Journal of Approximation Theory **113**, 59–90 (2001)
doi:10.1006/jath.2001.3624, available online at http://www.idealibrary.com on IDEAL®

Curvelets and Curvilinear Integrals[1]

Emmanuel J. Candès and David L. Donoho

Department of Statistics, Stanford University, Stanford, California 94305-4065, U.S.A.
E-mail: emmanuel@stat.stanford.edu, donoho@stat.stanford.edu

Communicated by Amos Ron

Received June 14, 2000; accepted in revised form June 21, 2001

Let $\mathscr{C}(t): I \mapsto \mathbf{R}^2$ be a simple closed unit-speed C^2 curve in $\mathbf{R}^2$ with normal $\vec{n}(t)$. The curve $\mathscr{C}$ generates a distribution Γ which acts on vector fields $\vec{v}(x_1, x_2)$: $\mathbf{R}^2 \mapsto \mathbf{R}^2$ by line integration according to

$$\Gamma(\vec{v}) = \int \vec{v}(\mathscr{C}(t)) \cdot \vec{n}(t)\, dt.$$

We consider the problem of efficiently approximating such functionals. Suppose we have a vector basis or frame $\Phi = (\vec{\phi}_\mu)$ with dual $\Phi^* = (\vec{\phi}^*_\mu)$; then an m-term approximation to Γ can be formed by selecting m terms (μ_i: $1 \leqslant i \leqslant m$) and taking

$$\tilde{\Gamma}_m(\vec{v}) = \sum_{i=1}^{m} \Gamma(\vec{\phi}^*_{\mu_i})\, [\vec{v}, \vec{\phi}_{\mu_i}].$$

Here the μ_i can be chosen adaptively based on the curve $\mathscr{C}$.

We are interested in finding a vector basis or frame for which the above scheme yields the highest-quality m-term approximations. Here performance is measured by considering worst-case error on vector fields which are smooth in an L^2 Sobolev sense:

$$Err(\Gamma, \tilde{\Gamma}_m) = \text{supp}\{|\Gamma(\vec{v}) - \tilde{\Gamma}_m(\vec{v})| : \|Div(\vec{v})\|_2 \leqslant 1\}.$$

We establish an isometry between this problem and the problem of approximating objects with edges in L^2 norm. Starting from the recently-introduced tight frames of scalar curvelets, we construct a vector frame of curvelets for this problem. Invoking results on the near-optimality of scalar curvelets in representing objects with edges, we argue that vector curvelets provide near-optimal quality m-term approximations. We show that they dramatically outperform both wavelet and Fourier-based representations in terms of m-term approximation error.

[1] This research was supported by NSF grant DMS 95-05152 and by NSF grant DMS 98-72890.

0021-9045/01 $35.00

The m-term approximations to Γ are built from terms with support approaching more and more closely the curve $\mathscr{C}$ with increasing m; the terms have support obeying the scaling law $width \approx length^2$.

Comparable results can be developed, with additional work, for scalar curvelet approximation in the case of scalar integrands

$$\mathscr{I}(f) = \int f(\mathscr{C}(t))\, dt.$$

Key Words: wavelets; curvelets; ridgelets; m-term approximation.

1. INTRODUCTION

1.1. *Point Evaluation*

Consider the Dirac distribution δ on $\mathbf{R}$, which acts on smooth functions to give point evaluation:

$$\langle \delta, f \rangle = f(0). \tag{1}$$

It is well-known that δ can be represented at least formally as a series in various bases and frames. For example, as δ is supported in $[-\pi, \pi)$ we may use sinusoids to represent it, getting

$$\begin{aligned} \delta &= \frac{1}{\sqrt{2\pi}} \sum_k \langle \delta, e^{ik\theta} \rangle\, e^{ik\theta} \\ &= \frac{1}{\sqrt{2\pi}} \sum_k e^{ik\theta}, \end{aligned} \tag{2}$$

where the equality has an appropriate distributional interpretation. All mathematical scientists are familiar with this representation, but this familiarity may have dulled our sensitivity to a key point. This representation of δ is extremely problematic, stemming from the many non-localized terms of equal intrinsic size. The two sides of Eq. (2) can only balance owing to a truly heroic cancellation of the terms on the right.

In another basis or frame, the representation of the Dirac distribution can be fundamentally better behaved. If we use nice orthonormal wavelets of compact support to represent δ, we have

$$\delta = \sum_I \langle \delta, \psi_I \rangle\, \psi_I \tag{3}$$

$$= \sum_I \psi_I(0)\, \psi_I, \tag{4}$$

where, for comparability with the Fourier example, the $(\psi_I)_I$ are taken as periodic wavelet orthobasis for $L^2[-\pi, \pi)$, and the indices I range, as usual, through the dyadic intervals $[k/2^j, (k+1)/2^j)$. Although the equality must still be interpreted in a distributional sense, it is much less problematic than in the Fourier case. In fact, Daubechies wavelets obey, for appropriate c_1 and c_2,

$$|\psi_I(x)| \leqslant 2^{j/2} 1_{\{c_1 I\}}(x) \cdot c_2,$$

where $c_1 I$ denotes the dilation of I by a factor of c_1 about its midpoint. Hence, the series is actually sparse. For example, there are only a finite number of nonzero terms at any $x \neq 0$. Moreover, the sum can be stratified by scale index j,

$$\sum_I \psi_I(0)\, \psi_I(x) = \sum_j \sum_k \psi_{j,k}(0)\, \psi_{j,k}(x)$$

and there will be at most C_3 nonzero terms at each level j.

The density of the Fourier representation and the sparsity of the wavelet representation are reflected in the effectiveness of the corresponding bases at representing δ by m-term approximations. Let $W_2^1(C)$ denote the ball of functions in $L^2(-\pi, \pi)$ obeying

$$\int_{-\pi}^{\pi} f(t)^2\, dt + \int_{-\pi}^{\pi} f'(t)^2\, dt \leqslant C^2.$$

Think of δ as defined as a distribution acting on such functions f as in (1), and consider m-term approximations in the Fourier basis

$$\tilde{\delta}_m^F = \sum_{\ell=1}^{m} c_\ell e^{ik_\ell \theta}$$

and in the wavelet basis

$$\tilde{\delta}_m^W = \sum_{\ell=1}^{m} a_\ell \psi_{I_\ell}.$$

Evaluate performance of such approximations by the worst-case error which either approach can incur on a function $f \in W_2^1(C)$. Formally,

$$Err_m(\delta, \text{Wavelets}) = \sup_{f \in W_2^1(C)} \inf_{I_1, \ldots, I_m} |\langle \delta, f \rangle - \tilde{\delta}_m^W(f; I_1, \ldots, I_m)|.$$

$$Err_m(\delta, \text{Fourier}) = \sup_{f \in W_2^1(C)} \inf_{k_1, \ldots, k_m} |\langle \delta, f \rangle - \tilde{\delta}_m^F(f; k_1, \ldots, k_m)|.$$

A simple calculation shows

$$Err_m(\delta, \textsc{Fourier}) \asymp m^{-1/2}, \qquad m \to \infty,$$

which exhibits rather slow decay with m, while for a certain $C_4 > 0$,

$$Err_m(\delta, \textsc{Wavelets}) \asymp \exp(-(m/C_4)), \qquad m \to \infty,$$

which exhibits much faster exponential decay.

In short, the wavelet basis gives radically better m-term approximations to δ than does the Fourier basis. The story is the same for every other point evaluation functional δ_x. In a sense, the wavelet basis is ideal for sparse representation of point evaluations.

1.2. *Curvilinear Integrals*

Let now $\mathscr{C}: I \mapsto \mathbf{R}^2$ denote a simple closed curve in the unit square, of finite length, with two continuous derivatives, and unit speed parametrization. Let $\vec{n}(t)$ denote the unit normal vector to $\mathscr{C}(t)$ at time t. Associated with the curve $\mathscr{C}(t)$ is the linear functional Γ acting on smooth vector fields $\vec{v}(x_1, x_2)$ via

$$\Gamma(\vec{v}) = \int \vec{v}(\mathscr{C}(t)) \cdot \vec{n}(t)\, dt.$$

Just as the Dirac distribution was supported on a point, Γ is supported on the curve $\mathscr{C}$.

This functional has a well-known interpretation from vector calculus. If $\vec{v}$ measures a fluid velocity, then Γ measures the net flux across $\mathscr{C}$ per unit time. One could consider alternate curve $\leftrightarrow$ functional correspondences, such as the scalar functional $\int f(\mathscr{C}(t))\, dt$, but it appears that the results are qualitatively similar—see Section 8 below. The analysis turns out to be particularly straightforward and insightful for Γ.

Just as we asked previously for an optimal approximation to δ, we can now ask for an optimal m-term approximation to Γ. Suppose, given a vector orthobasis or tight frame Φ, we construct an m-term approximation to Γ by

$$\tilde{\Gamma}_m(\vec{v}) = \sum_{i=1}^{m} \Gamma(\vec{\phi}_{\mu_i})\, [\vec{v}, \vec{\phi}_{\mu_i}],$$

where we understand $[\vec{v}, \vec{w}] = \sum_j \langle v_j, w_j \rangle$, with $\langle\, , \rangle$ denoting the inner product of L^2.

In analogy to the one-dimensional case we could consider Fourier and Wavelets orthobases for vector functions with components in $L^2[-\pi, \pi)^2$, getting m-term approximants $\tilde{\Gamma}_m^F$ and $\tilde{\Gamma}_m^W$. To measure quality of approximation, we compare performance on vector fields with component functions in the ball $W_2^1(C)$ of functions in $L^2(-\pi, \pi)^2$ obeying

$$\|f\|_2^2 + \sum_{j=1}^{2} \left\| \frac{\partial}{\partial t_j} f \right\|_2^2 \leqslant C^2.$$

As a measure of performance, we can set

$$Err_m(\Gamma, \text{WAVELETS}) = \sup_{u_j \in W_2^1(C)} \inf_{I_1, \ldots, I_m} |\Gamma(\vec{v}) - \tilde{\Gamma}_m^W(\vec{v}, I_1, \ldots, I_m)|,$$

and similarly for $Err_m(\Gamma, \text{FOURIER})$.

In Section 6 below, we show that

$$Err_m(\Gamma, \text{FOURIER}) \asymp m^{-1/4}, \qquad m \to \infty, \tag{5}$$

which exhibits rather slow decay with m, while

$$Err_m(\Gamma, \text{WAVELETS}) \asymp m^{-1/2}, \qquad m \to \infty. \tag{6}$$

In short wavelets are better than Fourier in representing curves—though the advantage is now far less dramatic than it was for representing points. In fact, the performance of wavelets, although better than Fourier methods, is not very good. One wonders if there isn't something even better than wavelets for representing curves.

1.3. *Curvelets*

In this paper, we will show that the rate of m-term approximation of Γ made available by previously-known bases, such as wavelets and Fourier methods, can be substantially improved. The authors have recently constructed in [6] a new tight frame for functions in $L^2(\mathbf{R}^2)$ called a frame of curvelets. Its original purpose was to represent objects with discontinuities along C^2 curves.

We construct a vector frame of curvelets based on the principle of biorthogonal decomposition of the gradient operator. We use that frame to approximate curve-supported functionals Γ and show that, for each C^2 curve $\mathscr{C}$, there is a sequence of m-term approximations constructed using vector curvelets and obeying, for each $\delta > 0$,

$$Err_m(\Gamma, \text{VECTOR CURVELETS}) = O(m^{-1+\delta}), \qquad m \to \infty.$$

This is substantially better than the rate available from Wavelets and Fourier methods.

Moreover, it appears that this rate is essentially optimal. No basis or frame can achieve an essentially faster rate of convergence than m^{-1} uniformly on all such curves $\mathscr{C}$.

1.4. *Contents*

Section 2 gives a rapid exposition of the curvelets construction. Section 3 shows how to construct vector curvelets giving a biorthogonal decomposition of the gradient operator. Section 4 exhibits an isometry showing that approximating Γ using vector curvelets is identical to approximating images with edges using scalar curvelets. Section 5 deploys this isometry by invoking results we have proven elsewhere to determine the minimax behavior of the error *Err*

$$\max_{\mathscr{C}} \min_{\mu_1, \ldots, \mu_m} Err(\Gamma, \tilde{\Gamma}_m).$$

Section 6 compares the curvelet approximation to wavelet and Fourier approximations. Section 7 discusses interpretations of these results. Section 8 shows that qualitatively similar results are available by similar methods for the case of scalar integrands.

2. CURVELET CONSTRUCTION

We now very briefly describe the curvelet construction. There is a difference at large scales between this construction and the one given in [6].

2.1. *Ridgelets*

The theory of ridgelets was developed in the Ph.D. Thesis of Emmanuel Candès (1998). In that work, Candès showed that one could develop a system of analysis based on ridge functions

$$\psi_{a,b,\theta}(x_1, x_2) = a^{-1/2}\psi((x_1 \cos(\theta) + x_2 \sin(\theta) - b)/a). \tag{7}$$

He introduced a continuous ridgelet transform $R_f(a, b, \theta) = \langle \psi_{a,b,\theta}(x), f \rangle$ with a reproducing formula and a Parseval relation. He showed how to construct frames, giving stable series expansions in terms of a special discrete collection of ridge functions. The approach was general, and gave ridgelet frames for functions in $L^2[0,1]^d$ in all dimensions $d \geqslant 2$—For further developments, see [4, 5].

[9] showed that in two dimensions, by heeding the sampling pattern underlying the ridgelet frame, one could develop an orthonormal set for $L^2(\mathbf{R}^2)$ having the same applications as the original ridgelets. The ortho ridgelets are indexed using $\lambda = (j, k, i, \ell, \varepsilon)$, where j indexes the ridge scale, k the ridge location, i the angular scale, and ℓ the angular location; ε is a gender token. Roughly speaking, the ortho ridgelets look like pieces of ridgelets (7) which are windowed to lie in discs of radius about 2^i; $\theta_{i,\ell} = \ell/2^i$ is roughly the orientation parameter, and 2^{-j} is roughly the thickness.

The ortho-ridgelets have a concrete definition in the Fourier domain. Let $(\psi_{j,k}(t)\colon j \in \mathbf{Z}, k \in \mathbf{Z})$ be an orthonormal basis of Meyer wavelets for $L^2(\mathbf{R})$ [13], [14, Engl. Transl. p. 75], and let $(w^0_{i_0\ell}(\theta), \ell = 0, ..., 2^{i_0}-1; w^1_{i,\ell}(\theta), i \geqslant i_0, \ell = 0, ..., 2^i-1)$ 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 $i \geqslant i_0$ [14, Engl. Transl. p. 113]. 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 $x \in \mathbf{R}^2$ in the frequency-domain

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}} (\hat{\psi}_{j,k}(|\xi|)\, w^\varepsilon_{i,\ell}(\theta) + \hat{\psi}_{j,k}(-|\xi|)\, w^\varepsilon_{i,\ell}(\theta+\pi))/2. \tag{8}$$

Here the indices run as follows: $j, k \in \mathbf{Z}$, $\ell = 0, ..., 2^{i-1}-1$; $i \geqslant i_0$, and, if $\varepsilon = 0$, $i = \max(i_0, j)$, while if $\varepsilon = 1$, $i \geqslant \max(i_0, j)$. Notice the restrictions on the range of i, ℓ. Let Λ denote the set of all such indices λ.

2.2. *Multiscale Ridgelets*

Think of ortho ridgelets as objects which have a "length" of about 1 and a "width" which can be arbitrarily fine. The multiscale ridgelet system renormalizes and transports such objects, so that one has a system of elements at all lengths and all finer widths.

The construction begins with a smooth partition of energy function $w(x_1, x_2) \geqslant 0$, $w \in C_0^\infty([-1, 1]^2)$ obeying $\sum_{k_1,k_2} w^2(x_1-k_1, x_2-k_2) \equiv 1$. Define a transport operator, so that with index Q indicating a dyadic square $Q = (s, k_1, k_2)$ of the form $[k_1/2^s, (k_1+1)/2^s) \times [k_2/2^s, (k_2+1)/2^s)$, by $(T_Q f)(x_1, x_2) = f(2^s x_1 - k_1, 2^s x_2 - k_2)$. The *Multiscale Ridgelet* with index $\mu = (Q, \lambda)$ is then

$$\psi_\mu = 2^s \cdot T_Q(w \cdot \rho_\lambda).$$

In short, one transports the normalized, windowed orthoridgelet.

Letting $\mathcal{Q}_s$ denote the dyadic squares of side 2^{-s}, we can define the subcollection of *Monoscale Ridgelets* at scale s:

$$\mathcal{M}_s = \{(Q, \lambda)\colon Q \in \mathcal{Q}_s, \lambda \in \Lambda\}.$$

It is immediate from the orthonormality of the ridgelets that each system of monoscale ridgelets makes a tight frame, in particular obeying the Parseval relation

$$\sum_{\mu \in \mathcal{M}_s} \langle \psi_\mu, f \rangle^2 = \|f\|_{L^2}^2.$$

It follows that the dictionary of multiscale ridgelets at all scales, indexed by

$$\mathcal{M} = \bigcup_{s \geqslant 1} \mathcal{M}_s$$

is not frameable, as we have energy blow-up:

$$\sum_{\mu \in \mathcal{M}} \langle \psi_\mu, f \rangle^2 = \infty. \tag{9}$$

The Multiscale Ridgelets dictionary is simply too massive to form a good analyzing set. It lacks inter-scale orthogonality–$\psi_{(Q,\lambda)}$ is not typically orthogonal to $\psi_{(Q',\lambda')}$ if Q and Q' are squares at different scales and overlapping locations. In analyzing a function using this dictionary, the repeated interactions with all different scales causes energy blow-up (9).

The construction of curvelets solves this problem by in effect disallowing the full richness of the Multiscale Ridgelets dictionary. Instead of allowing all different combinations of "lengths" and "widths", we allow only those where *width* $\approx$ *length*^{2}.

2.3. *Subband Filtering*

Our remedy to the "energy blow-up" (9) is to decompose f into subbands using standard filterbank ideas. Then we assign one specific monoscale dictionary $\mathcal{M}_s$ to analyze one specific (and specially chosen) subband.

We define coronae of frequencies $|\xi| \in [2^{2s}, 2^{2s+2}]$, and subband filters D_s extracting components of f in the indicated subbands; a filter P_0 deals with frequencies $|\xi| \leqslant 1$. The filters decompose the energy exactly into subbands:

$$\|f\|_2^2 = \|P_0 f\|_2^2 + \sum_s \|D_s f\|_2^2.$$

The construction of such operators is standard [17, 12, 13]; the coronization oriented around powers 2^{2s} is nonstandard—and essential for us. Explicitly, we build a sequence of filters Φ_0 and $\Psi_{2s} = 2^{4s}\Psi(2^{2s}\cdot)$, $s = 0, 1, 2, \ldots$ with the following properties: Φ_0 is a lowpass filter concentrated near frequencies $|\xi| \leqslant 1$; Ψ_{2s} is bandpass, concentrated near $|\xi| \in [2^{2s}, 2^{2s+2}]$; and we have

$$|\hat{\Phi}_0(\xi)|^2 + \sum_{s \geqslant 0} |\hat{\Psi}(2^{-2s}\xi)|^2 = 1, \qquad \forall \xi.$$

Hence, D_s is simply the convolution operator $D_s f = \Psi_{2s} * f$.

2.4. *Definition of Curvelet Transform*

Assembling the above ingredients, we are able to sketch the definition of the Curvelet transform. We let M' consist of M merged with the collection of integral triples $(s, k_1, k_2, \varepsilon)$ where $s \leqslant 0$, (s, k_1, k_2) indexes coarse scale dyadic squares in the plane of side $2^{-s} \geqslant 1$, $\varepsilon \in \{01, 10, 11\}^2$ is a gender indicator.

The curvelet transform is a map $L^2(\mathbf{R}^2) \mapsto \ell^2(\mathbf{M}')$, yielding curvelet coefficients $(\alpha_\mu: \mu \in M')$. These come in two types.

At *coarse scales* we have wavelet coefficients.

$$\alpha_\mu = \langle W_{s, k_1, k_2, \varepsilon}, P_0 f \rangle, \qquad \mu = (s, k_1, k_2) \in M' \backslash M,$$

where each $W_{s, k_1, k_2, \varepsilon}$ is a Meyer wavelet, while at *fine scale* we have Multi-scale Ridgelet coefficients of the bandpass filtered object:

$$\alpha_\mu = \langle D_s f, \psi_\mu \rangle, \qquad \mu \in M_s, \; s = 1, 2, \ldots.$$

Note well that for $s > 0$, each coefficient associated to scale 2^{-s} derives from the subband filtered version of $f - D_s f$—and not from f.

Several properties are immediate:

- Tight frame:

$$\|f\|_2^2 = \sum_{\mu \in M'} |\alpha_\mu|^2.$$

- Existence of coefficient representers (Frame elements): There are $\gamma_\mu \in L^2(\mathbf{R}^2)$ so that

$$\alpha_\mu \equiv \langle f, \gamma_\mu \rangle.$$

- L^2 Reconstruction formula:

$$f = \sum_{\mu \in M'} \langle f, \gamma_\mu \rangle \, \gamma_\mu.$$

• Formula for frame elements: for $s \leqslant 0$, $\gamma_\mu = P_0 W_{s,k_1,k_2,\varepsilon}$, while for $s > 0$,

$$\gamma_\mu = D_s \psi_\mu, \qquad \mu \in M_s.$$

In short, fine-scale curvelets are obtained by bandpass filtering of Multiscale Ridgelets coefficients where the *passband* is rigidly linked to the *scale* of spatial localization.

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

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

Note: it is at this last step that our 2^{2s} coronization scheme comes fully into play.

• Oscillatory nature. Both for $s > 0$ and $s \leqslant 0$, each frame element has a Fourier transform supported in an annulus away from 0.

3. VECTOR CURVELET FRAMES FOR ∇

We define now a pair of vector frames $\vec{w}_\mu^{\pm}$ associated to biorthogonal decomposition of the grad operator ∇. Define the sequence of multipliers

$$\kappa_s = \begin{cases} 2^{-s} & s < 0 \\ 2^{-2s} & s \geqslant 0, \end{cases}$$

and derive the vector frame $(\vec{w}_\mu^{+})$ from the curvelet frame by differentiation

$$\vec{w}_\mu^{+}(x) = \kappa_s \cdot \nabla \gamma_\mu. \tag{10}$$

This makes sense because curvelet frame elements are smooth and of rapid decay.

Define each member of the dual vector frame $(\vec{w}_\mu^{-})$ componentwise in the frequency domain, with jth component

$$\widehat{w_{\mu,j}^{-}}(\xi) = \kappa_s^{-1} \cdot \hat{\gamma}_\mu(\xi) \cdot \frac{i\xi_j}{\|\xi\|^2}, \qquad \xi \in \operatorname{supp}(\hat{\gamma}_\mu(\xi)). \tag{11}$$

This makes sense because each γ_μ omits the origin from its support in the frequency domain. Defining the divergence $Div(\vec{v}) = \sum_j (\partial/\partial x_j)\, v_j$, we then have

$$\kappa_s \cdot Div(\vec{w}_\mu^-) = \gamma_\mu, \tag{12}$$

which bears comparison with (10).

We will also need the definition of the Riesz transforms [16] R_j for $j = 1, 2$:

$$R_j(f)\,(x) = \frac{1}{(2\pi)^2} \int \hat{f}(\xi)\, \frac{i\xi_j}{\|\xi\|}\, e^{i\xi' x}\, d\xi.$$

These are bounded operators of $L^2(\mathbf{R}^2)$ which obey the Pythagorean relation

$$\sum_j \|R_j(f)\|_2^2 = \|f\|_2^2.$$

DEFINITION 3.1. We say that a vector field $\vec{v}$ with components v_j in $L^2(\mathbf{R}^2)$ for $j = 1, 2$ is **irrotational** if the components obey

$$\|R_1(v_2) - R_2(v_1)\|_2 = 0.$$

The irrotational vector fields have a special structure:

LEMMA 3.1. *Given a vector field $\vec{v}$ whose components arise according to*

$$v_j = R_j(V), \qquad j = 1, 2,$$

where V is an $L^2(\mathbf{R}^2)$ function, then $\vec{v}$ is irrotational with $L^2(\mathbf{R}^2)$ components.

Given an irrotational vector field $\vec{v}(x_1, x_2)$ with $L^2(\mathbf{R}^2)$ components, there is a scalar function $V \in L^2(\mathbf{R}^2)$ with

$$v_j = R_j(V), \qquad j = 1, 2. \tag{13}$$

Moreover, we have the Pythagorean relation

$$\sum_j \|v_j\|_2^2 = \|V\|_2^2. \tag{14}$$

Proof. To see the first half, substitute into the definition of irrotationality, getting:

$$\|R_1(v_2) - R_2(v_1)\|_2 = \|R_1(R_2(V)) - R_2(R_1(V))\|_2.$$

Now as the R_i are Fourier multipliers, they commute: $R_1 R_2 \equiv R_2 R_1$. Hence $\|R_1(R_2(V)) - R_2(R_1(V))\|_2 = 0$.

To see the second half, note that the Fourier transforms of components of an irrotational field obey

$$\xi_1 \cdot \hat{v}_2(\xi) = \xi_2 \cdot \hat{v}_1(\xi) \qquad \text{a.e.} \quad d\xi.$$

It follows that at a.e. ξ, the vector $(\hat{v}_1(\xi), \hat{v}_2(\xi))$ lies in the subspace spanned by (ξ_1, ξ_2). Letting $\vec{\rho}(\xi) = (\xi_1/\|\xi\|, \xi_2/\|\xi\|)$, and defining a.e. ξ the function

$$\hat{V}(\xi) = (\xi_1 \hat{v}_1(\xi) + \xi_2 \hat{v}_2(\xi))/\|\xi\|,$$

it follows that a.e. ξ we have

$$(\hat{v}_1(\xi), \hat{v}_2(\xi)) = \hat{V}(\xi) \cdot \rho(\xi).$$

The desired result (13) is just the same equation in the original domain. The Pythagorean relation is immediate. Q.E.D.

We note that the $\vec{w}_\mu^{\pm}$ are irrotational. Indeed, $\vec{w}_\mu^{+}$ arises as the gradient field of a scalar function, and any such field is irrotational: the Fourier representation

$$\left(\frac{\partial}{\partial x_j} f\right)^{\wedge} (\xi) = (i\xi_j)\, \hat{f}(\xi)$$

shows that V is given by $\hat{V}(\xi) = \hat{f}(\xi)\, \|\xi\|$.

As for $\vec{w}_\mu^{-}$ the Fourier domain formula (11) may be viewed as exhibiting $\vec{w}_\mu^{-}$ explicitly as such as function, with $\hat{V}(\xi) \propto \hat{\gamma}_\mu(\xi)/\|\xi\|$.

The following result shows that we can represent all irrotational fields with the $(\vec{w}_\mu^{\pm})$,

THEOREM 3.1. *The systems $(\vec{w}_\mu^{+})_\mu$ and $(\vec{w}_\mu^{-})_\mu$ are vector frames for the space of irrotational vector fields with components in $L^2(\mathbf{R}^2)$. For each choice of sign for $\pm$ we obtain a system obeying the almost orthogonality*

$$\left(\sum_j \left\| \sum_\mu a_\mu w_{\mu,j}^{\pm} \right\|_2^2\right)^{1/2} \leqslant C \cdot \left(\sum_\mu a_\mu^2\right)^{1/2} \tag{15}$$

along with the L^2 norm equivalence:

$$\sum_\mu [\vec{w}_\mu^{\pm}, \vec{v}]^2 \asymp \left(\sum_j \|v_j\|_2^2\right)^{1/2} \qquad \forall \textit{ irrotational fields } \vec{v}. \tag{16}$$

In addition the two vector frames are quasi-biorthogonal:

$$[\vec{w}_{\mu}^{+}, \vec{w}_{\mu'}^{-}] = 2^{s'-s} \cdot \langle \gamma_{\mu}, \gamma_{\mu'} \rangle, \qquad \mu, \mu' \in M, \tag{17}$$

where we understand

$$[\vec{w}_{\mu}^{+}, \vec{w}_{\mu'}^{-}] = \sum_{j} \langle w_{\mu,j}^{+}, w_{\mu',j}^{-} \rangle.$$

The proof of Theorem 1 is effectively a repeated application of homogeneous Fourier multiplier ideas. We begin with the biorthogonality (17), note that we have the Fourier side definition of $w_{\mu,j}^{+}$:

$$\widehat{w_{\mu,j}^{+}}(\xi) = \kappa_s \cdot \hat{\gamma}_{\mu}(\xi) \cdot i\xi_j, \qquad \forall \xi. \tag{18}$$

Hence, passing to the Frequency side,

$$\langle w_{\mu,j}^{+}, w_{\mu',j}^{-} \rangle = \frac{1}{(2\pi)^2} \int \widehat{w_{\mu,j}^{+}}(\xi)\, \overline{\widehat{w_{\mu',j}^{-}}(\xi)}\, d\xi \tag{19}$$

$$= \frac{1}{(2\pi)^2} \int \kappa_s \cdot \hat{\gamma}_{\mu}(\xi)\, i\xi_j \cdot \kappa_s^{-1} \cdot \overline{\hat{\gamma}_{\mu'}}(\xi)\, \frac{-i\xi_j}{\|\xi\|^2}\, d\xi \tag{20}$$

$$= \frac{1}{(2\pi)^2} \int \hat{\gamma}_{\mu}(\xi) \cdot \overline{\hat{\gamma}_{\mu'}(\xi)} \cdot \frac{\xi_j^2}{\|\xi\|^2}\, d\xi. \tag{21}$$

Hence,

$$\sum_{j} \langle w_{\mu,j}^{+}, w_{\mu',j}^{-} \rangle = \frac{1}{(2\pi)^2} \int \hat{\gamma}_{\mu}(\xi) \cdot \overline{\hat{\gamma}_{\mu'}(\xi)} \cdot \sum_{j} \frac{\xi_j^2}{\|\xi\|^2}\, d\xi \tag{22}$$

$$= \langle \gamma_{\mu}, \gamma_{\mu'} \rangle, \tag{23}$$

giving (17).

To get the frame properties (15)-(16), we use the fact that curvelets relate well to fractional powers of the Laplacian. The usual Laplacian $\Delta = \sum_{i=1}^{2} \partial^2/(\partial x_i^2)$ corresponds to the Fourier multiplier $(\Delta f)\hat{}\,(\xi) = -|\xi|^2 \cdot \hat{f}(\xi)$; it makes sense therefore to define fractional powers of the Laplacian by the Fourier multiplier

$$((-\Delta)^{\alpha} f)\hat{}\,(\xi) = |\xi|^{2\alpha} \cdot \hat{f}(\xi).$$

Following the article [7] we define, for a curvelet $\gamma_{\mu}(x_1, x_2)$, two *companions* $c_{\mu}^{\pm}(x_1, x_2)$ according to

$$c_{\mu}^{\pm} = \kappa_s^{\mp 1}(-\Delta)^{\pm 1/2}\, \gamma_{\mu},$$

where of course s refers to the scale index occupying the first slot $(s, k_1, k_2, j, k, i, \ell, \varepsilon)$ in the curvelet index μ. Because γ_μ is effectively concentrated in the frequency domain near $|\xi| \approx 2^{2s}$, we have $2^{2s}\,|\xi| \approx 1$ through the bulk of the frequency domain support of γ_μ and hence we anticipate $\|c_\mu^\pm\| \approx \|\gamma_\mu\|$. The following result can be proved along the lines of a similar result in [7].

THEOREM 3.2. *The systems $(c_\mu^+)_\mu$ and $(c_\mu^-)_\mu$ are frames for $L^2(\mathbf{R}^2)$. For either choice of sign $\pm$, one obtains a system with almost orthogonality*

$$\Big\| \sum_\mu a_\mu c_\mu^\pm \Big\|_2 \leqslant C \cdot \Big(\sum_\mu a_\mu^2 \Big)^{1/2}, \tag{24}$$

and with L^2 norm equivalence:

$$\sum_\mu \langle c_\mu^\pm, f\rangle^2 \asymp \|f\|_2^2. \tag{25}$$

In addition, they are quasi-biorthogonal:

$$\langle c_\mu^+, c_{\mu'}^-\rangle = 2^{s'-s} \langle \gamma_\mu, \gamma_{\mu'}\rangle. \tag{26}$$

To get now (15) form Theorem 2, we observe that

$$w_{\mu,j}^\pm = R_j(c_\mu^\pm), \qquad j = 1, 2. \tag{27}$$

From this and the Pythagorean relation for Riesz transforms we have

$$\sum_j \Big\| \sum_\mu a_\mu w_{\mu,j}^\pm \Big\|_2^2 = \Big\| \sum_\mu a_\mu c_\mu^\pm \Big\|_2^2.$$

Thus the frame relation (15) for $(\vec{w}_\mu^\pm)$ follows from (24) for $(c_\mu^\pm)$.

We now use Lemma 3.1 to get (16). Given a vector field $\vec{v}$ it furnishes an associated scalar field V with

$$\begin{aligned} \sum_\mu [\vec{w}_\mu^\pm, \vec{v}]^2 &= \sum_\mu \Big(\sum_j \langle R_j(c_\mu^\pm), v_j\rangle \Big)^2 \\ &= \sum_\mu \Big(\sum_j \langle R_j(c_\mu^\pm), R_j(V)\rangle \Big)^2. \end{aligned}$$

The Pythagorean relation for the Riesz transforms gives

$$\sum_j \langle R_j(c_\mu^\pm), R_j(V)\rangle = \langle c_\mu^\pm, V\rangle.$$

Substituting into the preceding display, we have an isometry between the $w_\mu^\pm$ coefficients of $\vec{v}$ and the $c_\mu^\pm$ coefficients of V:

$$\sum_\mu [\vec{w}_\mu^\pm, \vec{v}]^2 = \sum_\mu \langle c_\mu^\pm, V\rangle^2.$$

Finally,

$$\begin{aligned}\sum_\mu \langle c_\mu^\pm, V\rangle^2 &\asymp \|V\|_2^2 \\ &= \sum_j \|v_j\|_2^2,\end{aligned}$$

where we used the frame property (25) of the companions $(c_\mu^\pm)$ and the Pythagorean relation (14).

THEOREM 3.3. *We have the reproducing formula*

$$\nabla f = \sum_\mu [\nabla f, \vec{w}_\mu^-]\, \vec{w}_\mu^+ \tag{28}$$

valid for all sufficiently nice f, where we understand

$$[\nabla f, \vec{w}_\mu^-] = \sum_j \left\langle \frac{\partial}{\partial x_j} f, w_{\mu,j}^- \right\rangle.$$

Suppose that f is a finite superposition of γ_μ's. Evidently,

$$\nabla f = \sum_\mu \langle f, \gamma_\mu\rangle \nabla\gamma_\mu \tag{29}$$

$$= \sum_\mu \kappa_s^{-1} \cdot \langle f, \gamma_\mu\rangle \cdot \vec{w}_\mu^+. \tag{30}$$

We now show that

$$\kappa_s^{-1} \cdot \langle f, \gamma_\mu\rangle = [\nabla f, \vec{w}_\mu^-]. \tag{31}$$

We start from an integration-by-parts:

$$\left\langle \frac{\partial}{\partial x_j} f, w_{\mu,j}^- \right\rangle = -\left\langle f, \frac{\partial}{\partial x_j} w_{\mu,j}^- \right\rangle.$$

Passing to the frequency side, we have

$$\left\langle f, \frac{\partial}{\partial x_j} w^-_{\mu,j} \right\rangle = \frac{\kappa_s^{-1}}{(2\pi)^2} \int \hat{f}(\xi)\, \overline{i\xi_j \hat{\gamma}_\mu(\xi)}\, \frac{-i\xi_j}{\|\xi\|^2}\, d\xi \tag{32}$$

$$= \frac{\kappa_s^{-1}}{(2\pi)^2} \int \hat{f}(\xi)\, \overline{\hat{\gamma}_\mu(\xi)}\, \frac{\xi_j^2}{\|\xi\|^2}\, d\xi \tag{33}$$

$$= -\kappa_s^{-1} \cdot \langle R_j(f), R_j(\gamma_\mu)\rangle. \tag{34}$$

The Pythagorean relation for the Riesz transforms gives

$$\sum_j \langle R_j(f), R_j(\gamma_\mu)\rangle = \langle f, \gamma_\mu\rangle$$

and the proof is complete.

4. A VECTOR/SCALAR ISOMETRY

Let B be a region in $\mathbf{R}^2$ with smooth simple boundary curve $\mathscr{C}$. Let Γ be the corresponding linear functional. We will see below that in a distributional sense,

$$\Gamma = \sum \Gamma(\vec{w}^-_\mu)\, \vec{w}^+_\mu. \tag{35}$$

Suppose that we build an m-term approximation $\tilde{\Gamma}_m$ using terms $\mu_1, ..., \mu_m$:

$$\tilde{\Gamma}_m = \sum_{i=1}^m \Gamma(\vec{w}^-_{\mu_i})\, \vec{w}^+_{\mu_i}. \tag{36}$$

To see how good an approximation $\tilde{\Gamma}_m$ might be, we consider its maximal deviation from Γ over smooth vector fields:

$$Err(\Gamma, \tilde{\Gamma}_m) = \sup\{|\Gamma(\vec{v}) - \tilde{\Gamma}_m(\vec{v})| : \|Div(\vec{v})\|_2 \leqslant 1\}.$$

It turns out that there is an isometry linking performance of m-term vector curvelet approximation to Γ with performance of an m-term scalar curvelet approximation to B.

THEOREM 4.1. *Let $\mu_1, \ldots, \mu_m$ be given. Consider the m-term approximation to $f = 1_B$ by curvelets, formed according to*

$$\tilde{f}_m = \sum_{i=1}^{m} \langle 1_B, \gamma_{\mu_i} \rangle \gamma_{\mu_i}.$$

Let $\tilde{\Gamma}_m$ be as in (36). *Then*

$$Err(\Gamma, \tilde{\Gamma}_m) = \|f - \tilde{f}_m\|_2.$$

In this section, we develop the proof of this theorem, which is really the application of two specific isometries.

4.1. *Gauss-Green Theorem*

Let B be a region in the plane with C^2 boundary, and let ϕ be a smooth function. Then

$$\int_B \Delta\phi = \int_{\partial B} \vec{n} \cdot \nabla\phi.$$

Here ∂B is the boundary of B, and n is the normal field along the boundary.

Reinterpret this as follows. Suppose B is a region with C^2 boundary ∂B and that $\mathscr{C}(t)$ is a unit speed parametrization of the boundary. Let Γ be the corresponding linear functional

$$\Gamma(\vec{v}) = \int \vec{v}(\mathscr{C}(t)) \cdot \vec{n}(t)\, dt.$$

Then

$$\langle 1_B, \Delta\phi \rangle = \Gamma(\nabla\phi). \tag{37}$$

In short, for smooth functions ϕ, integrals over B can be related to the functional Γ applied to related functions of $\nabla\phi$.

Recall the formal identity from vector calculus $\Delta = \nabla \cdot \nabla$, which can be stated correctly for nice f and g as

$$\langle f, \Delta g \rangle = -[\nabla f, \nabla g].$$

From a modern functional analysis perspective, we may use a distributional version of this, where f is not smooth, to rewrite (37), giving

$$\Gamma(\nabla\phi) = \langle 1_B, \Delta\phi \rangle = -[\nabla 1_B, \nabla\phi],$$

identifying $\Gamma = -\nabla 1_B$ in a distributional sense. Applying now (28) from Theorem 3, we get (35).

4.2. *Evaluation Isometry*

Let $\vec{W}$ denote a vector field on $\mathbf{R}^2$ and consider the seminorm

$$|||\vec{W}||| = \sup\{[\vec{W}, \vec{v}] : \|Div(\vec{v}\|_2 \leqslant 1\}.$$

This is well-defined on smooth vector fields of compact support having zero mean value.

It is a simple exercise in integration by parts—again reducing to the formal identity $\Delta = \nabla \cdot \nabla$—to see that we have the isometry

$$|||\nabla\phi||| = \|\phi\|_2, \tag{38}$$

valid whenever ϕ is smooth and of compact support.

4.3. *m-term Approximation Isometry*

We are now able to prove Theorem 4.

Owing to the definition of the norm $|||\cdot|||$, we may write

$$Err(\Gamma, \tilde{\Gamma}_m) = |||\Gamma - \tilde{\Gamma}_m|||.$$

Owing to (37) and the definition of $\vec{w}_\mu^-$, we can write

$$\Gamma(\vec{w}_\mu^-) = \kappa_s^{-1}\langle 1_B, \gamma_\mu\rangle.$$

Applying this,

$$\begin{aligned} |||\Gamma - \tilde{\Gamma}_m||| &= \left|\left|\left|\nabla 1_B - \sum_{i=1}^m \Gamma(\vec{w}_{\mu_i}^-)\, w_{\mu_i}^+\right|\right|\right| \\ &= \left|\left|\left|\nabla 1_B - \sum_{i=1}^m \Gamma(\vec{w}_{\mu_i}^-)\, \kappa_s \cdot \nabla\gamma_{\mu_i}\right|\right|\right| \\ &= \left|\left|\left|\nabla\left(1_B - \sum_{i=1}^m \Gamma(\vec{w}_{\mu_i}^-)\, \kappa_s \gamma_{\mu_i}\right)\right|\right|\right| \\ &= \left\|1_B - \sum_{i=1}^m \langle 1_B, \gamma_{\mu_i}\rangle\, \gamma_{\mu_i}\right\|_2, \end{aligned}$$

the last step invoking (38).

5. EXPLOITING THE VECTOR/SCALAR ISOMETRY

The problem of approximating an indicator 1_B by curvelets when B has a C^2 boundary has recently been studied by Candès and Donoho (1999). They showed the following.

THEOREM 5.1 [6]. *For $m > 0$ define the m-term approximation to $f = 1_B$ as follows. Let $\mu_1, \ldots, \mu_m$ denote the indices of the m largest-amplitude curvelet coefficients of f. Define an m-term approximation to f by*

$$\tilde{f}_m \sum_{i=1}^{m} \langle f, \gamma_{\mu_i} \rangle \gamma_{\mu_i}.$$

Then if B is a subset of the unit square with a simple C^2 boundary curve $\mathscr{C}$, and the C^2 norm of $t \mapsto \mathscr{C}(t)$ is at most A, then

$$\|f - \tilde{f}_m\|_2 \leqslant C(A) \cdot \log(m)^{3/2} m^{-1},$$

where $C(A)$ denotes a constant depending only on A.

Combining this with our isometry, we have the following.

THEOREM 5.2. *Let $\mathscr{C}$ be a simple C^2 unit-speed curve, and suppose the C^2 norm of $t \mapsto \mathscr{C}(t)$ is at most A. Let Γ be the corresponding curve integral. For each $m > 0$, let $\mu_i, \ldots, \mu_m$ denote the indices of the m largest-amplitude vector curvelet coefficients of Γ (i.e. the m-largest amplitude $\Gamma(\vec{w}_\mu)$). Define an m-term approximation to Γ by*

$$\tilde{\Gamma}_m = \sum_{i=1}^{m} \Gamma(\vec{w}_{\mu_i}^{-}) \vec{w}_{\mu_i}^{+}.$$

Then

$$Err(\Gamma, \tilde{\Gamma}_m) \leqslant C(A) \cdot \log(m)^{3/2} m^{-1},$$

where $C(A)$ denotes a constant depending only on A.

6. COMPARISON: FOURIER AND WAVELET REPRESENTATIONS

To place Theorem 5.2 in perspective, we now sketch comparable constructions for sinusoids and for wavelets, justifying the results of the introduction.

6.1. *Vector Fourier Representation*

Let the domain of interest be $[-\pi, \pi)^2$; with $\mathbf{k}$ denoting the pair (k_1, k_2), let $\mathbf{K}$ denote the collection of *nonzero pairs*. With $e_{\mathbf{k}}(t)$ denoting the complex exponential $e^{i(k_1 t_1 + k_2 t_2)}/(2\pi)$, the collection $(e_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}}$ is an orthonormal set, but not a basis, as it is missing the constant function $e_{0,0}$.

With $\|\mathbf{k}\| = (k_1^2 + k_2^2)^{1/2}$, let $\vec{E}_{\mathbf{k}}(t_1, t_2)$ be a vector field on $[-\pi, \pi)^2$, defined by

$$\vec{E}_{\mathbf{k}}(t) = e_{\mathbf{k}}(t) \cdot \mathbf{k}/\|\mathbf{k}\|, \qquad \mathbf{k} \in \mathbf{K}.$$

Then we have

$$Div(\vec{E}_{\mathbf{k}}) = \|\mathbf{k}\| \cdot e_{\mathbf{k}},$$

and

$$\nabla e_{\mathbf{k}} = \|\mathbf{k}\| \cdot \vec{E}_{\mathbf{k}}$$

and

$$[\vec{E}_{\mathbf{k}}(t), \vec{E}_{\mathbf{k}'}(t)] = \delta_{\mathbf{k}, \mathbf{k}'}.$$

This system makes a tight frame giving a sort of diagonal representation of ∇:

$$\nabla f = \sum_k [\nabla f, \vec{E}_{\mathbf{k}}]\, \vec{E}_{\mathbf{k}}.$$

We will use the vanishing-mean system $(\vec{E}_{\mathbf{k}})_{\mathbf{k} \in \mathbf{K}}$ to represent Γ; we write tentatively

$$\Gamma = \sum_{\mathbf{k} \in \mathbf{K}} \Gamma(\vec{E}_{\mathbf{k}})\, \vec{E}_{\mathbf{k}}.$$

At first glance, this expansion seems to omit a necessary component at zero frequency $\mathbf{k} = (0, 0)$. However, this turns out not to be needed, because of the following.

LEMMA 6.1. *Let B be a region in the plane with boundary curve $\mathscr{C}$, and let Γ denote the corresponding curvilinear integral. Then for any constant vector field $\vec{c}$,*

$$\Gamma(\vec{c}) = 0.$$

The lemma is quite clear from the net flux interpretation of Γ. As $Div(\vec{c}) = 0$, the net flux into the region B must vanish.

For the m-term approximation

$$\tilde{\Gamma}_m = \sum_{i=1}^{m} \Gamma(\vec{E}_{\mathbf{k}(i)})\, \vec{E}_{\mathbf{k}(i)},$$

we can again derive the isometry

$$Err(\Gamma, \tilde{\Gamma}_m) = \|f_0 - \tilde{f}_m\|_2,$$

where $f_0 = 1_B - \langle 1_B, e_{0,0} \rangle e_{0,0}$ is a zero-mean version of the indicator of B and where

$$\tilde{f}_m = \sum_{i=1}^{m} \langle f_0, e_{\mathbf{k}(i)} \rangle e_{\mathbf{k}(i)}.$$

The claim (5) of the introduction follows from a simple observation: Let B be a region with smooth C^2 boundary having nonvanishing curvature. Then the Fourier coefficients obey $\hat{f}_{\mathbf{k}} \asymp \|\mathbf{k}\|^{-3/2}$ as $\|\mathbf{k}\| \to \infty$. The typical example of this is given by the indicator of a disk, whose Fourier representation involves Bessel functions. It follows that, in general, there are $O(R^2)$ coefficients of size $\geqslant R^{-3/2}$. Now as the system of complex exponentials is orthonormal, the best m-term approximation is built from the m terms having the m largest amplitudes. It follows that

$$\|f - \tilde{f}_m\|_2 \asymp m^{-1/4}.$$

This establishes (5).

6.2. *Vector Wavelet Representation*

Now let the domain be $\mathbf{R}^2$ and consider an orthobasis of Meyer wavelets, $(\psi_{s, k_1, k_2, \varepsilon} : s, k_1, k_2 \in \mathbf{Z})$, where s is a scale index, k_1, k_2 are position indices and ε is a bivariate gender indicator. For short, we put $I = (s, k_1, k_2, \varepsilon)$.

We define associated vector fields via Fourier multiplier methods as in Section 3. With $\kappa_s = 2^{-s}$, we let

$$\vec{\phi}_I^+ = \kappa_s \cdot \nabla \psi_I$$

and we define the vector field $\vec{\phi}_I^-$ componentwise by

$$(\vec{\phi}_I^-)_j = \kappa_s^{-1} \cdot R_j((-\Delta)^{-1/2} \psi_I),$$

with R_j the jth Riesz transform and $(-\Delta)^{-1/2}$ the appropriate fractional Laplacian.

This pair of systems again makes a pair of biorthogonal frames, and we may write formally

$$\Gamma = \sum_I \Gamma(\vec{\phi}_I^-) \vec{\phi}_I^+.$$

For the m-term approximation

$$\tilde{\Gamma}_m = \sum_{i=1}^{m} \Gamma(\vec{\phi}^{-}_{I(i)})\, \vec{\phi}^{+}_{I(i)},$$

we can again derive the isometry

$$Err(\Gamma, \tilde{\Gamma}_m) = \|f - \tilde{f}_m\|_2,$$

where $f = 1_B$ and where

$$\tilde{f}_m = \sum_{i=1}^{m} \langle f, \psi_{I(i)} \rangle\, \psi_{I(i)}.$$

The claim (6) of the introduction follows from a simple observation: Let B be a region with smooth C^2 boundary having nonvanishing curvature. Then there are order 2^s wavelet coefficients at scale 2^{-s} which correspond to spatial locations "on the curve", and these wavelet coefficients are of size $\approx 2^{-s}$. It follows that, in general, there are $O(N)$ coefficients of size $\geqslant c \cdot N^{-1}$. Now as the system of wavelets is orthonormal, the best m-term approximation is built from the m terms having the m largest amplitudes. It follows that

$$\|f - \tilde{f}_m\|_2 \asymp m^{-1/2}.$$

This establishes (6).

7. DISCUSSION

7.1. *Optimality of These Results*

We know, as explained in [6, 8], that the result quoted in Theorem 5 is near optimal. That is, no well-posed system of representation (even allowing substantial adaptation) can do better than the rate $1/m$ in approximating objects with edges; and the curvelet system does essentially this well (except for the log terms).

This seems to imply that the best possible rate of m-term approximation of curvilinear integrals Γ cannot converge faster than a $1/m$ rate in general. However, what we have shown is actually not quite that strong.

Unfortunately the edges/curvilinear integrals isometry does not quite settle the question of optimal m-term approximation of Γ. For a comprehensive answer, we would need an independent lower bound for the problem of m-term approximation of Γ. Without this there remains the possibility that some method of direct approximation to Γ could be invented to which the isometry cannot be applied, and for which no corresponding method of approximation of f existed, and for which lower bounds for approximation of f would not be relevant.

7.2. *Eulerian-Lagrangian Perspective*

Our results are of most interest in connection with approximating a family of integrals, rather than a single integral. Suppose we have a sequence of vector fields $\vec{v}_i$, and a collection of integrals $\Gamma_1, \dots, \Gamma_N$ which we may like to evaluate, and which are of quite general form. For example, we might have a sequence of evolving curves C_n, $n = 1, \dots, N$. Then an interesting strategy for this evaluation would be to expand both the vector fields and the integrals into the vector curvelet frame and exploit the coefficient sparsity of the Γ_n to calculate the integrals rapidly.

Thus we are adopting a single coordinate system for representing these curvilinear integrals; this is a kind of Eulerian perspective. In contrast, an approach which attempted to build a representation specifically driven by the curves $\mathscr{C}_n$ would be Lagrangian. Our opinion is that there may be little advantage to doing so; compare [8] for a clearer exposition on this point.

7.3. *Interpretation of the Expansion*

The vector curvelet expansion synthesizes the singular functional using terms obeying the scaling law *width* $\approx$ *length*^{2}, clustering more and more tightly about the curve at fine scales. An m-term approximation represents Γ as a smooth vector field peaking very strongly in the vicinity of the curve $\mathscr{C}$. The functional $\tilde{\Gamma}_m(\vec{v})$ therefore uses information about $\vec{v}$ not just on the curve on $\mathscr{C}$ but also in a neighborhood about the curve; this neighborhood shrinks as successively more terms are included in the approximation.

8. SCALAR INTEGRANDS

Until now, we have focused on curvilinear integrals associated with vector integrands. The same methods can give results for scalar integrals acting on functions,

$$\mathscr{I}(f) = \int f(\mathscr{C}(t))\,dt,$$

where again $\mathscr{C}(t)$ is a simple unit-speed C^2 curve bounding a closed region B, which is contained entirely inside $[-1, 1]^2$.

In this setting, suppose we have a pair of dual frames with synthesizing elements (ϕ_μ) and analyzing elements (φ_μ). We consider methods which build m-term approximations of the form

$$\tilde{\mathscr{I}}_m(f) = \sum_{i=1}^{m} \mathscr{I}(\varphi_{\mu_i})\, \phi_{\mu_i}, \tag{39}$$

where the μ_i can be chosen adaptively based on the curve $\mathscr{C}$. We are interested in finding a basis or frame which, when used in the above scheme, yields the highest-quality m-term approximations. We measure performance according to

$$Err(\mathscr{I}, \tilde{\mathscr{I}}_m) = \sup\{|\mathscr{I}(f) - \tilde{\mathscr{I}}_m(f)| : \|\Delta f\|_2 \leqslant 1\}.$$

In short, we are asking that the m-term approximation reproduce the integral accurately for all sufficiently smooth functions, where smoothness is measured by the L^2 size of the Laplacian.

8.1. *Isometry with L^2 Curvelet Approximation*

In analogy with the vector case, we can establish an isometry between this problem and a problem of approximation of a piecewise smooth object with a singularity along a curve. Given a curve $\mathscr{C}$, our plan is to formally associate a corresponding continuous function $H(x_1, x_2)$ which we use curvelets to approximate in L^2. Then, we observe that there is a formal isometry, showing that when we approximate H well in L^2 norm by curvelets, then we approximate $\mathscr{I}$ well by dual curvelets according to the *Err* metric. Making this work out completely and not just formally in this case will require some adjustments to the approximation scheme (39).

The key steps in this process are as follows.

• *Frame construction.* Starting with curvelets (γ_μ), we can build a pair of frames in a fashion similar to the c_μ functions. With the powers of the Laplacian $(-\Delta)^\alpha$ defined in the obvious way, set

$$\phi_\mu^+ = 2^{-4s}(-\Delta)\,\gamma_\mu,$$
$$\phi_\mu^- = 2^{+4s}(-\Delta)^{-1}\,\gamma_\mu.$$

The resulting systems obey inequalities exactly paralleling those in Theorem 2. With constants $\kappa_s = 2^{-4s}$ at positive s and 2^{-s} at negative s, they make a biorthogonal decomposition of the Laplacian Δ:

$$\Delta f = \sum_\mu \kappa_s^{-1} \cdot \langle \Delta f, \phi_\mu^- \rangle\, \phi_\mu^+.$$

• *Induced potential.* Given the curve $\mathscr{C}$, consider the object H defined on R^2 by

$$H(x) = \int \log(\|x - \mathscr{C}(s)\|)\, ds.$$

This is a C^∞ function on $R^2 \setminus \text{Image}(\mathscr{C})$, which, as $\log(r)$ is the fundamental solution of Laplace's equation, satisfies

$$\Delta H = \mathscr{I},$$

in the distributional sense. In fact, H is the electrostatic potential associated with a uniform distribution of charges on $\mathscr{C}$. For more information on potential theory and Laplace's equation, see e.g. [1, 15]. We wish to exploit the fact that H solves this PDE, and the fact that our error measure Err involves the same Laplacian Δ, to obtain an isometry of m-term approximation problems.

• *Moment matching.* Unfortunately, the potential H is not in L^2: it has growth $O(\log(\|x\|))$ at ∞; see the example below. We desperately need the L^2 property for our approach to make sense. To obtain this, we construct a special function G "matching" the asymptotic growth of H and subtract it off. To get G, we construct g with these properties:

— g is C^∞ and of rapid decay.

— g is in the span of the curvelets (γ_λ) for $s < 0$. Essentially, it is a combination of wavelets at coarse scales only.

— g matches the low-order moments of $\mathscr{I}$: for every linear polynomial $\pi(x) = a + bx_1 + cx_2$,

$$\mathscr{I}(\pi) = \langle g, \pi \rangle.$$

We then define G as the solution to

$$\Delta G = g;$$

this may be obtained from the convolution $\log(r) * g$. It is easily shown that G is C^∞ and of slow growth at ∞. In fact, it has asymptotic growth properties snatching H, so that $H - G$ is in L^2. (Of course much more is true).

• *Compensation.* Define now the compensated object

$$H^0 = H - G.$$

This belongs to L^2; in fact, it is of rapid decay at ∞ and is C^∞ away from the curve $\mathscr{C}$. Moreover,

$$\Delta H^0 = \mathscr{I} - g = \mathscr{I}^0, \text{ say.}$$

Hence, while $\mathscr{I}$ does not correspond to an L^2 object, a smooth perturbation $\mathscr{I}^0$ does.

• *m-term Approximation to* H^0. Now we use curvelets to approximate H^0. In fact, H^0 is a highly regular object, C^∞ except on the curve $\mathscr{C}$, where it exhibits a cusp singularity. Enumerate the curvelets as γ_{μ_i} in order of decreasing curvelet coefficients of H^0; then consider the m-term approximation:

$$\tilde{H}_m^0 = \sum_{i=1}^{m} \langle H^0, \gamma_{\mu_i} \rangle \gamma_{\mu_i}.$$

• *m-term approximation to* $\mathscr{I}^0$. Now we use dual curvelets to approximate $\mathscr{I}^0$. Take the coefficients $\langle \mathscr{I}^0, \phi_\mu^- \rangle \kappa_s$ arranged in decreasing magnitude order; and define

$$\tilde{\mathscr{I}}_m^0 = \sum_{i=1}^{m} \langle \mathscr{I}^0, \phi_{\mu_i}^- \rangle \phi_{\mu_i}^+.$$

• *Correspondence.* We note that $\tilde{\mathscr{I}}_m^0 = \Delta \tilde{H}_m^0$; the point is simply that the dual curvelets were constructed so that (1) the same terms get selected as the m terms in the approximation

$$\langle \Delta f, \phi_\mu^- \rangle \kappa_s = \langle f, \gamma_\mu \rangle$$

and (2) the corresponding terms agree,

$$\langle \Delta f, \phi_\mu^- \rangle \phi_\mu^+ = \langle f, \gamma_\mu \rangle \Delta \gamma_\mu,$$

giving term-by-term equality of an m-term sum.

• *Isometry.* We finally have

$$\begin{aligned}
\|H^0 - \tilde{H}_m^0\|_2 &= \sup\{\langle H^0 - \tilde{H}_m^0, f \rangle : \|f\|_2 \leqslant 1\} \\
&= \sup\{\langle H^0 - \tilde{H}_m^0, \Delta F \rangle : \|\Delta F\|_2 \leqslant 1\} \\
&= \sup\{\langle \Delta(H^0 - \tilde{H}_m^0), F \rangle : \|\Delta F\|_2 \leqslant 1\} \\
&= \sup\{\langle \mathscr{I}^0 - \tilde{\mathscr{I}}_m^0), F \rangle : \|\Delta F\|_2 \leqslant 1\} \\
&= Err(\mathscr{I}^0, \tilde{\mathscr{I}}_m^0). \qquad (40)
\end{aligned}$$

To the same extent that we can approximate H^0 by curvelets, we can approximate the perturbed curvilinear integral $\mathscr{I}^0$ by operator-biorthogonal curvelets, with the same number of terms and the same error.

• *$m+1$-term approximation to $\mathscr{I}$.* Now we translate (40) into a result about approximating $\mathscr{I}$ itself. Consider the $m+1$-term approximant

$$\tilde{\mathscr{I}}_{m+1} = g + \sum_{i=1}^{m} \langle \mathscr{I}^0, \phi^-_{\mu_i} \rangle \, \phi^+_{\mu_i}. \tag{41}$$

Now obviously

$$Err(\mathscr{I}, \tilde{\mathscr{I}}^0_m) = Err(\mathscr{I}, \tilde{\mathscr{I}}_{m+1}).$$

Hence, the $m+1$-term approximation to $\mathscr{I}$ has an error identical to the m term approximation to $\mathscr{I}^0$.

To make this approximation more familiar, note that at fine scales

$$\langle \mathscr{I}^0, \phi^-_\mu \rangle = \langle \mathscr{I}, \phi^-_\mu \rangle, \qquad \forall s > 0.$$

Hence, if we define coefficients

$$a_\mu = \begin{cases} \langle \mathscr{I}, \phi^-_\mu \rangle, & s > 0 \\ \langle \mathscr{I} - g, \phi^-_\mu \rangle, & s \leqslant 0, \end{cases}$$

the approximation scheme (41) $\tilde{\mathscr{I}}_{m+1}$ has the form

$$\tilde{\mathscr{I}}_{m+1} = g + \sum_{i=1}^{m} a_{\mu_i} \phi^+_{\mu_i}.$$

It involves an m-term approximation where coefficients at coarse scales are lightly modified. To summarize,

$$\|H^0 - \tilde{H}^0_m\|_2 = Err(\mathscr{I}, \tilde{\mathscr{I}}_{m+1}). \tag{42}$$

To the same extent that we can approximate the object H^0 by curvelets, we can approximate the perturbed curvilinear integral $\mathscr{I}$ by operator-biorthogonal curvelets, with the same number of terms and the same error.

We conclude that *whatever the rate of L^2-approximation to H^0 using curvelets, we have the same rate of Err- approximation to $\mathscr{I}$ by dual curvelets*— provided we expand our approximation scheme to have the form (41).

The study of L^2 approximation to H^0 using curvelets is not covered by the existing analysis in [6], which focused on objects with discontinuities along curves. For a general curve $\mathscr{C}$, H^0 is actually continuous, and is in fact Lipschitz. The object H^0 has a singularity along $\mathscr{C}$, but it is not a simple discontinuity; instead ∇H^0 has a discontinuity across $\mathscr{C}$. It turns out that the techniques of [6] can yield results in this setting.

8.2. *An Example: The Circle*

Let our curve traverse the unit circle: $\mathscr{C}(t) = (\cos(t), \sin(t))$ defined on $t \in [0, 2\pi)$. Then

$$H(x) = \log^+ \|x\|,$$

and the gradient is defined everywhere off the unit circle, with

$$\nabla H(x) = \nabla \log^+ \|x\| = \begin{cases} 0 & \|x\| < 1, \\ x/\|x\|^2 & \|x\| > 1. \end{cases}$$

Let now $H_j(x) = \frac{\partial}{\partial x_j} H(x)$ denote the jth component of ∇H, for $j = 1, 2$. There exists a well-defined function $\tilde{H}_j$ which is $C^\infty(\mathbf{R}^2)$ and such that

$$H_j(x) = 1_{\{\|x\| > 1\}} \cdot \tilde{H}_j(x).$$

In short, *component H_j has a representation as a C^∞ function which has been mutilated by multiplication by the indicator of a disc.*

Windowed versions of such mutilated objects were studied in the article [6], and the techniques developed there give immediately information about the sparsity properties of the curvelet coefficients of windowed versions of H_j. These in turn can be used to infer sparsity properties of the curvelet coefficients of H^0.

The proof of Theorem 5.1 actually yields the following conclusion.

THEOREM 8.1. *Let B be a region with C^2 boundary curve. Let $f = 1_{B^c} \cdot \phi$ where ϕ is a C^2 function of compact support. Fix $p > 2/3$. The curvelet coefficients at scale $s > 0$ obey the inequality*

$$\left(\sum_{M_s} |\alpha_\mu|^p \right)^{1/p} \leqslant C_p, \qquad \forall s > 0.$$

We now make a remark about the similar sparsity properties of curvelets and dual curvelets expansions. It essentially follows from the fact that smooth Fourier multipliers transform a curvelet frame element—i.e. a curvelet-atom—into an object with sparse curvelet coefficients—i.e. a curvelet-molecule. We omit the proof.

THEOREM 8.2. *Suppose that $\vec{v}$ is an irrotational vector field with L^2 components of compact support. Suppose we expand each component v_j in a curvelet expansion, and the curvelet coefficients $A_{\mu,j}$ at scale $s > 0$ obey the inequality*

$$\left(\sum_{M_s} |A_{\mu,j}|^p\right)^{1/p} \leqslant C_p, \qquad \forall s > 0.$$

Then the vector curvelet coefficients

$$\alpha_\mu = [\vec{w}_\mu^-, \vec{v}]$$

obey also the inequality

$$\left(\sum_{M_s} |\alpha_\mu|^p\right)^{1/p} \leqslant C'_p, \qquad \forall s > 0.$$

Now we apply these results on sparsity of representation of ∇H^0 to infer properties of H. We use for our tool the biorthogonal expansion of ∇H^0. With all definitions as in Section 3 above:

$$\kappa_s \cdot [\nabla H^0, \vec{w}_\mu^-] = \langle H^0, \gamma_\mu \rangle, \qquad \mu \in M'.$$

It follows from this identity that summability properties of the vector curvelet analysis of ∇H^0 give summability properties of the curvelet coefficients of H. Combining this with Theorem 8.2 gives

COROLLARY 8.1. *Let $H(x) = \log^+(\|x\|)$ as above. The curvelet coefficients of H^0 at scale $s > 0$ obey the inequality*

$$\left(\sum_{M_s} |2^{2s} \cdot \alpha_\mu|^p\right)^{1/p} \leqslant C_p, \qquad \forall s > 0.$$

From this, and some elementary analysis relating ℓ^p norms to m-term approximations, we get the following:

COROLLARY 8.2. *Let $H(x) = \log^+(\|x\|)$ as above. For $m > 0$ define the m-term approximation to H^0 as follows. Let $\mu_1, \ldots, \mu_m$ denote the indices of the m largest-amplitude curvelet coefficients of H^0. Define*

$$\tilde{H}_m^0 = \sum_{i=1}^{m} \langle H^0, \gamma_{\mu_i} \rangle \, \gamma_{\mu_i}.$$

Then for each $\delta > 0$,

$$\|H^0 - \tilde{H}^0_m\|_2 \leqslant C_\delta \cdot m^{-3+\delta}, \qquad m \geqslant 1.$$

COROLLARY 8.3. *Let* $\mathscr{I}$ *be as above. For* $m > 0$ *define the* $m+1$*-term approximation to* $\mathscr{I}$ *as above. Then for each* $\delta > 0$,

$$Err(\mathscr{I}, \tilde{\mathscr{I}}_{m+1}) \leqslant C_\delta \cdot m^{-3+\delta}, \qquad m \geqslant 1.$$

8.3. *General Case*

Our analysis of the case $H(x) = \log^+(\|x\|)$, immediately suggests the following:

Conjecture. For a typical C^2 curve bounding a nice region B, the rate of best m-term curvelet approximation to H^0 is $O(m^{-3+\delta})$ for each $\delta > 0$.

The key point is to use potential theory to observe that the case of a circle should be rather typical.

If we consider ∇H when the potential H derives from a general C^2 curve, we are studying what is well known in potential theory as the "gradient of a single-layer potential".

Existing literature of single-layer potentials shows that quite generically, the component of the gradient normal to the curve $\mathscr{C}$ will have a step discontinuity across the curve $\mathscr{C}$. Consider for example, Coifman and Meyer, [15], Chapter 12, "Potential Theory in Lipschitz Domains". Away from the discontinuity, the gradient will obey uniform C^2 smoothess estimates; so each component of the gradient exhibits qualitatively the same properties which we used for the case of the circle. A rigorous proof of our conjecture would of course require a formalization of the above observations into suitable estimates.

In short, results for approximation of curvilinear integrals follow immediately from bounds for approximation of objects with singularities along C^2 curves, when the singularity across the curve $\mathscr{C}$ is not a discontinuity but instead has a discontinuous gradient.

8.4. *Comparison to Fourier and Wavelets*

We can compare the above result to rates for Fourier and Wavelet approximation.

For Wavelet approximation, note that for compactly supported wavelets, one can construct wavelet Riesz Bases for the Laplacian in the obvious way, and then use them to build isometrics between H and $\mathscr{I}$ using wavelets rather than curvelets. One then builds $m+1$-term dual wavelet approximations $\tilde{\mathscr{I}}^W_{m+1}$ to $\mathscr{I}$ by wavelets, which correspond in a natural way to orthonormal wavelet approximations to H^0. Proceeding in this fashion, (or, in fact, by direct calculation) we get the rate result

$$Err(\mathscr{I}, \tilde{\mathscr{I}}^W_m) \asymp m^{-3/2}. \tag{43}$$

The main calculation underlying this result yields the following conclusion. There are order 2^s wavelet coefficients of object H^0 at scale 2^{-s} which correspond to spatial locations "on the singularity", and these wavelet coefficients are of size $\approx 2^{-2s}$. It follows that, in general, there are $O(N)$ coefficients of size $\geqslant c \cdot N^{-2}$. Now as the wavelets are orthonormal, the best m-term approximation to H^0 is built from the m terms having the m largest amplitudes. It follows that

$$\|H^0 - \tilde{H}^{0,W}_m\|_2 \asymp m^{-3/2}.$$

This leads to (43).

Our isometry approach does not work well with the Fourier series basis for $L^2[-\pi, \pi)^2$, as H^0 is not supported in $[-\pi, \pi)^2$. However, direct calculations give the following result for a certain $m+1$-term dual Fourier approximation $\tilde{\mathscr{I}}^F_{m+1}$ to $\mathscr{I}$, analogous to (41):

$$Err(\mathscr{I}, \tilde{\mathscr{I}}^F_m) \asymp m^{-3/4}. \tag{44}$$

We observe again the pattern that Curvelet decompositions achieve roughly twice the approximation rate of wavelet decompositions and roughly four times the rate of Fourier decompositions.

REFERENCES

1. L. V. Ahlfors, "Complex Analysis," McGraw–Hill, New York, 1966.
2. E. J. Candès, Harmonic analysis of neural networks, *Appl. Comput. Harmon. Anal.* **6** (1999), 197–218.
3. E. J. Candès, "Ridgelets: Theory and Applications," Ph.D. thesis, Department of Statistics, Stanford University, 1998.
4. E. J. Candès, Monoscale ridgelets for image compression and denoising, manuscript, 1999.
5. E. J. Candès and D. L. Donoho, Ridgelets: the key to high-dimensional intermittency?, *Philos. Trans. Roy. Soc. London* **357** (1999), 2495–2509.
6. E. J. Candès and D. L. Donoho, Curvelets, manuscript, 1999, available at `http://www-stat.stanford.edu/~donoho/Reports/1999/curvelets.pdf`.
7. E. J. Candès and D. L. Donoho, "Recovering Edges in Ill-Posed Inverse Problems: Optimality of Curvelet Frames," Technical Report 2000-16, Department of Statistics, Stanford University, 2000.

8. E. J. Candès and D. L. Donoho, Curvelets: a surprisingly effective nonadaptive representation of objects with edges, *in* "Curve and Surface Fitting: Saint-Malo 1999" (A. Cohen, C. Rabut, and L. L. Schumaker, Eds.), Vanderbilt University Press, Nashville, TN, 2000.
9. D. L. Donoho, Orthonormal ridgelets and linear singularities, *SIAM J. Math. Anal.*, to appear.
10. M. Frazier and B. Jawerth, Decomposition of Besov spaces, *Indiana Univ. Math J.* **34** (1985), 777–799.
11. M. Frazier and B. Jawerth, A discrete transform and decomposition of distribution spaces, *J. Func. Anal.* **93** (1990), 34–170.
12. M. Frazier, B. Jawerth, and G. Weiss, "Littlewood–Paley Theory and the Study of Function Spaces," NSF-CBMS Regional Conf. Series in Mathematics, Vol. 79, American Math. Soc., Providence, RI, 1991.
13. P. G. Lemarié and Y. Meyer, Ondelettes et bases hilbertiennes, *Rev. Mat. Ibero-americana* **2** (1986), 1–18.
14. Y. Meyer, "Ondelettes et Opérateurs: I. Ondelettes," Hermann, Paris, 1990.
15. Y. Meyer and R. R. Coifman, "Ondelettes et Opérateurs: III. Opérateurs multilinéaires," Hermann, Paris, 1991.
16. E. M. Stein, "Singular Integrals and Differentiability Properties of Functions," Princeton University Press, Princeton, NY, 1970.
17. M. Vetterli and J. Kovacevic, "Wavelets and Subband Coding," Prentice–Hall, Englewood Cliffs, NJ, 1995.