New Multiscale Transforms, Minimum Total Variation Synthesis: Applications to Edge-Preserving Image Reconstruction

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

December 2001

Abstract

This paper describes newly invented multiscale transforms known under the name of the ridgelet [6] and the curvelet transforms [9, 8]. These systems combine ideas of multiscale analysis and geometry. Inspired by some recent work on digital Radon transforms [1], we then present very effective and accurate numerical implementations with computational complexities of at most $N \log N$.

In the second part of the paper, we propose to combine these new expansions with the Total Variation minimization principle for the reconstruction of an object whose curvelet coefficients are known only approximately: quantized, thresholded, noisy coefficients, etc. We set up a convex optimization problem and seek a reconstruction that has minimum Total Variation under the constraint that its coefficients do not exhibit a large discrepancy from the the data available on the coefficients of the unknown object.

We will present a series of numerical experiments which clearly demonstrate the remarkable potential of this new methodology for image compression, image reconstruction and image 'de-noising.'

Keywords. Ridgelets, curvelets, wavelets, pseudo-polar FFT, Slant Stack, Radon transform, trigonometric interpolation, total-variation, edges, edge-preserving tomography/deconvolution, thresholding.

Acknowledgments. We would like to acknowledge fruitful conversations with David L. Donoho. This work was partially supported by an Alfred P. Sloan Fellowship.

1 Introduction

1.1 The Multiscale/Multiresolution Revolution

The title of this special issue is "Image and Video Coding Beyond Standards." Claiming that the 'standards,' at least in image coding, have evolved over the last few years is quite an understatement. In fact, the last two decades or so witnessed an extraordinary revolution in the way we process, store, visualize and transmit digital data; a revolution one might call *the triumph of the multiscale approach.*

The pioneering work of Burt and Adelson [5] on Laplacian pyramids, of Galand [19] on quadrature mirror filters, and of Grossman and Morlet [20] paved the way to the powerful concept of multiresolution, which today is –perhaps– best known under the generic name of wavelets. Multiresolution provides us with tools to describe mathematical objects such as functions, signals and more generally datasets at different level of resolutions. The celebrated orthogonal wavelet pyramids of Mallat and Meyer [24] and Daubechies' construction of compactly supported wavelets [14] are among the greatest accomplishments of this generation of researchers who moved the field away from classical Fourier analysis.

Today, multiscale/multiresolution ideas permeate many fields of contemporary science and technology. In signal processing, they led to convenient tools to navigate through large datasets, to transmit compressed data rapidly, to remove noise from signals and images, and to identify crucial transient features in such datasets. By now, wavelets have made their way into our livelihood as they were included in JPEG 2000, the new still-picture compression standard, and are being considered for future generations of standard video coders such as the 'MPEG family.'

1.2 New Multiscale Geometric Transforms

In this paper, we will describe some ideas which take the multiscale approach in a whole new direction, perhaps justifying the title "Beyond Standards." Indeed, Candès and Donoho [10, 9, 7] argue that the concept of multiscale is much broader than what classical multiresolution ideas or wavelets imply. They developed new systems of representation which are very different from wavelet-like systems. These systems combine ideas of multiscale analysis with ideas of geometric features and structures.

For instance, a newly invented representation is the ridgelet system [6] which allows the representation of arbitrary bivariate functions by superpositions of elements of the form

$$a^{-1/2}\psi((x_1\cos\theta + x_2\sin\theta - b)/a); \tag{1.1}$$

ψ is an oscillatory univariate function (a one one-dimensional wavelet if you wish), $a > 0$ is a scale parameter, θ is an orientation parameter, and b is a location scalar parameter. Thus, ridgelets occur at all scales, locations, and orientations. Unlike wavelets, ridgelets are non-local as they are infinitely elongated in the codirection θ. In the context of image processing where one would restrict objects to 'live' on the unit square, ridgelet functions are roughly of unit length and have arbitrary width, so that one would have all possible aspect ratios. If we think of wavelets as 'fat' points, then we may view ridgelets as 'fat' lines.

Motivated by the problem of finding efficient representation of objects with discontinuities along curves and of compression of image data, Candès and Donoho introduced yet another representation system, the curvelet transform. Like ridgelets, curvelets occur at all scales, locations, and orientations. However, whereas ridgelets have unit length, curvelets

occur at all dyadic lengths and exhibit an anisotropy increasing with decreasing scale like a power law. Section 2 will introduce the curvelet transform. In short, curvelet 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. [8, 9] build tight frames of curvelets (γ_μ) obeying

$$f = \sum_\mu \langle f, \gamma_\mu \rangle \gamma_\mu, \tag{1.2}$$

and

$$||f||^2_{L_2(\mathbb{R}^2)} = \sum_\mu |\langle f, \gamma_\mu \rangle|^2. \tag{1.3}$$

1.3 Curvelets and Image Coding

New transforms may be very significant for practical concerns. For instance, the potential for sparsity of wavelet expansions led the way to very successful applications in areas such as signal/image compression or denoising and feature extraction/recognition. The numerical examples presented in this paper will make clear that the curvelet transform is very relevant for image processing applications. Beyond those examples, there is now substantial amount of evidence supporting our claim:

- *Sparse Representations by Curvelets.* The curvelet representation is far more effective for representing objects with edges than wavelets or more traditional representations. In fact, [8] proves that curvelets provide optimally sparse representations of C^2 objects with C^2 edges. In general, improved sparsity leads to improved compression performance, at least at high compression rates. Hence, transform coders based on the quantization of curvelet coefficients may benefit from provably superior asymptotic properties.

- *Sparse Component Analysis.* In computer vision, there has been an interesting series of experiments whose aim is to describe the 'sparse components' of images. Of special interest is the work of Field and Olshausen [27] who set up a computer experiments for *empirically* discovering the basis that best represents a database of 16 by 16 image patches. Although this experiment is limited in scale, they discovered that the best basis is a collection of needle shaped filters occurring at various scales, locations and orientations. The reader will find a stimulating discussion about the connection between data analysis and harmonic analysis in [17]. We will not detail this connection and simply wonder, with Donoho, at the striking resemblance between curvelets, which derive from mathematical analysis, and these empirical basis elements arising from data analysis.

- *Numerical Experiments.* Donoho and Huo [21] study sparse decompositions of images in a dictionary of waveforms including wavelet bases and multiscale ridgelets. They apply the Basis Pursuit (BP) in this setting and obtain sparse syntheses by solving

$$\min \|a\|_{\ell_1} \quad \text{subject to } f = \sum_m a_m \varphi_m,$$

  where (φ_m) is our collection of waveforms. BP gives an 'equal' chance to every member of the dictionary and yet, Donoho and Huo observe that BP preferably

selects multiscale ridgelets over wavelets, except for possibly the very coarse scales and the finest scale. This experiment seems to indicate that multiscale ridgelets are better for representing image data than pre-existing mathematical representations. There is a natural connection with our previous point which showed that empirically selected bases do not look like wavelets.

1.4 New Partial Synthesis Rules: Total-Variation Minimization

Let T be a linear transformation such as the Fourier, wavelet, ridgelet or curvelet transforms, for example, or the amalgamation of a few of these. Suppose that one is given partial information about f in the sense that only a subset of the coefficients Tf are known or approximately known. To make things concrete, we might consider a 'compression scenario' where the partial information consists in the value of the m-largest coefficients, possibly subject to quantization error. The classical approach to partial synthesis consists in setting the other coefficients to zero and inverting the transform T.

There are, of course, more sophisticated approaches. For instance, in the case where T is the wavelet transform, [4] develops a wavelet embedded predictive image coder; the idea is to use statistical properties of natural scenes to predict the absolute value of neighboring wavelet coefficients. In a somewhat similar spirit, [13] attempts to model the joint probability density of the wavelet coefficients using Hidden Markov Models and exploit this model for efficient partial synthesis.

In this paper, we introduce a very different idea based on the minimization of a complexity penalty. Of special interest is the Total Variation norm $\|f\|_{BV}$, which is roughly equal to the integral of the Euclidean norm of the gradient. We introduce some notations. We let $(Tf)_{\mu\in M}$ be the coefficient sequence of an object f and let M' denote a subset of the coefficient index M. Suppose we are given possibly approximate values $\tilde{\theta}_\mu$ of $(Tf)_\mu$, for $\mu \in M'$. We propose the following rule for partial synthesis:

$$\min \|g\|_{BV} \quad \text{subject to} \quad |\theta_\mu(g) - \tilde{\theta}_\mu| \le e_\mu,\ \mu \in M'. \tag{1.4}$$

In a nutshell, given a vector of tolerance (e_μ) – some of the e_μ's may be zero – we seek a solution f^* with minimum Total Variation norm whose coefficients $\theta_\mu(g)$ are within e_μ of θ_μ. In passing, note that equation (1.4) may be seen as a very interesting special case of the more general rule

$$\min J(g), \quad \text{subject to} \quad Tg \in \mathcal{C}_f, \tag{1.5}$$

where $J(g)$ is a functional measuring the complexity of the fit, e.g. $\|g\|_{BV}$, and $\mathcal{C}_f$ is a convex set of constraints depending on available information about f, e.g. $|Tg - \tilde{\theta}| \le e$.

Reconstruction based on (1.4) were briefly suggested in [29] and have been independently introduced by Durand and Froment [18] in the context of wavelet-based signal denoising. This paper will deploy (1.4) in our newly constructed geometric multiscale systems, namely ridgelets and curvelets. We will present a series of preliminary experiments which will clearly demonstrate the potential for image compression, image reconstruction and 'de-noising.'

1.5 Contents

Section 2 below will review the basic components of both the ridgelet and curvelet constructions. In Section 3, we will present a strategy for developing approximate digital transforms which can be used effectively on n by n Cartesian arrays. Section 4 will discuss new partial synthesis rules introduced in Section 1.4. In Section 5, we will present applications in the

area of image reconstruction and specifically edge-preserving tomography. Finally, we will close with some remarks and future directions for research.

2 The Ridgelet Transform

There are several ridgelet transforms. For instance, Candès introduced both a continuous and a discrete ridgelet transform. Letting $\psi_{a,\theta,b}(x)$ being a ridgelet as in (1.1), he defines a ridgelet coefficient by

$$\mathcal{R}f(a,\theta,b) = \int f(x_1,x_2)\psi_{a,\theta,b}(x_1,x_2)\,dx_1dx_2.$$

Then provided that ψ is oscillatory, [6] shows that we have a stable and concrete reproducing formula just as we have a stable and concrete reproducing formula for the wavelet transform. He then develops a discretization of the continuous transform and develops ridgelet frames. We give an overview of the main components of the discrete transform.

- Dyadic scale: $a = 2^{-j}$, $j \geq 0$.
- Dyadic location: $b = k \cdot 2^{-j}$, $k \in \mathbb{Z}$.
- Angular resolution proportional to scale: $\theta = 2\pi\delta \cdot \ell \cdot 2^{-j}$.

In an image processing context, this work suggests taking line-like measurements on image data and shows how to combine these measurements to reconstruct an image.

In dimension 2, Donoho [16] introduced a new orthonormal basis whose elements he called 'orthonormal ridgelets.' We quote from [10]: "Such a system can be defined as follows: let $(\psi_{j,k}(t) : j \in \mathbb{Z}, k \in \mathbb{Z})$ be an orthonormal basis of Meyer wavelets for $L^2(\mathbb{R})$ [22], 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

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}}(\hat{\psi_{j,k}}(|\xi|)w^\varepsilon_{i,\ell}(\theta) + \hat{\psi_{j,k}}(-|\xi|)w^\varepsilon_{i,\ell}(\theta+\pi))/2\ . \tag{2.1}$$

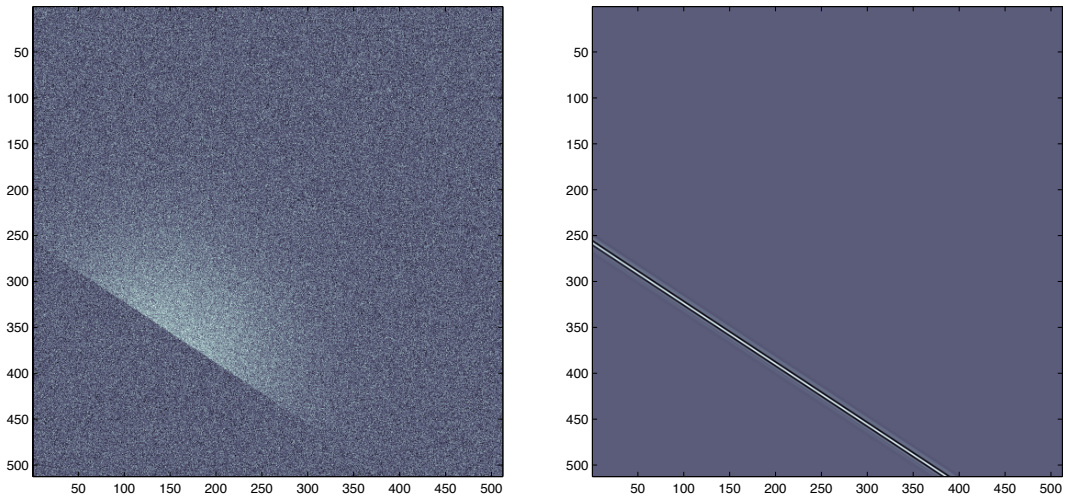
Here the indices run as follows: $j,k \in \mathbb{Z}$, $\ell = 0,\ldots,2^{i-1}-1$; $i \geq i_0$, $i \geq j$. Notice the restrictions on the range of ℓ and on i. Let λ denote the set of all such indices λ. It turns out that $(\rho_\lambda)_{\lambda\in\Lambda}$ is a complete orthonormal system for $L^2(\mathbb{R}^2)$."

Figure 1 represents an object and that ridgelet which at a given scale, best correlate with this object. Observe that the ridgelet is aligned with the discontinuity.

3 The Curvelet Transform

We now briefly discuss the curvelet frame; for more details, see [8]. The construction combines several ingredients, which we briefly review

- *Ridgelets*, a method of analysis very suitable for objects which are discontinuous across straight lines.



(a) Noisy Object

(b) Ridgelet

Figure 1: A Ridgelet

- *Multiscale Ridgelets*, a pyramid of analyzing elements which consists of ridgelets renormalized and transported to a wide range of scales and locations.
- *Bandpass Filtering*, a method of separating an object out into a series of disjoint scales.

The last section introduced ridgelets and we first briefly discuss multiscale ridgelets and bandpass filtering. We then describe the combination of these three components. There is a difference between this construction and the one given in [8] at large scales.

3.1 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 $w(x_1, x_2) \geq 0$, $w \in C_0^\infty([-1,1]^2)$ 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 $\mathcal{Q}_f$ denote the dyadic squares of side 2^{-s}, we can define the subcollection of *Monoscale Ridgelets* at scale s:

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

It is immediate from the orthonormality of the ridgelets that each system of monoscale ridgelets makes 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

$$\mathcal{M} = \cup_{s \geq 1} \mathcal{M}_s$$

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

$$\sum_{\mu \in \mathcal{M}} \langle \psi_\mu, f \rangle^2 = \infty. \tag{3.1}$$

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 (3.1).

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 $width \approx length^2$.

3.2 Subband Filtering

Our remedy to the 'energy blow-up' (3.1) is to decompose f into subbands using standard filterbank ideas. Then we assign one specific monoscale dictionary $\mathcal{M}_f$ to analyze one specific (and specially chosen) subband.

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

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

The construction of such operators is standard [32]; 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 $|\xi| \leq 1$; Ψ_{2s} is bandpass, concentrated near $|\xi| \in [2^{2s}, 2^{2s+2}]$; and we have

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

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

3.3 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, e) where $s \leq 0$, $e \in \{0, 1\}$, indexing all dyadic squares in the plane of side $2^s > 1$.

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

At *coarse scales* we have wavelet coefficients.

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

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

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

Note well that for $s > 0$, each coefficient associated to scale 2^{-s} derives from the subband filtered version of f – $\Delta_s f$ – 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 $\gamma_\mu \in L^2(\mathbb{R}^2)$ 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 $s \leq 0$, $\gamma_\mu = P_0 \phi_{s,k_1,k_2}$, while for $s > 0$,
$$\gamma_\mu = \Delta_s \psi_\mu, \qquad \mu \in \mathcal{Q}_s. \tag{3.2}$$
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 Δ_s); (2) the nonnegligible curvelets obey $length \approx 2^{-s}$ while $width \approx 2^{-2s}$. In short, the system obeys approximately the scaling relationship
$$width \approx length^2.$$
Note: it is at this last step that our 2^{2s} coronization scheme comes fully into play.

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

4 Digital Transforms

4.1 Ridgelet and Radon Transforms

There is an intimate relationship between the Radon transform and the ridgelet transform. In some sense, ridgelet analysis is a kind of wavelet analysis in the Radon domain.

Let $\psi_{a,\theta,b}$ be a ridgelet

$$\psi_{a,\theta,b}(x_1, x_2) = a^{-1/2}\psi\left(-x_1 \sin\theta + x_2 \cos\theta - b)/a\right) \tag{4.1}$$

and define a ridgelet coefficient by

$$\mathcal{R}f(a, \theta, b) = \int f(x_1, x_2)\psi_{a,\theta,b}(x_1, x_2)\, dx_1 dx_2. \tag{4.2}$$

A simple change of variables shows that

$$\mathcal{R}f(a, \theta, b) = \int Rf(t, \theta)a^{-1/2}\psi((t - b)/a)\, dt,$$

where Rf is the Radon transform [15] defined by

$$Rf(t, \theta) = \int f(x_1, x_2)\delta(-x_1 \sin\theta + x_2 \cos\theta - t)\, dx_1 dx_2. \tag{4.3}$$

In short, the ridgelet transform is precisely the application of a 1-dimensional wavelet transform to the slices of the Radon transform where the angular variable θ is constant and t is varying.

A similar connection exists with the orthonormal ridgelet transform as well. Letting Δ^+ be the fractional derivative operator

$$\Delta^+ h = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\omega|^{1/2}\hat{h}(\omega)e^{i\omega t}\, d\omega, \tag{4.4}$$

we introduce the fractionally differentiated Radon transform defined as

$$\tilde{R} = (\Delta^+ \otimes I)R, \tag{4.5}$$

where R is the Radon transform (4.3), and with Δ^+ acting on t and the identity I on θ. It is well-known [15] that $\tilde{R}$ is an isometry

$$[\tilde{R}f, \tilde{R}g] = \langle f, g\rangle, \tag{4.6}$$

where the notation $[\cdot, \cdot]$ denote the normalized inner product in $L_2(dt\, d\theta)$

$$[F, G] = \frac{1}{4\pi}\int F(t, \theta)\overline{G(t, \theta)}\, dtd\theta.$$

Let (ρ_λ) be an orthobasis of ridgelets (2.1) and W_λ be the tensor wavelet basis

$$W_\lambda(t, \theta) = (\psi_{j,k} \otimes w^\epsilon_{i,\ell})(t, \theta), \quad \lambda = (j, k; i, \ell, \epsilon), \tag{4.7}$$

where the range of λ is as in (2.1). With these notations, it is an easy calculation to show that

$$\tilde{R}\rho_\lambda = P_{\mathcal{R}}W_\lambda,$$

where $P_{\mathcal{R}}W_\lambda$ is the antipodally-symmetrized version of W_λ

$$(P_{\mathcal{R}}W_\lambda)(t, \theta) = \left(W_\lambda(t, \theta) + W_\lambda(-t, \theta + \pi)\right)/2.$$

Then

$$\theta_\lambda := \langle f, \rho_\lambda\rangle = [\tilde{R}f, P_{\mathcal{R}}W_\lambda] \tag{4.8}$$

In addition, we may pass the operator $(\Delta^+ \otimes I)$ to $P_{\mathcal{R}} W_\lambda$ and obtain

$$\theta_\lambda = [Rf, (\Delta^{1/4} \otimes I) P_{\mathcal{R}} W_\lambda].$$

This gives

$$\theta_\lambda = [Rf, (\Delta^{1/4} \otimes I) P_{\mathcal{R}} W_\lambda] = [Rf, P_{\mathcal{R}} W_\lambda^+], \tag{4.9}$$

with

$$W_\lambda^+ = \psi_{j,k}^+ \otimes w_{i,\ell}^\epsilon, \quad \psi_{j,k}^+ = \Delta^+ \psi_{j,k}. \tag{4.10}$$

The identities (4.8)– (4.9) also justify our interpretation of ridgelet analysis as some kind wavelet analysis in the Radon domain.

4.2 Strategy

Following upon the previous section, a sensible strategy for a digital ridgelet transform would involve the development of digital wavelet and Radon transforms. Once, we have a digital ridgelet transform, digital curvelet transforms follow rather easily, see [30]. Digital wavelet transforms are now routine. However, developing a *highly accurate* and fast digital Radon transform is challenging as making sense of geometrical concepts such as line integrals on discrete Cartesian arrays is not straightforward. Indeed, these difficulties together with the importance of the Radon transform in many problems of scientific interest spurred an extensive literature on this subject.

Many of the proposed methods rely on the Projection Slice Theorem which states that the Radon transform may be obtained by inverting the 2-dimensional Fourier transform along radial lines

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

Continuing on this line of thinking, one would need to introduce a collection Ξ of 'polar samples' and evaluate the Fourier transform

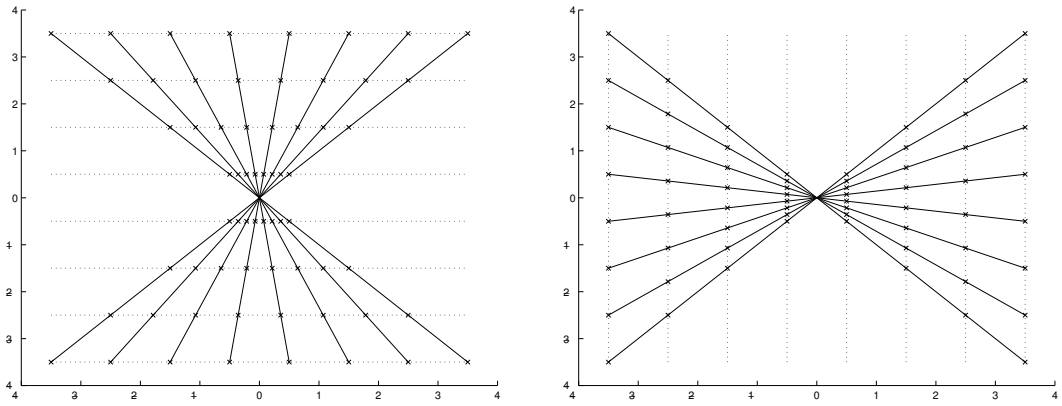
$$F(\xi) = \sum_{n_1, n_2} f(n_1, n_2) e^{-i(\xi_1 n_1 + \xi_2 n_2)}$$

at each point $\xi \in \Xi$. For a polar grid of size N, a naive evaluation would require of the order of N^2 operations which is prohibitive for large values of N. Another issue is that image data are often given a Cartesian grids and classical polar grids may not be well-adapted to this data structure.

In this paper, we follow a recent approach developed in [1] which goes by the name of *Pseudo-Polar FFT*. The Pseudo-Polar FFT defines a Rectopolar grid which is adapted to Cartesian grids and introduces an interpolation scheme allowing a fast evaluation of the Fourier transform on that grid.

4.3 Pseudo-Polar FFT

The approach [1] starts with a Rectopolar grid also known under the name of the *Concentric Squares* grid. Given a Cartesian array, (k_1, k_2), $-n/2 \le k_1, k_2 < n/2$, we define $2n$ radial lines joining pairs of symmetric points from the boundary of the square. Figure 2 shows radial lines in Rectopolar coordinates; we will refer to these as Rectopolar lines and it will be convenient to distinguish between basically vertical and horizontal lines as illustrated on our Figure. We then introduce the Rectopolar grid $\xi_{\ell,m}^p$ where ℓ indexes Rectopolar lines, m the position on that line, and p is a gender token equal to 1 if the Rectopolar line is basically horizontal, and 2 otherwise. Define



(a) Basically Vertical Lines

(b) Basically Horizontal Lines

Figure 2: Rectopolar Lines

- $\xi^1_{\ell,m}$ as the intersection of the ℓ^{th} 'vertical' Rectopolar line with the m^{th} horizontal line (see 2(a)), and
- $\xi^2_{\ell,m}$ is the intersection of the ℓ^{th} 'horizontal' Rectopolar line with the m^{th} vertical line (see 2(b)).

The meaning of the index ℓ is as follows: the coordinates of the points $\xi^p_{\ell,m}$ are given by

$$\begin{cases} \xi^1_{\ell,m} = (2\ell m/n, m), \\ \xi^2_{\ell,m} = (m, 2\ell m/n), \end{cases} \quad -n/2 \le \ell, m < n/2. \tag{4.11}$$

From now on, we will let Ξ denote the full Rectopolar grid. This grid is pictured on (3).

Obviously, the sampled angular directions are not equispaced; the sampling is a little finer near the 'corners.' To be specific, let $\hat{f}(\xi)$ be the Fourier transform of $f(x)$ and put

$$\hat{f}^p_{\ell,m} = \hat{f}(2\pi\xi^p_{\ell,m}/n).$$

With these notations and for $p = 1$, say, $\hat{f}^1$ are sampled values of $\hat{f}(-\lambda\tan\theta, \lambda)$ on the Cartesian grid

$$\lambda = m,\ \tan\theta = \ell/n, \quad -n/2 \le \ell, m < n/2.$$

A similar statement applies to $\hat{f}^2$. The sampling is irregular in θ but equispaced in the slope $\tan\theta$.

4.4 Interpolation

In this section, we work with Cartesian array $f(n_1, n_2)$, $0 \le n_1, n_2 \le n-1$. The FFT of $(f(n_1, n_2))$ gives Fourier samples $\hat{f}(k_1, k_2)$ on the Cartesian grid $k = (k_1, k_2), -n/2 \le k_1, k_2 < n/2$ and we now need to address the problem of resampling the Fourier transform on the Pseudo-Polar grid Ξ. Interpolating the Fourier transform is a delicate matter as this is a very oscillatory object.

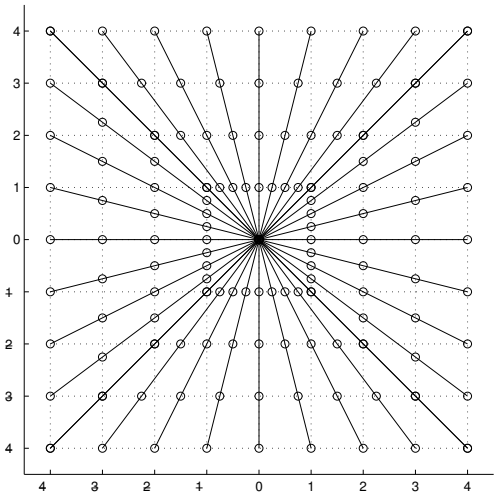


Figure 3: Rectopolar Grid

We view $\hat{f}(k_1, k_2)$ as samples from the trigonometric polynomial F defined by

$$F(\omega_1, \omega_2) = \sum_{n_1=0}^{n-1} \sum_{n_2=0}^{n-1} f(n_1, n_2) e^{-i(\omega_1 n_1 + \omega_2 n_2)}. \tag{4.12}$$

Then

$$\hat{f}(k_1, k_2) = \sum_{n_1=0}^{n-1} \sum_{n_2=0}^{n-1} f(n_1, n_2)\, e^{-i2\pi(k_1 n_1 + k_2 n_2)/n} = F(2\pi k_1/n, 2\pi k_2/n). \tag{4.13}$$

Consider the restriction of this two-dimensional trigonometric polynomial to the line $\omega_2 = 2\pi \cdot m/n$

$$P_m(\omega) := F(\omega, \frac{2\pi m}{n}). \tag{4.14}$$

Of course, P_m is a trigonometric polynomial of degree n which might be expressed as

$$P_m(\omega) = \sum_{u=0}^{n-1} c_u\, e^{-iu\omega}.$$

Now, because P_m is the restriction of F, its coefficients are given by

$$c_{n_1} = \sum_{n_2=0}^{n-1} f(n_1, n_2) e^{-i\frac{2\pi m}{n} n_2}, \quad 0 \le n_1 \le n-1.$$

We then need to evaluate this polynomial at the points

$$\omega_\ell = \alpha \cdot \ell, \quad \alpha = (2\pi/n) \cdot 2m/n, \ -n/2 \le \ell < n/2,$$

see (4.11). This gives

$$\hat{f}^1_{\ell,m} = P_m(\alpha \cdot \ell) = \sum_{u=0}^{n-1} c_u e^{-i\alpha u \ell}.$$

In short, our interpolation calls for the resampling of a one-dimensional trigonometric polynomial at a sampling rate, which is below the Nyquist rate, except at the 'borders' $m = -n/2$.

Introduce the Fractional Fourier Transform (FrFT) of a vector x

$$\hat{x}_k = (\mathcal{F}_\alpha x)_k = \sum_{j=0}^{n-1} x_j e^{-i\alpha\, j\, k}, \quad 0 \le k \le n-1.$$

For fixed p and m, the sample values of $(\hat{f}^1_{\ell,m})_\ell$ may, of course, be interpreted as an FrFT. Fix m, then

$$\hat{f}^1_{k-n/2,m} = (\mathcal{F}_\alpha d)_k, \quad 0 \le k \le n-1,$$

where d is the vector

$$d_u = c_u\, e^{i\alpha u n/2}, \quad 0 \le u \le n-1.$$

Bailey and Swartztrauber [2] have proposed an algorithm to compute the FrFT of a vector of length n in $O(n \log n)$ multiplications and additions. Therefore, for each horizontal line or vertical line m, the evaluation of $\hat{f}^p_{\ell,m}$ requires a number of multiplications and additions of the order $n \log n$. There are a total of $2n$ such lines and, therefore, the complexity of the Pseudo-Polar FFT is $N \log N$ for a Cartesian array of size $N = n^2$.

This approach is that taken in [1] and we reiterate its main property; the Pseudo-Polar FFT gives an exact evaluation of F (4.12) at each $\xi \in \Xi$.

4.5 Interpretation: Slant Stack Transform

Define the Slant Stack transform of an object f as follows:

$$(Sf)(t,\theta) = \int f(u, t + u\tan\theta)\, du, \quad -\pi/4 \le \theta \le \pi/4. \tag{4.15}$$

That is, the Slant Stack is a line integral along $\mathcal{L}_{t,\theta} = \{(u, t + u\tan\theta),\ u \in \mathbb{R}\}$. There is a similar definition for θ in the range $[\pi/4, 3\pi/4]$, namely,

$$(Sf)(t,\theta) = \int f(u\cot\theta + t, u)\, du, \quad \pi/4 \le \theta \le 3\pi/4. \tag{4.16}$$

Not surprisingly there is a relationship between the one dimensional Fourier transform of S and the two-dimensional Fourier transforms of f. Indeed, the following equality holds for $\theta \in [-\pi/4, \pi/4]$

$$(Sf)(t,\theta) = \frac{1}{2\pi} \int\int \hat{f}(-\lambda\tan\theta, \lambda) e^{i\lambda t}\, d\lambda. \tag{4.17}$$

In other words, inverting the one dimensional Fourier transform along the Rectopolar line $(\lambda\tan\theta, \lambda)$, $\lambda \in \mathbb{R}$ gives the Slant Stack. Of course, the Slant Stack transform is also directly related to the Radon transform as

$$Sf(t,\theta) = \cos\theta \cdot Rf(t\cos\theta, \theta).$$

We let T denote the Pseudo-Polar FFT which maps a Cartesian array $(f(n_1, n_2))$ into $\hat{f}^p_{\ell,m} = F(\xi^p_{\ell,m})$ (4.12) and $\mathcal{F}_1^{-1}$ be the one-dimensional inverse Fourier transform applied to each of the $2n$ Rectopolar line. The identity (4.17) suggests that the composition $\mathcal{F}_1^{-1}\, T$ is a discrete analog of the Slant Stack transform for Cartesian arrays. In fact, it is possible to define a discrete Slant Stack S_n transform with this property

$$S_n = \mathcal{F}_1^{-1}\, T. \tag{4.18}$$

4.6 Discrete Slant Stack

Put

$$(S_n f)(u,\theta) = \sum_u \tilde{f}(u, v + u\tan\theta)$$

where for a fixed $u \in \{-n/2,\ldots,n/2-1\}$, the function $\tilde{f}(u,t)$, $t \in \mathbb{R}$ interpolates the discrete vector $f(u,v)$, $v = -n/2,\ldots,n/2-1$. Let u be fixed and let $g(-n/2)$, $g(-n/2+1),\ldots,g(n/2-1)$ be the sample values of $f(u,v)$. We pad this vector with two zero-vectors of size $n/2$ each, thereby defining $g(v) = 0$ for $v \in \{-n,\ldots -n/2-1\}$ and $v \in \{n/2,\ldots n-1\}$. Let $G(t)$, $-n \le t < n$ be that trigonometric polynomial of degree $2n-1$ such that

$$G(v) = g(v), \quad v = -n,\ldots,n-1.$$

Then G is given by

$$G(t) = \sum_k \hat{g}_k e^{i\pi kt/n}$$

or equivalently,

$$G = \sum_v g(v) K_{2n}(t-v)$$

where K_{2n} is the Dirichlet kernel

$$K_{2n}(t) = e^{-i\pi t/2n} \frac{\sin \pi t}{2n \sin(\pi t/2n)}.$$

Then [1] shows that the identity (4.18) holds with this special choice of trigonometric interpolant.

Note that we used zero-padding to avoid periodic effects. Had we ignored this, our trigonometric interpolation would warp lines around the square as it assumes periodicity. This point is illustrated on Figure 4(a).

The padding gives a rectangular array of size n by $2n$ or $2n$ by n depending on whether we consider integrals on lines whose angle θ with the horizontal axis is in the range $\theta \in [-\pi/4, \pi, 4]$ or $\theta \in [\pi/4, 3\pi, 4]$. Define the discrete Slant Stack grid as follows

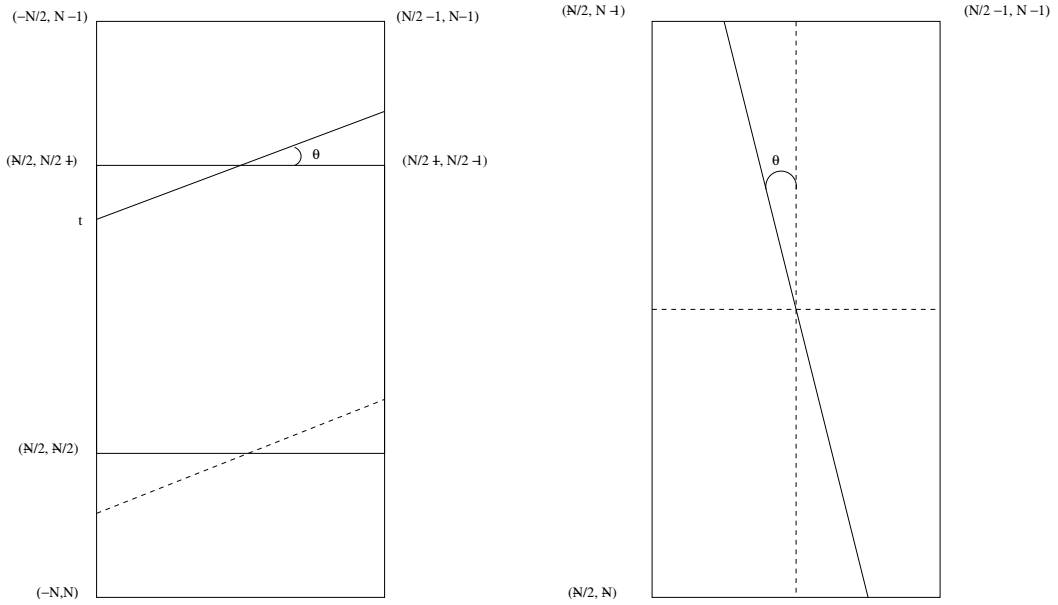
$$t_k = k,\ -n \le k < n, \quad \tan\theta_\ell = \ell/n,\ -n/2 \le \ell < n/2. \tag{4.19}$$

Then letting S_n be the Discrete Slant Stack which maps

$$f(n_1,n_2) \mapsto (S_n f)(t_k, \theta_\ell),$$

[1] proves that S_n obeys (4.18). Figure 4 provides a graphical representation of the identity (4.18): for each θ_ℓ, taking the inverse FFT along the Rectopolar lines (we have $2n$ equispaced points) gives the Slant Stack for $t = -n,\ldots,n-1$. In short, the discrete Slant Stack S_n takes a Cartesian array of size n by n and returns an array of size $2n$ by $2n$ in two batches of size $2n$ by n each.

On Figure 4, we chose to display 'long' rectangles merely reflecting the size of the zero-padded Cartesian arrays. We hope that this does not introduce confusion in the reader's mind as this figure seems to indicate that we only consider slopes in the range $[-1/2, 1/2]$ and $(-\infty,-2) \cup (2,\infty)$. The correct geometrical interpretation of the Pseudo-Polar FFT



(a) Slant Stack (b) Polar FFT

Figure 4: Slant Stack and Polar FFT

is that we oversample the Rectopolar grid by a factor 2, namely, we evaluate $F(\omega_1, \omega_2)$ at the samples

$$\begin{cases} (\omega_1, \omega_2) = \left(\frac{\pi}{n} \cdot \frac{\ell m'}{n}, \frac{\pi}{n} \cdot m'\right), \\ (\omega_1, \omega_2) = \left(\frac{\pi}{n} \cdot m', \frac{\pi}{n} \cdot \frac{\ell m'}{n}\right), \end{cases} \qquad -n/2 \le \ell < n/2,\ -n \le m' < n. \tag{4.20}$$

In comparison, the sampling suggested by Figure 3 is of the type $(\omega_1, \omega_2) = (2\pi/n \cdot \ell m/n, 2\pi n \cdot m)$, where $-n/2 \le \ell, m < n/2$.

4.7 Wavelet Transforms

If one wishes to calculate a table of discrete ridgelet coefficients as in (4.2), we simply apply a discrete wavelet transform to the columns of $S_n f$ which gives the array of coefficients $\alpha = (\alpha^p_{j,k}(\ell))$; this is an array of size $2n$ by $2n$ in two batches of size $2n$ by n each.

The discrete implementation of the orthonormal ridgelet transform proceeds a little differently. With the notations of the last section we compute the 'fractionally' differentiated Slant Stack

$$S^+_n f = \mathcal{F}^{-1} D T f,$$

where D and T are respectively the density weighting and Pseudo-Polar FFT matrices. We then apply discrete wavelet transforms both to the rows (t-variable) and columns (θ-variable) of $S^+_n f$. The result is an array of coefficients $\alpha = (\alpha^p_{j_1,k_1;j_2,k_2})$ with (j_1, k_1) and (j_2, k_2) denoting the scale and location variables of the wavelet transforms in the t-variable and θ-variable respectively.

4.8 Pseudo-Ridgelets

By substituting the Radon transform with the Slant Stack, we slightly changed the definition of the ridgelet transform. The discrete transform computes a discrete analog of the pseudo-ridgelet transform defined as follows:

$$\begin{aligned}\widetilde{\mathcal{R}}f(a,\theta,b) &= \cos\theta \cdot \int f(x_1,x_2)\, a^{-1/2}\psi\left(\frac{-x_1\tan\theta + x_2 - b}{a}\right) dx_1 dx_2 \\ &= \int Sf(t,\theta)a^{-1/2}\psi((t-b)/a)\, dt,\end{aligned}$$

for $\theta \in (-\pi/4, \pi/4)$ and likewise for $\theta \in (\pi/4, 3\pi/4)$. In other words, with $\psi_{a_\theta,\theta,b}$ as in (4.1), we have

$$\widetilde{\mathcal{R}}f(a,\theta,b) = \sqrt{\cos\theta} \cdot \langle f, \psi_{a_\theta,\theta,b}\rangle, \quad a_\theta = a/\cos\theta.$$

We close this section with an important point. The discrete 'pseudo-ridgelets,' i.e. the Riesz representers of the discrete transform, are not ridge functions. They are not of the form $\rho(a_1 n_1 + a_2 n_2)$.

4.9 Inversion

We now briefly discuss the strategy for inverting the Pseudo-Polar FFT. A naive approach would consist in interpolating the sample values of $P_m(\alpha\cdot\ell)$ back to $P_m(\ell)$; that is, one would determine the coefficients of P_m and apply the one-dimensional FFT to find $P_m(\ell)$. The problem with this approach is that it is ill-posed. In fact, this is not really an interpolation, but rather, an extrapolation problem. For small values of α, tiny perturbations of the samples $P(\alpha\ell)$ near the origin may cause large variations away from the origin.

Let L be the Cartesian to polar conversion which maps $\hat{f}(k_1,k_2)$ into $\hat{f}^p_{\ell,m}$ so that with the notations of this section, $T = L\,\mathcal{F}_2$, where $\mathcal{F}_2$ is the two-dimensional FFT. The linear mapping L corresponds to a change of coordinates and has eigenvalues decaying like a power-law. Indeed, the continuous analog maps $\hat{f}$ into $g(r,\theta) = \hat{f}(r\cos\theta, r\sin\theta)$. Multiplying g with the square root of the Jacobian, i.e. $r^{1/2}$, gives an isometry and we then introduce a density weighting matrix D which mimics this renormalization. Recall that for samples $\hat{f}^p(\ell,m)$, m may be thought of as a radial index and ℓ an angular or slope index. Put

$$\tilde{f}^p(\ell,m) = \hat{f}^p(\ell,m)\sqrt{w^p_{\ell,m}}, \quad w^p_{\ell,m} = \begin{cases} \frac{|m|}{2n^2}, & m \neq 0 \\ \frac{1}{8n^2}, & m = 0 \end{cases} \tag{4.21}$$

We refer the reader to [1] for a discussion of these weights. In the vocabulary of numerical linear algebra, D is a preconditioning matrix so that DL is nearly an isometry.

We then invert the Cartesian to Polar conversion $y = Lx$ using a classical iterative numerical algorithm. Starting with $x^{(0)} = 0$, we inductively define

$$\begin{aligned} r^{(n)} &= D(y - Lx^{(n-1)}) \\ x^{(n)} &= \lambda^{(n)} L^* r^{(n)} + x^{(n-1)} \end{aligned}$$

The matrix DL is well conditioned, see the discussion in [1]. In our experiments and for arrays with dimensions ranging from 32 by 32 to 1024 by 1024, we need about 4 iterations to achieve a relative error of order 10^{-4}.

5 New Nonlinear Synthesis Rules

5.1 Osher's Criticism

Roughly speaking, there are two main lines of thinking in the field of mathematical image processing. The first is inspired by recent development in Computational Harmonic Analysis (CHA) such as wavelets, wavelet packets, local cosine bases, etc. The second school has a very different viewpoint, namely, that of partial differential equations (PDE). In recent years, following upon the work of Mumford and Shah, the latter school pioneered new concepts and processing methods which became widely popular. We single out anisotropic diffusion, mean curvature motion, and Total-Variation denoising to name just a few of these ideas. Morel and Solimini [26] give an excellent review of the main results and developments in this field. Our intention is not to oppose both the PDE and CHA approaches. Rather, we merely observe that there exist two distinct sources of inspiration.

Despite their great achievements, image processing methods based on CHA have been thus far received with skepticism by the latter group. This is especially true in signal or image denoising. In fact, we have often heard that Total-Variation approaches are vastly superior to CHA methods for recovering edges from noisy data. For instance, Osher and his colleagues tend to oppose CHA methods arguing that, to put it mildly, *everything that has to do with computational harmonic analysis tends to produce artificial oscillations near discontinuities* even though the original signal/image may be flat on both sides of the discontinuity; they would refer to this phenomenon as a 'pseudo-Gibbs phenomenon.' The terminology 'side-band effect' seems more appropriate.

We will not argue with this objection. Rather, we will develop nonlinear rules for partial synthesis which are not subject to the above phenomenon. These nonlinear rules combine CHA ideas and functionals in common use in the PDE community.

5.2 Minimum Total Variation Synthesis

Roughly speaking, the Total Variation of a function is the integral of the Euclidean norm of the gradient

$$\|f\|_{BV} = \int |\nabla f(x)|\, dx. \tag{5.1}$$

For discrete data $f_{i,j}$, $1 \le i, j \le n$, this would take the form

$$\|f\|_{BV} = \sum_{i,j} \sqrt{|(\delta_1 f)_{i,j}|^2 + |(\delta_2 f)_{i,j}|^2},$$

where D_1 is the finite difference $(\delta_1 f)_{ij} = f_{i,j} - f_{i-1,j}$ and similarly for δ_2.

Letting T be a linear transform, we propose solving

$$\min_f \|g\|_{BV} \quad \text{subject to} \quad |(Tg - b)_\mu| \le e_\mu. \tag{5.2}$$

From now on, we will take T to be the curvelet transform as to make the discussion more concrete.

As suggested in the introduction we may deploy this synthesis rule in a variety of contexts. We give two examples:

- *Compression.* Here, we think of b as being quantized values of Tf. We then use (5.2) to synthesize a 'decompressed' image. In this case, the vector of tolerance would

depend on the properties of the Quantizer. For instance, in the case of a scalar quantizer with quantization q, we may want to use something like $e = q/2$, or q.

- *Noise Removal.* We wish to recover an object f from noisy data

$$y_{i,j} = f_{i,j} + \sigma z_{i,j}, \quad z_{i,j} \ i.i.d. \ N(0,1). \tag{5.3}$$

Here we think of b as the noisy coefficients of f, $b = Ty$, with $b_\mu \sim N(\theta_\mu, \sigma_\mu^2)$. We may use a thresholding rule to identify a set M' of significant coefficients, $|y_\mu| \geq \lambda\sigma_\mu$, say, and set a vector of tolerance $e_\mu = \alpha_\mu \cdot \sigma_\mu$, where α_μ may take different values depending on whether or not the index μ corresponds to a significant coefficient; e.g. $\alpha_\mu = 1$ for $\mu \in M'$ and $\alpha_\mu = \lambda$ for $\mu \notin M'$.

In [28], the authors proposed a minimum Total Variation approach for denoising image data. Given data of the form (5.3), they suggest solving

$$\min_g \|g\|_{TV} \quad \text{subject to} \quad \|g - y\|_{\ell_2}^2/n^2 = \sigma^2. \tag{5.4}$$

The approach (5.2) is very different and combines thresholding rules and minimum variation ideas. Unlike classical thresholding rules, it does not set the nonsignificant coordinates to zero. Instead, (5.2) will typically input small values to cancel oscillations near the discontinuities and eliminate ripples while preserving the strong features of the image such as edges.

The formulation (5.2) was suggested in [29]. Other researchers proposed to combine Total Variation with the wavelet transform, see [18] and more recently [23]. For related ideas although in a very different direction, check [12].

5.3 A Tantalizing Perspective

Consider the following model of images with edges: $\mathcal{F}$ is the class of C^2 objects which may be discontinuous along C^2 edges. Then, the thresholding of curvelet coefficients is provably optimal for estimating objects of that class as this procedure yields near-minimax rates of convergence over $\mathcal{F}$, see [11] for details.

Conjecture 1. *The combined curvelet and minimum Total Variation approach (5.2) is asymptotically nearly optimal over $\mathcal{F}$.*

We view the challenge of proving/disproving Conjecture 1 as intellectually significant. Indeed, if Conjecture 1 were true, this would establish the existence of a procedure enjoying nearly optimal statistical properties, and giving visually pleasing practical results at the same time.

In a different direction, however, we have accumulated evidence pointing to the fact that unlike curvelet-based procedures, TV-denoising is far from being optimal in such setting which leads to a second conjecture.

Conjecture 2. *The Minimum Total Variation approach (5.4)–[28] achieves markedly suboptimal asymptotic rates in that setting. These rates are comparable with those attainable by wavelet shrinkage procedures.*

5.4 Current Implementation Strategy

Although the minimization problem (5.2) is convex, one must frankly admit that it looks daunting. The issue is the dimension of the problem; (5.2) is an optimization in $\mathbb{R}^{n^2}$ with

of the order of $n^2 \log n$ constraints. When $n = 512$, say, these dimensions are indeed very large!

A common approach for solving problems of the type (5.2) is to the so-called gradient descent method with projection; one iteratively alternates unconstrained minimization problems and orthogonal projections onto the set of feasible constraints. This approach is realistic whenever the transform T is orthogonal so that the projection may be easily determined, see [18]. When T is non-orthogonal, this projection will, in general, be the solution of a Quadratic Problem (QP) and each iteration may be very expensive. More generally, we will also have to rule out many other iterative methods such as Interior Points methods because of the sheer size of the problem and the cost of evaluating the constraints Tg.

5.5 Dual Problem

Our approach for solving (5.2) or better said, approximately solving (5.2), will rely on duality theory. Let $\mathcal{L}$ be the Lagrangian

$$\mathcal{L}(f, \Lambda) = \|f\|_{BV} + \Lambda^T K f, \tag{5.5}$$

where Λ is a vector of Lagrange multipliers

$$\Lambda = (\Lambda^{\pm}) \geq 0, \ Kf = \pm(Tf - b - e).$$

The dual problem is of the form

$$\sup_{\Lambda} \mathcal{G}(\Lambda) \tag{5.6}$$

where

$$\mathcal{G}(\Lambda) = \inf_{f} \mathcal{L}(f, \Lambda). \tag{5.7}$$

For strictly feasible constraints, strong duality holds and one way to go about it is to find a saddle point (f^*, Λ^*) of the Lagrangian $\mathcal{L}(f, \Lambda)$

$$\inf_{f} \sup_{\Lambda} \mathcal{L}(f, \Lambda) = \mathcal{L}(f^*, \Lambda^*) = \sup_{\Lambda} \inf_{f} \mathcal{L}(f, \Lambda) = \mathcal{G}(\Lambda^*). \tag{5.8}$$

The Usawa algorithm approaches this problem with an iterative procedure. From $\Lambda_0 \geq 0$, inductively define

$$\mathcal{L}(f_n, \Lambda_n) = \min_{f} \mathcal{L}(f, \Lambda_n) \tag{5.9}$$

$$\Lambda_{n+1} = P(\Lambda_n + \delta K f_n), \tag{5.10}$$

where P is the projection on the positive cone, $P\Lambda = \max(\Lambda, 0)$, and δ is a positive parameter.

The Usawa algorithm is, in fact, a gradient 'descent' with projection algorithm applied to the dual problem. The word 'descent' is to be taken cautiously as one seeks to maximize –and not minimize– the dual function $\mathcal{G}$. In our case, this procedure is relatively simple as our optimization problem reduces to a sequence of unconstrained optimization problems (5.9). Unconstrained Total-Variation minimization problems such as (5.9) are now classical. We will apply this idea to a modified version of the Lagrangian.

From our viewpoint, the dual approach presents several advantages: first, the projection step is straightforward; second, it only requires one evaluation of the curvelet transform per iteration; and third, we never need to invert the curvelet transform.

5.6 Practical Issues

5.6.1 Stability and Convergence

Various theorems are known about the convergence of the Usawa algorithm to the saddle-point of the Lagrangian. Most of these results, however, assume α-convexity of the functional to minimize. The Total Variation norm is not α-convex as it grows only linearly at infinity.

In addition to convergence issues, we also need to address stability. Indeed, the unconstrained minimization problem

$$\min_f \mathcal{L}(f, \Lambda) = \min_f \|f\|_{BV} + \Lambda^T K f$$

may be unstable in the following sense: there might be directions along which the decay of the linear constraints is faster than the growth of the Total-Variation functional!

To get around these difficulties, we proposed to penalize the Lagrangian with a small quadratic term. We substitute the sequence of unconstrained problems (5.7) with a sequence of penalized minimization problems

$$\min_{f_n} \mathcal{L}(f, \Lambda_n) + \gamma_n \|f\|^2, \tag{5.11}$$

where γ_n is a sequence of positive parameters decaying to 0. The penalty restores α-convexity. Under suitable conditions on the sequence γ_n, the Usawa algorithm with (5.11) in place of (5.7) can be shown to converge to a saddle-point (f^*, Λ^*) of the Lagrangian (5.5), see [3].

5.6.2 Burn-in

The speed of convergence depends, of course, upon the initial choice of parameters Λ_0. The closer Λ_0 to Λ^* the better. Finding good guesses of the initial parameter values is more an art than anything else, although there are some guiding principles.

Let $J(f)$ be the Total Variation norm $\|f\|_{BV}$. At the saddle-point, the derivative of the Lagrangian with respect to f vanishes

$$(\nabla_f \mathcal{L})(f^\star, \Lambda^\star) = 0$$

and, therefore,

$$-(\nabla J)(f^\star) = K^* \Lambda^\star. \tag{5.12}$$

Let f_0 be an initial guess of the solution, e.g. obtained after applying the inverse transform to b (5.2). In the applications considered, f_0 might be the object obtained after inversion of the quantized coefficients, or the reconstruction obtained via classical thresholding. The optimum f^* is in some sense close to f_0 and (5.12) suggests finding Λ such that

$$-(\nabla J)(f_0) = K^* \Lambda = T^* \Lambda^+ - T^* \Lambda^-$$

as this would –hopefully– give a parameter value 'close' to $\Lambda^\star$. We use the near orthogonality of T and set

$$\Lambda_0^+ - \Lambda_0^- = -T(\nabla J)(f_0), \quad \Lambda_0^+ = \max(-T(\nabla J)(f_0), 0). \tag{5.13}$$

In our numerical experiments, the choice (5.13) turns out to be a good guess although it is possible to refine it further by using a priori knowledge about the side-band effects we wish to eliminate and analyzing the curvelet transform of $(\nabla J)(f_0)$. Because of space limitations, we do not detail these refinements.

5.6.3 Practice

From a practical viewpoint only four or five iterations are truly needed if one selects a reasonable initial guess. In our numerical experiments, we did not observe much change in the sequence (f_n) after the first five iterations. After only one or two iterations, most of the wrinkles (side-band effects) near the discontinuities are removed while the sharpness of edges is preserved.

6 Numerical Experiments

6.1 Ridgelet Partial Reconstruction

The first experiment compares the quality of wavelet, ridgelet and TV-post-processed ridgelet partial reconstructions. In this experiment, the original figure is a picture of a 512 by 512 sliced-Gaussian. We use only the 100 largest coefficients for each reconstruction. Unlike the wavelet reconstruction, the ridgelet reconstruction of the discontinuity is near-perfect. On the ridgelet picture, however, one can still distinguish oscillations near the edge. The minimum Total Variation ridgelet reconstruction erases those side-bands and enhances the 'decompressed' object. The ridgelet-TV reconstruction and the original figure are nearly indistinguishable.

We may term the remaining approximation error 'pixelization error.' This a kind of error which does not have anything to do with real physical phenomena but rather, occurring because of issues such as discretization and pixelization.

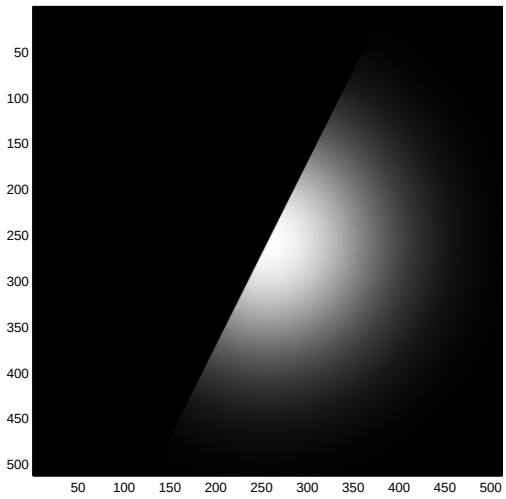
6.2 Noisy Radon Inversion

The second experiment is about reconstructing an image from noisy Radon data. This is a problem of considerable interest in the literature of medical imaging (tomography).

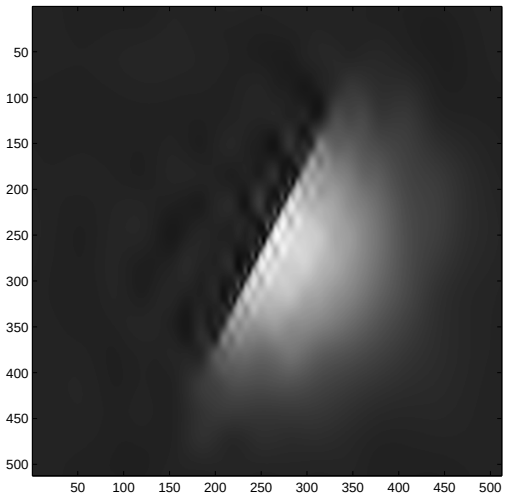
The data for this example are simulated (we do not have access to real datasets). We start with the famous Logan-Shepp phantom and calculate the sinogram, i.e. the discrete Radon transform of the phantom. We use the tools introduced in Section 4 to compute this sinogram. The sinogram is then contaminated with Gaussian white noise. Inverting the discrete Radon transform gives the noisy picture shown in 7(a). The noise is colored; that is, noisy pixel values are correlated and the noise level increases with frequency.

We then apply level dependent thresholding to the wavelet and curvelet coefficients. The noise level in each coordinate is estimated by Monte-Carlo simulations. The choice of the threshold parameter is 3 times the noise level except for the last dyadic subband where it is 4 times the noise level. We also applied cycle-spinning to remove some artifacts which are well-known to occur with wavelet thresholding procedures. Cycle spinning is a kind of translation invariant thresholding rule; this technique computes several individual reconstructions by applying shifts to the noisy data and averages them out –after applying the reverse shifts, of course. We applied cycle-spinning in both cases. (To be exact, the wavelet reconstruction is an average over 64 shifts while we used only 16 shifts for the curvelet reconstruction.) The results are presented on Figures 7, 8 and 9. Finally, these figures also display the combined curvelet and minimum TV reconstruction; this reconstruction does not involve any cycle-spin. With the notations of Section 5, we used a vector of tolerance with $e_\mu = \sigma_\mu$ for the significant coefficients and $e_\mu = 4\sigma_\mu$ for the others.

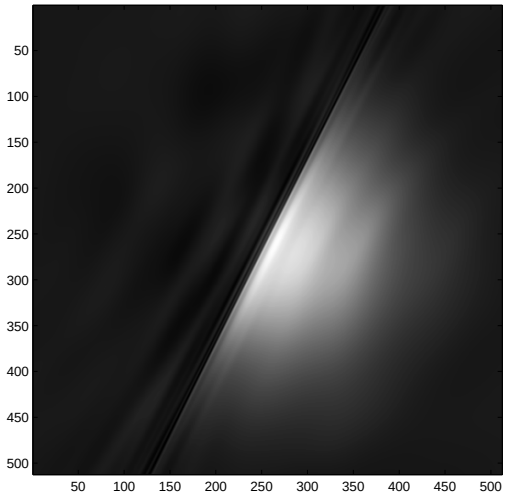
Details are shown on Figures 8 and 9. Edges are very sharp in both the curvelet and the curvelet-TV reconstructions. The curvelet-TV reconstruction is free of artifacts and



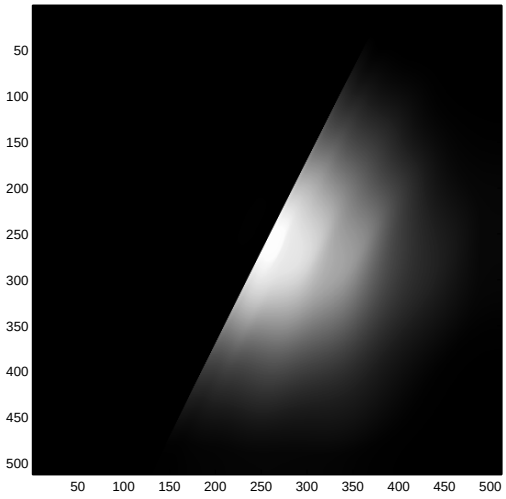
(a) Original



(b) 100 Wavelets

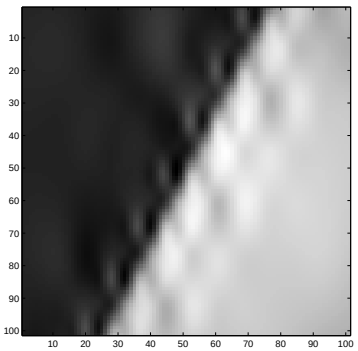


(c) 100 Ridgelets

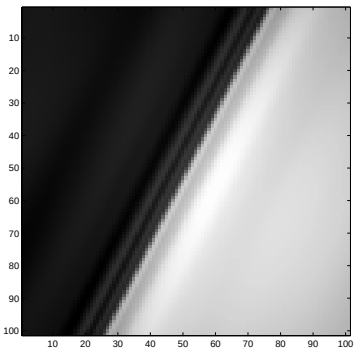


(d) 100 Ridgelets and TV

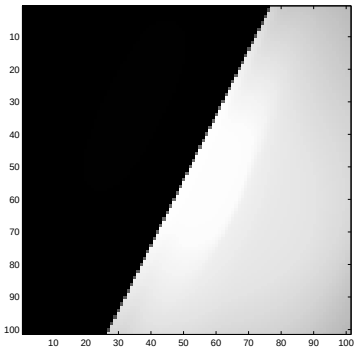
Figure 5: Partial Reconstructions of a Sliced-Gaussian



(a) 100 Wavelets



(b) 100 Ridgelets



(c) 100 Ridgelets and TV

Figure 6: Details

	Wavelets	Curvelets	Curvelets and TV
PSNR	32.23	34.46	35.2

Table 1: Table of PSNR values for the recontruction of the Logan-Shepp phantom.

side-bands are removed as evidenced on Figure 10, which shows the middle horizontal scanline of the phantom. The noisy and reconstructed scanlines are plotted on Figures 10(a), 10(b) and 10(c). Figure 10(d) shows a superposition of both the true and the curvelet-TV reconstructed scanlines. The reconstructed line is very close to the truth.

Figures 7, 8 and 9 show some residual plots. These pictures clearly demonstrate that curvelets are able to extract meaningful signal at very high-frequency. This is possible because unlike wavelets, curvelets correlate very well with signal structures (edges) at high frequency, yielding improved signal to noise ratio.

Finally, we also report the PSNR for each method which we interpret as an 'objective' measure of performance. The values are given in the table below. Curvelet reconstructions exhibit a higher PSNR. We report here a PSNR of 35.2 because it is that of the results presented in the various figures. We would like to point out that for other iteration counts and initial choice of parameters, we have been able to obtain PSNRs as high as 37.

6.3 Denoising

After these encouraging results on synthetic examples, the last experiment is about 'denoising' a real photograph. The original image is contaminated with Gaussian white noise and the processing steps are nearly the same as in the previous experiments (same parameter values). We do not apply cycle-spinning. The results and details are shown on Figures 11 and 12, respectively.

In some sense, this numerical example downplays the objection raised near the beginning of Section 5.1. It is certainly true that on sharp geometrical objects such as a white square on black background, one might perceive spurious oscillations when applying CHA methods. However, on more, complicated and realistic imagery such as photographs with textures, these artifacts are often barely visible.

7 Perspective

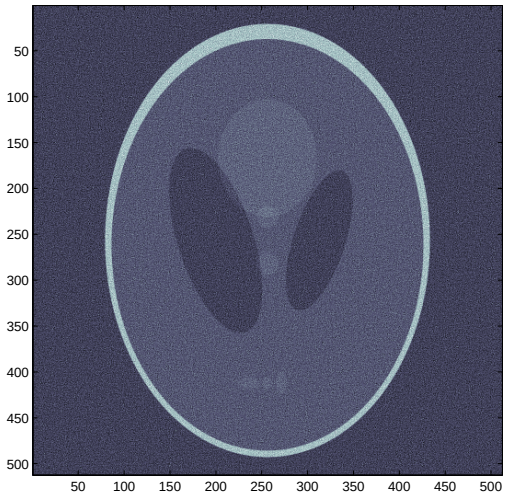
7.1 Second Generation of Curvelets

We believe that there is a simpler way to construct tight frames of curvelets. This construction does not use ridgelets.

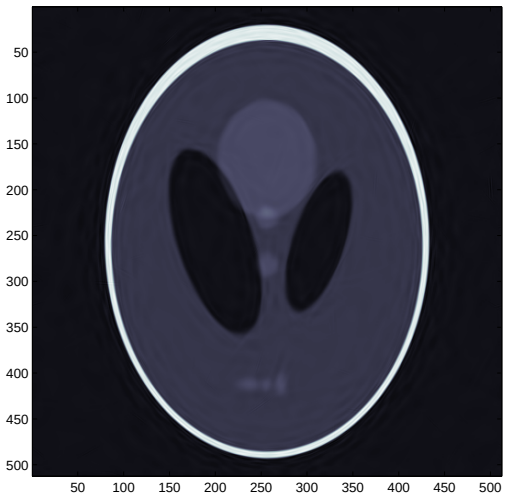
7.1.1 Tight Frames

For each $j \geq 0$, consider a family of orthogonal and compactly supported windows $\chi_{j,\ell}$ obeying

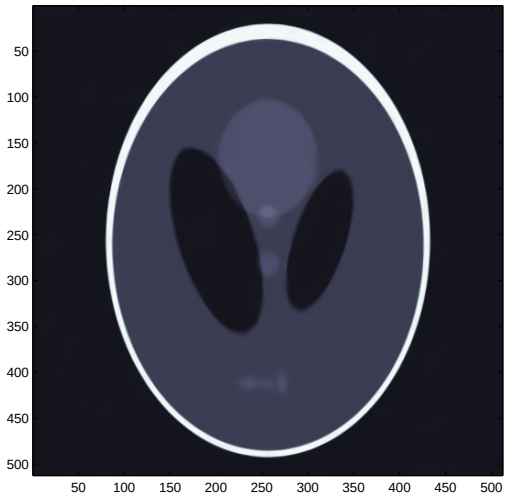
$$|\chi_0(\xi)|^2 + \sum_{j,\ell} |\chi_{j,\ell}(\xi)|^2 = 1. \tag{7.1}$$



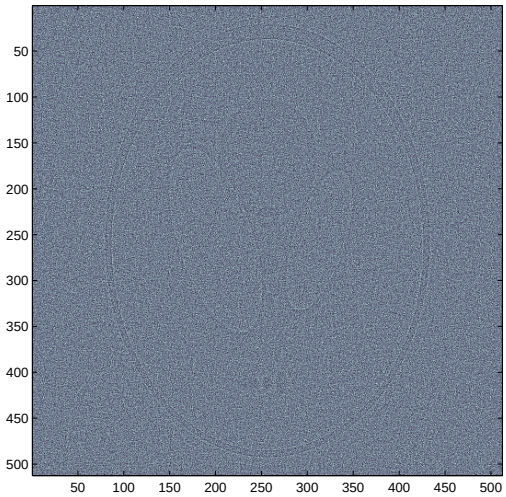
(a) Noisy Phantom



(b) Curvelets

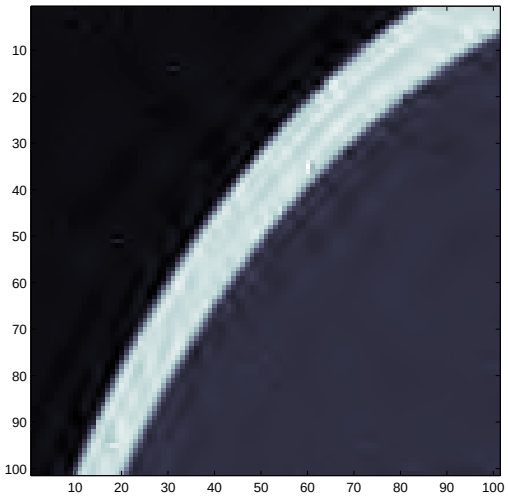


(c) Curvelets and TV

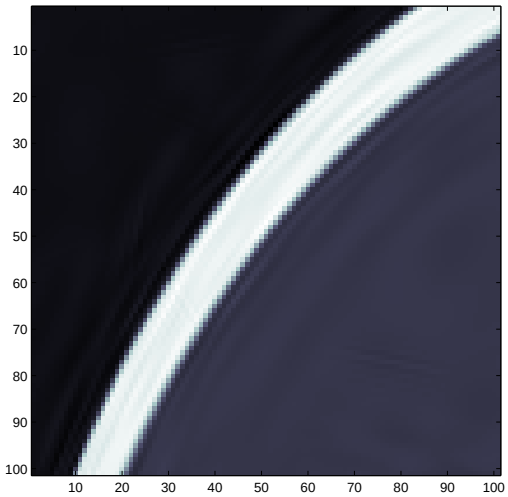


(d) Curvelets and TV: Residuals

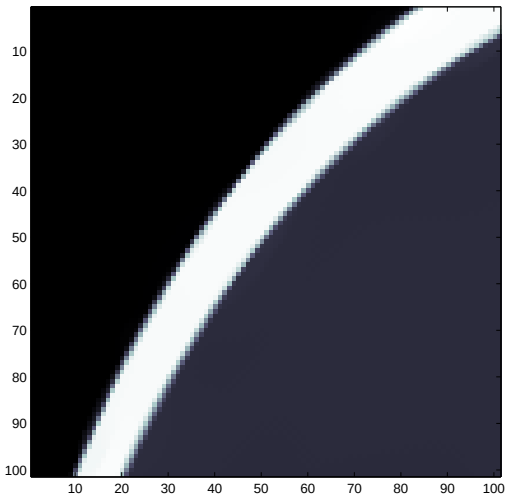
Figure 7: Reconstruction of the Logan-Shepp Phantom



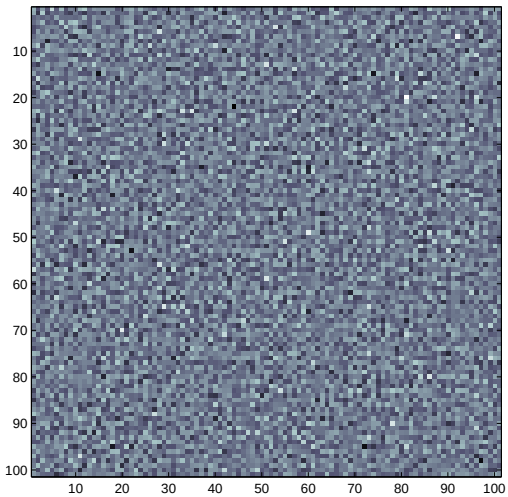
(a) Wavelets



(b) Curvelets

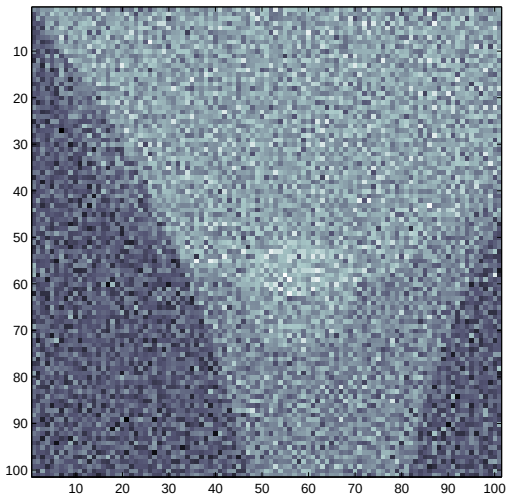


(c) Curvelets and TV

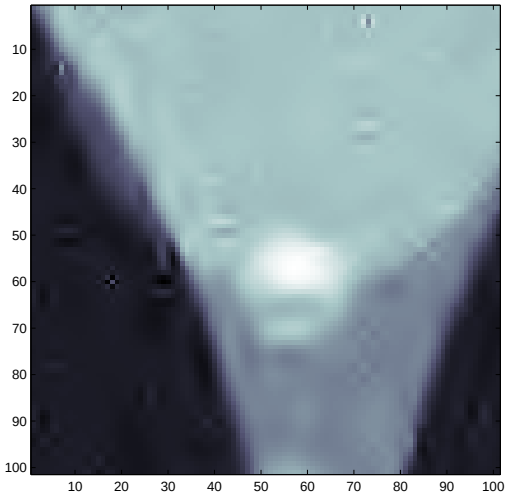


(d) Curvelets and TV: Residuals

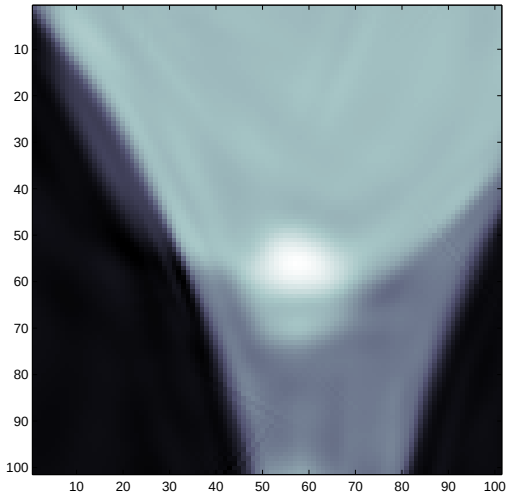
Figure 8: Details



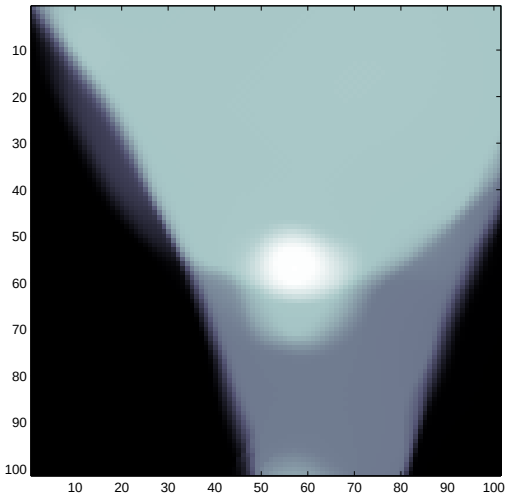
(a) Noisy



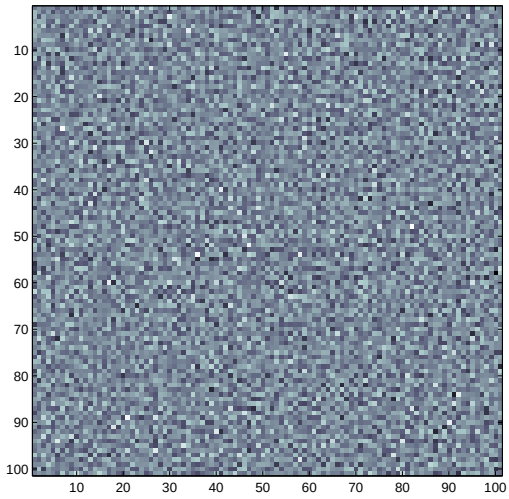
(b) Wavelets



(c) Curvelets

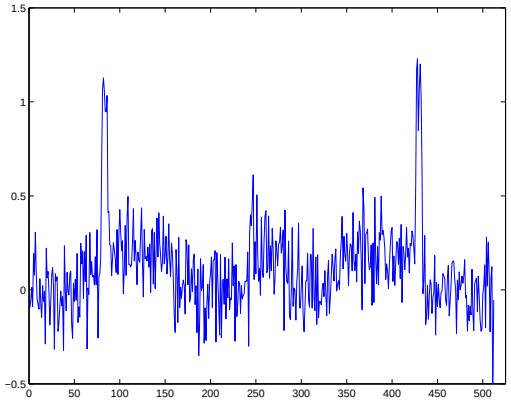


(d) Curvelets and TV

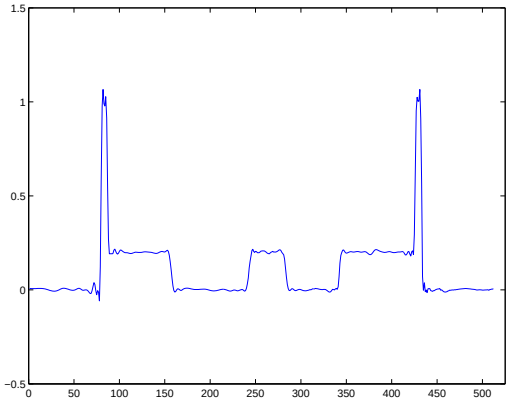


(e) Curvelets and TV: Residuals

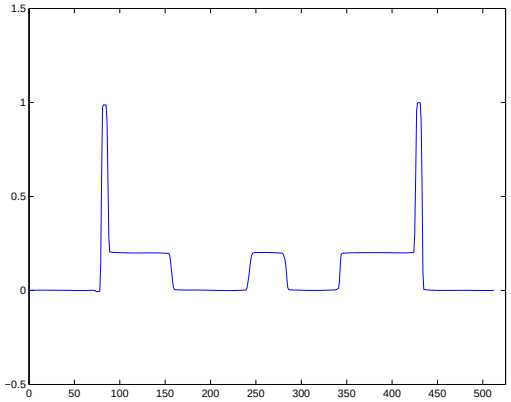
Figure 9: Details



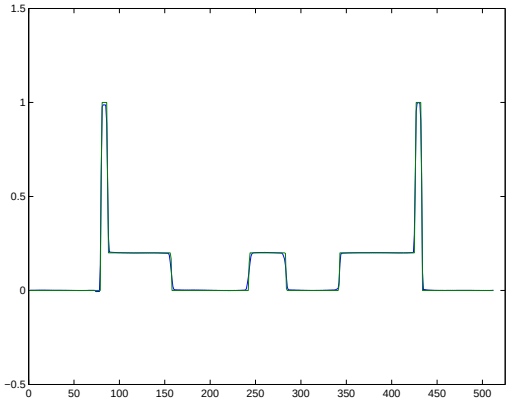
(a) Noisy Scanline



(b) Curvelets

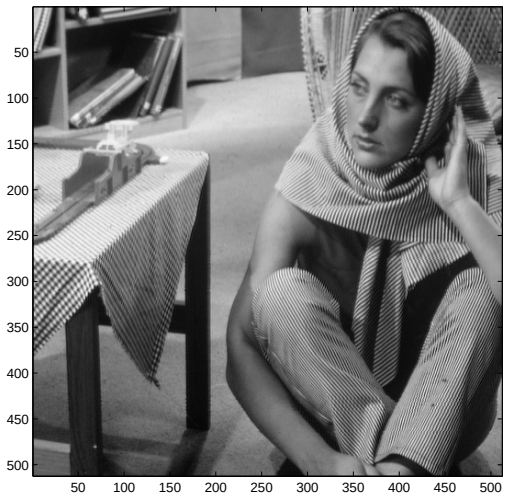


(c) Curvelets and TV

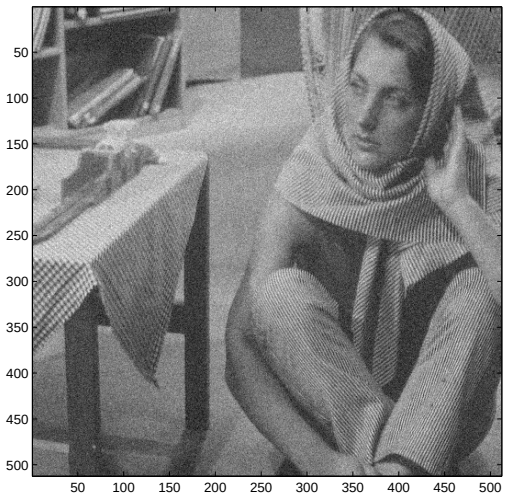


(d) True Scanline and Curvelets and TV Reconstruction

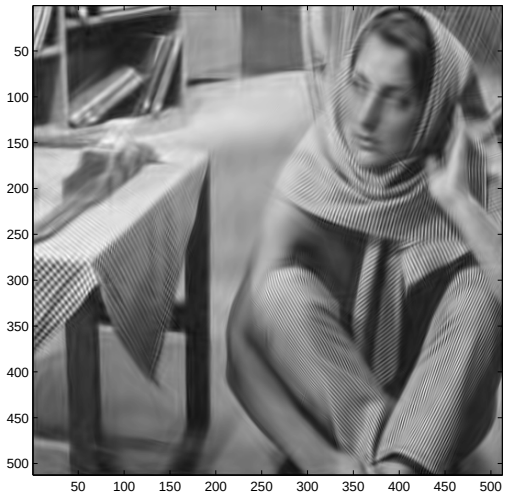
Figure 10: Scanline Plots



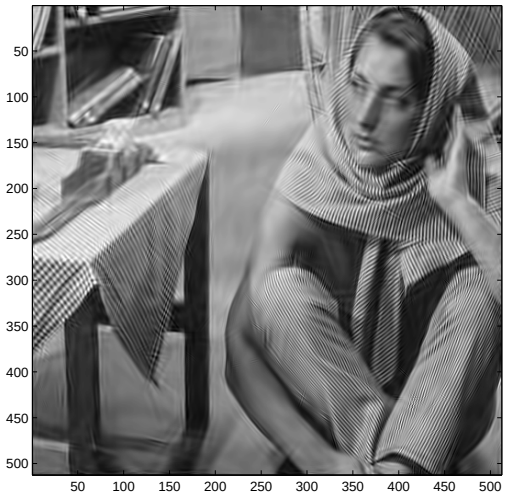
(a) Original



(b) Noisy

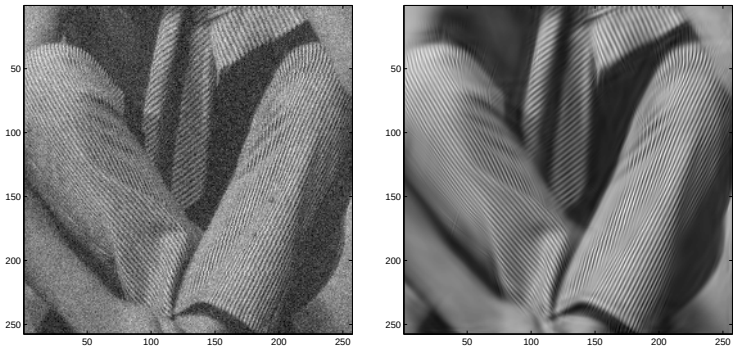


(c) Curvelets

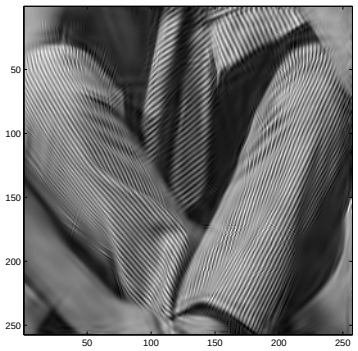


(d) Curvelets and TV

Figure 11: Denoising Barbara



(a) Noisy Detail (b) Curvelets



(c) Curvelets and TV

Figure 12: Details of Barbara

We will use the windows $\chi_{j,\ell}$ to localize the Fourier transform near wedges of length 2^{2j} and width 2^j. In polar coordinates we let

$$\chi_{j,\ell}(\xi) = w(2^{-2j}|\xi|)\,\nu(2^j\theta - \pi\ell),$$

and put $\chi_0(\xi) = w_0(|\xi|)$. Then, the decomposition (7.1) would hold provided

$$|w_0(|\xi|)|^2 + \sum_j |w(2^{-2j}|\xi|)|^2 = 1, \tag{7.2}$$

and

$$\sum_{\ell=0}^{2^{j+1}-1} |\nu(2^j\theta - \pi\ell)|^2 = 1. \tag{7.3}$$

As for Meyer wavelets [25], we will assume that w is supported on $[2\pi/3, 8\pi/3]$ and ν on $[-5\pi/6, 5\pi/6]$, say. The window $\chi_{j,0}$ is then localized near the 'horizontal' wedge

$$[\pi 2^{2j} \le |\xi| \le \pi 2^{2(j+1)}] \times [-\pi/2\cdot 2^{-j} \le \theta \le \pi/2\cdot 2^{-j}]. \tag{7.4}$$

For each ℓ, $\chi_{j,\ell}$ is obtained from $\chi_{j,0}$ by applying a rotation. It is not difficult to check that for each $j \ge 1$, the support of $\chi_{j,0}$ is contained in a rectangle of length $2\pi\delta_1 2^{2j}$ and width $2\pi\delta_2 2^j$ with $\delta_1 = 1 + O(2^{-j})$ and $\delta_2 = 10\pi/9$.

This type of tiling of the frequency plane was introduced by Fefferman and used very successfully thereafter to study properties of Fourier Integral Operators, see Chapter 9 in [31] and references therein.

For the special pair $(j,0)$, we introduce the Cartesian lattice $k = (\delta_1^{-1}\cdot k_1 2^{-2j}, \delta_2^{-1}\cdot k_2 2^{-j})$ with δ_1, δ_2 as before. Then for each pair $J = (j,\ell)$, we let k^J be the lattice rotated by an angle equal to $\theta_J = \pi\cdot\ell\cdot 2^{-j}$. Formally,

$$k^J = R_J\, k, \quad R_J = \begin{pmatrix} \cos\theta_J & -\sin\theta_J \\ \sin\theta_J & \cos\theta_J \end{pmatrix}.$$

Observe that with these notations, $\chi_{j,\ell}(\xi) = \chi_{j,0}(R_J^*\xi)$.

Define

$$\hat\gamma_\mu(\xi) = \frac{1}{\sqrt{\delta_1\delta_2}}\, 2^{-3j/2}\, e^{-i\langle k^J,\xi\rangle}\chi_{j,\ell}(\xi) \quad \mu = (j,\ell,k). \tag{7.5}$$

With the same notation as in Section 3, we also define coarse scale curvelets $\hat\gamma_{\mu_0}$ as being proportional to the modulated blobs $e^{-ik\xi}\chi_0(\xi)$. Then with obvious notations, $(\gamma_{\mu'})_{\mu'\in M'}$ is a tight frame:

$$\sum_{\mu'} |\langle f, \gamma_{\mu'}\rangle|^2 = \|f\|^2_{L_2(\mathbb{R}^2)}, \quad f = \sum_{\mu'}\langle f,\gamma_{\mu'}\rangle\gamma_{\mu'}. \tag{7.6}$$

This merely follows from the fact that $\frac{1}{2\pi\sqrt{\delta_1\delta_2}} e^{ik^J\xi}$ is an orthonormal basis of square integrable functions defined over a rectangle containing the support of $\chi_{j,\ell}$. Then,

$$\sum_{k^J} |\langle f,\gamma_\mu\rangle|^2 = \frac{1}{(2\pi)^4}\sum_{k^J}|\langle \hat f, \hat\gamma_\mu\rangle|^2 = \frac{1}{(2\pi)^2}\int |\hat f(\xi)|^2\,|\chi_{j,\ell}(\xi)|^2 d\xi,$$

and (7.6) follows from (7.1).

In a more refined construction, one may want to merge symetric wedges, namely, $\chi_{j,\ell}$ and $\chi_{j,\ell+2^j}$ thanks to the symmetry $\chi_{j,\ell}(-\xi) = \chi_{j,\ell+2^j}(\xi)$. Details will appear in a forthcoming paper.

7.1.2 Space-Side Picture

Next, we let $\theta_{j,\ell}$ be the inverse Fourier transform of $\frac{1}{\sqrt{\delta_1\delta_2}}2^{-3j}\chi_{j,\ell}(\xi)$. Then

$$\gamma_\mu(x) = 2^{3j/2}\,\theta_{j,\ell}(x - k_J). \tag{7.7}$$

The relationship $\chi_{j,\ell}(\xi) = \chi_{j,0}(R_J^*\xi)$ gives

$$\gamma_\mu(x) = 2^{3j/2}\,\theta_{j,0}(R_J^*x - k).$$

Now, the envelope of $\theta_{j,0}$ is concentrated near a vertical ridge of length about 2^{-j} and width 2^{-2j}. Define $\theta^{(j)}$ by

$$\theta_{j,0}(x) = \theta^{(j)}(D_j x)$$

where D_j is the diagonal matrix

$$D_j = \begin{pmatrix} 2^{2j} & 0 \\ 0 & 2^j \end{pmatrix}. \tag{7.8}$$

In other words, the envelope $\theta^{(j)}$ is supported near a disk of radius about one, and owing to the fact that $\chi_{j,0}$ is supprted away from the axis $\xi_1 = 0$, $\theta^{(j)}$ oscillates along the horizontal direction. In short, $\theta^{(j)}$ 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. Equation (7.7) becomes

$$\gamma_\mu(x) = 2^{3j/2}\,\theta^{(j)}(D_j(R_J^*x - k)). \tag{7.9}$$

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. We list a few of their properties:

- The *Parabolic Scaling* (7.8) yields an *Anisotropy Scaling Law*: the system is well-localized in space and obeys approximately the relationships

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

  and, therefore,

  $$width \approx length^2.$$

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

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

In short, this system exhibit all the geometrical and multiscale features of the curvelet transform. We believe that this is new system is an alternative to the curvelet construction reviewed in Section 3. This leads to a last conjecture.

Conjecture 3. *This new system and 'classical' curvelets enjoy nearly the same optimality properties for representing C^2 images with C^2 edges [9, 8].*

We will conclude this section with a brief summary of the main points of the construction of new tight frames of curvelets:

- 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 13.

- We use oriented local Fourier bases on each wedge.

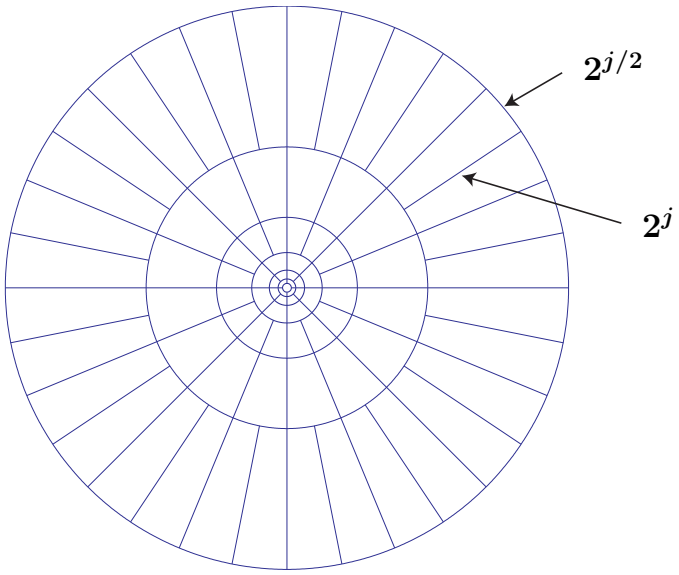


Figure 13: Split Every Other Two

7.2 Ridgelet Packets?

The previous section paves the way to a wealth of new multiscale systems. By arbitrary segmentation of dyadic coronae, we can design tight frames with arbitrary aspect ratios at arbitrary scales. Splitting at every scale would essentially give tight frames of ridgelets, at every other scales, tight frames of curvelets, and no splitting essentially yields wavelet frames.

The reader familiar with the literature on wavelet packets, for example, may see that there is an opportunity to develop trees of tight frames by applying Recursive Dyadic Partitioning ideas to each frequency corona. Along this line of research, it would be of interest to develop fast algorithms à la Coifman and Wickerhauser for searching sparse decompositions in these trees.

7.3 3 Dimensions

This paper focused on new multiscale representations of bivariate functions and their potential for practical applications in image processing. However, most of the ideas, tools and algorithms we presented in this paper may easily be extended to higher dimensions, and especially the tree-dimensional case. We may develop 3D ridgelets, 3D curvelets and 3D digital transforms. This may lead to the development of convenient tools to represent, analyze, store and transmit 3 dimensional datasets. The situation where the third dimension is time as in digital video would be of special interest.

7.4 Epilogue

We have shown some of our numerical results to prominent researchers in the PDE-based image processing community. These results have been received with a lot of enthusiasm and we take this as a sign of encouragement.

References

[1] A. Averbuch, D. L. Donoho, R. R. Coifman, M. Israeli, and J. Walden. Fast slant stack: A notion of radon transform for data on a cartesian grid which is rapidly computable, algebraically exact, geometrically faithful, and invertible. http://www-stat.stanford.edu/~donoho/Reports/index.html.

[2] D. H. Bailey and P. N. Swartztrauber. The fractional fourier transform and applications. *SIAM Review*, 33:389–404, 1991.

[3] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, Massachussets, 1999.

[4] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans Image Processing*, 8(12):1688–1701, 1999.

[5] P. J. Burt and A. E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.

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

[7] E. J. Candès and D. Donoho. Curvelets, multiresolution representation, and scaling laws. In *SPIE conference on Signal and Image Processing: Wavelet Applications in Signal and Image Processing VIII, San Diego*, 2000.

[8] E. J. Candès and D. L. Donoho. *Curvelets*. Manuscript. `http://www-stat.stanford.edu/ donoho/Reports/1998/curvelets.zip`, 1999.

[9] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In L. L. Schumaker et al., editor, To appear *Curves and Surfaces*, Vanderbilt University Press, 1999. Nashville, TN.

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

[11] E. J. Candès and D. L. Donoho. Recovering edges in ill-posed inverse problems: Optimality of curvelet frames. Technical report, Department of Statistics, Stanford University, 2000. To appear *Ann. Statist.*

[12] T. F. Chan and H. M. Zhou. Optimal construction of wavelet coefficients using total variation regularization in image compression. Technical report, UCLA, 2000. CAM Technical Report.

[13] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46:886–902, 1998.

[14] I. Daubechies. Orthogonal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics*, 41:909–996, 1988.

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

[16] D. L. Donoho. *Orthonormal ridgelets and linear singularities*. Technical report, Department of Statistics, Stanford University, 1998. Submitted for publication.

[17] D. L. Donoho. Sparse components analysis and optimal atomic decomposition. Technical report, Department of Statistics, Stanford University, 1998.

[18] S. Durand and J. Froment. Artifact free signal denoising with wavelets. In *Proceedings of ICASSP 2001*, 2001. IEEE 26th International Conference on Acoustics, Speech, and Signal Processing.

[19] D. Esteban and C. Galand. Application of quadrature mirror filter to split-band voice coding schemes. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pages 191–195, 1977.

[20] J. Grossman, A. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15:723–736, 1984.

[21] X. Huo. *Sparse Image Representation via Combined Transforms*. PhD thesis, Stanford Univesity, August 1999.

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

[23] F. Malgouyres. A unified framework for image restoration. Technical report, UCLA, 2001. CAM Technical Report.

[24] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

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

[26] J. M Morel and S. Solimini. *Variational Methods in Image Segmentation*, volume 14 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Boston, 1995.

[27] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[28] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica D*, 60:259–268, 1992.

[29] 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.

[30] J.L. Starck, E. Candès, and D.L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 2000. to appear.

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

[32] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice Hall, Englewood Cliffs, NJ, 1995.