# Ridgelets and Their Derivatives: Representation of Images with Edges

Emmanuel J. Candes

**Abstract.** This paper reviews the development of several recent tools from computational harmonic analysis. These new systems are presented under a coherent perspective, namely, the representation of bivariate functions that are singular along smooth curves (edges). First, the representation of functions that are smooth away from straight edges is presented, and ridgelets will be shown to provide near optimal nonlinear approximations to these objects. Motivated by the limitations of the ridgelet methodology, new representation systems, namely, monoscale ridgelets and curvelets – both of which use the ridgelet transform as a building block – will be introduced. Curvelets are shown to provide concrete and constructive optimal nonlinear approximations to smooth functions with twice differentiable singularities. In addition these approximations are obtained simply by thresholding the curvelet series.

## §1. Introduction

Throughout the sciences, sparse representations of classes of objects are often sought because of the well-known applications of sparsity to problems ranging from data compression and statistical estimation to feature detection. Indeed, finding sparse representations together with rapid algorithms to compute them is one of the main objectives of a rapidly growing field, computational harmonic analysis (CHA). In this paper, we will argue that CHA has not really addressed the problem of efficiently representing smooth multivariate functions with sharp discontinuities, like smooth images with edges. Motivated by this gap in the literature, we present a collection of new representation tools that efficiently represent smooth functions that are singular along curves. Here, the tone is expository; details may be found in the cited references. In this paper, attention is restricted to the two-dimensional situations although extensions to higher dimensions exist, or are anticipated.

Saint-Malo Proceedings
XXX, XXX, and Larry L. Schumaker (eds.), pp. 1–10.


ISBN 1-xxxxx-xxx-x.

### The Wavelet Miracle

One of the most appealing features of wavelet systems is their ability to provide efficient representations of spatially inhomogeneous functions, i.e., functions that may be discontinuous, spiky, etc. In Mallat's words "bases of smooth wavelets are the best bases for representing objects composed of singularities, when there may be an arbitrary number of singularities, which may be located in all possible spatial positions" [8]. For instance, on the unit interval define

$$f(t) = H(t - t_0)\, g(t), \quad t \in [0, 1], \tag{1}$$

where $H$ is the Heavyside $H(t) = 1_{\{t>0\}}$ and $g$ is a smooth arbitrary function with compact support and finite Sobolev norm $\|g\|_{W_2^s}$ (see [1] for the classical definition of $L_2$ Sobolev norms). Then, the number of Fourier coefficients of $f$ exceeding $1/n$ in absolute value is bounded below by $c \cdot n$, regardless of the degree of smoothness of $f$ away from the singular point $t_0$. This means that a lot of different terms are needed to obtain good partial reconstructions; keeping the $n$ largest terms in the Fourier series gives only an $L_2$ error of approximation of order $n^{-1/2}$. (Throughout the paper, it will always be implicit that the error is measured in the $L_2$ norm.) In contrast, the sparsity of the wavelet coefficient sequence of $f$ is in some sense the same as if $f$ were not singular. In effect, the number of wavelet coefficients exceeding $1/n$ is bounded by $C\, n^{2/(2s+1)}$ giving rates of approximation of order $n^{-s}$ corresponding to the nonlinear bandwidth of $W_2^s$ Sobolev balls. This remarkable adaptivity property is what we call the "wavelet miracle."

### The Curse

Unfortunately, wavelets can deal with point-like singularities, but are seriously challenged by line-like singularities in dimension two. Let us for instance consider the object

$$f(x_1, x_2) = H(x_1 \cos\theta_0 + x_2 \sin\theta_0 - t_0)\, g(x_1, x_2) \quad (x_1, x_2) \in [0, 1]^2, \tag{2}$$

where, again, $g$ is a bivariate function taken from the Sobolev space $W_2^s$; $f$ is singular on the line $x_1 \cos\theta_0 + x_2 \sin\theta_0 = t_0$, but smooth otherwise. Then, the number of wavelet coefficients exceeding $1/n$ is now of the order $n$. Hence, partial $n$-term wavelet reconstructions will only converge at a rate $n^{-1/2}$, regardless of the almost everywhere degree $s$ of smoothness. The edge limits the speed of convergence. This result is intuitively not very surprising as wavelet bases are made of local isotropic oscillatory bumps at various scales, and are not adapted to represent long elongated structures like edges.

This clearly raises an important question: in two dimensions (and, more generally, in arbitrary $d$ dimensions) can we develop a representation enjoying the same adaptivity features as wavelets in dimension one?

## §2. Ridgelets and Linear Singularities

In [3], Candes introduced a new tiling of the frequency plane that led to the construction of ridgelet frames. We say that a collection $(\varphi_n)$ is a frame of a Hilbert space $H$ if there exist two constants $A, B > 0$ such that for any element of $H$, we have

$$A \|f\|_H^2 \leq \sum_n |\langle f, \varphi_n \rangle_H|^2 \leq B \|f\|_H^2.$$

When $A = B$, the frame is said to be tight. A collection $(\varphi_n)$ that verifies the frame property is of course complete and there is a very concrete way to reconstruct $f$ from the datum of its coefficients $(\langle f, \varphi_n \rangle_H)$. Generalities about frames can be found in [11].

Let $\psi$ be a univariate oscillatory function and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$. The ridgelet frame $\psi_{j,\ell,k}$ is a collection of ridge functions given by

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

in the frequency domain [3] ($\delta$ denotes the dirac distrbution).

Donoho [9] modified the ridgelet construction by essentially replacing the discretization of the angular variable with a periodic wavelet transform resulting in an orthonormal basis. He called these new basis elements orthonormal ridgelets. In the remainder of this paper, we make the choice of the orthonormal ridgelets although all the results and constructions that follow would hold true if one were to use 'pure ridgelets.'

As stated in [9]: Such a system can be defined as follows: let $(\psi_{j,k}(t) : j \in \mathbb{Z}, k \in \mathbb{Z})$ be an orthonormal basis of Meyer wavelets for $L^2(\mathbb{R})$ [12], and let $(w^0_{i_0,\ell}(\theta),\ \ell = 0, \dots, 2^{i_0} - 1;\ w^1_{i,\ell}(\theta),\ i \geq i_0,\ \ell = 0, \dots, 2^i - 1)$ be an orthonormal basis for $L^2[0, 2\pi)$ made of periodized Lemarié scaling functions $w^0_{i_0,\ell}$ 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

$$\hat{\rho}_\lambda(\xi) = |\xi|^{-\frac{1}{2}}(\hat{\psi}_{j,k}(|\xi|)w^\varepsilon_{i,\ell}(\theta) + \hat{\psi}_{j,k}(-|\xi|)w^\varepsilon_{i,\ell}(\theta + \pi))/2. \tag{3}$$

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

$$f = \sum_\lambda \langle f, \rho_\lambda \rangle \rho_\lambda.$$

Ridgelets turn out to be optimal for representing functions with linear singularities. Indeed, let us consider the template (2). The following theorem is proved in [4].

**Theorem 1.** *Let $g \in W_2^s(\mathbb{R}^2)$ and $f(x_1, x_2) = H(x_1 \cos\theta_0 + x_2 \sin\theta_0 - t_0)\, g(x_1, x_2)$. Then the sequence $(\alpha_\lambda = \langle f, \rho_\lambda \rangle)$ of orthonormal ridgelet coefficients of $f$ satisfies*

$$\#\{|\alpha_\lambda| \geq 1/n\} \leq C\, n^{1/(s+1)}\, \|g\|_{W_2^s}$$

*for some constant $C$ not depending on $f$. As a consequence, the $n$-term approximation $f_n$ – obtained by keeping the terms corresponding to the $n$ largest coefficients in the ridgelet expansion – satisfies*

$$\|f - f_n\| \leq C\, n^{-s/2}\, \|g\|_{W_2^s}.$$

Hence, the theorem states that we obtain a rate of approximation as if the object were not singular, simply by thresholding the orthonormal ridgelet expansion. Whereas the singularity caused partial wavelet reconstructions to converge very slowly, its effect on the approximation rate of truncated ridgelet series is 'harmless.'

## §3. Ridgelets and Curved Edges.

Theorem 1 considered linear singularities and it seems natural to ask whether similar results will hold if one replaced the singularity along a straight line with one along an arbitrary curve $\gamma$. To simplify our exposition, consider the simple case of a singular function defined on the unit square by

$$f(x_1, x_2) = g(x_1, x_2)\, 1_{\{x_2 \leq \gamma(x_1)\}}, \tag{4}$$

where $g$ is a smooth function and $\gamma$ is smooth curve. Then the ridgelet coefficient sequence of such an object is in general not sparse:

$$\#\{\lambda, |\alpha_\lambda| \geq 1/n\} \geq c\, n.$$

Thus, the speed of convergence of the best $n$-term ridgelet approximation is only of order $n^{-1/2}$. It is interesting to observe that the degree of approximation of both wavelet and ridgelet partial reconstructions is the same, although they correspond to radically different systems of representation. Ridgelets are elongated and directional, whereas wavelets are isotropic and local.

The limitations that we presented in this section motivate the refinements and new tools that we are about to introduce.

## §4. Monoscale Ridgelets

The approach developed in this section builds on Theorem 1. The idea here is to take advantage of the optimal representation of linear singularities by localizing the ridgelets. A detailed exposition is provided in [5].

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

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

The following details a way of making up such an orthonormal partition: take a $C^\infty$ univariate window $\nu$ supported in $[-3/4, 3/4]$ such that $\nu(t) = 1$ on $[-1/2, 1/2]$; define $v_Q = \nu(2^s x_1 - k_1)\, \nu(2^s x_2 - k_2)$; and renormalize the windows $v_Q$ with

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

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

Define the rescaling operator $T_Q g$ by

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

which is an isometry of $L_2$. Throughout this section, $s$ is arbitrary but fixed. Monoscale ridgelets are defined as follows: let $\rho_\lambda$ be an orthonormal ridgelet basis and define

$$\psi_{Q,\lambda}(x_1, x_2) = w_Q(x_1, x_2)(T_Q \rho_\lambda)(x_1, x_2);$$

the collection

$$\{\psi_{Q,\lambda}, Q \in \mathcal{Q}_s, \lambda \in \Lambda\} \tag{5}$$

is what we call the monoscale ridgelet dictionary.

It is easy to check that the monoscale ridgelet dictionary is a tight frame of $L_2(\mathbb{R}^2)$ as we have a Parseval relationship

$$\|f\|_2^2 = \sum_{Q \in \mathcal{Q}_s} \sum_{\lambda} \langle f, \psi_{Q,\lambda} \rangle^2.$$

Standard arguments show that we then have the decomposition

$$f = \sum_{Q \in \mathcal{Q}_s} \sum_{\lambda} \langle f, \psi_{Q,\lambda} \rangle \psi_{Q,\lambda}, \tag{6}$$

with equality holding in an $L_2$ sense.

We add an "extra layer of coarse scale coefficients" to eliminate various artifacts. Consider a standard multiresolution analysis that is adapted to the unit square [7] so that the set of translates $\{2^s\, \varphi(2^s \cdot - k)\}$, $k = (k_1, k_2)$, $k_i = 0, 1, \ldots, 2^s - 1$ is orthonormal. Let $P_0$ be the orthogonal projector onto $V_s$, the span of the $\varphi_{s,k}$'s; i.e.,

$$P_0 f := \sum_k \langle f, \varphi_{s,k} \rangle \varphi_{s,k} := \sum_k \beta_{s,k} \varphi_{s,k}. \tag{6}$$

The following Pythagorean relationship holds:

$$\|f\|_2^2 = \|P_0 f\|_2^2 + \|(I - P_0)f\|_2^2. \tag{7}$$

Finally, define the coefficients

$$\alpha_{s,\mu} = \langle Rf, \psi_{Q,\lambda}\rangle \quad \mu = (Q, \lambda), \quad Q \in \mathcal{Q}_s, \lambda \in \Lambda. \tag{8}$$

**Definition 1.** *The monoscale ridgelet transform with base scale $s$ is the mapping from functions $f \in L_2(\mathbb{R}^2)$ to the amalgamation of coefficients $(\beta_{s,k})$ and $(\alpha_{s,\mu})$.*

Note that we again have a partial isometry

$$\|f\|_2^2 = \sum_k |\beta_{s,k}|^2 + \sum_\mu |\alpha_{s,\mu}|^2,$$

thanks to the Pythagorean relationship (7).

Let us return now to the main theme of this paper, and study the efficiency of monoscale ridgelets to represent objects that are singular along curves. Suppose that one is interested in constructing an $n$-term approximation of the function $f$ in (4). Without loss of generality, we will suppose that $n$ is of the form $n = 2^{2J+1}$. We simply expand $f$ in the monoscale ridgelet dictionary (5) *with $s = J$ as a choice of base scale*; that is, we define the $n$-term approximation by

$$f_n = P_0 f + R_{n/2} f, \tag{9}$$

where $R_{n/2}f$ is the partial reconstruction of the residual $Rf$ obtained by keeping the terms corresponding to the $n/2 = 2^{2J}$ largest coefficients $\alpha_{J,\mu}$.

It is interesting to observe that the choice of the base scale $s$ of the monoscale dictionary depends on the number $n$ of terms we wish to keep in the approximant. We have the following result [5]:

**Theorem 2.** *Let $g \in W_2^s(\mathbb{R}^2)$ and $f(x) = g(x)\, 1_{\{x_2 \le \gamma(x_1)\}}$, with $\gamma$ being three times differentiable. Let $f_n$ be the $n$-term approximation defined by (9). Then,*

$$\|f - f_n\|_2 \le C \max(n^{-s/2}, n^{-3/4}).$$

This simple approximation scheme provides optimal rates of convergence as long as $s \le 3/2$; that is, approximation bounds as if $f$ were not singular. In some sense, one is allowed to say that unlike wavelets, ridgelets can be adapted to provide efficient representations of curved singularities. There is a critical value $s = 3/2$ of the smoothness parameter, however, beyond which the method saturates; as $s$ increases, the approximation rate is blocked at $n^{-3/4}$. Nevertheless, this represents already a substantial improvement over wavelet approximations whose convergence rates are blocked at $n^{-1/2}$.

Better results are theoretically possible. For instance, let $\mathcal{F}(C)$ be a model of smooth images with twice differentiable edges defined as follows:

$$\mathcal{F}(C) = \{f : f \text{ satisfies (4) with } \|g\|_{W_2^2} \le C \text{ and } \|\gamma\|_{\dot{C}^2} \le C\}.$$

The condition $\|\gamma\|_{\dot{C}^2} \le C$ states that the homogeneous Hölder norm of order 2 is bounded by $C$. In other words, $\gamma$ is differentiable and its first derivative satisfies the Lipschitz condition $|\gamma'(u) - \gamma'(v)| \le C\,|u - v|$. For this class of objects, it can be shown that there are reasonable ways of constructing approximations converging at the rate $n^{-1} \log n$. Monoscale ridgelets do not attain this optimal rate.

## §5. Curvelets and Curved Singularities

The curvelet transform – introduced by Candes and Donoho in [6] – is the last of the representation tools that we will review. Whereas the monoscale ridgelet transform involved taking ridgelet coefficients with a fixed base scale $s$, the curvelet transform spans all possible scales $s \ge 0$. A useful slogan is that the curvelet transform is obtained by filtering and then applying a multiscale ridgelet transform. The multiscale ridgelet dictionary is the collection of the monoscale dictionaries at all possible scales $s \ge 0$; i.e.,

$$\{\psi_\mu := \psi_{Q,\lambda},\ s \ge 0, Q \in \mathcal{Q}_s, \lambda \in \Lambda\}. \tag{10}$$

The curvelet transform requires the use of a sequence of filters that we now describe. Let $\Phi_0$ and $\Psi_{2s}$, $s = 0, 1, 2, \dots$ satisfy the following properties:

- $\Phi_0$ is a lowpass filter and is concentrated at frequencies $|\xi| \le 2$;
- $\Psi_{2s}$ is bandpass and concentrated at frequencies $|\xi| \in [2^{2s-1}, 2^{2s+3}]$;
- the filters satisfy

$$|\hat{\Phi}_0(\xi)|^2 + \sum_{s \ge 0} |\hat{\Psi}_{2s}(\xi)|^2 = 1.$$

Existence and constructions of such filters are well-known. The last relationship implies that the transformation of $f$ into a bank of functions

$$f \mapsto (P_0 f = \Phi_0 * f, \Delta_0 f = \Psi_0 * f, \Delta_1 f = \Psi_1 * f, \dots, \Delta_s f = \Psi_{2s} * f, \dots)$$

is a partial isometry in the sense that

$$\|f\|_2^2 = \|P_0 f\|_2^2 + \sum_{s \ge 0} \|\Delta_s * f\|_2^2.$$

Equipped with both a multiscale ridgelet dictionary and a sequence of filters, define the *curvelet coefficient* $\alpha_\mu$ of $f$ by

$$\alpha_\mu = \langle \Delta_s f, \psi_{Q,\lambda} \rangle, \quad Q \in \mathcal{Q}_s, \lambda \in \Lambda. \tag{11}$$

Thus, the coefficient $\alpha_\mu$ is interpreted as the multiscale ridgelet coefficient of a piece of $f$ containing information at frequencies near $2^{2s}$. We would like to point out that there is a quadratic scaling relationship between the scale $2^s$ of the multiscale ridgelet and the frequency content, localized around the corona of radius $2^{2s}$, of the piece that is analyzed. This relationship is the key feature of the curvelet transform.

We proceed a little bit differently for the piece of $f$ containing information at low frequencies $P_0 f$. Recall the orthogonal collection of Lemarié-Meyer scaling functions $V_k(x_1, x_2) = V(x_1 - k_1, x_2 - k_2)$, for $k = (k_1, k_2) \in \mathbb{Z}^2$. We make the choice of a base scale so that $\hat{V}_0(\xi) = 1$ for $|\xi| \leq 4/3$; and we make sure that the span of the translates $V_k$ contains the range of the projector $P_0 f$. We define the coarse scale curvelet coefficients by

$$\beta_k = \langle P_0 f, V_k \rangle, \quad k \in \mathbb{Z}^2.$$

It will be more convenient to use a single notation to index the set of curvelet coefficients; the notation $M'$ will stand for the union of $M$ and $k \in \mathbb{Z}^2$. When $\mu \in M' \setminus M$, we let $\alpha_\mu = \beta_k$.

**Definition 1.** *The curvelet transform is the mapping that associates the coefficients sequence $\alpha_\mu$, $\mu \in M'$ to an arbitrary square integrable function $f$.*

We will call curvelets those elements $\sigma_\mu$ defined by

$$\sigma_\mu = \Delta_s \psi_{Q,\lambda}, \quad Q \in \mathcal{Q}_s, \lambda \in \Lambda, \tag{12}$$

with an obvious modification for the piece corresponding to the low frequencies, $\sigma_\mu = P_0 V_k$.

The collection of curvelets is then a tight frame for $L_2(\mathbb{R}^2)$

$$\|f\|_2^2 = \sum_{\mu \in M'} \langle f, \sigma_\mu \rangle^2, \tag{13}$$

and, of course, we have the decomposition

$$f = \sum_{\mu \in M'} \langle f, \sigma_\mu \rangle \sigma_\mu \tag{14}$$

with equality in an $L_2$ sense.

Let $f_n$ be the truncated $n$-term curvelet series

$$f_n = \sum_{\mu \in M'} \alpha_\mu 1_{\{|\alpha_\mu| \geq |\alpha|_{(n)}\}} \sigma_\mu. \tag{15}$$

The following theorem is proved in [6].

**Theorem 3.** *Let $g \in W_2^s(\mathbb{R}^2)$ and $f(x) = g(x)\, 1_{\{x_2 \leq \gamma(x_1)\}}$, with $\gamma$ being two times differentiable. Let $f_n$ be the $n$-term approximation (15). Then,*

$$\|f - f_n\|_2 \leq C\, n^{-1} (\log n)^{1/2}.$$

Again, we have a very concrete procedure that achieves rates of approximation that cannot be fundamentally improved. A detailed discussion about the optimality of this result is in [6].

## §6. Conclusion

In this paper, we presented a connected set of ideas originating in the ridgelet transform and culminating in the curvelet transform. We have shown how these representations provide efficient representations of objects that are singular along curves. These tools, however, may have several other potential applications.

Because of space limitations, we set aside questions related to the practicability of these new methods. We would like to point out that fast algorithms have been developed to implement the ridgelet, monoscale ridgelet and curvelet transform. We will report on the numerical aspects of these transforms in a separate paper.

**Acknowledgments.** The author is especially grateful to David Donoho for many fruitful discussions. This research was supported by National Science Fundation grant DMS 98–72890 (KDI) and grant DMS 95–05151 and by AFOSR MURI 95–P49620–96–1–0028

## References

1. Adams, R. A., *Sobolev spaces*, New York, Academic Press, 1975.
2. Candes, E. J., Ridgelets: theory and applications, Ph.D. thesis, Department of Statistics, Stanford University, 1998.
3. Candes, E. J., Harmonic analysis of neural networks, *Appl. Comput. Harmonic Anal.* **6** (1999) 197–218.
4. Candes, E. J., On the representation of mutilated Sobolev functions, Tech. Report, Department of Statistics, Stanford University, 1999.
5. Candes, E. J., Monoscale ridgelets for the representation of images with edges, Tech.Report, Department of Statistics, Stanford University, 1999.
6. Candes, E. J., and D. L. Donoho, Curvelets, Tech. Report, Department of Statistics, Stanford University, 1999.
7. Cohen, A. and I. Daubechies, and P. Vial, Wavelets on the interval and fast wavelet transforms, *Appl. Comput. Harmonic Anal.* **1** (1993), 54–81.
8. Donoho, D. L., Unconditional bases are optimal bases for data compression and for statistical estimation, *Appl. Comput. Harmonic Anal.* **1** (1993), 100–115.
9. Donoho, D. L., Orthonormal ridgelets and linear singularities, Tech. Report, Department of Statistics, Stanford University, 1998.
10. Daubechies, I., Orthonormal bases of compactly supported wavelets, *Commun. Pure Appl. Math.* **41** (1988), 909–996.
11. Daubechies, I., *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
12. Lemarié, P. G., and Y. Meyer, Ondelettes et bases Hilbertiennes, *Rev. Mat. Iberoamericana* **2** (1986), 1–18.
13. Meyer Y., *Ondelettes et Opérateurs*, Hermann, 1990.

Department of Statistics
Stanford University
Stanford, CA 94305-4065
emmanuel@stat.stanford.edu
http://www-stat.stanford.edu/∼emmanuel/