Curvelets and Fourier Integral Operators

Emmanuel Candès , Laurent Demanet

November 18th, 2002

Abstract

A recent body of work introduced new tight-frames of *curvelets* [3, 4] to address key problems in approximation theory and image processing. This paper shows that curvelets essentially provide optimally sparse representations of Fourier Integral Operators.

Dedicated to Yves Meyer on the occasion of his 65th birthday.

1 Introduction

This paper is concerned with the representation of a large class of operators, namely, Fourier Integral Operators (FIOs) in the newly introduced tight frames of *curvelets* [3, 4]. Curvelets are a new multiscale construction for representing bivariate functions and were originally introduced in connection with central problems in approximation theory and statistical estimation; since then, curvelets have also made their way in image processing as an alternative to other classical image representations. Recall that a collection of functions $(f_\mu)_\mu$ is said to be a tight frame for $L_2(\mathbb{R}^2)$ if it obeys the Parseval relation

$$\sum_\mu |\langle f, f_\mu \rangle|^2 = \|f\|^2_{L_2(\mathbb{R}^2)}, \quad \forall f \in L_2(\mathbb{R}^2). \tag{1}$$

This relation implies, by standard arguments, that we have available the reconstruction formula

$$f = \sum_\mu \langle f, f_\mu \rangle f_\mu, \tag{2}$$

with equality holding in an L_2 sense.

To introduce the concept of representation of a linear operator, suppose we are given a linear transformation T acting on square integrable functions $f \in L_2(\mathbb{R}^2)$

$$Tf(x) = \int K(x, y) f(y)\, dy. \tag{3}$$

Alternatively, we may want to think of T via its action on the frame elements f_μ and introduce the infinite matrix

$$T_F(\mu, \mu') = \langle f_\mu, T f_{\mu'} \rangle; \tag{4}$$

in other words, T_F maps the coefficients of an object f into those of Tf. Needless to say, the datum of the matrix $T_F(\mu, \mu')$ completely specifies the operator T as an operator from $L_2(\mathbb{R}^2)$ to itself (extensions beyond L_2 are of course generally possible) since $(f_\mu)_\mu$ is a tight frame for $L_2(\mathbb{R}^2)$.

1.1 Sparsity

Here, we are interested in finding a representation that would provide optimally sparse representations of a wide class of operators in common use in mathematical and numerical analysis. We would like to emphasize that we are not interested in finding the sparsest possible representation of any given operator T of interest. Rather we are interested in finding a single representation that would arguably be nearly the best *simultaneously* over a wide range of operators.

To illustrate this matter, consider the representation of translation-invariant operators. Fourier representations arguably provide very sparse representations as they actually diagonalize such operators. However, this property is very fragile and Fourier analysis is not suitable for a wider class of interesting transformations. Following upon the work of Littlewood and Paley, work in modern harmonic analysis developed new time-scale representations perhaps best known under the name of *wavelet transforms* which were proved to *almost* diagonalize a much richer class of operators, namely, pseudo-differential operators and certain types of Calderón-Zygmund (CZO) [8]. In a nutshell, a CZO is an operator whose kernel K is (1) singular along the diagonal but (2) smooth away from the diagonal. Just as trigonometric systems provide the best representations of translation-invariant operators, wavelets provide, in some sense, fundamentally correct representations of CZOs which facilitate their study and manipulation.

The potential for sparsity is of course wide-ranging and because of space limitations, we shall only discuss its implications for scientific computing. Here, sparsity may allow the design of fast matrix multiplication and/or fast matrix inversion algorithms. For instance, because translation-invariant operators are diagonal in Fourier bases and that one has available a fast algorithm for computing discrete Fourier transforms (FFT), it is possible to apply a vector of size N to such a matrix in $O(N \log N)$ operations. (Consider the extensive literature on the use of Fourier methods (the so-called spectral methods) in numerical analysis.) Further, [1] made a big splash by showing how to use the wavelet transform to compute certain types of singular integrals in a number of operations of the order of $C(\epsilon) \cdot N \log N$ where $C(\epsilon)$ is a constant depending upon the desired accuracy ϵ.

1.2 Fourier Integral Operators

An operator T is said to be a Fourier Integral Operator (FIO) if it is of the form

$$Tf(x) = \int e^{i\Phi(x,\xi)} a(x, \xi) \hat{f}(\xi) \, d\xi. \tag{5}$$

Here Φ is a phase function and a is an amplitude which we suppose obey the following standard assumptions [12]:

- the phase $\Phi(x, \xi)$ is C^∞, homogeneous of degree 1 in ξ, i.e. $\Phi(x, \lambda\xi) = \lambda\Phi(x, \xi)$ for $\lambda > 0$, and with $\Phi_{x\xi} = \nabla_x \nabla_\xi \Phi$, obeys the nondegeneracy condition
$$|\det \Phi_{x\xi}(x, \xi)| > c > 0, \tag{6}$$
uniformly in x and ξ;

- the amplitude a is a symbol of order m, which means that a is C^∞, and obeys
$$|\partial_\xi^\alpha \partial_x^\beta a(x, \xi)| \leq C_{\alpha\beta}(1 + |\xi|)^{m-\alpha}. \tag{7}$$

It is known that both Fourier and wavelet bases do not provide sparse representations of FIOs.

2 Curvelets

By now, there are several types of curvelet frames [3, 4] and we now briefly discuss the curvelet frame as introduced in [5, 4]. We let μ be the triple (j, ℓ, k): here, $j = 0, 1, 2, \dots$ is a scale parameter; $\ell = 0, 1, \dots 2^{\lfloor j/2 \rfloor} - 1$ is an orientation parameter ($\lfloor x \rfloor$ is the integer part of x); and $k = (k_1, k_2)$, $k_1, k_2 \in \mathbb{Z}$ is a translation parameter. Introduce

1. the *parabolic scaling* matrix $D_j = \text{diag}(2^j, 2^{\lfloor j/2 \rfloor})$ which is diagonal and whose entries are 2^j and $2^{\lfloor j/2 \rfloor}$;

2. the *rotation angle* $\theta_J = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \cdot \ell$, with J indexing the scale/angle pair $J = (j, \ell)$;

3. and the *translation parameter* $k_\delta = (k_1 \cdot \delta_1, k_2 \cdot \delta_2)$ (δ_1, δ_2 are some universal numerical quantities, e.g. $\delta_1 = 14/3$ and $\delta_2 = 10\pi/9$, see [4] for details).

With these notations, we define curvelets as functions of $x \in \mathbb{R}^2$ by
$$\gamma_\mu(x) = 2^{3j/4} \gamma_{(j)} \left(D_j R_{\theta_J} x - k_\delta \right). \tag{8}$$

The profile $\gamma_{(j)}$ actually depends on the scale parameter j but in a non essential way; for each j, $\gamma_{(j)}$ is smooth and oscillatory in the horizontal direction and bell-shaped (nonoscillatory) along the vertical direction and is well-localized in space; e.g. for each $m = 0, 1, \dots$, $\gamma_{(j)}(x)$ and its derivatives obey $|\gamma_{(j)}(x)| \leq C_m \cdot (1 + |x|)^{-m}$, uniformly in j.

The frequency-side description of a curvelet is equally important to understand our main results. In the frequency domain, curvelets are compactly supported and each element $\hat{\gamma}_\mu$ is localized near the symmetric wedge
$$W_J = \{\pm\xi,\ 2^j \leq |\xi| \leq 2^{j+1},\ |\theta - \theta_J| \leq \pi \cdot 2^{-\lfloor j/2 \rfloor}\}, \tag{9}$$

i.e. curvelets are supported inside symmetric wedges of length about 2^j and width about $2^{j/2}$. This explains their oscillatory nature: at scale 2^{-j}, a curvelet is a little needle whose envelope is a specified 'ridge' of effective length $2^{-j/2}$ and width 2^{-j}, and which displays an oscillatory behavior across the main 'ridge'.

As in wavelet theory, we also have coarse scale elements which are of the form $\varphi_{k_1,k_2}(x) = \varphi(x - k_\delta)$, $k_1, k_2 \in \mathbb{Z}$, i.e. translates of a waveform $\varphi(x_1, x_2)$ that we shall take to be bandlimited and rapidly decaying. Augmented with this layer of coarse scale elements, the system $(\gamma_\mu)_\mu$ obeys the Parseval relation (1) and the reproducing formula (2).

3 Microlocal Correspondence

A distinguished feature of the curvelet transform is that the action of an FIO on curvelet elements is in some sense very "simple." Roughly speaking, a curvelet γ_μ is mapped into another curvelet at a corresponding index $t(\mu)$. To be more specific, an FIO induces a mapping $\mu \mapsto t(\mu)$ with the property that the significant coefficients of $T\gamma_\mu$ are located at indices μ' near the index $t(\mu)$ (the next section will introduce a notional distance on our index set).

There are many ways to establish a formal index correspondence and we only present a possible approach. Let γ_μ be a curvelet with scale 2^{-j_μ}, with codirection θ_μ and location x_μ (which we may formally define as e.g. $x_\mu = (D_j R_{\theta_J})^{-1} k_\delta$) and T be an FIO with phase Φ. Set $\xi_\mu = (\cos\theta_\mu, \sin\theta_\mu)$ to be the unit vector in the direction θ_μ—so that in frequency, $\hat{\gamma}_\mu$ is localized near $\{\xi, |\xi/|\xi| - \xi_\mu| \le \pi \cdot 2^{-\lfloor j/2 \rfloor}\}$—and define

$$\phi_\mu(x) = \Phi_\xi(x, \xi_\mu), \quad \text{and} \quad y_\mu = \phi_\mu^{-1}(x_\mu), \tag{10}$$

(note that ϕ_μ is a diffeomorphism). Letting A_μ be the derivative with respect to the space variable of $\phi_\mu(x)$ evaluated at the point y_μ, i.e. $A_\mu = \Phi_{x\xi}(y_\mu, \xi_\mu)$, we then define τ_μ by

$$\tau_\mu = A_\mu^T \xi_\mu / \|A_\mu^T \xi_\mu\|.$$

With these notations, we introduce the index mapping t defined as follows: $\mu' = t(\mu)$ with (1) $j_{\mu'} = j_\mu$, (2) $\xi_{\mu'}$ is the closest point to τ_μ on our discrete lattice, and (3) $x_{\mu'}$ is the closest point to y_μ on the Cartesian lattice induced by the pair $(j_{\mu'}, \theta_{\mu'})$. Although there exist more sophisticated mappings, our microlocal correspondence provides a simple description which is sufficient for our exposition.

4 Main result

We begin by introducing a notional distance ω between pairs of indices (μ, μ'):

$$|\omega(\mu,\mu')| = 2^{|j_\mu - j_{\mu'}|} \cdot \left(1 + \min(2^{j_\mu}, 2^{j_{\mu'}}) \left[|\theta_\mu - \theta_{\mu'}|^2 + |x_\mu - x_{\mu'}|^2 + |\langle \xi_\mu, x_\mu - x_{\mu'} \rangle|\right]\right). \tag{11}$$

We see that ω increases as the distance between the scale, angular, and location parameters increases. Note that the extra term $|\langle \xi_\mu, x_\mu - x_{\mu'} \rangle|$ induces a non-Euclidean

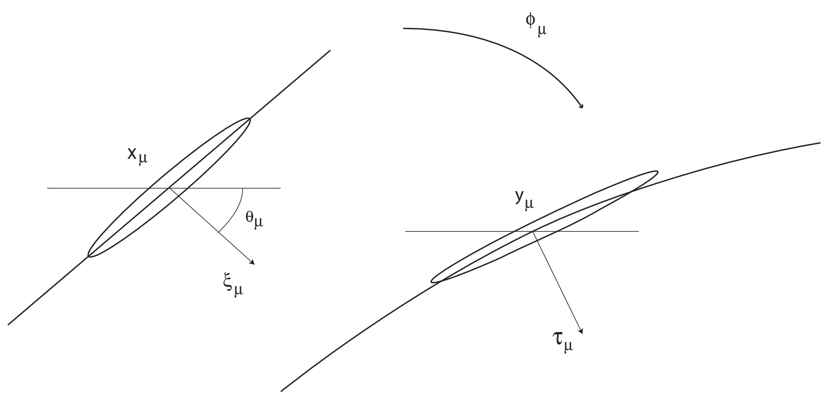

Figure 1: Microlocal correspondence and associated notation. In some, sense, a curvelet localized near x_μ and with codirection ξ_μ, is mapped into a curvelet localized near y_μ and with codirection τ_μ.

notion of distance between x_μ and $x_{\mu'}$. Equipped with this definition, we may now state the main results of this paper.

Theorem 4.1 *Let T be a FIO with a symbol of order $m = 0$ so that T is a bounded L_2-operator which obeys the above assumptions. Then for each $N \geq 0$, the matrix entries in a curvelet frame obey*

$$|T_F(\mu, \mu')| \leq C_N \cdot |\omega(\mu, t(\mu'))|^{-N}, \tag{12}$$

for some constant $C_N > 0$.

The specialist will see how to adapt the result to arbitrary symbol orders. A corollary of this results is as follows:

Theorem 4.2 *Under the same assumptions of Theorem 4.1, the matrix T maps boundedly ℓ_p into ℓ_p for every $0 < p \leq \infty$. For $p \leq 1$, this says that*

$$\|T\|^p_{\ell_p \to \ell_p} = \sup_{\mu'} \sum_{\mu} |T_{\mu,\mu'}|^p \leq C^p. \tag{13}$$

We give an equivalent formulation of (13). Letting (a_μ) be either a row or a column of A, and let $|a|_{(n)}$ be the n-largest entry of the sequence $|a_\mu|$, then for each $r > 0$, $|a|_{(n)}$ obeys

$$|a|_{(n)} \leq C_r \cdot n^{-r}. \tag{14}$$

with a constant which does not depend on the row or column index. In short, the row or column entries of the matrix A_0 decay nearly exponentially, i.e. faster than any negative polynomial.

The above two theorems say that the curvelet matrix is both sparse and well-organized. Roughly speaking, if we think about the wavelet matrix of a CZO as being *almost* diagonal, then we may think of the curvelet matrix of an FIO as being *almost* a permutation. Mathematically speaking, consider the approximate or 'compressed' operator T^B with at most B elements per row and column—those indices which are the closest to $t(\mu)$ in the sense of (11)—and setting the others to zero. Then an application of Schur's lemma shows that T^B would obey

$$\|T - T^B\|_{L_2 \to L_2} \le C_m \cdot B^{-m}, \tag{15}$$

for each $m > 0$. In plain English, whereas the Fourier or wavelet representations are dense, the estimate (15) says that in the curvelet domain, there are very sparse matrices which are arbitrarily close to the action of fairly general FIOs.

5 Why does this work?

Because of space constraints, we cannot possibly give a proof of Theorems 4.1 and 4.2. However, this section presents a heuristic argument which explains why we may expect an FIO to be sparse in a curvelet frame.

Let γ_μ be a fixed curvelet. With the same notations as before, we decompose the phase of our FIO as

$$\Phi(x,\xi) = \Phi_\xi(x,\xi_\mu) \cdot \xi + \delta(x,\xi), \quad \phi_\mu(x) = \Phi_\xi(x,\xi_\mu). \tag{16}$$

In effect, the above decomposition 'linearizes' the frequency variable and is classical, see [9, 12]. With these notations, we may rewrite the action of T on our curvelet γ_μ as

$$(T\gamma_\mu)(x) = \int e^{i\phi_\mu(x)\cdot\xi} e^{i\delta(x,\xi)} a(x,\xi) \hat{\gamma}_\mu(\xi)\, d\xi. \tag{17}$$

Now for a fixed value of the parameter μ, we introduce the decomposition

$$T = T_{2,\mu} T_{1,\mu},$$

where with $b(x,\xi) = e^{i\delta(\phi_\mu^{-1}(x),\xi)} a(\phi_\mu^{-1}(x),\xi)$,

$$(T_{1,\mu} f)(x) = \int e^{ix\cdot\xi} b(x,\xi) \hat{f}(\xi)\, d\xi, \quad (T_{2,\mu} f)(x) = f(\phi_\mu(x)). \tag{18}$$

In effect, this decomposition allows the separate study of the nonlinearities in frequency ξ and space x in the phase function Φ. The point is that both $T_{1,\mu}$ and $T_{2,\mu}$ are sparse in a curvelet tight frame—only for very different reasons.

- Because the frequency support of a curvelet γ_μ is effectively a thin dyadic rectangle of sidelengths about 2^j in the direction ξ_μ and about $2^{j/2}$ in the orthogonal direction, the "phase perturbation" δ is small over such rectangles. Therefore, the term $b(x,\xi)$ is essentially nonoscillatory over the support of $\hat{\gamma}_\mu$ and $T_{1,\mu}$ behaves like a pseudo-differential (local) operator. The output $T_{1,\mu}\gamma_\mu$ is then essentially a curvelet at the same scale, orientation and location. Note that this frequency localization idea is known in the literature as the the Second Dyadic Decomposition, see [12] and references therein.

- To understand why a smooth change of coordinates is a sparse mapping, we need to use the spatial localization of curvelet elements. Roughly speaking, a curvelet is an oscillatory needle with length about $2^{-j/2}$ and width 2^{-j} and qualitatively, it is clear that a warped curvelet would also be just that; i.e. an oscillatory needle whose width approximately equals the square of its length. Mathematically speaking, [2] proves that the 'warped curvelet' $\gamma_\mu \circ \phi_\mu$ is nearly a curvelet in the sense that the coefficients of its expansion are ℓ_p-summable for any $p > 0$.

There is a very interesting phenomenon which occurs here and we now highlight. Instead of curvelets, we may want to consider general scaling matrices of the form $D_j = \text{diag}(2^j, 2^{j\alpha})$, $0 \leq \alpha \leq 1$. We would then obtain tight frames whose elements would be needles with length about $2^{-j\alpha}$ and width 2^{-j}. We could then consider representing an FIO with basis elements exhibiting such arbitrary scaling ratios.

- T_1 is sparse if and only if the frequency support of our anisotropic elements are supported near elongated rectangles with a scaling ratio obeying $\alpha \leq 1/2$.
- While T_2 is sparse if and only if the effective support of our anistropic elements is not too elongated and obeys $\alpha \geq 1/2$.

To fix ideas, suppose on the one hand that $\alpha = 1$, which essentially gives tight frames of wavelets. Then in a wavelet tight-frame, $T_{2,\mu}$ would be sparse but $T_{1,\mu}$ would not because wavelets do not have a sufficiently fine frequency localization. On the other hand, suppose that $\alpha = 0$ which essentially gives tight frames of ridgelets. Then $T_{1,\mu}$ would be sparse but not $T_{2,\mu}$ as a warped ridgelet does not look like another ridgelet. The parabolic scaling $\alpha = 1/2$ is the only scaling for which both operators are sparsified *simultaneously*.

6 Discussion

While working on this project, we became aware of the work of Smith [10, 11] which addresses topics such as the description of Hardy spaces for FIOs and the construction of parametrices for nonsmooth second-order linear wave equations. Especially, [11] alludes to estimates similiar to those developed in Theorem 4.1 although we have not been able to find proofs of such results, see also [6]. We find the connection with this line of research in pure harmonic analysis nevertheless stimulating. Our agenda is of course very different here and corresponds to the viewpoint of Computational Harmonic Analysis: namely, we are interested in a remarkable mathematical statement which says that curvelets provide optimally sparse representations of FIOs. As we mentioned earlier, the consequences are far reaching and we now briefly explore some of them.

The development and study of FIOs is motivated by the connection with the field of partial differential equations. It is well-known that FIOs are, in some sense, almost solution operators of linear hyperbolic systems [13]. Therefore, the results presented here raise a tantalizing perspective: *for large classes of hyperbolic PDEs, the solution operator is sparse in a curvelet frame,* while at the same time, such operators are of course known to be dense in Fourier bases or in classical muliscale systems such as wavelets. Work in progress attempts to exploit this feature to develop fast multiscale solvers (based on fast digital curvelet transforms) for linear hyperbolic problems. We hope to report on this in a future publication.

This research was supported by a National Science Foundation grant DMS-0140540 and by an Alfred P. Sloan Fellowship. E. C. thanks the Institute for Pure and Applied Mathematics at UCLA and especially Mark Green and Eilish Hathaway for their warm hospitality.

References

[1] G. Beylkin, R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms, *Comm. on Pure and Appl. Math.*, **44** (1991), 141-183

[2] E. Candès, L. Demanet, Curvelets, warpings and optimally sparse representations of Fourier Integral Operators, *Manuscript*, (2002).

[3] E. Candès, D. Donoho, Curvelets - a suprisingly effective nonadaptive representation for objects with edges, in *Curves and Surface Fitting : Saint-Malo 1999*, A. Cohen, C. Rabut, L. Schumaker (eds.), Vanderbilt University Press, Nashville, 2000, 105-120

[4] E. Candès, D. Donoho, New tight Frames of curvelets and optimal representations of objects with C^2 singularities, *Submitted*, `http://www.acm.caltech.edu/~emmanuel/publications.html`, (2002)

[5] E. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction, *Signal Processing*, **82** (2002), 1519–1543.

[6] A. Cordoba, C. Fefferman, Wave packets and Fourier integral operators, *Comm. PDE's*, **3(11)** (1978), 979-1005

[7] C. Fefferman, A note on spherical summation multipliers, *Israel J. Math.* **15** (1973) 44-52

[8] Y. Meyer, *Ondelettes et Opérateurs: II. Opérateurs de Calderón Zygmund*, Hermann, Paris, 1990.

[9] A. Seeger, C. Sogge, E. Stein, Regularity properties of Fourier integral operators, *Annals of Math.* **134** (1991), 231-251

[10] H. Smith, A Hardy space for Fourier Integral Operators, *J. Geom. Anal.* **7** (1997)

[11] H. Smith, A parametrix construction for wave equations with $C^{1,1}$ coefficients, *Ann. Inst. Fourier*, Grenoble, **48**, 3 (1998), 797-835

[12] E. Stein, *Harmonic Analysis*, Princeton University Press, Princeton NJ, 1993

[13] F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators*, Plenum press, 1982, 2 volumes.