis is then considerably simpler. Easier manipulation is certainly highly desirable but we would like to emphasize that the alteration is uncompromising; every published 'curvelet result' would hold true with our new system and we are simply not aware of any significant mathematical result –starting with the main result of this paper– which would hold true for one system and not for the other.

Second, the curvelet 99 transform is in some sense a lapped transform as it involves spatial localization with multiscale windows. In practical settings, to overcome blocking effects, one would need to use overlapping windows, thereby, increasing the redundancy of a digital implementation. The new curvelet transform, however, does not exhibit this phenomenon and suggests a new digital implementation which shall be discussed briefly in Section 9.

In short, we believe that our new tight frames yields a system which improves upon the original construction while obeying its philosophy.

1.9 Inspiration and Relation to Other Work

Underlying our work is the inspiration of the original curvelet transform as already discussed. Of interest here, however, is the connection between applied harmonic analysis and a central problem in approximation theory which is new. Indeed, this paper gives the first proof of the optimality result for otherwise smooth objects with edges although results like Theorem 1.2 have been claimed without proofs elsewhere [7].

The ideas underlying the curvelet transform are also loosely related with the theory of affine wavelets. Several researchers [1, 26] have proposed to study the decompositions of objects as superposition of affine wavelets of the form $\psi(Ax + b)$, where $A \in GL(\mathbb{R}^2)$ and $b \in \mathbb{R}^2$. This literature is mainly about continuous transformation and is connected to the theory of square-integrable group representations [11]. Here, we suggest studying representations where A is of the form $D_j R_{\theta_\ell}$, with D_j a parabolic scaling and rotation matrices with an angular step proportional to the square-root of the scale. First, the particular geometry of curvelets does not allow the identification of the parameterization with a linear group representation. And second, guided by the theory of wavelets, we are especially interested in obtaining discrete representations, namely tight frames, with provably optimal approximation properties.

There are deep connections between the curvelet transform and ideas from the field of mathematical analysis. In the seventies, Fefferman [23, 36] studied the boundedness of Riesz spherical means and introduced the so-called Second Dyadic Decomposition (SDD). The SDD is a principle for localizing objects in the frequency plane which goes beyond the classical Littlewood-Paley theory [25]. In fact, curvelets imply a tiling of the frequency plane which is that suggested by SDD, see Section 2 and Figure 3 for details. We would also like to remark that in the early nineties, SDD proved to be a very useful tool for the study of Fourier Integral Operators, see [36] and references therein.

Finally, we recently became aware of the work of Do and Vetterli on *contourlets* which is also directly inspired by the curvelet transform. We will comment on this line research in the discussion section.

2 Second Generation of Curvelets

This section introduces new tight frames we shall call 'curvelets.' Unlike the original curvelet transform [7], this construction does not use ridgelets.

2.1 Scale/Angle Localization

For each pair (j,ℓ), $j \geq 0$ and $\ell = 0, 1, 2, \ldots, 2^j - 1$, we let $\nu_{j,\ell}$ be the angular window $\nu_{j,\ell}(\theta) = \nu(2^j\theta - \pi\ell)$. Note that for $\ell = 0, 1, \ldots, 2^j - 1$, $\nu_{j,\ell}(\theta + \pi) = \nu_{j,\ell+2^j}(\theta)$. Then define the symmetric window $\chi_{j,\ell}(\xi)$ in the polar coordinates system by

$$\chi_{j,\ell}(\xi) = w(2^{-2j}|\xi|)\ \left(\nu_{j,\ell}(\theta) + \nu_{j,\ell}(\theta+\pi)\right). \tag{2.1}$$

Here, we will assume that ν is an even, C^∞ angular window which is supported on $[-\pi, \pi]$ and obeys

$$|\nu^2(\theta)|^2 + |\nu^2(\theta - \pi)|^2 = 1, \quad \theta \in [0, 2\pi), \tag{2.2}$$

where in the above equation, it is understood that we take the 2π-periodization of the function ν, see Figure 2. It is not hard to deduce from our assumptions that for each $j \geq 0$,

$$\sum_{\ell=0}^{2^{j+1}-1} |\nu(2^j\theta - \pi\ell)|^2 = 1, \tag{2.3}$$

where again we have assumed 2π-periodization of the translates $\nu(2^j\theta - \pi\ell)$.

As for the radial window, we will suppose that w is compactly supported and obeys

$$|w_0(t)|^2 + \sum_{j\geq 0} |w(2^{-2j}t)|^2 = 1, \quad t \in \mathbb{R}. \tag{2.4}$$

A possible choice is to select w as in the construction of Meyer wavelets [28, 30]. With v a C^∞ window whose support is included in $[2\pi/3, 8\pi/3]$, Meyer introduces the partition of unity

$$|v_0(t)|^2 + \sum_{j\geq 0} |v(2^{-j}t)|^2 = 1, \quad t \geq 0;$$

here, v_0 is a C^∞ window which is identically equal to one on $[0, 2\pi/3)$ and vanishes on $[4\pi/3, \infty)$. Define w as

$$|w(t)|^2 = |v(t)|^2 + |v(t/2)|^2, \quad w_0(t) = v_0(t). \tag{2.5}$$

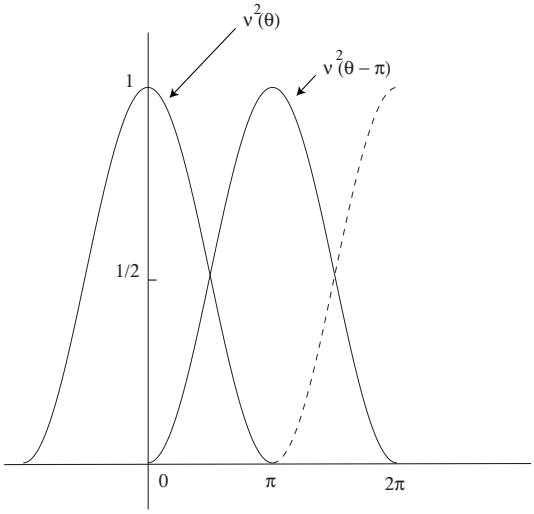

Figure 2: Basic angular window.

Then w obeys (2.4). Note that w is smoothly increasing on the interval $[2\pi/3, 4\pi/3]$, is constant and equal to one on $[4\pi/3, 8\pi/3]$ and smoothly decreasing on $[8\pi/3, 16\pi/3]$. In the remainder of this paper, we will assume this special choice of window.

Put $\chi_0^2(\xi) = w_0^2(|\xi|) + w^2(|\xi|)^2$. For $j \geq 1$, $\nu_{j,\ell}(\theta)$ and $\nu_{j,\ell}(\theta + \pi)$ have non-overlapping supports and, therefore, (2.3) and (2.4) give that the family $(\chi_{j,\ell})$ is a family of orthogonal and compactly supported windows in the sense that

$$|\chi_0(\xi)|^2 + \sum_{j\geq 1} \sum_{\ell=0}^{2^{j-1}} |\chi_{j,\ell}(\xi)|^2 = 1. \tag{2.6}$$

We will use such windows to localize the Fourier transform near symmetric wedges of length about 2^{2j} and width about 2^j. Indeed, $\chi_{j,\ell}$ is localized near the symmetric wedge

$$W_{j,\ell} = \{\pm\xi,\ 2^{2j} \leq |\xi| \leq 2^{2(j+1)},\ |\theta - \pi \cdot \ell \cdot 2^{-j}| \leq \frac{\pi}{2} 2^{-j}\}, \tag{2.7}$$

and note that for each ℓ, $\chi_{j,\ell}$ is obtained from $\chi_{j,0}$ by applying a rotation. Figure 3 gives a graphical representation of theses wedges and associated tiling.

2.2 New Tight Frames of Curvelets

We now introduce some notations that we will use throughout the remainder of this article. We put J to be the pair of indices $J = (j, \ell)$, $j \geq 0, \ell = 0, 1, \ldots, 2^j - 1$ and let $\theta_J = \pi \cdot \ell \cdot 2^{-j}$. Next, we let M_J denote the set of coefficients $\mu = (j, \ell, k)$ with a fixed value of the scale/angle pair $J = (j, \ell)$.

For each $j \geq 1$, the support of $w(2^{-2j}|\xi|)v(2^j\theta)$ is contained in the rectangle $R_j = I_{1j} \times I_{2j}$ where

$$I_{1j} = \{\xi_1,\ t_j \leq \xi_1 \leq t_j + L_j\}, \quad I_{2j} = \{\xi_2,\ |\xi_2| \leq l_j/2\};$$

R_j is symmetric around the axis $\theta = 0$. We will write the length L_j and width l_j as $L_j = \delta_1 \pi 2^{2j}$ and $l_j = \delta_2 2\pi 2^j$. It is not difficult to verify that our assumptions about

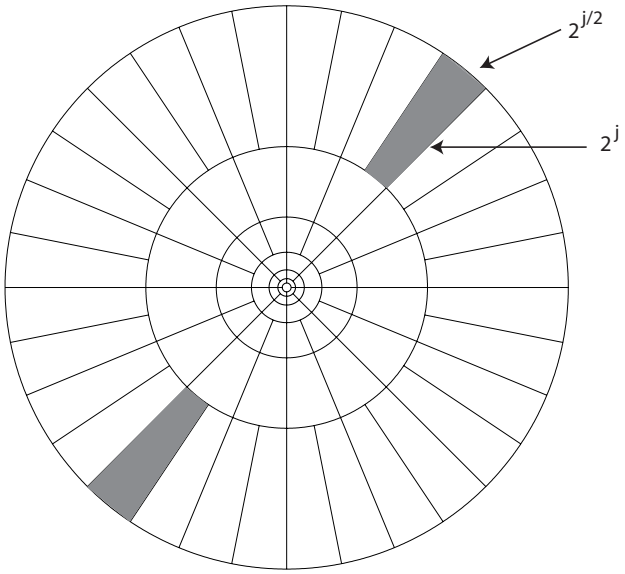

Figure 3: Curvelet Tiling of the Frequency Plane. In the frequency domain, curvelets are supported near symmetric 'parabolic' wedges. The shaded area represents such a generic wedge.

localizing windows imply that δ_1 and δ_2 obey $\delta_1 = 14/3(1 + O(2^{-j}))$ and $\delta_2 = 10\pi/9$ respectively.

We let $\tilde{I}_{1j}$ be $\pm I_{1j}$ and set $\tilde{R}_j = \tilde{I}_{1j} \times I_{2j}$. It is well-known that $e^{\mathbf{i}\pi(k_1+1/2)\xi_1/L_j}/\sqrt{2L_j}$, $k_1 \in \mathbb{Z}$, is an orthobasis for $L_2(\tilde{I}_{1j})$. Since $e^{\mathbf{i}2\pi k_2\xi_2/l_j}/\sqrt{l_j}$ is an orthobasis for $L_2(I_{2j})$, the sequence $(u_{j,k})_{k\in\mathbb{Z}^2}$ defined as

$$u_{j,k}(\xi_1,\xi_2) = \frac{2^{-3j/2}}{2\pi\sqrt{\delta_1\delta_2}} e^{\mathbf{i}(k_1+1/2)2^{-2j}\xi_1/\delta_1}\, e^{\mathbf{i}k_2 2^{-j}\xi_2/\delta_2}, \quad k_1,k_2 \in \mathbb{Z}, \tag{2.8}$$

is then an orthobasis for $L_2(\tilde{R}_j)$.

We are now in position to introduce curvelets using the frequency-domain definition. Letting R_{θ_J} be the rotation by θ_J, we define

$$\hat{\gamma}_{\mu'}(\xi) = (2\pi)\cdot\chi_J(\xi)u_{j,k}(R^*_{\theta_J}\xi), \quad \mu' = (j,\ell,k). \tag{2.9}$$

With the same notation as in Section 1, we also define coarse scale curvelets $\hat{\gamma}_{\mu_0}(x) = (2\pi)\cdot\chi_0(\xi)u_k(\xi)$ where $u_k(\xi) = (2\pi\delta_0)^{-1}\cdot e^{\mathbf{i}(k_1\xi_1/\delta_0 + k_2\xi_2/\delta_0)}$. Here, δ_0 is chosen small enough for $(u_k)_{k\in\mathbb{Z}^2}$ to be an orthobasis for L_2 functions with a compact support containing that of χ_0, e.g. $\delta_0 = 32/3$.

Observe that

$$\sum_{\mu\in M_J} |\langle F,\hat{\gamma}_\mu\rangle|^2 = (2\pi)^2\cdot\int |F(\xi)|^2|\chi_J(\xi)|^2\,d\xi$$

since by construction $(u_{jk}(R^*_{\theta_J}\xi))_k$ is an orthobasis over the support of χ_J. It then follows from (2.6) that for any $F \in L_2(\mathbb{R}^2)$,

$$\sum_{\mu} |\langle F,\hat{\gamma}_\mu\rangle|^2 = (2\pi)^2\cdot\|F\|^2_{L_2}$$

and, therefore, $\hat{\gamma}_\mu$ is a tight frame for $L_2(\mathbb{R}^2)$. In conclusion, the Plancherel formula gives that $(\gamma_\mu)_{\mu\in M}$ obeys

$$\sum_\mu |\langle f, \gamma_\mu\rangle|^2 = \|f\|^2_{L_2(\mathbb{R}^2)}. \tag{2.10}$$

This last equality says $(\gamma_\mu)_{\mu\in M}$ is a tight frame and standard arguments imply that the decomposition (1.7) holds.

We would like to remark that the construction presented here was rapidly introduced by Candès and Guo in [9]. Since the redaction of that paper, Candès became aware of the work of Smith. In [32], Smith introduces a tight frame which is nearly identical to that described above for the purpose of studying parametrices of general hyperbolic equations.

2.3 Space-Side Picture

The point of our construction is that curvelets are real-valued objects. Indeed, let γ_j be the inverse Fourier transform of $\frac{2^{-3j/2}}{\sqrt{\delta_1\delta_2}}\chi_{j,0}(\xi)e^{\mathbf{i}\frac{2^{-2j}\xi_1}{2\delta_1}}$. This function is real-valued and

$$\gamma_{j,0,k}(x) = \gamma_j(x_1 - 2^{-2j}k_1/\delta_1, x_2 - 2^{-j}k_2/\delta_2).$$

Now, the envelope of γ_j is concentrated near a vertical ridge of length about 2^{-j} and width 2^{-2j}. Define $\gamma^{(j)}$ by

$$\gamma_j(x) = 2^{3j/2}\gamma^{(j)}(D_j x)$$

where D_j is the diagonal matrix

$$D_j = \begin{pmatrix} 2^{2j} & 0 \\ 0 & 2^j \end{pmatrix}. \tag{2.11}$$

In other words, the envelope $\gamma^{(j)}$ is supported near a disk of radius about one, and owing to the fact that $\chi_{j,0}$ is supported away from the axis $\xi_1 = 0$, $\gamma^{(j)}$ oscillates along the horizontal direction. In short, $\gamma^{(j)}$ resembles a 2-dimensional wavelet of the form $\psi(x_1)\varphi(x_2)$ where ψ and φ are respectively father and mother-gendered wavelets. Let k_δ be the Cartesian grid $(k_1/\delta_1, k_2/\delta_2)$. With these notations,

$$\gamma_{j,0,k}(x) = 2^{3j/2}\gamma_j(D_j x - k_\delta).$$

and the relationship $\gamma_{j,\ell,k}(\xi) = \gamma_{j,0,k}(R^*_{\theta_J}\xi)$ gives

$$\gamma_\mu(x) = 2^{3j/2}\,\gamma^{(j)}(D_j R^*_{\theta_J} x - k_\delta). \tag{2.12}$$

Hence, we defined a tight frame of elements which are obtained by anisotropic dilations, rotations and translations of a collection of unit-scale oscillatory blobs. Curvelets occur at all dyadic lengths and exhibit an anisotropy increasing with decreasing scale like a power law; curvelets obey a scaling relation which says that the width of a curvelet element is about the square of its length; $width \sim length^2$. Conceptually, we may think of the curvelet transform as a multiscale pyramid with many directions and positions at each length scale, and needle-shaped elements (or 'fat' segments) at fine scales.

2.4 Split At Every at Other Scale

Variations about the definition are of course possible. For instance, note that in the frequency plane, our tight-frames are supported near coronae of the form $\{\pi 2^{2j} \leq |\xi| \leq \pi 2^{2(j+1)}\}$. This coronization is non-standard; these are not dyadic coronae as in wavelet theory. It is, of course possible to adapt the construction and define dyadic 'curvelets' by choosing windows of the form

$$\chi_{j,\ell}(\xi) = w(2^{-j}|\xi|) \left(\nu_{\lfloor j/2 \rfloor,\ell}(\theta) + \nu_{\lfloor j/2 \rfloor,\ell}(\theta+\pi)\right). \tag{2.13}$$

The exact same construction works with this choice of windows and we conclude this section with a brief summary of the main points of the curvelet transform:

- We decompose the frequency domain into dyadic annuli $|x| \in [2^j, 2^{j+1})$.
- We decompose each annulus into wedges $\theta = \pi\ell \cdot 2^{-j/2}$. That is, we *divide at every other scale* as shown on Figure 3.
- We use oriented local Fourier bases on each wedge.

Important remark. In the remainder of this paper, we will assume this special choice so that at scale 2^{-j}, curvelets have length about $2^{-j/2}$ and width 2^{-j} and which in the frequency plane live near the dyadic subband $|\xi| \in [\pi \cdot 2^j, \pi \cdot 2^{j+1}]$. We find this choice more consistent with the standard literature which emphasizes partitions of the frequency plane near dyadic subbands instead of coronae of the form $[\pi \cdot 2^{2j}, \pi \cdot 2^{2j+2}]$.

3 Geometry and Tilings: Ridgelet Packets

The previous section makes clear that there is a general machinery for designing tight frames. Instead of considering smooth segmentations of each subband into a fixed number of wedges, i.e. roughly $2^{j/2}$, we might consider arbitrary dyadic segmentations and thereby design tight frames with arbitrary aspect ratios at arbitrary scales. When the number of segmentations is (1) independent of scale, we essentially obtain tight frames of wavelet-like elements or steerable wavelets [31] (2) increasing like $1/\sqrt{\text{scale}}$, we obtain tight frames of curvelets, and (3) increasing like 1/scale, we obtain tight frames of ridgelets [2, 6].

In [24], the authors followed this organization principle and constructed a family of tight frames they call *ridgelet packets*. One problem with this work is that the family of tight frames or orthobases they exhibit is missing a key ingredient, *a translation parameter*, which may limit their applicability. In response to this, the ideas we exposed in the previous section have of course an explicit translation index.

4 Why Does This Work?

Theorem 1.2 claims that the curvelet coefficients of an object which is singular along a C^2 curve but otherwise smooth decay at nearly the rate $n^{-3/2}$. This section presents a heuristic argument which explains the 3/2 exponent.

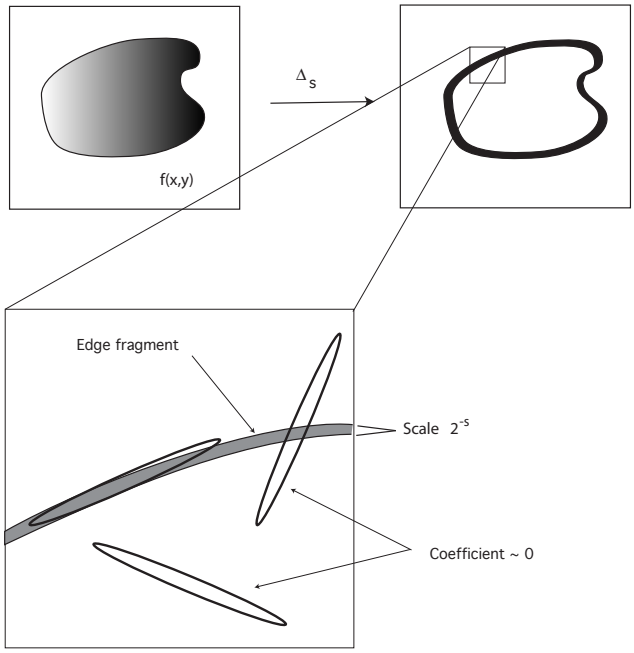

Figure 4: Schematic decomposition of a subband. The top figures represents an object with an edge and that same object after applying a bandpass filter which keeps details at scale 2^{-j}. The bottom picture represents a bandpassed edge fragment together with the three types of curvelets.

4.1 Heuristic Argument

Curvelets are not compactly supported. However, at scale 2^{-j}, they are of rapid decay away from a 'ridge' of length about $2^{-j/2}$ and width 2^{-j} so that we can talk about such a ridge as being their effective support. With this in mind, curvelet coefficients come in essentially three types.

1. Type A. Those curvelets whose essential support does not overlap with the discontinuity.

2. Type B. Those curvelets whose essential support overlap with the discontinuity but are not tangent to the singularity.

3. Type C. Those curvelets which overlap with the singularity and are nearly tangent to the singularity.

These three types are schematically represented on Figure 4. We will argue that coefficients of type A and B are in some sense 'negligible.'

First, coefficients of type A do not 'feel' the singularity and are basically those one would collect if we were to analyze a banal smooth, i.e. C^2, function. The decay exponent of these coefficients is $3/2$ which is that of the coefficients of an arbitrary C^2 function–a fact which will be formally established in Section 8. Note that this decay rate is that one might actually expect as this is also the rate of other classical expansions such as Fourier or wavelet series. Therefore, from the point of view of smooth C^2 functions, curvelets are

as good as Fourier or wavelet bases. To understand this phenomenon, observe that at scale 2^{-j}, curvelet coefficients isolate details of length 2^{-j} –to employ a terminology borrowed from the wavelet literature. Let Δ_j be a bandpass filter which extracts frequencies near the dyadic subband $2^j \le |\xi| \le 2^{j+1}$ and is identically equal to one over the support of $\hat{\gamma}_\mu$. From $\Delta_j(\gamma_\mu) = \gamma_\mu$, it follows that

$$\theta_\mu = \langle \Delta_j f, \gamma_\mu \rangle.$$

Effectively, the bandpass object $\Delta_j f$ is nearly vanishing everywhere except along a ridge of width about 2^{-j} –the width of the filter Δ_j– and whose spatial position of course coincides with that of the underlying edge. Figure 4 schematically represents this bandpassed image. Then coefficients of type A are negligible simply because $\Delta_j f$ nearly vanishes over the essential support of those curvelets.

Second, coefficients of type B are negligible because of the finer frequency localization of curvelet elements. Consider a portion of a bandpassed edge as illustrated in Figure 5 which has been spatially localized with a smooth window of radius about $2^{-j/2}$. In the frequency domain, the bandpassed edge fragment is supported near a wedge whose orientation is orthogonal to that of the edge. This is interesting because we have seen that in the frequency domain, curvelets are supported near dyadic wedges of length 2^j and width $2^{j/2}$, where again the orientations of such wedges are normal to the spatial orientation of our curvelets. Figure 5 represents the essential support of the bandpassed edge fragment and that of a curvelet. Then unless the curvelet orientation is nearly parallel to the edge, these wedges are disjoint and associated coefficients are small.

In short, coefficients of type A are small because the *spatial* supports of the edge and of the curvelet do not overlap whereas coefficients of type B are small because their *frequency* support are disjoint. This 'microlocalization' is what actually explains the sparsity of curvelet expansions of objects with edges.

We now focus our attention on the last group of coefficients, namely, coefficients of type C. The singularity is a C^2 curve of finite length and it is clear that for a fixed scale 2^{-j}, there are at most $O(2^{-j/2})$ coefficients of such type. We now estimate the size of each coefficient of type C. We have

$$|\theta_\mu| = |\langle f, \gamma_\mu \rangle| \le \|f\|_{L_\infty} \cdot \|\gamma_\mu\|_{L_1}.$$

Curvelets are L_2 normalized so that $\|\gamma_\mu\|_{L_2} \le 1$ and essentially supported in a box of side-length $2^{-j/2}$ and width 2^{-j}. Therefore, they obey

$$\|\gamma_\mu\|_{L_1} \le B \cdot 2^{-3j/4},$$

uniformly over the index μ. (Note that this easily and rigorously follows from the definitions (2.9) or (2.12).) Since f is a bounded function, the coefficients θ_μ then verify the a priori estimate

$$|\theta_\mu| \le B \cdot 2^{-3j/4} \cdot \|f\|_{L_\infty}. \tag{4.1}$$

To summarize, at each scale 2^{-j}, we have $O(2^{j/2})$ coefficients of type C which are bounded by $C \cdot 2^{-3j/4}$. Assuming that the other coefficients (of type A and B) are negligible, the nth largest coefficient $|\theta|_{(n)}$ is then bounded by

$$|\theta|_{(n)} \le C \cdot n^{-3/2}.$$

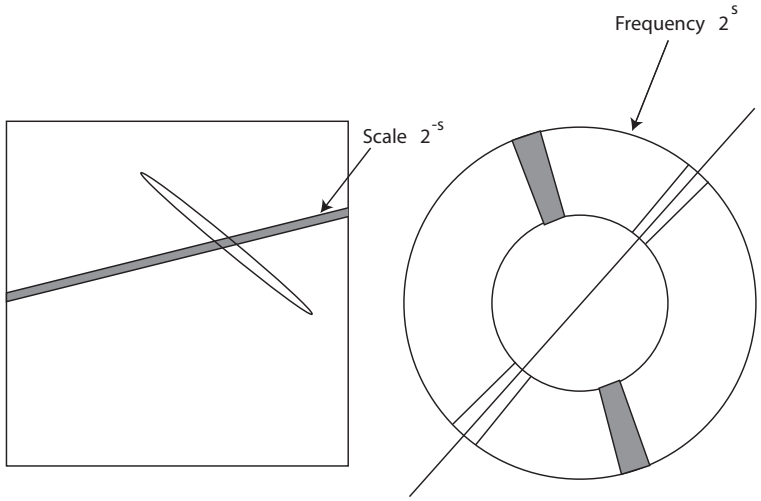

Figure 5: Microlocal behavior. The left figure is a spatial representation of a bandpassed edge fragment and of a curvelet (of type B) while the right figure portrays its frequency representation. The shaded area represents the essential frequency support of the bandpassed edge fragment while the curvelet is supported on the wedge centered around the radial line.

Further, note that the above decay would also give the $O(n^{-2})$ convergence rate for the nonlinear n-term approximation f_n defined by keeping the n largest term in the curvelet expansion as in Theorem 1.3. Indeed, f_n would obey

$$\|f - f_n\|_{L_2}^2 \leq \sum_{m>n} |\theta|_{(m)}^2 \leq C \cdot n^{-2}.$$

4.2 Necessary Refinements

The above arguments only suggest why we may expect 3/2-exponent and a careful proof should take into account several important facts.

- First, curvelets are not compactly supported and, therefore, it is inaccurate to claim that curvelets of type A do not feel the singularity – only their rapid spatial decay will control the 'edge effect' as the distance between the edge curve and the center of the curvelets increases.

- Second, it is inaccurate to claim that the frequency support of an edge does not overlap with that of a curvelet if they are not parallel. A rigorous argument should articulate this fact and quantify the overlap. In some sense, curvelet coefficients decay as the angle between their orientation and that of the edge increases; quantifying this phenomenon with the best possible accuracy is the central part of the proof.

5 Architecture of the Proof

In this section, we prove Theorem 1.2, our main result. However, the proof relies on a key estimate which is the object of a separate section.

With the notations of Section 2, we let M_j be the set of indices (j, ℓ, k), $\ell = 0, 1, \ldots$ and $k \in \mathbb{Z}^2$ so that $(\gamma_\mu)_{\mu \in M_j}$ is the set of all curvelets at scale 2^{-j}. Abusing notation slightly, let θ_j denote the subsequence of coefficients $(\theta_\mu)_{\mu \in M_j}$.

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

$$|\theta|_{w\ell_p} = \sup_{n>0} n^{1/p} |\theta|_{(n)}. \tag{5.1}$$

There are other equivalent definitions; for instance the weak-ℓ_p norm may also be defined as

$$\sup_{\epsilon>0} \#\{n,\, |\theta_n| > \epsilon\} \cdot \epsilon^p.$$

Note that the latter definition shows that the weak-ℓ_p norm obeys $|\theta|_{w\ell_p} \le \|\theta\|_{\ell_p}$. Equipped with this definition, the main result of this section is as follows.

Theorem 5.1 *The sequence θ_j obeys*

$$\|\theta_j\|_{w\ell_{2/3}} \le C, \tag{5.2}$$

for some constant C independent of scale.

Fix the scale parameter j. To analyze the coefficient sequence of an object f at a given scale 2^{-j}, we first smoothly localize this function near dyadic squares with a prescribed radius. We define a partition of unity $(w_Q)_{Q \in \mathcal{Q}}$

$$\sum_{Q \in \mathcal{Q}} w_Q(x) = 1,$$

so that with the index Q indicating a dyadic square of the form $Q = [k_1/2^{j/2}, (k_1+1)/2^{j/2}) \times [k_2/2^{j/2}, (k_2+1)/2^{j/2})$, by $w_Q = w(2^{j/2}x_1 - k_1, 2^{j/2}x_2 - k_2)$. Here w is a nonnegative C^∞ function vanishing outside of the square $[-1, 1]$, say. We use this partition to smoothly localize the function f near each dyadic square Q and define f_Q by

$$f_Q = f \cdot w_Q.$$

Note how the scale of the dyadic squares depends upon the scale 2^{-j}; we use dyadic squares of sidelength about the length of curvelets. For $\mu \in M_j$ and each dyadic square Q, define θ_Q to be the curvelet coefficient sequence of f_Q, i.e.

$$\theta_{Q,\mu} = \langle f_Q, \gamma_\mu \rangle, \quad \mu \in M_j.$$

Note the restriction on μ, namely, $\mu \in M_j$. Our strategy is simply to establish a series of results about the sparsity of the curvelet coefficient sequence θ_Q and combine them to derive our claim (5.2).

5.1 Partition of Dyadic Squares

The sequences θ_Q of course exhibit a very different behavior depending on whether or not the edge curve has a nonempty intersection with the support of w_Q. Accordingly, we partition the collection of dyadic squares $\mathcal{Q}$ into two sets $\mathcal{Q}^0$ and $\mathcal{Q}^1$ and define $\mathcal{Q}^0$ to be the collection of those squares such that the edge curve intersects with the support of w_Q. Clearly, the cardinality of $\mathcal{Q}^0$obeys

$$|\mathcal{Q}^0| \le A_0 \cdot 2^{j/2}, \tag{5.3}$$

for some constant A_0 independent of scale. Note that since we assume f to be compactly supported, there is a maximum of $2^j + 4 \cdot 2^{j/2}$ squares for which f_Q is possibly nonvanishing. We prove two results.

Theorem 5.2 *Let Q be a dyadic square such that $Q \in \mathcal{Q}^0$. The curvelet coefficient sequence θ_Q of f_Q obeys*

$$\|\theta_Q\|_{w\ell_{2/3}} \le C \cdot 2^{-3j/4}, \tag{5.4}$$

for some constant C independent of Q.

Theorem 5.3 *Let Q be a dyadic square such that $Q \in \mathcal{Q}^1$. The curvelet coefficient sequence θ_Q of f_Q obeys*

$$\|\theta_Q\|_{w\ell_{2/3}} \le C \cdot 2^{-3j/2},$$

for some constant C independent of Q. Actually, the stronger inequality $\|\theta_Q\|_{\ell_{2/3}} \le C \cdot 2^{-3j/2}$ also holds.

5.2 Proof of Theorem 5.1

The proof of Theorem 5.1 is a simple consequence of Theorems 5.2 and 5.3. Recall the p-triangle inequality for weak-ℓ_p, $p \le 1$,

$$\|a + b\|^p_{w\ell_p} \le \|a\|^p_{w\ell_p} + \|b\|^p_{w\ell_p}.$$

Since $\theta_j = \sum_Q \theta_Q$, we have

$$\|\theta_j\|^{2/3}_{w\ell_{2/3}} \le \sum_Q \|\theta_Q\|^{2/3}_{w\ell_{2/3}} \le |\mathcal{Q}^0| \cdot \sup_{\mathcal{Q}_0} \|\theta_Q\|^{2/3}_{w\ell_{2/3}} + |\mathcal{Q}^1| \cdot \sup_{\mathcal{Q}_1} \|\theta_Q\|^{2/3}_{w\ell_{2/3}}$$

The claim follows from Theorems 5.2 and 5.3 together with the earlier observation $|\mathcal{Q}^0| \le A_0 \cdot 2^{j/2}$ and $|\mathcal{Q}^1| \le \cdot 2^j + 4 \cdot 2^{j/2}$.

5.3 Proof of Theorem 1.2

The proof of Theorem 1.2 now easily follows from Theorem 5.1. Indeed, observe that, on the one hand, the latter theorem established that

$$\#\{\mu \in M_j,\, |\theta_\mu| > \epsilon\} \le C \cdot \epsilon^{-2/3},$$

and on the other, a previous section argued that there exists a constant B with the property

$$|\theta_\mu| = |\langle f, \gamma_\mu \rangle| \le B \cdot 2^{-3j/4} \cdot \|f\|_{L_\infty}.$$

As a consequence, there is a scale j_ϵ such that for each $j \geq j_\epsilon$, $|\theta_\mu| < \epsilon$. Formally,

$$B \cdot 2^{-3j/4} \cdot \|f\|_{L_\infty} < \epsilon \Rightarrow \#\{\mu \in M_j,\, |\theta_\mu| > \epsilon\} = 0;$$

thus, the number of scales j such that $\#\{\mu \in M_j,\, |\theta_\mu| > \epsilon\}$ is possibly nonzero is bounded by

$$3/4 \cdot \left(\log_2(\epsilon^{-1}) + \log_2(\|f\|_\infty) + \log_2(B)\right) \leq \log_2(\epsilon^{-1}),$$

for ϵ sufficiently small. We then showed that

$$\#\{\mu \in M,\, |\theta_\mu| > \epsilon\} \leq \sum_j \#\{\mu \in M_j,\, |a_\mu| > \epsilon\} \leq C \cdot \epsilon^{-2/3} \cdot \log(\epsilon^{-1}),$$

which is what we sought. Theorem 1.2 is proved. ∎

5.4 The Coarse Scales

The careful reader will point out that we have not treated the coarse scale coefficients. At coarse scales, curvelets are of the form $\varphi(x_1 - k_1 \cdot \delta, x_2 - k_2 \cdot \delta)$. Since φ is of rapid decay, i.e. for each $m \geq 0$, φ obeys $\varphi(x) \leq C_m(1 + |x|)^{-m}$, and f is supported on $[0,1]^2$, standard arguments give that for each $m \geq 0$, these coarse scale coefficients obey

$$|\theta_{k_1,k_2}| \leq C_m(1 + |k|)^{-m},$$

for some constant C_m. Hence, their ℓ_p summability is a not an issue.

6 Fourier Analysis of Edge Fragments

6.1 Edge Fragments

Suppose we are given an object with an edge along a C^2 curve. We window the object, multiplying by $w_Q(x) = w(2^{j/2}x - k)$, where w is smooth and compactly supported with support included in $[-1,1]^2$. We then translate the domain so that the resulting object is supported near the origin in a set contained in the square $-2^{-j/2} \leq x_1, x_2 \leq 2^{-j/2}$. We will call the result an *edge fragment*.

We suppose that the scale 2^{-j} is small enough such that over the support of w_Q, the edge curve may be parameterized as a graph either of the form $(x_1, x_2 = E(x_1))$ or $(x_1 = E(x_2), x_2)$. Indeed, the edge has a very simple interaction with dyadic squares at sufficiently fine scales; for $j \geq j_0$, the sidelength $2^{-j/2}$ of a square is too short to prevent one of the aforementioned parameterizations. (Here, j_0 may be a function of the maximum curvature of our edge curve.) Assume without loss of generality that the latter parameterization holds; then an edge fragment is a function of the form

$$f(x_1, x_2) = w(2^{j/2}x_1, 2^{j/2}x_2)g(x_1, x_2)1_{\{x_1 \geq E(x_2)\}}. \tag{6.1}$$

To make things concrete, we suppose that the edge goes through the origin and that at this point, its tangent is pointing in the vertical direction, i.e.

$$E(0) = 0, \qquad E'(0) = 0. \tag{6.2}$$

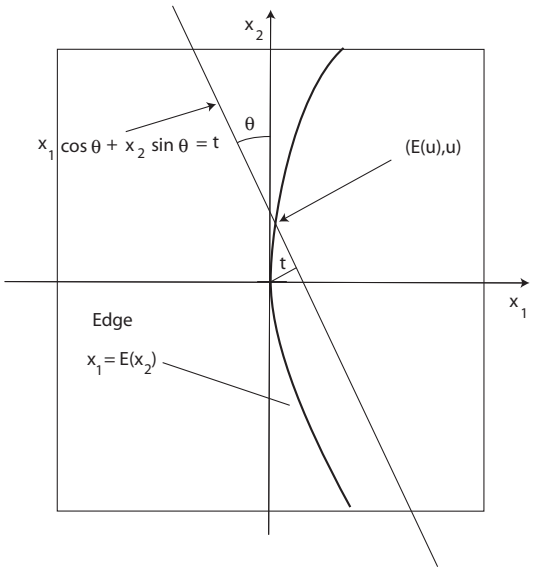

Figure 6: Schematic representation of an edge fragment and associated notation.

We would like to emphasize that this is not a loss of generality and that nothing in the below arguments depends on this specific assumption; see the discussion at the end of this section. It follows that E deviates little from zero:

$$\sup_{|x_2|\le 2^{-j/2}} |E(x_2)| \le \frac{1}{2} \cdot \sup_{|x_2|\le 2^{-j/2}} |E''(x_2)| \cdot 2^{-j}.$$

In this sense, the edge curve $(E(x_2), x_2)$, $|x_2| \le 2^{-j/2}$, is very nearly straight. Figure 6 gives a sketch of an edge fragment.

6.2 Fourier Analysis

We wish to study the localization of the Fourier transform of an edge fragment. Because the singularity is nearly vertical, it is quite clear that the Fourier transform of the edge fragment will have slow decay along the horizontal axis $\xi = (\xi_1, 0)$. In this section, however, we wish to understand the decay of the Fourier transform along radial lines of the form $(\lambda\cos\theta, \lambda\sin\theta)$, $\lambda \in \mathbb{R}$, as θ moves away from the singular co-direction $\theta = 0, \pi$. Our goal is to quantify this decay at a finite distance of the origin, namely, for $|\lambda| \sim 2^j$.

Theorem 6.1 *Let I_j be a dyadic interval $[\pi \cdot 2^{j-\alpha}, \pi \cdot 2^{j+\beta}]$ with $\alpha, \beta \in \{0, 1, 2, 3\}$. The Fourier transform of the edge fragment obeys*

$$\int_{|\lambda|\in I_j} |\hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \cdot 2^{-2j} \cdot (1 + 2^{j/2}|\sin\theta|)^{-5}. \tag{6.3}$$

We briefly discuss the relevance of this theorem for our problem. Curvelets are compactly supported near parabolic wedges in the frequency plane and Theorem 6.1 quantifies the frequency localization of an edge fragment as it gives bounds on the 'energy' an edge fragment puts on each such parabolic wedge. With f an edge fragment and χ_J a frequency window as in Section 2, (6.3) gives

$$\int |\hat{f}(\xi)|^2 |\chi_J(\xi)|^2 d\xi \le C \cdot 2^{-3j/2} \cdot (1 + 2^{j/2}|\sin\theta_J|)^{-5}. \tag{6.4}$$

Such bounds are sharp and cannot be improved. In a nushell, (6.4) controls the size of the coefficients for a fixed value of the scale/angle pair J.

Let F be the renormalization of an edge fragment to the unit square

$$F(x) = f(2^{j/2}x), \quad x \in [-1,1]^2. \tag{6.5}$$

Note that F is of the form

$$F(x) = w(x)g(2^{-j/2}x)1_{\{x_1 \geq E_j(x_2)\}}, \quad E_j(x_2) = 2^{j/2}E(2^{-j/2}x_2). \tag{6.6}$$

The edge curve E_j is again nearly straight. The Fourier transform of F is of course given by $\hat{F}(\xi) = 2^{-j}\hat{f}(2^{-j/2}\xi)$, and, therefore, (6.3) is equivalent to

$$\int_{|\lambda| \in 2^{-j/2} \cdot I_j} |\hat{F}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \leq C \cdot 2^{-j/2}(1 + 2^{j/2}|\sin\theta|)^{-5}. \tag{6.7}$$

From now on, we will use the 'nicer' notation $|\lambda| \sim 2^{j/2}$ to indicate $|\lambda| \in 2^{-j/2} \cdot I_j = [\pi \cdot 2^{j/2-\alpha}, \pi \cdot 2^{j/2+\beta}]$.

The Radon transform provides a convenient tool to study integrals of the type (6.3) and (6.7). We recall that the Radon transform of an object f is the collection of line integrals indexed by $(\theta, t) \in [0, 2\pi) \times \mathbb{R}$ given by

$$Rf(\theta, t) = \int f(x_1, x_2)\delta(x_1\cos\theta + x_2\sin\theta - t)\, dx_1 dx_2, \tag{6.8}$$

where δ is the Dirac distribution. The Radon transform is linked to the polar Fourier transform of an object f because the Projection Slice Theorem states that the Fourier transform along radial lines may be obtained by applying the 1-dimensional Fourier transform to the slices of the Radon transform

$$\hat{f}(\lambda\cos\theta, \lambda\sin\theta) = \int Rf(\theta, t)e^{-i\lambda t}dt.$$

6.3 The Radon Transform of an Edge Fragment

In this section, we will view $RF(t, \theta)$ as a function of the variable t while θ will merely play the role of a parameter. From now on, the subscript t shall denote partial derivatives with respect to the t variable.

Lemma 6.2 *Set $\delta = 2^{-j/2} \cdot \|E''\|_{L_\infty}$ and assume $|\sin\theta| \geq \max(2^{-j/2}, 2\delta)$. The Radon transform $RF(\cdot, \theta)$ is twice differentiable and admits the following decomposition*

$$(RF)_{tt}(t, \theta) = F^0(t, \theta) + F^1(t, \theta);$$

F^0 obeys

$$\|F^0\|_{L_2}^2 \leq C \cdot 2^{-j} \cdot |\sin\theta|^{-5}, \tag{6.9}$$

and F^1 is differentiable and obeys

$$\|(F^1)_t\|_{L_2}^2 \leq C \cdot |\sin\theta|^{-5}. \tag{6.10}$$

The Radon transform RF is the line integral along $\mathcal{L}_{t\theta} = \{(x_1, x_2),\ x_1 \cos\theta + x_2 \sin\theta - t = 0\}$ which may or may not intersect with the edge $\mathcal{E} = \{(E_j(u), u),\ |u| \le 1\}$. Note that an intersection point is the solution of

$$E_j(u)\cos\theta + u\sin\theta = t. \tag{6.11}$$

Recall our assumption $|\sin\theta| \ge 2\delta$. The function $u \mapsto E_j(u)\cos\theta + u\sin\theta$ is then strictly monotone and we let $a(\theta) = E_j(-1)\cos\theta - \sin\theta$, and $b(\theta) = E_j(1)\cos\theta + \sin\theta$. We denote by $I(\theta)$ the interval with endpoints $a(\theta), b(\theta)$; that is $I(\theta)$ is the range of $E_j(u)\cos\theta + u\sin\theta$ as u varies in the interval $[-1, 1]$. Observe that the size of this interval obeys

$$|I(\theta)| \le 2|\sin\theta| + 2\delta \le 3|\sin\theta|. \tag{6.12}$$

Lemma 6.3 *Let $E_j(x) \in C^2[-1,1]$ with $E_j(0) = 0, E_j'(0) = 0$, $\|E_j''\|_{L^\infty[-1,1]} \le \delta$. For $|\sin\theta| > 2\delta$,*

(1) Each line $\mathcal{L}_{t\theta}$ intersects $\mathcal{E}$ in at most one point.

(2) The intersection is empty if $t \notin I(\theta)$.

(3) Each $x_2 \in [-1,1]$ generates a point $(E_j(x_2), x_2)$ which is the intersection $\mathcal{L}_{t\theta} \cap \mathcal{E}$ for exactly one value of $t \in I(\theta)$.

We let $u(t,\theta)$ denote the value of x_2 named in part (3) of this lemma. Viewing this as a function of (t,θ), we observe the following behavior.

Lemma 6.4 *Let $|\sin\theta| \ge 2\delta$. For each $t \in I(\theta)$, the function $u(t,\theta)$ is defined by (6.11). It is C^2, with partial derivatives*

$$\begin{aligned} u_t &= (\sin\theta + E_j'(u)\cos\theta)^{-1}, \\ u_{tt} &= E_j''(u)\cos\theta/(\sin\theta + E_j'(u)\cos\theta)^3. \end{aligned}$$

Note that the partial derivatives then obey the following estimates:

$$|u_t| \le 2\cdot|\sin\theta|^{-1}, \tag{6.13}$$

$$|u_{tt}| \le 8\delta\cdot|\sin\theta|^{-3}. \tag{6.14}$$

These lemmas are elementary and we omit the proofs. Figure 6 gives a graphical indication of some of the objects just described.

We now let F^θ be the function obtained by composing F with the rotation by an angle θ, namely,

$$F^\theta(x) = F(x_1\cos\theta - x_2\sin\theta, x_1\sin\theta + x_2\cos\theta).$$

With these notations the Radon transform of F is given by

$$(RF)(t,\theta) = \int_{-\infty}^{\infty} F^\theta(t,u)\,du.$$

Define $(t, a(t,\theta))$ to be the coordinates of the point $(E_j(u), u)$ in the orthogonal coordinates system rotated by an angle θ; $a = -E_j(u)\sin\theta + u\cos\theta$. Set $G(x) = g(2^{-j/2}x)\cdot w(x)$ so that $F(x) = G(x)\cdot 1_{\{x_1 \ge E_j(x_2)\}}$. For $t \in I(\theta)$, the Radon transform of F is then given by

$$(RF)(t,\theta) = \int_{-\infty}^{a} G^\theta(t,u)\,du. \tag{6.15}$$

while for $t \notin I(\theta)$, the same expression holds but with an integral of the form $\int_{-\infty}^{\infty}$.

The Radon transform RF is twice differentiable with respect to the t-variable and for $t \in I(\theta)$, we calculate

$$(RF)_t(t,\theta) = a_t \cdot G^\theta(t,a) + \int_{-\infty}^{a} G_1^\theta(t,u)\,du,$$

where the subscript 1 (resp. 2) indicates differentiation with respect to the first (resp. second) variable. Further,

$$\begin{aligned}(RF)_{tt} &= a_{tt}G^\theta(t,a) + 2a_tG_1^\theta(t,a) + (a_t)^2G_2^\theta(t,a) + \int_{-\infty}^{a} G_{11}^\theta(t,u)\,du \\ &= T_1 + 2T_2 + T_3 + T_4\end{aligned}$$

while or $t \notin I(\theta)$, the second derivative is simply given by

$$(RF)_{tt} = \int_{-\infty}^{\infty} G_{11}^\theta(t,u)\,du.$$

Checking the differentiability of RF at the endpoints of $I(\theta) = (a(\theta), b(\theta))$ is a not an issue since $G(t,a)$ is identically zero for t in the neighborhood of both $a(\theta)$ and $b(\theta)$. In other words, both calculations agree for t near $a(\theta)$ or $b(\theta)$.

The proof then consists in expressing each term T_m as a sum of terms $T_{m,n}$ obeying either (6.9) or being further differentiable with respect to t and with a derivative obeying (6.10). Our calculations only use the following two facts: first, a is supported on the interval $I(\theta)$ which is of length at most $3|\sin\theta|$ and obeys $|a_t| \le C \cdot |\sin\theta|^{-1}$ and $|a_{tt}| \le C \cdot 2^{-j/2} \cdot |\sin\theta|^{-3}$; second, letting D be either $\partial/\partial x_1$ or $\partial/\partial x_2$, $|D^m(g(2^{-j/2}x))| \le C_m \cdot 2^{-jm/2}$. In the remainder of the proof, we will abuse notations and let g actually denote the rescaled object $g(2^{-j/2}x)$.

Consider T_1. This function obeys $\|T_1\|_{L_\infty} \le C \cdot 2^{-j/2} \cdot |\sin\theta|^{-3}$ and is supported in $I(\theta)$. Therefore,

$$\|T_1\|_{L_2(\mathbb{R})}^2 \le C \cdot 2^{-j} \cdot |\sin\theta|^{-5}.$$

Consider T_2. Express T_2 as

$$T_2 = a_t(g_1 w + w_1 g) = T_{2,1} + T_{2,2};$$

$T_{2,1}$ is supported on $I(\theta)$ and obeys $\|T_{2,1}\|_{L_\infty} \le C \cdot 2^{-j/2} \cdot |\sin\theta|^{-1}$. Hence, $\|T_{2,1}\|_{L_2(\mathbb{R})}^2 \le C \cdot 2^{-j} \cdot |\sin\theta|^{-1}$ and, therefore, (6.9). $T_{2,2}$ is differentiable and

$$(T_{2,2})_t = a_{tt}w_1 g + a_t(w_{11}g + w_1 g_1) + (a_t)^2(w_{12}g + w_1 g_2).$$

Similar arguments give $\|(T_{2,2})_t\|_{L_2(\mathbb{R})}^2 \le C \cdot 2^{-j} \cdot |\sin\theta|^{-5}$, i.e. (6.10).

Consider T_3. Express T_3 as

$$T_3 = (a_t)^2(g_2 w + w_2 g) = T_{3,1} + T_{3,2}$$

Then $T_{3,1}$ is supported $I(\theta)$ and obeys $\|T_{3,1}\|_{L_\infty} \le C \cdot 2^{-j/2} \cdot |\sin\theta|^{-2}$. Hence, $\|T_{3,1}\|_{L_2(\mathbb{R})}^2 \le C \cdot 2^{-j} \cdot |\sin\theta|^{-1}$ and, therefore, (6.9). $T_{3,2}$ is differentiable and

$$(T_{3,2})_t = 2a_{tt}a_t(w_2 g) + (a_t)^2(w_{12}g + w_1 g_2) + (a_t)^3(w_{22}g + w_2 g_2).$$

Our basic arguments now give

$$\|(T_{3,2})_t\|^2_{L_2(\mathbb{R})} \leq C \cdot (1 + 2^{-j}|\sin\theta|^{-2}) \cdot |\sin\theta|^{-5},$$

and, therefore, (6.10) since we assumed $|\sin\theta| \geq 2^{-j/2}$.

At last, consider T_4.

$$\begin{aligned} T_4 &= \int_{-\infty}^{a} (g_{11}w + 2g_1w_1 + gw_{11})\,du \\ &= T_{4,1} + 2T_{4,2} + T_{4,3}. \end{aligned} \tag{6.16}$$

Then $|T_{4,1}| \leq C \cdot 2^{-j}$ and therefore $\|T_{4,1}\|^2 \leq C \cdot 2^{-2j}$. Likewise, $|T_{4,2}| \leq C \cdot 2^{-j/2}$ and therefore $\|T_{4,2}\|^2 \leq C \cdot 2^{-j}$. Finally $T_{4,3}$ is differentiable and its derivative is given by

$$(T_{4,3})_t = a_t g w_{11} + \int_{-\infty}^{a} gw_{111} + g_1w_{11}\,du = T_{4,3,1} + T_{4,3,2}.$$

The function $T_{4,3,1} = a_t g w_{11}$ is supported on $I(\theta)$ and obeys $\|T_{4,3,1}\|_{L_\infty} \leq C \cdot |\sin\theta|^{-1}$. Hence, $\|T_{4,3,1}\|^2_{L_2} \leq C \cdot |\sin\theta|^{-1}$. As far as the other term is concerned, $\|T_{4,3,2}\|_{L_\infty} \leq C$ and hence $\|T_{4,3,1}\|^2_{L_2} \leq C$.

For the sake of completeness, we just briefly mention how the proof adapts when the integration line has an empty intersection with the edge curve. In this case, the second derivative $(RF)_{tt}$ is simply given by

$$(Rf)_{tt} = \int_{-\infty}^{\infty} g_{11}w + 2g_1w_1 + gw_{11}\,du = T_{4,1} + 2T_{4,2} + T_{4,3}.$$

The estimated we collected for both $T_{4,1}$ and $T_{4,2}$ of course still hold. As far as the third term is concerned, we have

$$(T_{4,3})_t = \int_{-\infty}^{\infty} gw_{111} + g_1w_{11}\,du$$

and it is then clear that $\|(T_{4,3})_t\|_{L_\infty} \leq C$ and consequently $\|(T_{4,3})_t\|^2_{L_2} \leq C$. This finishes the proof of Lemma 6.2. ∎

Suppose the object we wish to analyze is now of the form $\tilde{F}(x) = x_1^{m_1}F(x)$, where F is an edge-fragment as before and m_1 a nonnegative integer. Then $\tilde{F}$ is of course an edge fragment and therefore obeys the decomposition (6.9)-(6.10). However, these estimates can be improved upon (the reason why we need sharper bounds will become apparent below). Indeed, observe that along the edge curve $\mathcal{E}$, $x_1^{m_1}F(x)$ obeys

$$|x_1^{m_1}F(x)| \leq C \cdot 2^{-jm_1/2}, \quad x \in \mathcal{E}.$$

This is because E deviates little from zero and obeys $\sup_{|x_2|\leq 1}|E_j(x_2)| \leq \delta/2$, and thus for $x \in \mathcal{E}$, $|x_1| \leq C \cdot 2^{-j/2}$.

Corollary 6.5 *Let F be an edge fragment as in Lemma 6.2 and consider $\tilde{F}(x) = x_1^{m_1}F(x)$. Then $(R\tilde{F})_{tt}$ admits the following decomposition:*

$$(R\tilde{F})_{tt} = F_0 + F_1 + F_2;$$

F^0 obeys

$$\|F^0\|^2_{L_2} \leq C \cdot 2^{-jm_1} \cdot 2^{-j} \cdot |\sin\theta|^{-5} + C \cdot 2^{-2j}; \tag{6.17}$$

F^1 is differentiable and obeys

$$\|(F^1)_t\|_{L_2}^2 \leq C \cdot 2^{-jm_1} \cdot |\sin\theta|^{-5} + C \cdot 2^{-j}; \tag{6.18}$$

F_2 is twice differentiable and obeys

$$\|(F^2)_{tt}\|_{L_2}^2 \leq C. \tag{6.19}$$

The proof proceeds as that of Lemma 6.2 and we write

$$(R\tilde{F})_{tt} = T_1 + T_2 + T_3 + T_4$$

For each $i = 1, 2, 3$, T_i may be expressed as a sum of terms such that each verifies either (6.9) or (6.10) –but for an additional multiplicative factor 2^{-jm_1} since they all involve the value of the edge fragment along the edge curve. This is the content of (6.17)–(6.18).

Now write T_4 as before, i.e. (6.16). For $T_{4,1}$, we use the same estimate as before, namely $\|T_{4,1}\|^2 \leq C \cdot 2^{-2j}$ and, therefore, obeys (6.17). For $T_{4,2}$, we observe that this term is differentiable and the derivative is given by

$$(T_{4,2})_t = a_t(g_1 w_1) + \int_{-\infty}^{a} g_1 w_{11} + g_{11} w_1 \, du = T_{4,2,1} + T_{4,2,2}$$

The first term $T_{4,2,1} = a_t g_1 w_1$ obeys $\|T_{4,2,1}\|_{L_\infty} \leq C \cdot 2^{-j/2} \cdot 2^{-jm_1/2} \cdot |\sin\theta|^{-1}$ and as a consequence $\|T_{4,2,1}\|_{L_2}^2 \leq C \cdot 2^{-jm_1} \cdot 2^{-j} \cdot |\sin\theta|^{-1}$, which is acceptable for (6.18). The second term $T_{4,2,2}$ verifies $\|T_{4,2,2}\|_{L_\infty} \leq C \cdot 2^{-j/2}$ and then $\|T_{4,2,2}\|_{L_2}^2 \leq C \cdot 2^{-j}$ which is also acceptable. Finally

$$(T_{4,3})_t = a_t(g w_{11}) + \int_{-\infty}^{a} g_1 w_{11} + g w_{111} \, du = T_{4,3,1} + T_{4,3,2} + T_{4,3,3}.$$

Similar arguments give $\|T_{4,3,1}\|_{L_2}^2 \leq C \cdot 2^{-jm_1} \cdot |\sin\theta|^{-1}$, which is acceptable and $\|T_{4,3,2}\|_{L_2}^2 \leq C \cdot 2^{-j}$ which is also acceptable. Simple calculations show that the derivative of $T_{4,3,3}$ obeys (6.19) which concludes the proof. ∎

6.4 Proof of Theorem 6.1

The proof of Theorem 6.1 is now one step away. To establish (6.7), observe that the Fourier transform of $(RF)_{tt}$ is $-\lambda^2 \hat{F}(\lambda\cos\theta, \lambda\sin\theta)$ which then gives the decomposition

$$-\lambda^2 \hat{F}(\lambda\cos\theta, \lambda\sin\theta) = \hat{F}_0(\lambda\cos\theta, \lambda\sin\theta) + \hat{F}_1(\lambda\cos\theta, \lambda\sin\theta).$$

Now, it follows from Lemma 6.2 that

$$\int |\hat{F}_0(\lambda\cos\theta, \lambda\sin\theta)|^2 d\lambda \leq \|F_0(\cdot,\theta)\|_{L_2(\mathbb{R})}^2 \leq C \cdot 2^{-j} \cdot |\sin\theta|^{-5}.$$

Likewise, since $i\lambda\hat{F}_1(\lambda\cos\theta, \lambda\sin\theta)$ is the Fourier transform of the derivative of $F_1(\cdot,\theta)$, the bound (6.10) gives

$$\int |\lambda|^2 |\hat{F}_1(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \leq C \cdot |\sin\theta|^{-5}$$

and

$$\int_{|\lambda| \sim 2^{j/2}} |\hat{F}_1(\lambda\cos\theta, \lambda\sin\theta)|^2 \leq C \cdot 2^{-j} \cdot |\sin\theta|^{-5}.$$

(We recall that $|\lambda| \sim 2^{j/2}$ means $|\lambda| \in [\pi \cdot 2^{j/2-\alpha}, \pi \cdot 2^{j/2+\beta}]$). Therefore, we proved that

$$\int_{|\lambda|\sim 2^{j/2}} |\hat{F}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \cdot 2^{-3j} \cdot |\sin\theta|^{-5}, \tag{6.20}$$

as claimed.

To conclude the proof of the theorem, we need to address the decay of the Fourier transform in the directions $(\cos\theta, \sin\theta)$, for $|\sin\theta| \le \max(2^{-j/2}, 2\delta)$; that is, for directions which are nearly normal to the singularity.

Write the edge fragment as $F(x) = F_0(x) + \epsilon(x)$ where $F_0(x) = 1_{\{x_1 \ge \delta\}} g(2^{-j/2}x)w(x)$ and $\epsilon(x) = F(x) - F_0(x)$; observe that F_0 is much like F except that we substituted the edge curve with a vertical straight line and that the difference function ϵ is supported in a vertical strip whose width is at most δ. Then write $\hat{F} = \hat{F}_0 + \hat{\epsilon}$. The object $\hat{F}_0$ is the Fourier transform of a smooth function with a straight discontinuity and $\hat{F}_0(\lambda, 0)$ decays like $1/|\lambda|$ as $|\lambda| \to \infty$ and obeys

$$\int_{|\lambda|\sim 2^{j/2}} |\hat{F}_0(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \cdot 2^{-j/2}.$$

We refer the reader to [3] for a proof, although this is an elementary fact. Next, for each θ obeying $|\sin\theta| \le \max(2^{-j/2}, 2\delta)$, note that the Radon transform $R\epsilon(\cdot, \theta)$ is L_∞-bounded and supported in an interval of at most δ and thus obeys $\|R\epsilon(\cdot,\theta)\|^2_{L_2(\mathbb{R})} \le C \cdot \delta \le C \cdot 2^{-j/2}$. Therefore, $\int_{|\lambda|\sim 2^{j/2}} |\hat{\epsilon}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \cdot 2^{-j/2}$ and this last estimate proves that for each θ obeying $|\sin\theta| \le \max(2^{-j/2}, 2\delta)$,

$$\int_{|\lambda|\sim 2^{j/2}} |\hat{F}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C \cdot 2^{-j/2}.$$

This last inequality together with (6.20) finish the proof of the theorem. ∎

To establish the sparsity of curvelet coefficients of an edge fragment, it will prove to be useful to develop bounds on the derivatives of the Fourier transform of an edge fragment.

Corollary 6.6 *Suppose that f is an edge fragment as in (6.1). For each $m = (m_1, m_2)$, $m_1, m_2 = 0, 1, 2, \ldots$, let D^m be the mixed derivative $\partial_1^{m_1}\partial_2^{m_2}$. Then the derivative of the Fourier transform of an edge fragment obeys*

$$\int_{|\lambda|\in I_j} |D^m \hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 d\lambda \le C_m \cdot 2^{-j|m|} \cdot 2^{-jm_1} \cdot I_j(\theta) + C_m \cdot 2^{-j|m|} \cdot 2^{-5j}, \tag{6.21}$$

with $I_j(\theta)$ as in Theorem 6.1, i.e. $I_j(\theta) = 2^{-2j}(1 + 2^{j/2}|\sin\theta|)^{-5}$.

Consider $i^m D^m \hat{f}$. This object is the Fourier transform of $x^m f(x)$ which we may rewrite as

$$x^m f(x) = 2^{-j|m|/2} g(x) w_m(2^{j/2}x) 1_{\{x_1 \ge E_j(x_2)\}}, \quad w_m(x) = x^m w(x).$$

In other words, $x^m f(x) = 2^{-j|m|/2} f_m(x)$ where $f_m(x)$ is an edge fragment. Therefore, $D^m \hat{f}$ obeys

$$\int_{2^j \le \lambda \le 2^{j+1}} |D^m \hat{f}(\lambda\cos\theta, \lambda\sin\theta)|^2 \, d\lambda \le C_m \cdot 2^{-j|m|} \cdot 2^{-2j}(1 + 2^{j/2}|\sin\theta|)^{-5}.$$

This is of course a naive upper-bounds and we already argued that we had available better estimates, i.e. (6.17)–(6.19). Arguments identical to those developed above would turn the size estimates (6.17)–(6.19) into (6.21). The proof is a mere repeat of that of Theorem 6.1 and is omitted.

6.5 Arbitrary Edge Curves

As mentioned earlier, Theorem 6.1 does not depend upon the assumption that the edge obeys (6.2) and yields a more general result. Consider a typical edge curve $\mathcal{E}$ (and associated typical edge fragment) such that the point $x_0 = (x_{0,1}, x_{0,2}) \in \mathcal{E}$ and that at that point the tangent is pointing in the direction $(\sin\theta_0, \cos\theta_0)$. Then the edge fragment (6.1) would obey Theorem 6.1 with of course $|\sin(\theta - \theta_0)|$ in place of $|\sin\theta|$ in the right-hand side of (6.3).

Let R_0 be the rotation by the angle θ_0. Technically speaking, although a typical edge fragment f is not of the form $f(x) = f_0(R_0(x - x_0))$ with f_0 a 'standard' edge fragment because of the windowing (6.1), our size estimates (6.3)–(6.21) behave as if this were the case. Let us explain.

Corollary 6.7 *Consider a typical edge fragment f as descibed above then its Fourier transform may be expressed as*

$$\hat{f}(\xi) = e^{-ix_0\cdot\xi}\, \hat{f}_0(R_0\xi),$$

where $\hat{f}_0$ obeys Theorem 6.1 and Corollary 6.6.

We omit the proof of this intuitive corollary as this is a mere repreat of the arguments presented above.

7 Curvelet Analysis of Edge Fragments

Let M_J denote the set of coefficients $\mu = (j, \ell, k)$ with a fixed value of the scale/angle pair $J = (j, \ell)$. In the previous section, we developed a key inequality (6.4) which gives a very precise bound on the ℓ_2 norm of the curvelet coefficients of an edge fragment f. Indeed,

$$\sum_{\mu\in M_J} |\theta_\mu|^2 = \int |\hat{f}(\xi)|^2 |\chi_J(\xi)|^2 d\xi \le C \cdot 2^{-3j/2} \cdot (1 + 2^{j/2}|\sin\theta_J|)^{-5}. \tag{7.1}$$

For a fixed J, set $\ell_J = 1 + 2^{j/2}|\sin\theta_J|$ and let $N_J(\epsilon)$ be the number of indices $\mu \in M_J$ such that $|\theta_\mu| > \epsilon$. Roughly speaking, at scale 2^{-j} and for a fixed orientation J, there are only about $O(\ell_J)$ curvelets whose support overlaps significantly with the edge curve $\mathcal{E}$. Assuming that the other coefficients are negligible, it would follow from

$$N_J(\epsilon) \cdot \epsilon^2 \le \sum_{\mu\in M_J} |\theta_\mu|^2$$

together with (7.1) that

$$N_J(\epsilon) \le C \cdot \min\left(\ell_J,\, 2^{-3j/2} \cdot \epsilon^{-2} \cdot \ell_J^{-5}\right).$$

Recall that $\theta_J = 2\pi \cdot 2^{-j/2} \cdot \ell$, $\ell = 0, 1, \ldots, 2^{j/2} - 1$ or we may assume a slightly different parameterization and set $\ell = -2^{j/4}, \ldots, 2^{j/4} - 1$. For $\theta \in [-\pi/2, \pi/2]$, $2|\theta|/\pi \le |\sin\theta| \le |\theta|$ and, therefore,

$$N_J(\epsilon) \le C \cdot \min(1 + |\ell|,\, 2^{-3j/2} \cdot \epsilon^{-2} \cdot (1 + |\ell|)^{-5}).$$

Hence

$$\sum_{\ell} N_J(\epsilon) \le C \cdot 2^{-j/2} \cdot \epsilon^{-2/3},$$

which is (5.4). This is of course a rough sketch aimed at quantitatively explaining why the coefficients of an edge fragment obey (5.4). Indeed, we assumed that most coefficients were 'negligible' and a rigorous proof must of course quantify the size of these individuals.

7.1 Proof of Theorem 5.2

Recall the frequency domain definition of a curvelet (2.9): $\hat{\gamma}_\mu(\xi) = (2\pi)\chi_J(\xi)u_{j,k}(R^*_{\theta_J}\xi)$ with $u_{j,k}$ as in (2.8). We recall that $u_{j,k}(R^*_{\theta_J}\xi)$ is an orthogonal basis for $L_2(\Xi_J)$ where Ξ_J is a rectangle containing the support of χ_J. In the Fourier domain, curvelet coefficients are thus given by

$$\theta_\mu = \frac{1}{2\pi}\int \hat{f}(\xi)\chi_J(\xi)\overline{u_{j,k}(R^*_{\theta_J}\xi)}\,d\xi.$$

We now let D_1 be the partial derivative in the direction $(\cos\theta_J, \sin\theta_J)$ and D_2 be the derivative in the orthogonal direction, namely, $(-\sin\theta_J, \cos\theta_J)$. With ℓ_J as before, set $L = (1-(2^j/\ell_J)^2D_1^2)(1-2^jD_2^2)$. Then a simple calculation shows

$$L(u_{j,k}\circ R^*_{\theta_J}) = (1+\ell_J^{-2}(k_1+1/2)^2)^{-1}\cdot(1+k_2^2)^{-1}(u_{j,k}\circ R^*_{\theta_J}),$$

and integrating by parts gives

$$\theta_\mu = (1+\ell_J^{-2}(k_1+1/2)^2)^{-1}\cdot(1+k_2^2)^{-1}\cdot\int (L\hat{f}\chi_J)(\xi)\overline{u_{j,k}(R^*_{\theta_J}\xi)}\,d\xi. \tag{7.2}$$

Let $K = (K_1, K_2) \in \mathbb{Z}^2$ and define R_K to be the set of coefficients (k_1, k_2) such that $\ell_J^{-1}(k_1+1/2) \in [K_1, K_1+1)$ and $k_2 = K_2$. It then follows from the orthogonality property of the system $(u_{j,k}(R^*_{\theta_J}\xi))_k$ that

$$\sum_{k\in R_K}|\theta_\mu|^2 \le C\cdot(1+|K_1|^2)^{-2}\cdot(1+|K_2|^2)^{-2}\cdot\int|(L\hat{f}\chi_J)(\xi)|^2\,d\xi,$$

In the Appendix, we show that $(L\hat{f}\chi_J)$ obeys the same estimate as the edge fragment, namely,

$$\int|(L\hat{f}\chi_J)(\xi)|^2\,d\xi \le C\cdot 2^{-3j/2}(1+2^{j/2}|\sin\theta_J|)^{-5}. \tag{7.3}$$

In short, we have available the following bound

$$\sum_{k\in R_K}|\theta_\mu|^2 \le C\cdot L_K^{-2}\cdot 2^{-3j/2}(1+2^{j/2}|\sin\theta_J|)^{-5}, \quad L_K = (1+|K_1|^2)(1+|K_2|^2). \tag{7.4}$$

The rest of the proof mimics those estimates introduced at the beginning of this section. For a fixed orientation J, let $N_{J,K}(\epsilon)$ be the number of indices $\mu \in M_J$ such that $k \in R_K$ and $|\theta_\mu| > \epsilon$. $N_{J,K}(\epsilon)$ is of course bounded by $|R_K| \le \ell_J$ and obeys

$$N_{J,K}(\epsilon) \le C\cdot\min(\ell_J,\, 2^{-3j/2}\cdot L_K^{-2}\cdot\epsilon^{-2}\cdot\ell_J^{-5})$$

as this follows from (7.4). Put $\epsilon_{j,K} = \epsilon\cdot L_K\cdot 2^{3j/4}$. Then the same calculations as before now give

$$\sum_{|J|=j}N_{J,K}(\epsilon) \le C\cdot(\epsilon_{j,K})^{-2/3} = C\cdot 2^{-j/2}\cdot\epsilon^{-2/3}\cdot L_K^{-2/3}.$$

Since $\sum_{K\in\mathbb{Z}^2}L_K^{-2/3} \le A$, we proved that

$$|\{\mu\in M_j, |\theta_\mu| > \epsilon\}| \le C\cdot 2^{-j/2}\cdot\epsilon^{-2/3}.$$

This finishes the proof of Theorem 5.2. ∎

7.2 Arbitrary Edge Fragments

The previous calculations assumed that the edge fragment obeys (6.2). We hope that it is clear that nothing in the curvelet transform crucially depends upon this specific location and orientation. To see how the argument effortlessly adapts to arbitrary edge fragments, we follow Section 6.5. Recall that the Fourier transform of a typical edge fragment f (the edge curve has location x_0 and orientation θ_0) is of the form $\hat{f}(\xi) = e^{-ix_0\cdot\xi}\,\hat{f}_0(R_0\xi)$ where $\hat{f}_0$ obeys Theorem 6.1 and Corollary 6.6, see Corollary 6.7.

We introduce some notations and let $u_{J,k}$ be the function defined by $u_{j,k}(R^*_{\theta_J}x)$. Then

$$\begin{aligned}\theta_\mu &= \frac{1}{2\pi}\int \hat{f}_0(R_0\xi)e^{-ix_0\cdot\xi}\chi_J(\xi)\overline{u_{J,k}(\xi)}\,d\xi \\ &= \frac{1}{2\pi}\int \hat{f}_0(\xi)\chi_J(R_0^*\xi)\overline{u_{J,k-k_0}(R_0^*\xi)}\,d\xi;\end{aligned}$$

observe the shifted index $k-k_0$ with k_0 defined as $k_0 = R^*_{\theta_J}x_0$. In effect, the angle θ_J of the curvelet has also been shifted, i.e. by θ_0. Since $\hat{f}_0$ obeys Theorem 6.1 and Corollary 6.6, all of our estimates remain the same but for a shift in the parameters of the curvelet transform. For instance, (7.1) becomes

$$\sum_{\mu\in M_J}|\theta_\mu|^2 = \int |\hat{f}(\xi)|^2|\chi_J(R_0^*\xi)|^2d\xi \le C\cdot 2^{-3j/2}\cdot(1+2^{j/2}|\sin(\theta_J-\theta_0)|)^{-5}.$$

Next, letting $\ell_J = 1+2^{j/2}|\sin(\theta_J-\theta_0)|$ and setting $L = (1-(2^j/\ell_J)^2D_1^2)(1-2^jD_2^2)$ with D^1 being now the partial derivative in the direction $(\cos(\theta_J-\theta_0),\sin(\theta_J-\theta_0))$ would give

$$\theta_\mu = (1+\ell_J^{-2}(k_1-k_{0,1}+1/2)^2)^{-1}\cdot(1+(k_2-k_{0,2})^2)^{-1}\cdot\int (L\hat{f}\chi_J(R_0^*\cdot))(\xi)\overline{u_{J,k}(R_0^*\xi)}\,d\xi$$

just as before (7.2) and of course the left-hand side of (7.3) would be replaced by $C\cdot 2^{-3j/2}(1+2^{j/2}|\sin(\theta_J-\theta_0)|)^{-5}$. From here, it is clear how the rest of arguments would then proceed.

7.3 Coarse Scale Edge Fragments

We would like to conclude this section by remarking that strictly speaking, an edge fragment assumes that the scale is fine enough so that $j \ge j_0$ where j_0 is simply related to the maximum curvature of the edge curve. In all rigor, we have therefore not covered the cases where $j < j_0$ although these are trivial since they morally involve only a finite number of coefficients. In the next section, we will prove that for any edge fragment f, θ_Q obeys $\|\theta_Q\|_{\ell_{2/3}} \le 2^{j/2}$ which then establishes Theorem 5.2 for $j < j_0$.

8 Curvelet Analysis of Smooth Functions

8.1 Proof of Theorem 5.3

We begin with a lemma.

Lemma 8.1 *Let Ξ_j be a dyadic corona of the form $\pi\cdot 2^{j-\alpha} \le |\xi| \le \pi\cdot 2^{j+\beta}$ with $\alpha,\beta = 0,1,2,3$. Suppose that $Q\in\mathcal{Q}_1$, then f_Q obeys*

$$\int_{\Xi_j}|\hat{f}_Q(\xi)|^2\,d\xi \le C\cdot 2^{-5j}. \tag{8.1}$$

Proof of Lemma. The function $f_Q = gw_Q$ is supported in a square of sidelength $2 \cdot 2^{-j/2}$, is twice differentiable and its second partial derivative with respect to x_1, say, is given by

$$(f_Q)_{11} = g_{11}w_Q + 2g_1(w_Q)_1 + gw_{11} = T_1 + 2T_2 + T_3.$$

1. The function T_1 obeys $\|T_1\|_{L_\infty} \le C$ and, therefore, $\|T_1\|_{L_2}^2 \le C \cdot 2^{-j}$.
2. The function T_2 is differentiable and a trivial calculation gives $\|(T_2)_1\|_{L_\infty} \le C \cdot 2^j$; therefore, $\|(T_2)_1\|_{L_2}^2 \le C \cdot 2^j$.
3. The function T_3 is twice differentiable and a trivial calculation gives $\|(T_3)_{11}\|_{L_\infty} \le C \cdot 2^{2j}$; therefore, $\|(T_3)_{11}\|_{L_2}^2 \le C \cdot 2^{3j}$.

Let I_j be a dyadic interval of the form $[\pi \cdot 2^{j-\alpha}, \pi \cdot 2^{j+\beta}]$ with $\alpha, \beta = 0, 1, 2, 3$. Using arguments similar to those deployed in the proof of Theorem 6.1, we obtain that for each $n = 1, 2, 3$, $\hat{T}_n$ obeys

$$\int_{|\xi_1| \in I_j} \int_{\xi_2} |\hat{T}_n(\xi_1, \xi_2)|^2 \, d\xi_1 d\xi_2 \le C \cdot 2^{-j}.$$

Since $-\xi_1^2 \hat{f}_Q(\xi) = \hat{T}_1(\xi) + 2\hat{T}_2(\xi) + \hat{T}_3(\xi)$, we proved that

$$\int_{|\xi_1| \in I_j} \int_{\xi_2} |\hat{f}_Q(\xi_1, \xi_2)|^2 \, d\xi_1 d\xi_2 \le C \cdot 2^{-5j}. \tag{8.2}$$

Of course, a similar bound would hold with the dyadic strip $\{\xi, |\xi_2| \in I_j\}$ (instead of $\{\xi, |\xi_1| \in I_j\}$) as a domain of integration. ∎

The overall structure of the proof of Theorem 5.3 is now analogous to that of Theorem 5.2. We first turn the upper-bound (8.1) into a size estimate about the ℓ_2-norm of the coefficients of f_Q at a fixed scale 2^{-j}. Indeed, the sequence θ_Q obeys

$$\sum_{\mu \in M_j} |\theta_{Q,\mu}|^2 \le \int_{\Xi_j} |\hat{f}_Q(\xi)|^2 \, d\xi \le C \cdot 2^{-5j}. \tag{8.3}$$

We then turn this ℓ_2-estimate into an ℓ_p type of estimate. To do this, recall the interpolation inequality

$$\|\theta\|_{\ell_p} \le n^{1/p - 1/2} \|\theta\|_{\ell_2} \tag{8.4}$$

valid for arbitrary finite sequences of length n. Roughly speaking, at scale 2^{-j}, there are only about 2^j curvelets whose support overlap significantly with f_Q. Assuming that the other coefficients are negligible, (8.4) would give

$$\|\theta_Q\|_{\ell_p} \le 2^{j(1/p - 1/2)} \|\theta_Q\|_{\ell_2},$$

and, therefore,for $p = 2/3$, θ_Q would obey

$$\|\theta_Q\|_{\ell_{2/3}} \le 2^{-3j/2}$$

since $\|\theta_Q\|_{\ell_2} \le C \cdot 2^{-5j/2}$. This is the content of Theorem 5.3.

Proof of Theorem 5.3. In the Fourier domain, curvelet coefficients are again given by

$$\theta_\mu = \frac{1}{2\pi} \int \hat{f}(\xi) \chi_J(\xi) \overline{u_{j,k}(R_{\theta_J}^* \xi)} \, d\xi.$$

Set $L = (1 - 2^j\Delta)$ with Δ the usual Laplacian $\Delta = \sum_{i=1}^2 \partial^2/\partial\xi_i^2$. Then

$$L(u_{j,k} \circ R^*_{\theta_J}) = (1 + 2^{-j}(k_1 + 1/2)^2 + k_2^2)(u_{j,k} \circ R^*_{\theta_J}).$$

Hence, an integration by parts gives

$$\theta_\mu = (1 + 2^{-j}(k_1 + 1/2)^2 + k_2^2)^{-2} \cdot \int L^2(\hat{f}\chi_J)(\xi)\overline{u_{j,k}(R^*_{\theta_J}\xi)}\, d\xi.$$

Let $K = (K_1, K_2) \in \mathbb{Z}^2$ and define R_K to be the set of coefficients (k_1, k_2) such that $(k_1 + 1/2)2^{-j} \in [K_1, K_1 + 1)$ and $k_2 = K_2$. It then follows from the orthogonality property of the system $(u_{j,k}(R^*_{\theta_J}\xi))_k$ that

$$\sum_{k \in R_K} |\theta_\mu|^2 \le C \cdot (1 + |K|^2)^{-4} \cdot \int |L^2(\hat{f}\chi_J)(\xi)|^2\, d\xi.$$

We now sum this last inequality over all angular wedges $J = (j, \ell)$ at a fixed scale $|J| = j$ and obtain

$$\sum_{|J|=j} \sum_{k \in R_K} |\theta_\mu|^2 \le C \cdot (1 + |K|^2)^{-4} \cdot \int \sum_{|J|=j} |L^2(\hat{f}\chi_J)(\xi)|^2\, d\xi.$$

The Appendix develops a bound for the right-hand side of this last inequality, namely,

$$\int \sum_{|J|=j} |L^2(\hat{f}\chi_J)(\xi)|^2\, d\xi \le C \cdot 2^{-5j}. \tag{8.5}$$

The remainder of the proof mimics the argument presented above. We first observe that the number of terms in R_K is bounded by $1 + 2^{j/2}$ and thus the cardinality of the set indices $\mu \in M_j$ such that $k \in R_K$ and $\ell = 1, 2, \ldots 2^{j/2}$ is less or equal to $2^j + 2^{j/2}$. The interpolation inequality (8.4) gives

$$\sum_\ell \sum_{k \in R_K} |\theta_{Q,\mu}|^p \le C \cdot 2^{j(1-p/2)} \cdot 2^{-5jp/2} \cdot (1 + |K|^2)^{-2p}.$$

Since for $p > 1/2$, $\sum_{K \in \mathbb{Z}^2}(1 + |K|^2)^{-2p} \le A_p$, we proved that for each $p > 1/2$,

$$\sum_{\mu \in M_j} |\theta_{Q,\mu}|^p \le C \cdot 2^{j(1-3p)}.$$

In particular $\|\theta_Q\|_{\ell_{2/3}} \le C \cdot 2^{-3j/2}$. This finishes the proof of our Theorem. ∎

Remark. Let f_Q be an edge fragment. Then f_Q of course obeys $\|f_Q\|_{L_2} \le C \cdot 2^{-j/2}$ and, therefore, $\|\theta_Q\|_{\ell_2} \le C \cdot 2^{-j/2}$. The above analysis would then give $\|\theta_Q\|_{\ell_p} \le C \cdot 2^{j(1/p-1)}$, and in particular $\|\theta_Q\|_{\ell_{2/3}} \le C \cdot 2^{j/2}$ as claimed in the previous section.

8.2 Sparsity of Smooth Functions

The argument we presented above is general and shows that curvelets are just as effective as any other classical system for representing smooth objects (as claimed in Section 4). To make this statement precise, recall the defintion of the Sobolev norm of an object f:

$$\|f\|^2_{W_2^s} = \int |\hat{f}(\xi)|^2(1 + |\xi|^{2s})\, d\xi,$$

which is equivalent to $\|f\|^2_{L_2} + \|D^s f\|^2_{L_2}$, where D^s is the sth derivative of f proviso that s is an integer.

Theorem 8.2 *Suppose that $g \in W_2^s$, $s > 0$, and supp $g \subset [0,1]^2$. Let θ be the curvelet coefficient sequence of g. Then,*

$$\sum_{\mu} 2^{2js} |\theta_\mu|^2 \le C \cdot \|f\|^2_{W_2^s}, \tag{8.6}$$

and

$$\|\theta\|_{w\ell_p^*} \le C \cdot \|f\|_{W_2^s}, \quad 1/p^* = (s+1)/2. \tag{8.7}$$

We only sketch the proof of this result as this is an easy modification of the ideas we already exposed. With the same notations as before, by definition the Fourier transform obeys

$$\int_{\Xi_j} |\hat{f}(\xi)|^2 \, d\xi \le C \cdot 2^{-2js}.$$

which can be turned into the estimate

$$\sum_{\mu \in M_j} |\theta_\mu|^2 \le C \cdot 2^{-2js}.$$

This gives the first part of the result. Next, at scale 2^{-j}, we only have about 2^{2j} curvelets which interact with the unit square $[0,1]^2$. Ignoring the other coeffcients (they can be handled using rapid spatial decay of γ_μ), we have

$$\#\{\mu \in M_j, \quad |\theta_\mu| > \epsilon\} \le C \cdot \min(2^{2j}, \epsilon^{-2} \cdot 2^{-2js}).$$

Summing this inequality accros $j \ge 0$ gives

$$\#\{\mu \in M_j, \quad |\theta_\mu| > \epsilon\} \le C \cdot \epsilon^{-p^*}.$$

which proves the second part. ∎

Remark. The above result shows that suffices to prove Theorem 1.2 for objects f of the form $f = f_1 \cdot 1_B$. This follows from the fact that the curvelet coefficient sequence (θ_μ) of an arbitary object $f = f_0 + f_1 \cdot 1_B$ may be decomposed as follows

$$\theta_\mu = \langle f, \gamma_\mu \rangle = \langle f_0, \gamma_\mu \rangle + \langle f_2 \cdot 1_B, \gamma_\mu \rangle = \theta^0_\mu + \theta^1_\mu,$$

where θ^0 obeys (8.7), i.e for $s = 2$, $\|\theta^0\|_{w\ell_{2/3}} \le C$.

9 Discussion

This paper introduced new tight frames of curvelets and proved that curvelets provide optimally sparse representation of objects with singularities along C^2 edges. This result motivated the whole construction and in itself explains its appeal. Short of this result, the construction would be an interesting new multiscale architecture, but simply one among many possibilities.

In fact, just as any transform may be applied to a wide spectrum of problems, there is a range of possible applications of curvelet systems which is much wider than the types of approximation theoretic problems studied in this paper. In particular, we would like to point out applications in image processing [9, 34, 33], statistical estimation [8] and possible applications in partial differential equations and scientific computing. For instance,

some recent work [4] shows that some classical types of operators, namely Fourier Integral Operators, admit optimally sparse decompositions in curvelet frames.

Space limitations prevent extensive discussion of these applications. However, it is worth recalling that the sparsity estimates proven here are directly related to performance metrics in these applications. In all these applications, sparser expansions theoretically lead to better reconstructions and faster algorithms and so the estimates proven here form a central motivating factor in efforts towards applications.

10 Appendix

Proof of (7.3). With the notations of Section 7, note that for each pair $m = (m_1, m_2)$, $m_1, m_2 = 0, 1, 2, \ldots$ the mixed derivative of χ_J obeys

$$D_1^{m_1} D_2^{m_2} \chi_J(\xi) = O(2^{-jm_1} \cdot 2^{-jm_2/2}). \tag{10.1}$$

Next, from the definition the the partial derivatives D_1 and D_2, namely,

$$D_1 \hat{f} = \cos\theta_J \partial_1 \hat{f} + \sin\theta_J \partial_2 \hat{f}, \quad D_2 \hat{f} = -\sin\theta_J \partial_1 \hat{f} + \cos\theta_J \partial_2 \hat{f},$$

we deduce a formula for higher order derivatives

$$D_1^m \hat{f} = \sum_{\alpha+\beta=m} c_{\alpha,\beta} (\cos\theta_J)^\alpha (\sin\theta_J)^\beta \partial_1^\alpha \partial_2^\beta \hat{f},$$

and similarly for $D_2^m \hat{f}$. As in Section 2, assume that R_J is a symmetric rectangle containing the support of χ_J. Corollary 6.6 gives bounds on the L_2-norm of partial derivatives of the function $\hat{f}$, namely, letting $\ell_J = (1 + 2^{j/2} |\sin\theta_J|)$

$$\|\partial_1^\alpha \partial_2^\beta \hat{f}\|^2_{L_2(R_J)} \le C_{\alpha,\beta} \cdot 2^{-j(\alpha+\beta)} (2^{-\alpha j} 2^{-3j/2} \ell_J^{-5} + 2^{-5j}).$$

Now, from

$$\|D_1^m \hat{f}\|^2_{L_2(R_J)} \le c_m \cdot \sum_{\alpha+\beta=m} |\sin\theta_J|^{2\beta} \cdot \|\partial_1^\alpha \partial_2^\beta \hat{f}\|^2_{L_2(R_J)},$$

and bounds on each individual term $\|\partial_1^\alpha \partial_2^\beta \hat{f}\|_{L_2(R_J)}$, we then derive the size estimate

$$\|D_1^m \hat{f}\|^2_{L_2(R_J)} \le C_m \cdot 2^{-jm} \left(2^{-5j} + 2^{-jm} \cdot 2^{-3j/2} \cdot \ell_J^{2m-5} \right).$$

The above inequality used the relationship $|\sin\theta| \le 2^{-j/2} \cdot \ell_J$. For $m \le 2$, note that $2^{-5j} \le 2^{-j(m+3/2)} \cdot \ell_J^{2m-5}$ since $\ell_J \ge 1$ and, therefore,

$$\|D_1^m \hat{f}\|^2_{L_2(R_J)} \le C_m \cdot (2^j \ell_J^{-1})^{-2m} \cdot 2^{-3j/2} \cdot \ell_J^{-5}. \tag{10.2}$$

There is an analogous estimate for $D_2^m \hat{f}$ and indeed, similar calculations now give

$$\|D_2^m \hat{f}\|^2_{L_2(R_J)} \le C_m \cdot 2^{-jm} \cdot 2^{-3j/2} \cdot \ell_J^{-5}. \tag{10.3}$$

We now prove that

$$L(\hat{f}\chi_J) = (I - 2^{2j} \ell_J^{-2} D_1^2 - 2^j D_2^2 + 2^{3j} \ell_J^{-2} D_1^2 D_2^2)(\hat{f}\chi_J)$$

obeys (7.3). First, observe that

$$D_1^2(\hat{f}\chi_J) = (D_1^2\hat{f})\chi_J + 2(D_1\hat{f})(D_1\chi_J) + \hat{f}(D_1^2\chi_J).$$

Then (10.2), together with (10.1) gives

$$(2^j\ell_J^{-1})^2 \cdot \|D_1^2(\hat{f}\chi_J)\|_{L_2} \le C \cdot 2^{-3j/4} \cdot \ell_J^{-5/2}.$$

Second, (10.3), together with (10.1) gives

$$2^j \cdot \|D_2^2(\hat{f}\chi_J)\|_{L_2} \le C \cdot 2^{-3j/4} \cdot \ell_J^{-5/2}.$$

And third, it is easy to verify that similar calculations also give

$$(2^{3j}\ell_J^{-2}) \cdot \|D_1^2D_2^2(\hat{f}\chi_J)\|_{L_2} \le C \cdot 2^{-3j/4} \cdot \ell_J^{-5/2}.$$

Our claim (7.3) now follows from these last three inequalities.

Proof of (8.5). Let Δ be the usual Laplacian. We argue that for each each $m = 0, 1, 2, \ldots$, there is a constant C_m with the property

$$\int \sum_{|J|=j} |\Delta^m(\hat{f}\chi_J)(\xi)|^2\, d\xi \le C \cdot 2^{-2jm} \cdot 2^{-5j}. \tag{10.4}$$

Note that for $m = 0$, (10.4) holds because of Lemma 8.1. Further, (10.4) would give (8.5) since $L^2 = I - 2 \cdot 2^j\Delta + 2^{2j} \cdot \Delta^2$.

The proof will use the following basic fact about the window χ_J (see Section 2). For $\alpha = (\alpha_1, \alpha_2)$, we let D^α be the mixed derivative $\partial_1^{\alpha_1}\partial_2^{\alpha_2}$. Then for each pair α, the $D^\alpha\chi_J$'s obey

$$\sum_{|J|=j} |D^\alpha\chi_J(\xi)|^2 \le C_\alpha \cdot 2^{-j|\alpha|}, \tag{10.5}$$

for some constant C_α.

Now begin with

$$\Delta^m(\hat{f}\chi_J) = \sum_{|\alpha|+|\beta|=2m} c_{\alpha,\beta}\, D^\alpha\hat{f}D^\beta\chi_J.$$

We then use (10.5) and obtain

$$\int \sum_{|J|=j} |D^\alpha\hat{f}(\xi)|^2|D^\beta\chi_J(\xi)|^2\, d\xi \le C_\beta 2^{-j|\beta|} \int_{\pi\cdot 2^{j-1}\le|\xi|\le\pi\cdot 2^{j+2}} |D^\alpha\hat{f}(\xi)|^2\, d\xi.$$

Recall that here $f(x)$ is of the form $f(x) = g(x)w(2^{j/2}x)$ where w is a window with support in $[-1,1]^2$. Now $D^\alpha\hat{f}$ is the Fourier transform of $x^\alpha f(x)$. Write $x^\alpha f(x)$ as $2^{-j|\alpha|/2}g(x)w_\alpha(2^{j/2}x)$ where $w_\alpha(x) = x^\alpha w(x)$. The window w_α is of course C^∞, compactly supported in $[-1,1]^2$ and obviously Fourier transform of $g(x)w_\alpha(2^{j/2}x)$ obeys the decay estimate (8.1). Hence,

$$\int_{\pi\cdot 2^{j-1}\le|\xi|\le\pi\cdot 2^{j+2}} |D^\alpha\hat{f}(\xi)|^2\, d\xi \le C_\alpha \cdot 2^{-j|\alpha|} \cdot 2^{-5j}.$$

To conclude, for each α, β with $|\alpha| + |\beta| = 2m$, we proved that

$$\int \sum_{|J|=j} |D^\alpha\hat{f}(\xi)|^2|D^\beta\chi_J(\xi)|^2\, d\xi \le C_m \cdot 2^{-2jm} \cdot 2^{-5j},$$

and, therefore, (10.4) follows. ∎

References

[1] J. P. Antoine and R. Murenzi. Two-dimensional directional wavelets and the scale-angle representation. *Signal Processing*, 52:259–281, 1996.

[2] E. J. Candès. Harmonic analysis of neural netwoks. *Applied and Computational Harmonic Analysis*, 6:197–218, 1999.

[3] E. J. Candes. Ridgelets and the representation of mutilated Sobolev functions. *SIAM J. Math. Anal.*, 33:347–368, 2001.

[4] E. J. Candès and L. Demanet. Curvelets and Fourier integral operators. `http://www.acm.caltech.edu/~emmanuel/publications.html`, 2002.

[5] E. J. Candès and L. Demanet. Curvelets, warpings, and optimally sparse representations of Fourier integral operators. Submitted and available at `http://www.acm.caltech.edu/~emmanuel/publications.html`, 2002.

[6] E. J. Candes and D. L. Donoho. Ridgelets: the Key to Higher-dimensional Intermittency? *Phil. Trans. R. Soc. Lond. A.*, 357:2495–2509, 1999.

[7] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In C. Rabut A. Cohen and L. L. Schumaker, editors, *Curves and Surfaces*, pages 105–120, Vanderbilt University Press, 2000. Nashville, TN.

[8] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. *Ann. Statist.*, 30:784 –842, 2002.

[9] E. J. Candès and F. Guo. New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Processing*, 82:1519–1543, 2002.

[10] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[11] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys.*, 27:1271–1283, 1986.

[12] L. Demanet. Personal communication. 2002.

[13] M. N. Do. *Directional Multiresolution Image Representations*. PhD thesis, Swiss Federal Institute of Technology, Lausanne, November 2001.

[14] M. N. Do and M. Vetterli. Contourlets. In J. Stoeckler and G. V. Welland, editors, *Beyond Wavelets*. Academic Press, 2002. To appear.

[15] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1:100–115, 1993.

[16] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

[17] D. L. Donoho. Unconditional bases and bit-level compression. *Applied and Computational Harmonic Analysis*, 3:388–392, 1996.

[18] D. L. Donoho. Wedgelets: nearly-minimax estimation of edges. *Ann. Statist.*, 27:859–897, 1999.

[19] D. L. Donoho. Sparse components of images and optimal atomic decomposition. *Constr. Approx.*, 17:353–382, 2001.

[20] D. L. Donoho and M. R. Duncan. Digital curvelet transform: Strategy, implementation, experiments. Technical report, Stanford University, 1999.

[21] D. L. Donoho and I. M. Johnstone. *Empirical atomic decomposition*. Manuscript, 1995.

[22] D. L. Donoho, M Vetterli, R. A. DeVore, and I. Daubechies. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory*, 44:2435–2476, 1998.

[23] C. Fefferman. A note on spherical summation multipliers. *Israel J. Math.*, 15:44–52, 1973.

[24] A. G Flesia, H. Hel-Or, A. Averbuch, E. J. Candès, R. R. Coifman, and D. L. Donoho. Digital implementation of ridgelet packets. In J. Stoeckler and G. V. Welland, editors, *Beyond Wavelets*. Academic Press, 2002. To appear.

[25] M. Frazier, B. Jawerth, and G. Weiss. *Littlewood-Paley theory and the study of function spaces*, volume 79 of *NSF-CBMS Regional Conf. Ser. in Mathematics*. American Math. Soc., Providence, RI, 1991.

[26] P. Gressman, D. Labate, G. Weiss, and E. Wilson. Affine, quasi-affine and co-affine wavelets. In J. Stoeckler and G. V. Welland, editors, *Beyond Wavelets*. Academic Press, 2002. To appear.

[27] B. S. Kashin. Approximation properties of complete orthonormal systems. *Trudy Mat. Inst. Steklov.*, 353:187–191, 1985. English translation in Proc. Steklov Inst. Math. (1987).

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

[29] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[30] Y. Meyer. *Wavelets: Algorithms and Applications*. SIAM, Philadelphia, 1993.

[31] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms [or what's wrong with orthonormal wavelets]. *IEEE Trans. Information Theory, Special Issue on Wavelets*, 38:587–607, 1992.

[32] H. A. Smith. A parametrix construction for wave equations with $C^{1,1}$ coefficients. *Ann. Inst. Fourier (Grenoble)*, 48:797–835, 1998.

[33] J. L. Starck, E. J. Candès, and D. L. Donoho. Very high quality image restoration. In M. A. Unser eds. A. Aldroubi, A. F. Laine, editor, *Wavelet Applications in Signal and Image Processing IX*, Proc. SPIE 4478, 2001.

[34] J.L. Starck, E. Candès, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11:670–684, 2002.

[35] J.L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, Cambridge (GB), 1998.

[36] E. M. Stein. *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, volume 30. Princeton University Press, Princeton, N.J., 1993.

[37] D. Taubman. High performance scalable image compression with EBCOT. *IEEE Transactions on Image Processing*, 9:1158–1170, 2000.