Curvelets, Warpings, and Optimal Representations of Fourier Integral Operators

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

November 18, 2002

Abstract

This technical report is a companion to [4]. We prove that Fourier Integral Operators admit 'optimally' sparse representations in curvelet frames. We give all the necessary definitions and notations for the discussion to be self-contained.

Keywords. Curvelets, atomic decompositions, sparsity, second dyadic decomposition, Fourier Integral Operators.

Acknowledgments. 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.

1 Definitions and Notations

We adopt the following convention for the Fourier transform and its inverse,

$$\hat{f}(\xi) = \int e^{-ix\cdot\xi} f(x)\, dx,$$
$$f(x) = \frac{1}{(2\pi)^2} \int e^{ix\cdot\xi} \hat{f}(\xi)\, d\xi.$$

The inner product between two functions of $L^2(\mathbb{R}^n, \mathbb{C})$ is written $\langle f, g\rangle = \int f(x)\overline{g(x)}dx$. We denote inequalities up to a multiplicative constant by $A \le C \cdot B$ or $A = O(B)$. The constant C might vary from line to line and comes with or without subscripts to highlight its dependence on the relevant parameters. We use the multi-index formalism for derivatives and powers, $x^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ and $\partial_x^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \dots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$. In addition, $|\alpha| = \alpha_1 + \dots + \alpha_n$.

Fourier Integral Operators

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

$$Tf(x) = \int e^{i\Phi(x,\xi)} a(x, \xi) \hat{f}(\xi)\, d\xi. \tag{1.1}$$

Here Φ is a phase function and a is an amplitude which we suppose obey the following standard assumptions [21]:

- 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
$$|\det \Phi_{x\xi}(x, \xi)| > c > 0, \tag{1.2}$$
uniformly in x and ξ;
- the amplitude a is a symbol of order m, which means that a is C^∞, and obeys
$$|\partial_\xi^\alpha \partial_x^\beta a(x, \xi)| \le C_{\alpha\beta}(1 + |\xi|)^{m-\alpha}. \tag{1.3}$$

In the remainder of this paper, we will only consider symbols of order $m = 0$ although is it clear that our main result, namely, Theorem 5.3 adapts to any symbol order. Under these assumptions, it is known that T maps the Schwartz class $\mathcal{S}$ to $\mathcal{S}$ continuously and is bounded as an L_2-operator, see [21]. We may then express T in any orthonormal basis or tight frame of L_2,

$$\langle Tf, \gamma_\mu\rangle = \sum_{\mu'} \langle T\gamma_{\mu'}, \gamma_\mu\rangle \langle f, \gamma_{\mu'}\rangle.$$

In other words, the associated operator

$$T(\mu, \mu') = \langle f_\mu, Tf_{\mu'}\rangle; \tag{1.4}$$

maps the coefficients of an object f into those of Tf.

In this paper, we shall be interested in the decomposition of Fourier Integral Operators in curvelet tight frames of $L_2(\mathbb{R}^2)$. We would like to emphasize that there is nothing special about the dimension $d = 2$. It is clear that our definition of tight frames extends to arbitrary dimensions and that all of our main results, namely, Theorems 5.1, 5.2 and 5.3 would hold in that setting as well.

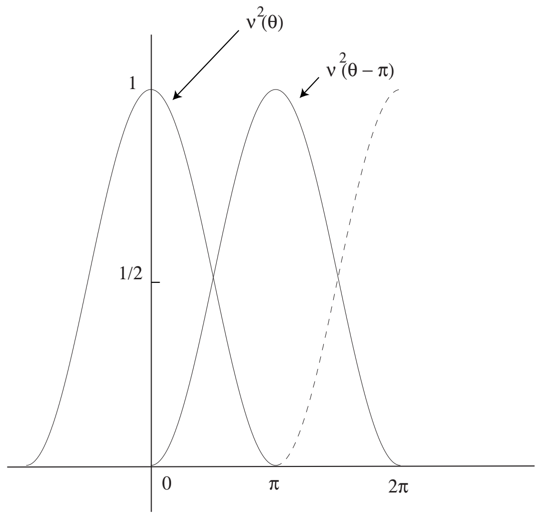

Figure 1: Basic angular window.

2 Second Generation of Curvelets

This section introduces tight frames. Unlike the original curvelet transform [5], this construction does not use ridgelets. Our exposition is taken from [6].

2.1 Scale/Angle Localization

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

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

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

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

where in the above equation, it is understood that we take the 2π-periodization of the function ν, see Figure 1. It is not hard to deduce from our assumptions that for each $j \geq 0$,

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

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

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

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

A possible choice is to select w as in the construction of Meyer wavelets [14, 15], namely C^∞ and supported on $[0,4\pi/3]$ and $[2\pi/3, 8\pi/3]$ respectively. In the remainder of this paper, we will assume this special choice of window.

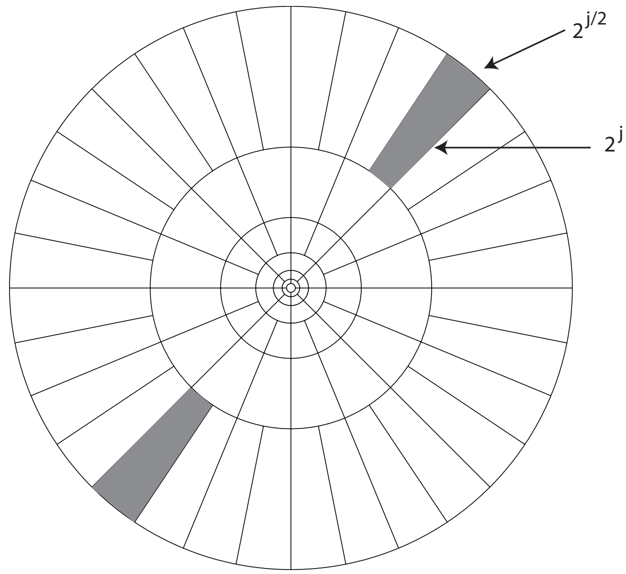

Figure 2: Curvelet Tiling of the Frequency Plane. In the frequency domain, curvelets are supported near symmetric 'parabolic' wedges. The shaded area represents such a generic wedge.

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

$$|\chi_0(\xi)|^2 + \sum_{j \geq 1} \sum_{\ell=0}^{2^{\lfloor j/2 \rfloor}-1} |\chi_{j,\ell}(\xi)|^2 = 1. \tag{2.5}$$

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

$$W_{j,\ell} = \{\pm\xi,\ 2^j \leq |\xi| \leq 2^{j+1},\ |\theta - \pi \cdot \ell \cdot 2^{-\lfloor j/2 \rfloor}| \leq \frac{\pi}{2} 2^{-\lfloor j/2 \rfloor}\}, \tag{2.6}$$

and note that for each ℓ, $\chi_{j,\ell}$ is obtained from $\chi_{j,0}$ by applying a rotation. This corresponds to splitting dyadic annuli at *every other scale*. Figure 2 gives a graphical representation of theses wedges and associated tiling.

2.2 New Tight Frames of Curvelets

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

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

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

R_j is symmetric around the axis $\theta = 0$. We will write the length L_j and width l_j as $L_j = \delta_1 \pi 2^j$ and $l_j = \delta_2 2\pi 2^{\lceil j/2 \rceil}$. It is not difficult to verify that our assumptions about localizing windows imply that δ_1 and δ_2 obey $\delta_1 = 2(1 + O(2^{-j/2}))$ and $\delta_2 = 10\pi/9$ respectively.

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

$$u_{j,k}(\xi_1, \xi_2) = \frac{2^{-3j/4}}{2\pi\sqrt{\delta_1\delta_2}} e^{\mathbf{i}(k_1+1/2)2^{-j}\xi_1/\delta_1}\, e^{\mathbf{i}k_2 2^{-\lceil j/2 \rceil}\xi_2/\delta_2}, \quad k_1, k_2 \in \mathbb{Z}, \tag{2.7}$$

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

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

$$\hat{\gamma}_{\mu'}(\xi) = (2\pi) \cdot \chi_J(\xi) u_{j,k}(R^*_{\theta_J}\xi), \quad \mu' = (j, \ell, k). \tag{2.8}$$

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

Observe that

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

since by construction $(u_{jk}(R^*_{\theta_J}\xi))_k$ is an orthobasis over the support of χ_J. It then follows from (2.5) that for any $F \in L_2(\mathbb{R}^2)$,

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

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

$$\sum_\mu |\langle f, \gamma_\mu \rangle|^2 = \|f\|^2_{L_2(\mathbb{R}^2)}. \tag{2.9}$$

This last equality says $(\gamma_\mu)_{\mu\in M}$ is a tight frame and standard arguments imply that any $f \in L_2(\mathbb{R}^2)$ can be expanded as

$$f = \sum_\mu \langle f, \gamma_\mu \rangle \gamma_\mu, \tag{2.10}$$

with equality holding in an L_2-sense.

We would like to remark that the construction presented here was rapidly introduced by Candès and Guo in [7]. Since the redaction of that paper, Candès became aware of the work of Smith. In [20], Smith introduces a tight frame which is nearly identical to that described above for the purpose of studying parametrices of second-order linear wave equations.

2.3 Space-Side Picture

The point of our construction is that curvelets are real-valued objects. Indeed, let γ_j be the inverse Fourier transform of $\frac{2^{-3j/4}}{\sqrt{\delta_1\delta_2}}\chi_{j,0}(\xi)e^{\mathbf{i}\frac{2^{-j}\xi_1}{2\delta_1}}$. This function is real-valued and

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

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

$$\gamma_j(x) = 2^{3j/4}\gamma^{(j)}(D_j x)$$

where D_j is the diagonal matrix

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

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

$$\gamma_{j,0,k}(x) = 2^{3j/4}\gamma_j(D_j x - k_\delta).$$

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

$$\gamma_\mu(x) = 2^{3j/4}\,\gamma^{(j)}(D_j R^*_{\theta_J} x - k_\delta). \tag{2.12}$$

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

We conclude this section with a brief summary of the main points of the curvelet transform:

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

2.4 Complex-Valued Curvelets

Although our definition implies that curvelets are real valued, we may also introduce complex-curvelets as follows: instead of tiling the frequency plane with pairs of symmetric wedges, we may actually consider tilings with single wedges. Specifically, we may choose windows of the form

$$\chi_{j,\ell}(\xi) = w(2^{-j}|\xi|)\,\nu_{j,\ell}(\theta). \tag{2.13}$$

Now, for $\ell = 0$, the support of this window is contained in the single rectangle $R_j = I_{1j} \times I_{2j}$. We then slightly modify our definitions of orthobases which are now $e^{\mathbf{i}2\pi k_1\xi_1/L_j}/\sqrt{L_j}$ for $L_2(I_{1j})$ and $e^{\mathbf{i}2\pi k_2\xi_2/l_j}/\sqrt{l_j}$ for $L_2(I_{2j})$. The sequence $(u_{j,k})_{k\in\mathbb{Z}^2}$ defined as

$$u_{j,k}(\xi_1,\xi_2) = \frac{2^{-3j/4}}{2\pi\sqrt{\delta_1\delta_2}} e^{\mathbf{i}k_1 2^{-j}\xi_1/\delta_1}\, e^{\mathbf{i}k_2 2^{-\lceil j/2\rceil}\xi_2/\delta_2}, \quad k_1, k_2 \in \mathbb{Z}, \tag{2.14}$$

is then an orthobasis for $L_2(R_j)$. The complex-valued curvelets are similarly defined with (2.8), and also constitute a tight frame.

3 Atomic Decompositions

As we will see later, to prove our main result and especially Theorem 5.2, it would be most helpful to work with tight frames of curvelet compactly supported in space. Unfortunately, it is unclear at this point how to construct such tight frames with nice frequency localization properties. However, there exist useful atomic decompositions with compactly supported curvelet-like atoms. We now explore such decompositions.

In this section, the notation $f_{a,\theta}$ refers to the function obtained from f after applying a parabolic scaling and rotation

$$f_{a,\theta}(x) = a^{3/4} f\left(D_a R_\theta x\right), \quad D_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix},$$

and where R_θ is the rotation matrix which maps the vector $(1,0)$ into $(\cos\theta, -\sin\theta)$. Note that this is an isometry as

$$\|f_{a,\theta}\|_{L_2} = \|f\|_{L_2}.$$

In [19], Smith proved the following result: let ϕ be a Schwartz function obeying $\hat{\varphi}(1,0) \neq 0$; then one can find another Schwartz function ψ, and a function $q(\xi)$ such that the following formula holds

$$q(\xi) \int_{a\geq 1} \hat{\phi}_{a,\theta}(\xi) \hat{\psi}_{a,\theta}(\xi)\, a da d\theta = r(\xi); \tag{3.1}$$

here r is a smooth cut-off function obeying

$$r(\xi) = \begin{cases} 1 & |\xi| \geq 2 \\ 0 & |\xi| \leq 1 \end{cases},$$

and q is a standard Fourier multiplier of order zero; that is, for each $m \geq 0$, there exists a constant C_m such that

$$|D^m q(\xi)| \leq C_m (1 + |\xi|^2)^{-m/2}.$$

This formula is useful because it allows us to express any object whose Fourier transform vanishes on $\{|\xi| \leq 2\}$ as a continuous superposition of curvelet-like elements. We now make some specific choices for ϕ. In the remainder of this section, we will take $\phi(x) = \psi(-x)$ and the function ψ of the form

$$\psi(x_1, x_2) = \psi^D(x_1)\, \varphi^D(x_2), \tag{3.2}$$

where both φ^D and ψ^D are compactly supported and obey

$$\operatorname{Supp} \varphi^D \subset [0,1], \quad \operatorname{Supp} \psi^D \subset [0,1].$$

We will assume that φ^D and ψ^D are C^∞ and that the function ψ^D has vanishing moments up to order D, i.e.

$$\int \psi^D(x_1) x_1^k \, dx_1 = 0, \quad k = 0, 1, \ldots, D. \tag{3.3}$$

For each $a \geq 1$, each $b \in \mathbb{R}^2$ and each $\theta \in [0, 2\pi)$, introduce

$$\psi_{a,\theta,b}(x) := \psi_{a,\theta}(x - b) = a^{3/4} \psi\left(D_a R_\theta (x - b)\right); \tag{3.4}$$

and given an object f, define coefficients by

$$\mathcal{R}(f)(a,b,\theta) = \int \overline{\psi}_{a,\theta,b}(x) f(x) dx. \tag{3.5}$$

Now, suppose for instance that $\hat{f}$ vanishes over $|\xi| \leq 2$, then (3.1) gives the exact reconstruction formula

$$f(x) = \int_{a \geq 1} \mathcal{R}(q(D)f))(a,b,\theta)\psi_{a,\theta,b}(x)\mu(dad\theta db), \tag{3.6}$$

with $\mu(dad\theta db) = adad\theta db$. In the remainder of this section, we will use the shorter notation $d\mu$ for $\mu(dad\theta db)$.

As is now well-established, the reproducing formula may be turned into a so-called 'atomic decomposition'. Not surprisingly, our atomic decomposition will just mimic the discretization of the curvelet frame as introduced in section 2. With the notations of that section, we introduce the cells Q_μ defined as follows: for $j \geq 0$, $\ell = 0, 1, \ldots 2^{\lfloor j/2 \rfloor} - 1$ and $k = (k_1, k_2) \in \mathbb{Z}^2$, the cell Q_μ is the collections of triples (a, θ, b) for which

$$2^j \leq a < 2^{j+1}, \quad |\theta - \theta_J| \leq \frac{\pi}{2} 2^{-\lfloor j/2 \rfloor}$$

and

$$D_j R_{\theta_J} b \in [k_1, k_1 + 1) \times [k_2, k_2 + 1).$$

Note that $\int_{Q_\mu} d\mu = 3\pi/2$ for j even, and 3π for j odd. We may then break the integral (3.6) into a sum of terms arising from different cells, namely,

$$f(x) = \sum_\mu \alpha_\mu m_\mu(x) \tag{3.7}$$

where

$$\alpha_\mu = \|\mathcal{R}(q(D)f)\|_{L_2(Q_\mu)}, \quad m_\mu(x) = \frac{1}{\alpha_\mu} \int_{Q_\mu} \mathcal{R}(q(D)f))(a,b,\theta)\psi_{a,\theta,b}(x)\, d\mu. \tag{3.8}$$

Of course, the decomposition (3.7) greatly resembles the tight frame expansion, compare (2.9). In particular, the atoms m_μ are curvelet-like in the sense that they share all the properties of the tight frame (γ_μ) –only they are compactly supported in space. In the remainder of the paper, we will call these elements *atoms.* Below are some crucial properties of these atoms.

Lemma 3.1. *Rewrite the atoms m_μ as $m_\mu(x) = 2^{3j/4} a^{(\mu)}(D_j R_{\theta_J} x - k)$ (note the resemblance with (2.12)). In other words, m_μ is obtained from $a^{(\mu)}$ after parabolic scaling, rotation, and translation. For all μ, the $a^{(\mu)}$'s verify the following properties.*

- *Compact support;*
$$Supp\ a^{(\mu)} \subset cQ. \tag{3.9}$$
- *Nearly vanishing moment along the horizontal axis; let $m = D/2$. Then for each $k = 0, 1, \ldots, m$, there is a constant C_m such that*
$$\int a^{(\mu)}(x_1, x_2) x_1^k \, dx_1 \leq C_m \cdot 2^{-j(m+1)}. \tag{3.10}$$
- *Regularity; for every $\alpha \in \mathbb{Z}_+^2$*
$$|D^\alpha a^{(\mu)}(x)| \leq c_\alpha. \tag{3.11}$$

In (3.10) and (3.11), the constants may be chosen independently of μ and f.

Proof of Lemma. By definition $a^{(\mu)}(x) = 2^{-3j/4} m_\mu\,(D_j R_{\theta_J} x - k)$ and, therefore,

$$\begin{aligned} a^{(\mu)}(x) &= \frac{1}{\alpha_\mu} \int (Rf)(a,\theta,b) a^{3/4} 2^{-3j/4} \psi(D_a R_\theta (R_{\theta_J}^{-1} D_j^{-1}(x+k) - b)\, d\mu \\ &= \frac{1}{\alpha_\mu} \int (Rf)(a,\theta,b) |A|^{1/2} \psi\,(A(x - (\beta - k))\, d\mu, \qquad (3.12) \end{aligned}$$

where $A = D_a R_\delta D_j^{-1}$ with $\delta = \theta - \theta_J$ and $\beta = D_j R_{\theta_J} b$.

Let us first verify the assertion about the support of $a^{(\mu)}$. Recall that over a cell Q_μ, $\beta \in [k_1, k_1 + 1) \times [k_2, k_2 + 1)$, and hence for all $b \in Q_\mu$, we have

$$\operatorname{Supp} \psi\,(A(x - (\beta - k))) \subset \operatorname{Supp} \psi(Ax) + [0,1]^2.$$

Next $\operatorname{Supp} \psi(Ax) \subset A^{-1}[0,1]^2$ with $A^{-1} = D^j R_{-\delta} D_a^{-1}$. It is not difficult to check that $A^{-1}[0,1]^2 \subset [c_1, c_2) \times [d_1, d_2)$ which then gives (3.9).

There are several ways to prove the property about nearly vanishing moments. A possibility is to show that the Fourier transform of $a^{(\mu)}$ is appropriately small in a neighborhood of the axis $\xi_1 = 0$. We choose a more direct strategy and show that

$$\left| \int \psi(A(x - \beta)) x_1^k\, dx_1 \right| \le C_m \cdot 2^{-j(m+1)}. \qquad (3.13)$$

uniformly over the $(a, \theta, b) \in Q_\mu$. The property (3.10) follows from this fact. Indeed,

$$\int a^{(\mu)}(x_1, x_2)\, x_1^k\, dx_1 = \frac{1}{\alpha_\mu} \int_{Q_\mu} Rf(a,\theta,b) d\mu \int |A|^{1/2} \psi(A(x-\beta)) x_1^k\, dx_1,$$

and the Cauchy-Schwarz inequality gives

$$\begin{aligned} \left| \int a^{(\mu)}(x_1, x_2)\, x_1^k\, dx_1 \right| &\le \frac{1}{\alpha_\mu} \cdot \|Rf\|_{L_2(Q_\mu)} \cdot \left(\int_{Q_\mu} \left| \int |A|^{1/2} \psi(A(x-\beta)) x_1^k\, dx_1 \right|^2 d\mu \right)^{1/2} \\ &= \left(\int_{Q_\mu} \left| \int |A|^{1/2} \psi(A(x-\beta)) x_1^k\, dx_1 \right|^2 d\mu \right)^{1/2}. \end{aligned}$$

The uniform bound (3.13) together with the fact that $\int_{Q_\mu} d\mu$ is either 3π or $3\pi/2$ gives (3.10).

We then need to establish (3.13). Let D be $\partial/\partial x_2$, recall that by assumption (3.2)–(3.3), we have that for all $x_2 \in \mathbb{R}$,

$$\int D^n \psi(x_1, x_2) x_1^k\, dx_1 = 0, \quad k = 0, 1, \ldots, D,$$

and more generally, for each $\alpha \neq 0$ and β

$$\int D^n \psi(\alpha x_1 + \beta, x_2) x_1^k\, dx_1 = 0, \quad k = 0, 1, \ldots, D. \qquad (3.14)$$

We shall use (3.14) to prove (3.13). Letting

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

and with the same notations as before, a simple calculation shows that $a_{21} = -\sqrt{a}2^{-j}\sin\delta$. As $a \le 2^{(j+1)}$ and $|\delta| \le \pi/2 \cdot 2^{-\lfloor j/2 \rfloor}$, we have

$$|a_{21}| \le c \cdot 2^{-j}. \tag{3.15}$$

We then write

$$\begin{aligned}\psi(Ax) &= \psi(a_{11}x_1 + a_{12}x_2, a_{21}x_1 + a_{22}x_2) \\ &= \sum_{n=0}^{N-1} D^n\psi(a_{11}x_1 + a_{12}x_2, a_{22}x_2)\frac{(a_{21}x_1)^n}{n!} + O((a_{21}x_1)^N)\end{aligned}$$

and, therefore,

$$\int \psi(Ax)x_1^k\,dx_1 = \sum_{n=0}^{N-1} \frac{a_{21}^n}{n!}\int D^n\psi(a_{11}x_1 + a_{12}x_2, a_{22}x_2)x_1^{n+k}\,dx_1 + O(a_{21}^N)$$

Fix $k \le D$ and pick $N = D - k + 1$ so that for $n = 0, 1, \ldots, N-1$, $n + k \le D$. By virtue of (3.14) all the integrals in the sum vanish and the only remaining term is $O(a_{21}^N)$ which because of (3.15) is $O(2^{-jN})$. As a consequence, setting $m = D/2$, we conclude that

$$\left|\int \psi(Ax)x_1^k\,dx_1\right| \le C_m \cdot 2^{-j(m+1)}, \quad k = 0, 1, \ldots m;$$

this is the content of (3.13).

Last, the regularity property is a simple consequence of the Cauchy Schwarz inequality;

$$\begin{aligned}\left|a^{(\mu)}(x_1, x_2)\right| &\le \frac{1}{\alpha_\mu}\int |Rf(a,\theta,b)||A|^{1/2}\|\psi\|_{L_\infty}\,d\mu \\ &\le \|\psi\|_{L_\infty} \cdot \frac{1}{\alpha_\mu}\|Rf\|_{L_2(Q_\mu)} \cdot \left(\int_{Q_\mu} |A|d\mu\right)^{1/2} \\ &= 2\sqrt{3\pi} \cdot \|\psi\|_{L_\infty}.\end{aligned}$$

These last inequalities used the facts that $|A| \le 4$ for $(a,\theta,b) \in Q_\mu$ and $\int_{Q_\mu} d\mu \le 3\pi$. Estimates for higher derivatives are obtained in exactly the same fashion –after differentiation of the integrand. This finishes the proof of the lemma. ∎

4 Curvelet Molecules

We introduce the notion of *curvelet molecule*; our objective, here, is to encompass under this name a wide collection of systems which share the same essential properties as the curvelets and curvelet atoms we have just introduced. In some sense, our formulation is inspired by the notion of 'vaguelettes' in wavelet analysis [17]. Our motivation for introducing this concept is the fact that operators of interest do not map curvelets into curvelets or curvelet atoms into curvelet atoms, but rather into these molecules. Note that the terminology 'molecule' is somewhat standard in the literature of harmonic analysis [13].

Definition 4.1. *A family of functions $(m_\mu)_\mu$ is said to be a family of curvelet molecules with regularity R iff (for $j > 0$) they may be expressed as*

$$m_\mu(x) = 2^{3j/4}a^{(\mu)}\left(D_j R_{\theta_J} x - k\right),$$

where for all μ, the $a^{(\mu)}$'s verify the following properties:

- *Smoothness and spatial localization: for each $|\beta| \le R$, and each $m = 0, 1, 2 \ldots$*

$$|\partial_x^\beta a^{(\mu)}(x)| \le C_m \cdot (1 + |x|)^{-m}. \tag{4.1}$$

- *Nearly vanishing moments: for each $n = 0, 1, \ldots, R$, there is a constant C_n such that*

$$|\hat{a}^{(\mu)}(\xi)| \le C_n \cdot \min(1, 2^{-j} + |\xi_1| + 2^{-j/2}|\xi_2|)^n. \tag{4.2}$$

Here, the constants may be chosen independently of μ so that the above inequalities hold uniformly over μ. There is of course an obvious modification for the coarse scale molecules which are of the form $a^{(\mu)}(x - k)$ with $a^{(\mu)}$ as in (4.1).

In short, the definition is similar to that introduced in Lemma (3.1) but for the requirement that $a^{(\mu)}$ be compactly supported. Indeed, the assumption about the vanishing moments (3.10) is nearly equivalent to (4.2). We find the formulation (4.2) to be more concrete as it gives precise size estimates of the Fourier transform of a molecule at low frequencies.

4.1 Interpretation

This definition implies a series of useful estimates. For instance, consider $\theta_J = 0$ so that R_{θ_J} is the identity (arbitrary molecules are obtained by rotations). Then, m_μ obeys

$$|m_\mu(x)| \le C_m \cdot 2^{3j/4} \cdot \Big(1 + |2^j x_1 - k_1| + |2^{j/2} x_2 - k_2|\Big)^{-m} \tag{4.3}$$

for each $m > 0$, and similarly for its derivatives

$$|\partial_x^\beta m_\mu(x)| \le C \cdot 2^{3j/4} \cdot 2^{(\beta_1 + \beta_2/2)j} \cdot \Big(1 + |2^j x_1 - k_1| + |2^{j/2} x_2 - k_2|\Big)^{-m}. \tag{4.4}$$

Another useful property is the almost vanishing moments property which says that in he frequency plane, a molecule is localized near the dyadic corona $\{2^j \le |\xi| \le 2^{j+1}\}$; $|\hat{m}_\mu(\xi)|$ obeys

$$|\hat{m}_\mu(\xi)| \le C \cdot 2^{-3j/4} \cdot \min(1, 2^{-j}(1 + |\xi|))^n, \tag{4.5}$$

which is valid for every $n < R$, which gives the frequency localization

$$|\hat{m}_\mu(\xi)| \le C \cdot 2^{-3j/4} \cdot |S_J(\xi)|^n \tag{4.6}$$

where for $J = (j, 0)$

$$S_J(\xi) = \min(1, 2^{-j}(1 + |\xi|)) \cdot (1 + |2^{-j}\xi_1| + |2^{-j/2}\xi_2|)^{-1}. \tag{4.7}$$

For arbitrary J, S_J is obtained from S_{J_0}, will $J_0 = (j, 0)$ by a simple rotation of angle θ_J, i.e. $S_{J_0}(R_{\theta_J}\xi)$.

In short, a curvelet molecule is a needle whose envelope is supported near a ridge of length about $2^{-j/2}$ and width 2^{-j} and which displays an oscillatory behavior across the ridge.

4.2 Near Orthogonality of Curvelet Molecules

Let μ and μ' be two indices corresponding to angular parameters θ_J (resp. $\theta_{J'}$) and location parameter x_μ (resp. x'_μ). Introduce the notional squared distance

$$d(\mu,\mu') = |\theta_J - \theta_{J'}|^2 + |x_\mu - x_{\mu'}|^2 + |\langle e_J, x_\mu - x_{\mu'}\rangle|, \tag{4.8}$$

where e_J is the codirection of the first curvelet molecule, i.e. $e_J = (\cos\theta_J, \sin\theta_J)$. (The expression may equivalently be written using $e_{J'}$ instead, as shown in the appendix) In (4.8) angles and angle differences are understood modulo π.

Curvelet molecules are not necessarily orthogonal to each other, [1] but in some sense, they are almost orthogonal.

Lemma 4.2. *Let $(m_\mu)_\mu$ and $(p_{\mu'})_{\mu'}$ be two families of curvelet molecules with regularity R. Define ω as*

$$\omega(\mu,\mu') = 2^{|j-j'|} \cdot \left(1 + 2^{\min(j,j')} d(\mu,\mu')\right). \tag{4.9}$$

Then for $j, j' \geq 0$,

$$|\langle m_\mu, p_{\mu'}\rangle| \leq C \cdot \omega(\mu,\mu')^{-n}. \tag{4.10}$$

for every $n \leq D(R)$ where $D(R)$ goes to infinity as R goes to infinity.

Proof. Throughout the proof of (4.10), it will be useful to keep in mind that $A \leq C \cdot (1+|B|)^{-m}$ for every $m \leq 2m'$ is equivalent to $A \leq C \cdot (1+B^2)^{-m}$ for every $m \leq m'$. Similarly, if $A \leq C \cdot (1+|B_1|)^{-m}$ and $A \leq C \cdot (1+|B_2|)^{-m}$ for every $m \leq 2m'$, then $A \leq C \cdot (1 + |B_1| + |B_2|)^{-m}$ for every $m \leq m'$. As noted earlier, the constants may vary from expression to expression.

As in Section 2, we let J be the scale/angle pair (j,ℓ). For notational convenience put $\Delta\theta = \theta_J - \theta_{J'}$ and $\Delta x = x_\mu - x_{\mu'}$. We abuse notation by letting m_J be the molecule $a^{(\mu)}(D_j R_{\theta_J} x)$, i.e. m_J is obtained from m_μ by translation to that it is centered near the origin. Put $I_{\mu\mu'} = \langle m_\mu, p_{\mu'}\rangle$. In the frequency domain, $I_{\mu\mu'}$ is given by

$$I_{\mu\mu'} = \int \hat{m}_J(\xi)\overline{\hat{p}_{J'}(\xi)}\, e^{-i(\Delta x)\cdot\xi}\, d\xi.$$

Put j_0 to be the minimum of j and j'. The Appendix shows that

$$\int |S_J(\xi)\, S_{J'}(\xi)|^n\, d\xi \leq C \cdot 2^{3j/4 + 3j'/4} \cdot 2^{-|j-j'|n} \cdot \left(1 + 2^{j_0}|\Delta\theta|^2\right)^{-n}. \tag{4.11}$$

Therefore, the frequency localization of the curvelet molecules (4.6) gives

$$\begin{aligned} \int |\hat{m}_J(\xi)|\, |\hat{p}_{J'}(\xi)|\, d\xi &\leq C \cdot 2^{-3j/4 - 3j'/4} \cdot \int |S_J(\xi)\, S_{J'}(\xi)|^n\, d\xi \\ &\leq C \cdot 2^{-|j-j'|n} \cdot \left(1 + 2^{j_0}|\Delta\theta|^2\right)^{-n}. \end{aligned} \tag{4.12}$$

This inequality explains the angular decay. A series of integrations by parts will introduce the spatial decay.

The partial derivatives of $\hat{m}_J$ obey

$$|\partial_\xi^\alpha \hat{m}_J(\xi)| \leq C \cdot 2^{-3j/4} \cdot 2^{-j(\alpha_1 + \frac{\alpha_2}{2})} \cdot |S_J(\xi)|^n.$$

[1] We do not know yet whether orthobases of curvelets exist.

Put Δ_ξ to be the usual Laplacian in ξ. Because $\hat{p}_{\mu'}$ is misoriented with respect to e_J, simple calculations show that

$$|\Delta_\xi \hat{p}_{J'}(\xi)| \leq C \cdot 2^{-3j'/4} \cdot 2^{-j'} \cdot |S_{J'}(\xi)|^n,$$
$$|\frac{\partial^2}{\partial \xi_1^2} \hat{p}_{J'}(\xi)| \leq C \cdot 2^{-3j'/4} \cdot (2^{-2j'} + 2^{-j'} |\sin(\Delta\theta)|^2) \cdot |S_{J'}(\xi)|^n.$$

Recall that for $t \in [-\pi/2, \pi/2]$, $2/\pi \cdot |t| \leq |\sin t| \leq |t|$, so we may just as well replace $|\sin(\Delta\theta)|$ by $|\Delta\theta|$ in the above inequality. Set

$$L = I - 2^{j_0} \Delta_\xi - \frac{2^{2j_0}}{1 + 2^{j_0} |\Delta\theta|^2} \frac{\partial^2}{\partial \xi_1^2},$$

On the one hand, for each k, $L^k(\hat{m}_J \overline{\hat{p}_{J'}})$ obeys

$$|L^k(\hat{m}_J \overline{\hat{p}_{J'}})(\xi)| \leq C \cdot 2^{-3j/4 - 3j'/4} \cdot |S_J(\xi)|^n \cdot |S_{J'}(\xi)|^n.$$

On the other hand

$$L^k e^{-i(\Delta x)\cdot\xi} = [1 + 2^{j_0} |\Delta x|^2 + \frac{2^{2j_0}}{1 + 2^{j_0}|\Delta\theta|^2} |\langle e_J, \Delta x\rangle|^2]^k e^{-i(\Delta x)\cdot\xi}.$$

Therefore, a few integrations by parts gives

$$|I_{\mu\mu'}| \leq C \cdot 2^{-|j-j'|n} \cdot \left(1 + 2^{j_0}|\theta_J - \theta_{J'}|^2\right)^{-n} \cdot \left(1 + 2^{j_0}|\Delta x|^2 + \frac{2^{2j_0}}{1 + 2^{j_0}|\Delta\theta|^2} |\langle e_J, \Delta x\rangle|^2\right)^{-n},$$

and then

$$|I_{\mu\mu'}| \leq C \cdot 2^{-|j-j'|n} \cdot \left(1 + 2^{j_0}(|\Delta\theta|^2 + |\Delta x|^2) + \frac{2^{2j_0}}{1 + 2^{j_0}|\Delta\theta|^2} |\langle e_J, \Delta x\rangle|^2\right)^{-n}.$$

One can simplify this expression by noticing that

$$(1 + 2^{j_0}|\Delta\theta|^2) + \frac{2^{2j_0} |\langle e_J, \Delta x\rangle|^2}{1 + 2^{j_0}|\Delta\theta|^2} \gtrsim \sqrt{1 + 2^{j_0}|\Delta\theta|^2} \frac{2^{j_0} |\langle e_J, \Delta x\rangle|}{\sqrt{1 + 2^{j_0}|\Delta\theta|^2}} = 2^{j_0} |\langle e_J, \Delta x\rangle|.$$

This yields equation (4.10) as required. ∎

Remark Assume that one of the two terms or both terms are coarse scale molecules, e.g. $p_{\mu'}$, then the decay estimate is of the form

$$|\langle m_\mu, p_{\mu'}\rangle| \leq C \cdot 2^{-jn} \cdot \left(1 + |x_\mu - x_{\mu'}|^2 + |\langle e_J, x_\mu - x_{\mu'}\rangle|\right)^{-n}.$$

For instance, if they are both coarse scale molecules, this would give

$$|\langle m_\mu, p_{\mu'}\rangle| \leq C \cdot \left(1 + |x_\mu - x_{\mu'}|\right)^{-n}.$$

The following result is a different expression for the almost-orthogonality, and will be at the heart of the sparsity estimates for FIO's.

Lemma 4.3. *Let $(m_\mu)_\mu$ and $(p_\mu)_\mu$ be two families of curvelet molecules with regularity R. Then for each $p > p^*$,*

$$\sup_\mu \sum_{\mu'} |\langle m_\mu, p_{\mu'} \rangle|^p \leq C_p.$$

Here $p^* \to 0$ as $R \to \infty$. In other words, for $p > p^*$, the matrix $I_{\mu\mu'} = (\langle m_\mu, p_{\mu'} \rangle)_{\mu,\mu'}$ acting on sequences (α_μ) obeys

$$\|I\alpha\|_{\ell_p} \leq C_p \cdot \|\alpha\|_{\ell_p}.$$

Proof. Put as before $j_0 = \min(j, j')$. The appendix shows that

$$\sum_{\mu \in M_{j'}} \left(1 + 2^{j_0}(d(\mu, \mu')\right)^{-np} \leq C \cdot 2^{2|j-j'|} \tag{4.13}$$

provided that $np > 2$. We then have

$$\sum_{\mu'} |I_{\mu\mu'}|^p \leq C \cdot \sum_{j' \in \mathbb{Z}} 2^{-2|j-j'|np} \cdot 2^{2|j-j'|} \leq C_p,$$

provided again that $np > 2$.

Hence we proved that for $p \leq 1$, I is a bounded operator from ℓ_p to ℓ_p. We can of course interchange the role of the two molecules and obtain

$$\sup_{\mu'} \sum_{\mu} |\langle m_\mu, p_{\mu'} \rangle|^p \leq C_p.$$

For $p = 1$, the above expression says that I is a bounded operator from ℓ_∞ to ℓ_∞. By interpolation, we then conclude that I is a bounded operator from ℓ_p to ℓ_p for every p. ∎

4.3 Curvelets, atoms and molecules

The picture of curvelet molecules in the frequency plane is exactly that of curvelets : two bumps located on opposite wedges, symmetric with respect to the origin. Obviously, curvelets obey the molecule properties for arbitrary degrees R of regularity. Further, complex-valued curvelets, as introduced in Section 2.4, are supported on a single wedge in the frequency plane and accordingly, they also verify the assumptions of curvelet molecules. They are indeed a very special case of molecules.

Needless to say that the atoms of Section 3 are curvelet molecules with spatial compact support, compare Lemma 3.1 with the definition of a molecule. We conclude this section with an important observation. It is of course possible to decompose a molecule into a series of atoms

$$m_\mu = \sum_{\mu'} \alpha^\mu_{\mu'} \rho_{\mu'}.$$

The coefficients would then obey the same estimate as in Lemma 4.2

$$|\alpha^\mu_{\mu'}| \leq C \cdot |\omega(\mu, \mu')|^{-n}, \tag{4.14}$$

and in particular, for each $p > 0$

$$\sup_\mu \sum_{\mu'} |\alpha^\mu_{\mu'}|^p < A_p.$$

This is briefly justified in the appendix.

5 Main Results

We decompose a Fourier Integral Operator T in a tight frame of curvelets $(\gamma_\mu)_\mu$ as in Section 2. We introduce the curvelet matrix

$$T_{\mu\mu'} = \langle T\gamma_{\mu'}, \gamma_\mu \rangle;$$

we wish to show that this matrix is sparse. Our main result (Theorem 5.3) will prove that for each $p > 0$, T obeys

$$\sup_{\mu'} \sum_\mu |T_{\mu\mu'}|^p \leq B_p.$$

Let γ_μ be a fixed curvelet with codirection θ_μ. 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 $\{\xi, |\xi/|\xi| - \xi_\mu| \leq \pi \cdot 2^{\lceil j/2 \rceil}\}$). With the same notations as in Section 9, we decompose the phase of our FIO as

$$\Phi(x,\xi) = \Phi_\xi(x,\xi_\mu)\cdot\xi + \delta(x,\xi), \quad \phi_\mu(x) = \Phi_\xi(x,\xi_\mu). \tag{5.1}$$

In effect, the above decomposition 'linearizes' the frequency variable and is classical, see [18, 21]. With these notations, we may rewrite the action of T on our curvelet γ_μ as

$$(T\gamma_\mu)(x) = \int e^{i\phi_\mu(x)\cdot\xi} e^{i\delta(x,\xi)} a(x,\xi)\hat{\gamma}_\mu(\xi)\, d\xi. \tag{5.2}$$

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

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

where

$$(T_{1,\mu} f)(x) = \int e^{ix\cdot\xi} b_\mu(x,\xi)\hat{f}(\xi)\, d\xi, \quad (T_{2,\mu} f)(x) = f(\phi_\mu(x)), \tag{5.3}$$

with $b_\mu(x,\xi) = e^{i\delta(\phi_\mu^{-1}(x),\xi)}$. 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.

Theorem 5.1. *Let $(\gamma_\mu)_\mu$ be a tight frame of complex-valued curvelets. T_1 maps this family of curvelets into curvelet molecules with arbitrary regularity R.*

The choice of the curvelet family being complex-valued in the above theorem is not essential. T_1 acting on real curvelets would give rise to *two* molecules, and keeping track of this fact in subsequent discussions would be unnecessarily heavy. In the real case it is clear that the structure and the sparsity of a FIO matrix can be recovered by expressing each real curvelet as a superposition of two complex curvelets.

Theorem 5.2. *Let $(\rho_\mu)_\mu$ be a family of curvelet atoms (compactly supported in space) with regularity R. There is an explicit mapping $\mu \mapsto t(\mu)$ acting on the indices $\mu \in M$, so that T_2 maps each curvelet atom ρ_μ into another atom $\tilde{\rho}_{t(\mu)}$ of the same regularity R.*

The latter theorem says that the 'warped' atom $\rho_\mu \circ \phi_\mu$ is an atom, only its scale, orientation, and location may have been changed. Section 7 will make explicit this index mapping t; at this point, we would like to emphasize that t is not necessarily one-to-one nor onto.

Note that Theorem 5.2 also means that a smooth warping preserves the sparsity of curvelet expansions, which is a result of independent interest.

Our main result follows and is mostly a consequence of Theorems 5.1 and 5.2.

Theorem 5.3. *Let T be a Fourier Integral Operator acting on functions of $\mathbb{R}^2$, as defined in Section 9 and $T_{\mu\mu'}$ be denote its matrix elements in a curvelet tight frame. Then with t the index mapping introduced in Theorem 5.2, the element $T_{\mu\mu'}$ obeys for each $n > 0$*

$$|T_{\mu\mu'}| \le C \cdot |\omega(\mu, t(\mu'))|^{-n},$$

where ω is as in (4.9). Moreover, for every $0 < p \le \infty$, $(T_{\mu\mu'})$ is bounded from ℓ^p to ℓ^p.

The remaining 3 sections are devoted to the proofs of Theorems 5.1, 5.2 and 5.3. The dependence of ϕ_μ upon μ is not essential in proving Theorems 5.1, 5.2 as the only property of interest is that the derivatives of ϕ_μ are bounded from above and below uniformly over μ (which follows from our assumptions about Φ). This is the reason why in Sections 6 and 7, we will drop this dependence on μ and work with a generic warping ϕ.

6 Proof of Theorem 5.1

We will assume without loss of generality that our curvelet γ_μ is centered near zero ($k = 0$) and is nearly vertical ($\theta_J = 0$).

Set $m_\mu = T_1\gamma_\mu$. We first show that m_μ obeys the smoothness and spatial localization estimate of a molecule (4.1). With the same notations as before, recall that m_μ is given by

$$m_\mu(x) = \int e^{ix\cdot\xi} b_\mu(x,\xi)\hat{\gamma}_\mu(\xi)\, d\xi, \quad b_\mu(x,\xi) = e^{i\delta(\phi^{-1}(x),\xi)} a(\phi^{-1}(x),\xi). \tag{6.1}$$

To study the spatial decay of $m_\mu(x)$, we introduce the differential operator

$$L_\xi = I - 2^{2j}\frac{\partial^2}{\partial\xi_1^2} - 2^j\frac{\partial^2}{\partial\xi_2^2},$$

and evaluate the integral (6.1) using an integration by parts argument. First, observe that

$$L_\xi^N e^{ix\cdot\xi} = \left(1 + |2^j x_1|^2 + |2^{j/2}x_2|^2\right)^N e^{ix\cdot\xi}.$$

Second, we claim that for every integer $N \ge 0$,

$$|L_\xi^N[b_\mu(x,\xi)\hat{\gamma}_\mu(\xi)]| \le C \cdot 2^{-3j/4}. \tag{6.2}$$

(The factor $2^{-3j/4}$ comes from the L^2 normalization of $\hat{\gamma}_\mu$.) This inequality is proved in the Appendix. Hence,

$$m_\mu(x) = \left(1 + |2^j x_1|^2 + |2^{j/2}x_2|^2\right)^{-N} \int L_\xi^N[b_\mu(x,\xi)\hat{\gamma}_\mu(\xi)]\, e^{ix\cdot\xi}.$$

Since $|L_\xi^N[b_\mu(x,\xi)\hat{\gamma}_\mu(\xi)]| \le C \cdot 2^{-3j/4}$ and is supported on a pair of symmetric dyadic rectangles R_J, of length about 2^j and width $2^{j/2}$, we then established that

$$|m_\mu(x)| \le C \cdot \frac{2^{3j/4}}{\left(1 + |2^j x_1|^2 + |2^{j/2}x_2|^2\right)^N}.$$

The derivatives of m_μ are essentially treated in the same way. Begin with

$$\begin{aligned} \partial_x^\alpha(e^{ix\cdot\xi}b_\mu(x,\xi)) &= \sum_{\beta+\gamma\le\alpha} \partial^\beta(e^{ix\cdot\xi})\,\partial^\gamma(b_\mu(x,\xi)) \\ &= \sum_{\beta+\gamma\le\alpha} \partial^\gamma(b_\mu(x,\xi))\,\xi^\beta e^{ix\cdot\xi} \end{aligned}$$

Therefore, the partial derivatives of m_μ are given by

$$(\partial_x^\alpha m_\mu)(x) = \sum_{\beta+\gamma\le\alpha} I_{\beta,\gamma}(x), \tag{6.3}$$

where

$$I_{\beta,\gamma}(x) = \int e^{ix\cdot\xi}\partial_x^\gamma(b_\mu(x,\xi))\xi^\beta\hat{\gamma}_\mu(\xi)\,d\xi. \tag{6.4}$$

First, observe that on the support of $\hat{\gamma}_\mu$, $|\xi|^\beta$ obeys $|\xi|^\beta \le C\cdot 2^{j\beta_1}\cdot 2^{j\beta_2/2}$. Second, the term $\partial_x^\gamma b(x,\xi)$ is of the same nature as $b_\mu(x,\xi)$ in the sense that it obeys all the same estimates as before. In particular, we claim that for every integer $N \ge 0$,

$$|L_\xi^N[\partial_x^\gamma b_\mu(x,\xi)\xi^\beta\hat{\gamma}_\mu(\xi)]| \le C\cdot 2^{-3j/4}\cdot 2^{j\beta_1}\cdot 2^{j\beta_2/2}. \tag{6.5}$$

Hence, the same argument as before gives

$$|I_{\beta,\gamma}(x)| \le C\cdot\frac{2^{3j/4}\cdot 2^{j\beta_1}\cdot 2^{j\beta_2/2}}{\left(1+|2^jx_1|^2+|2^{j/2}x_2|^2\right)^N}.$$

Now since $\beta\le\alpha$, we may conclude that

$$|(\partial_x^\alpha m_\mu)(x)| \le C\cdot\frac{2^{3j/4}\cdot 2^{j\alpha_1}\cdot 2^{j\alpha_2/2}}{\left(1+|2^jx_1|^2+|2^{j/2}x_2|^2\right)^N}.$$

This establishes the smoothness and localization property.

The above analysis shows that m_μ is a "ridge" of effective length $2^{-j/2}$ and width 2^{-j}; to prove that m_μ is a molecule, we now need to evidence its oscillatory behavior across the ridge. In other words, we are interested in the size of the Fourier transform at low frequencies (4.2)–(4.5).

Formally, the Fourier transform of m_μ is given by

$$\hat{m}_\mu(\xi) = \int\int e^{ix\cdot(\lambda-\xi)}b_\mu(x,\lambda)\hat{\gamma}_\mu(\lambda)\,dxd\lambda. \tag{6.6}$$

We should point out that because the amplitude b is not of compact support in x, the sense in which (6.6) holds is not obvious. This is a well-known phenomenon in Fourier analysis and a classical technique to circumvent such difficulties would be to multiply m_μ (or equivalently b_μ) by a smooth and compactly supported cut-off function $\chi(\epsilon x)$ and let ϵ tend to zero. We omit those details as they are standard.

Set $L = -i\partial_{x_1}$. To develop bounds on $|\hat{m}_\mu(\xi)|$, observe that

$$L^n e^{ix\cdot\lambda} = (\lambda_1)^n.$$

An integration by parts then gives

$$\hat{m}_\mu(\xi) = \int\int e^{ix\cdot\lambda}L^n\left(e^{-ix\cdot\xi}\,b_\mu(x,\xi)\right)\,\lambda_1^{-n}\hat{\gamma}_\mu(\lambda)\,dxd\lambda$$

Hence,

$$\hat{m}_\mu(\xi) = \sum_{m=0}^{n} c_m\,\xi_1^m\hat{F}_m(\xi),$$

where

$$F_m(x) = \int e^{ix\cdot\lambda} (\partial_{x_1}^{n-m} b(x,\xi))\, \lambda_1^{-n} \hat{\gamma}_\mu(\lambda)\, d\lambda.$$

Note that F_m is exactly of the same form as (6.4)—but with λ_1^{-n} instead of λ^β—and therefore, the exact same argument as before would give

$$|F_m(x)| \le C \cdot \frac{2^{3j/4} \cdot 2^{-jn}}{\left(1 + |2^j x_1|^2 + |2^{j/2} x_2|^2\right)^N}.$$

We then established

$$\|\hat{F}_m\|_{L_\infty} \le \|F\|_{L_1} \le C_m \cdot 2^{-3j/4} \cdot 2^{-jn},$$

which gives

$$|m_\mu(\xi)| \le C \cdot 2^{-3j/4} \cdot 2^{-jn} \cdot (1 + |\xi|^n),$$

as required. This finishes the proof of Theorem 5.1.

The careful reader will object that we did not study the case of coarse scale curvelets; it is obvious that coarse scale elements are mapped into coarse scale molecules and, here, the argument would not require the deployment of the sophisticated tools we exposed above. We omit the proof.

7 Warpings

7.1 Proof of Theorem 5.2

As mentioned earlier, curvelet atoms depend in a nonessential way upon the object f we wish to analyze and we shall drop this dependence in our notations. To prove Theorem 5.2, recall that we need to show that for each curvelet atom ρ_μ with regularity R, the 'warped' atom $\rho_\mu \circ \phi$ is also a curvelet atom, with the same regularity.

As in Section 3, we suppose our curvelet atom is of the form

$$\rho_\mu(x) = 2^{3j/4} a^{(\mu)}(D_j R_{\theta_J}(x - x_\mu)),$$

where $a^{(\mu)}$ obeys the conditions of Lemma 3.1. (Here, the location x_μ may be formally defined by $x_\mu = (D_j R_{\theta_J})^{-1} k_\delta$.) Define y_μ and A_μ by

$$y_\mu = \phi^{-1}(x_\mu), \quad \text{and} \quad A_\mu = (\nabla\phi)(y_\mu) \tag{7.1}$$

so that

$$\phi(y) = x_\mu + A_\mu(y - y_\mu) + h(y - y_\mu).$$

With these notations, it is clear that the warped atom $\rho_\mu \circ \phi$ will be centered near the point y_μ; that is,

$$\rho_\mu(\phi(y)) = 2^{3j/4} a^{(\mu)} \left(D_j R_{\theta_J}(A_\mu(y - y_\mu) + h(y - y_\mu))\right).$$

To simplify matters, we first assume that A_μ is the identity and show that $\rho_\mu \circ \phi$ is a curvelet atom with the same scale and orientation as ρ_μ. Later, we will see that in general, $\rho_\mu \circ \phi$ is an atom whose orientation depends upon A_μ, and whose scale may be taken to be the same as that of ρ_μ. Assume without loss of generality that $\theta_J = 0$ and $y_\mu = 0$ (statements for arbitrary orientations and locations are obtained in an obvious fashion) so that

$$\rho_\mu(\phi(y)) = 2^{3j/4} a^{(\mu)} \left(D_j(y + h(y))\right) = 2^{3j/4} b^{(\mu)}(D_j y), \tag{7.2}$$

with

$$b^{(\mu)}(y) = a^{(\mu)}\left(y + D^j h(D_j^{-1} y)\right).$$

The atom $a^{(\mu)}$ is supported over a square of sidelength about 1; likewise, $b^{(\mu)}$ is also compactly supported in a box of roughly the same size—uniformly over μ. We then need to derive smoothness estimates and show that $b^{(\mu)}$ obeys

$$|\partial^\alpha b^{(\mu)}(y)| \le C_\alpha, \quad |\alpha| \le R. \tag{7.3}$$

Over the support of $\rho_\mu \circ \phi$, $h = (h_1, h_2)$ deviates little from zero and obeys

$$|h_k(y)| \le M \cdot 2^{-j}, \quad |\partial_i h_k(y)| \le M \cdot 2^{-j/2}, \ i, k = 1, 2,$$

and for each α, $|\alpha| > 1$,

$$|\partial^\alpha h_k(y)| \le C_\alpha, \quad k = 1, 2. \tag{7.4}$$

This estimates holds, of course, uniformly over μ. It follows that for $|y_1|, |y_2| \le C$, the perturbation h obeys for each α

$$2^j \cdot |\partial^\alpha h_1(2^{-j} y_1, 2^{-j/2} y_2)| \le C_\alpha, \quad 2^{j/2} \cdot |\partial^\alpha h_2(2^{-j} y_1, 2^{-j/2} y_2)| \le C_\alpha. \tag{7.5}$$

The bound (7.3) is then a simple consequence of (7.5) together with the fact that all the derivatives of $a^{(\mu)}$ up to order R are bounded, uniformly over μ.

We now show that $\rho_\mu \circ \phi$ exhibits the appropriate behavior at low frequencies.

$$\begin{aligned} \widehat{\rho_\mu \circ \phi}(\xi) &= \int e^{-ix\cdot\xi} \rho_\mu(\phi(x))\, dx \\ &= \int e^{-i\phi^{-1}(x)\cdot\xi} \rho_\mu(x) \frac{dx}{|\det \nabla\phi|(\phi^{-1}(x))}. \end{aligned}$$

We will use the nearly vanishing moment property of ρ_μ. Set

$$S_\xi(x) = e^{-i\phi^{-1}(x)\cdot\xi} / |\det \nabla\phi|(\phi^{-1}(x));$$

note that over the support of ρ_μ and for each $n \le R$, we have available the following upper bound on the partial derivative of S_ξ

$$|\partial_1^n S_\xi(x)| \le C_n \cdot (1 + |\xi|^n).$$

Classical arguments give

$$\widehat{\rho_\mu \circ \phi}(\xi) = \sum_{k=0}^{n-1} \int \frac{\partial_1^k S_\xi(0, x_2)}{k!}\, dx_2 \int \rho_\mu(x_1, x_2) x_1^k \, dx_1 dx_2 + E, \tag{7.6}$$

where E is a remainder term obeying

$$|E| \le C_n \cdot 2^{-3j/4} \cdot 2^{-jn} \cdot \sup |\partial_1^n S_\xi(x)| \le C_n \cdot 2^{-3j/4} \cdot 2^{-jn} (1 + |\xi|^n). \tag{7.7}$$

The near-vanishing moment property gives that each term in the right-hand side of (7.6) obeys the estimate in (7.7). This proves that the Fourier transform of $\rho_\mu \circ \phi$ obeys

$$|\widehat{\rho_\mu \circ \phi}(\xi)| \le C_n \cdot 2^{-jn} (1 + |\xi|^n)$$

as required.

We now discuss the case where the derivative A_μ is arbitrary. In this case, (7.2) becomes

$$\rho_\mu(\phi(y)) = m_\mu(A_\mu y),$$

with

$$m_\mu(y) = 2^{3j/4} a^{(\mu)} \left(D_j(y + \tilde{h}(y))\right), \quad \text{and} \quad \tilde{h}(y) = h(A_\mu^{-1} y).$$

Our assumptions about FIOs guarantee that $|A_\mu^{-1}|$ is uniformly bounded and, therefore, it follows from the previous analysis that m_μ is a curvelet atom. As a consequence $\rho_\mu \circ \phi$ is a curvelet atom with the same regularity R since it is clear that bounded linear transformations of the plane map curvelet atoms into curvelet atoms.

7.2 Microlocal Correspondence

There are many ways to establish a formal index correspondence and we only present a possible approach. With the notations of Section 5, we have seen that the warping $\phi_\mu(x) = \Phi_\xi(x, \xi_\mu)$ maps the curvelet atom ρ_μ into $m_\mu(A_\mu(y - y_\mu))$ where m_μ is an atom centered at the origin and with the same scale 2^{-j_μ} and codirection ξ_μ as ρ_μ. Define τ_μ by

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

With these notations, the atom $\rho_\mu \circ \phi_\mu$ is another atom at about the same scale 2^{-j_μ} and with codirection τ_μ and location y_μ. We then may introduce the index mapping t (see Section 5) 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.

8 Proof of Theorem 5.3

Decompose T as $T_2 \circ T_1$ and let γ_{μ_0} be a fixed curvelet. First, Theorem 5.1 proved that $T_1 \gamma_{\mu_0}$ is a curvelet molecule m_{μ_0} which we will express as a superposition of curvelet atoms ρ_{μ_1}

$$T_1 \gamma_{\mu_0} = m_{\mu_0} = \sum_{\mu_1} \beta_0(\mu_1, \mu_0) \rho_{\mu_1}.$$

Second, for each μ_1, Theorem 5.2 shows that $T_2 \rho_{\mu_1}$ is a molecule $m_{t(\mu_1)}$ at the location $t(\mu_1)$, and

$$\langle \gamma_{\mu_2}, T_2 \rho_{\mu_1} \rangle = \beta_1(\mu_2, t(\mu_1)).$$

Hence,

$$\langle \gamma_{\mu_2}, T \gamma_{\mu_0} \rangle = \sum_{\mu_1} \beta_1(\mu_2, t(\mu_1)) \beta_0(\mu_1, \mu_0).$$

Of course, both β_0 and β_1 obey very special decay properties.

- By Theorem 5.1 and Lemma 4.2, $|\beta_0(\mu_1, \mu_0)| \le C_n \cdot \omega(\mu_1, \mu_0)^{-n}$ for arbitrarily large $n > 0$, provided that the selected atoms are regular enough.
- By Theorem 5.2 and Lemma 4.2, $|\beta_1(\mu_2, t(\mu_1))| \le C_n \cdot \omega(\mu_2, t(\mu_1))^{-n}$ for arbitrarily large $n > 0$, provided that the selected atoms are regular enough.

To prove Theorem 5.3, we then only need to establish that

$$\sum_{\mu_1} \omega(\mu_2, t(\mu_1))^{-n} \cdot \omega(\mu_1, \mu_0)^{-n} \leq C_n \cdot \omega(\mu_2, t(\mu_0))^{-n}. \tag{8.1}$$

We first argue that

$$\sum_{\mu_1} \omega(\mu_2, \mu_1)^{-n} \cdot \omega(\mu_1, \mu_0)^{-n} \leq C_n \cdot \omega(\mu_2, \mu_0)^{-n}. \tag{8.2}$$

Recall that $d(\mu, \mu') \sim d(\mu', \mu)$ where d is defined as follows: $d(\mu, \mu') = |\theta_J - \theta_{J'}|^2 + |x_\mu - x_{\mu'}|^2 + |\langle e_J, x_\mu - x_{\mu'}\rangle|$ and $\omega(\mu, \mu') = 2^{|j-j'|}\ (1 + 2^{\min(j,j')} d(\mu, \mu'))$. We closely follow the argument in [20] and use the following lemma which is shown in the Appendix.

Lemma 8.1.

$$d(\mu, \mu') \leq C \cdot \left(d(\mu, \mu'') + d(\mu'', \mu')\right).$$

Define I_{μ_1} by

$$\begin{aligned} I_{\mu_1} &:= \omega(\mu_2, \mu_1)^{-N} \cdot \omega(\mu_1, \mu_0)^{-N} \\ &= \left(2^{|j_2-j_1|+|j_1-j_0|}(1 + 2^{\min(j_2,j_1)} d(\mu_2, \mu_1))(1 + 2^{\min(j_0,j_1)} d(\mu_0, \mu_1))\right)^{-N}. \end{aligned} \tag{8.3}$$

To ease notations, put temporarily $a_0 = 2^{\min(j_0,j_1)}$, $a_2 = 2^{\min(j_2,j_1)}$, $d_{01} = d(\mu_0, \mu_1)$, and $d_{12} = d(\mu_2, \mu_1)$. We develop a lower bound on $(1 + a_2 d_{12})(1 + a_0 d_{01}) = 1 + a_2 d_{12} + a_0 d_{01} + a_2 d_{12} a_0 d_{01}$. We make three simple observations: first,

$$a_2 d_{12} + a_0 d_{01} \geq \min(a_2, a_0)(d_{12} + d_{01}) = A_0, \quad \text{and} \quad d_{12} + d_{01} \gtrsim d(\mu_0, \mu_2);$$

second,

$$a_2 d_{12} + a_0 d_{01} \geq \max(a_2 d_{12}, a_0 d_{01}) \geq \max(a_2, a_0) \min(d_{12}, d_{01}) = B_0;$$

and third

$$\begin{aligned} a_2 d_{12} a_0 d_{01} &= \max(a_2, a_0) \min(a_2, a_0) \max(d_{12}, d_{01}) \min(d_{12}, d_{01}) \\ &\geq \max(a_2, a_0) \min(a_2, a_0) \min(d_{12}, d_{01}) \frac{d_{12} + d_{01}}{2} = A_0 B_0 / 2. \end{aligned}$$

This gives

$$1 + a_2 d_{12} + a_0 d_{01} + a_2 d_{12} a_0 d_{01} \geq \frac{1}{2}\left(1 + A_0 + B_0 + A_0 B_0\right) \geq \frac{1}{2}(1 + A_0)(1 + B_0).$$

We replace the values of A_0, B_0 by their expression, use the relation $A_0 \geq d(\mu_0, \mu_2)$ and obtain

$$I_{\mu_1} \leq C \cdot 2^{-(|j_2-j_1|+|j_0-j_1|)N} \cdot \left(1 + 2^{\min(j_2,j_0,j_1)} d(\mu_2, \mu_0)\right)^{-N} \cdot (L_1)^{-N}$$

with

$$L_1 = 1 + \max(2^{\min(j_2,j_1)}, 2^{\min(j_0,j_1)}) \min(d_{01}, d_{12}).$$

Now observe that $2^{|j_2-j_1|} \cdot 2^{\min(j_2,j_0,j_1)} \geq 2^{\min(j_2,j_0)}$ and thus, we may remove the dependence on μ_1 in the factor $(1 + 2^{\min(j_2,j_0,j_1)} d(\mu_2, \mu_0))^{-N}$ at the expense of changing the power from $-N$ to $-N/2$, namely

$$I_{\mu_1} \leq C \cdot 2^{-(|j_2-j_1|+|j_0-j_1|)N/2} (1 + 2^{\min(j_2,j_0)} d(\mu_2, \mu_0)) \cdot (L_1)^{-N}.$$

Note that

$$\begin{aligned} L_1 &= \min\Big(1+\max(2^{\min(j_2,j_1)},2^{\min(j_0,j_1)})d_{12}, 1+\max(2^{\min(j_2,j_1)},2^{\min(j_0,j_1)})d_{01}\Big) \\ &\geq \min\Big(1+2^{\min(j_2,j_1)}d_{12}, 1+2^{\min(j_0,j_1)}d_{01}\Big) \end{aligned}$$

and, therefore,

$$\begin{aligned} (L_1)^{-N} &\leq \max\Big((1+2^{\min(j_2,j_1)}d_{12})^{-N}, (1+2^{\min(j_0,j_1)}d_{01})^{-N}\Big) \\ &\leq (1+2^{\min(j_2,j_1)}d_{12})^{-N} + (1+2^{\min(j_0,j_1)}d_{01})^{-N}. \end{aligned}$$

We now sum this quantity over μ_1. The proof of Lemma 4.3 shows that

$$\sum_{\mu_1 \in M_{j_1}} (1+2^{\min(j_2,j_1)}d(\mu_2,\mu_1))^{-N} \leq C \cdot 2^{2|j_2-j_1|},$$

with, of course, a similar bound for the other term. In short,

$$\sum_{\mu_1 \in M_{j_1}} I_{\mu_1} \leq C \cdot 2^{-(|j_2-j_1|+|j_0-j_1|)(\frac{N}{2}-2)} \cdot (1+2^{\min(j_2,j_0)}d(\mu_2,\mu_0))^{-N/2},$$

and thus

$$\sum_{\mu_1} I_{\mu_1} \leq C \cdot 2^{-|j_2-j_0|(N-4)/2} \cdot (1+2^{\min(j_2,j_0)}d(\mu_2,\mu_0))^{-(N-4)/2},$$

for sufficiently large N. This proves (8.2).

It remains to establish (8.1). To see why inequality (8.2) implies (8.1), observe that it follows from our definitions that $d(\mu, t(\mu')) \sim d(t^{-1}(\mu), \mu')$ and, therefore, $\omega(\mu, t(\mu')) \sim \omega(t^{-1}(\mu), \mu')$. (This follows from our assumption that the microlocal correspondence is nonsingular, i.e. $|\det \Phi_{x\xi}|$ is bounded from below.) Formally,

$$\begin{aligned} \sum_{\mu_1} \omega(\mu_2, t(\mu_1)) \cdot \omega(\mu_1,\mu_0)^{-n} &\leq C \cdot \sum_{\mu_1} \omega(t^{-1}(\mu_2),\mu_1) \cdot \omega(\mu_1,\mu_0)^{-n} \\ &\leq C \cdot \omega(t^{-1}(\mu_2),\mu_0)^{-n} \leq C \cdot \omega(\mu_2, t(\mu_0))^{-n}. \end{aligned}$$

which is what we thought to establish.

Cases involving coarse scale elements are treated similarly and we omit the proof. Theorem 5.3 is proved.

9 Conclusion

We proved that Fourier Integral Operators admit sparse representations in tight frames of curvelets. Our results are not sharp with respect to smoothness, localization or number of vanishing moments of these curvelets. We only discussed amplitudes and phases infinitely differentiable in both variables.

It is instructive to notice why wavelets or ridgelets cannot be expected to sparsify FIO's. We quote from [4]: "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.

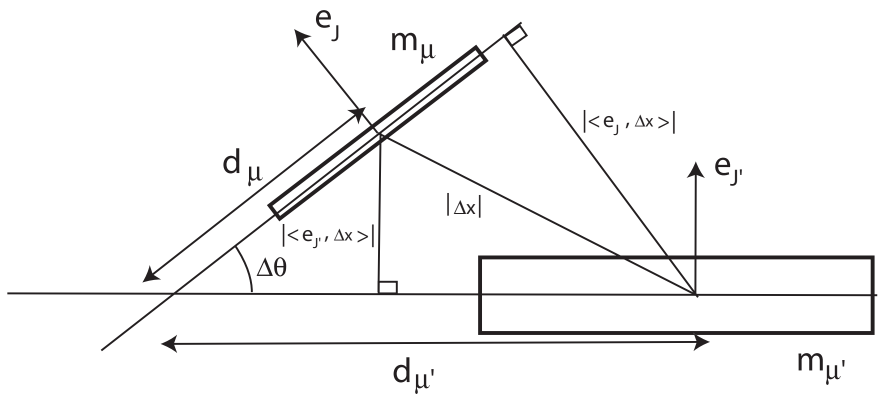

Figure 3: Relative position and orientation of two curvelet molecules in x-space. The rectangles indicate their essential support.

- 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*."

10 Appendix

Proof of the fact that $\omega(\mu, \mu') \sim \omega(\mu', \mu)$. The only thing we have to check is that formulating the estimate (4.8) through e_J or $e_{J'}$ is equivalent. It is sufficient to notice that

$$|\langle e_J, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2 \sim |\langle e_J, \Delta x\rangle| + |\langle e_{J'}, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2.$$

In order to justify the nontrivial inequality, use the law of cosines illustrated in Figure 3:

$$\begin{aligned} |\langle e_J, \Delta x\rangle|^2 + |\langle e_{J'}, \Delta x\rangle|^2 &= \sin^2|\Delta\theta|\,(d_\mu^2 + d_{\mu'}^2) \\ &= \sin^2|\Delta\theta|\,|\Delta x|^2 \pm 2|\langle e_J, \Delta x\rangle|\,|\langle e_{J'}, \Delta x\rangle| \cos|\Delta\theta| \\ &\leq \sin^2|\Delta\theta|\,|\Delta x|^2 + 2|\langle e_J, \Delta x\rangle|\,|\langle e_{J'}, \Delta x\rangle|. \end{aligned}$$

It follows that $||\langle e_J, \Delta x\rangle| - |\langle e_{J'}, \Delta x\rangle|| \leq C \cdot |\Delta\theta||\Delta x| \leq C \cdot (|\Delta\theta|^2 + |\Delta x|^2)$ and, therefore,

$$|\langle e_J, \Delta x\rangle| + |\langle e_{J'}, \Delta x\rangle| \leq C \cdot (2|\langle e_J, \Delta x\rangle| + |\Delta\theta|^2 + |\Delta x|^2).$$

Proof of the inequality (4.11). Assume without loss of generality that $J = J_0$. We may express $S_{J'}(\xi)$ as $S_{J_0'}(R_{\Delta\theta}\xi)$, with $\Delta\theta = \theta_J - \theta_{J'}$. We begin by expressing the integral in polar coordinates,

$$\xi_1 = r\cos\theta \qquad (R_{\Delta\theta}\xi)_1 = r\cos(\theta+\Delta\theta),$$
$$\xi_2 = r\sin\theta \qquad (R_{\Delta\theta}\xi)_2 = r\sin(\theta+\Delta\theta).$$

As we will see, the cosine factor is not crucial and we may just as well drop it. Consequently,

$$\int |S_J(\xi)\, S_{J'}(\xi)|^n\, d\xi \leq C\cdot \int_0^\infty r dr\, \frac{1}{[1+2^{-j}r]^N}\frac{1}{[1+2^{-j'}r]^N}\int_0^{2\pi} d\theta[1+a|\sin\theta|]^{-N}[1+a'|\sin(\theta+\Delta\theta)|]^{-N},$$

where $a = \frac{2^{-j/2}r}{1+2^{-j}r}$ and $a' = \frac{2^{-j'/2}r}{1+2^{-j'}r}$. This decoupling makes the problem of bounding the inner integral on the variable θ tractable. For example when $a > a' > 1$, following [17] p.56,

$$\int_{-\infty}^{\infty} d\theta[1+a|\theta|]^{-N}[1+a'|\theta+\Delta\theta|]^{-N} \leq C\cdot\frac{1}{a}\frac{1}{[1+a'|\Delta\theta|]^N}.$$

We get other estimates for other values and orderings of a and a'. The integral on r is then broken up into several pieces according to the values of a, a', j and j'. It is straightforward to show that each of these contributions satisfies the inequality (4.11).

Proof of the inequality (4.13). Without loss of generality, assume that $\mu = (j,0,0)$ so that the curvelet molecule m_μ is nearly vertical and centered near the origin. We recall that $\Delta\theta = \pi\cdot\ell'\cdot 2^{-\lfloor j'/2\rfloor}$, $\ell' = 0,1,\ldots 2^{\lfloor j'/2\rfloor}-1$, and $x_{\mu'} = R_{\theta_{J'}}D_{j'}^{-1}k'$, say. Then the sum of interest is

$$\sum_{\ell'=0}^{2^{\lfloor j'/2\rfloor}-1}\sum_{k'\in\mathbb{Z}^2}\left(1+2^{j_0}(|2^{-j'/2}l'|^2+|x_{\mu',2}|^2+|x_{\mu',1}|)\right)^{-np},$$

with $j_0 = \min(j,j')$. Note that and $\det D_{j'}^{-1} = 2^{-j-\lfloor j'/2\rfloor}$ and that this Riemann sum is bounded— up to a numerical multiplicative constant—by the corresponding integral:

$$\int_{\mathbb{R}^2}\frac{dx}{2^{-3j'/2}}\int_{\mathbb{R}}\frac{dy}{2^{-j'/2}}C_n\cdot[1+2^{j_0}(y^2+x_2^2+|x_1|)]^{-np} \leq C\cdot 2^{-2(j_0-j')} \leq C\cdot 2^{2|j-j'|}$$

provided that $np \geq 2$.

Proof of (4.14). Recall that

$$\alpha_{\mu'}^{\mu} = \left(\int_{Q_{\mu'}}|\mathcal{R}(q(D)\gamma_\mu)(a,b,\theta)|^2\,d\mu\right)^{1/2}.$$

The first thing to notice is that $q(D)\gamma_\mu$ is still a family of curvelet molecules, because $q(\xi)$ is a multiplier of order zero. Since $\psi_{a,\theta,b}$ also obeys the molecule properties, lemma 4.2 implies the corresponding almost-orthogonality condition. Integrating over $Q_{\mu'}$ does not compromise this estimate, as can be seen by applying the Cauchy-Schwarz inequality.

Proof of inequality (6.2). Derivatives of $\hat{\gamma}_\mu$ and a are treated using the following estimates.

$$|\partial_\xi^\alpha\hat{\gamma}_\mu(\xi)| \leq C_\alpha\cdot 2^{-3j/4}2^{-\alpha_1 j}2^{-\alpha_2 j/2}$$
$$|\partial_\xi^\alpha a(\phi^{-1}(x),\xi)| \leq C_\alpha\cdot 2^{-|\alpha|j} \qquad \text{on } W_\mu = \mathrm{supp}(\hat{\gamma}_\mu).$$

We now develop size estimates for the phase perturbation δ. Following closely the discussion in [21], p.407, we claim that on W_μ,

$$|\partial_\xi^\alpha \partial_x^\beta \delta(x,\xi)| \le C_{\alpha\beta} \cdot 2^{-\alpha_1 j} 2^{-\alpha_2 j/2}. \tag{10.1}$$

The derivations in x add no complications. Hence, assume that $\beta = 0$. As the above result (10.1) relies upon the homogeneity of the phase with respect to ξ, we recall a few useful facts about homogeneous functions of degree one:

$\Phi = \Phi_\xi \cdot \xi$ (Euler's theorem),

$\Phi_{\xi\xi} \cdot \xi = 0$ (differentiate the above relation),

$\partial_\xi^\alpha \Phi = O(|\xi|^{1-|\alpha|})$.

It follows from the definition that $\delta(x, \xi_1, 0) = 0$ and likewise $\frac{\partial \delta}{\partial \xi_2}(x, \xi_1, 0) = 0$. Thus for every n, $\frac{\partial^n \delta}{\partial \xi_1^n}(x, \xi_1, 0) = 0$ and $\frac{\partial}{\partial \xi_2} \frac{\partial^n \delta}{\partial \xi_1^n}(x, \xi_1, 0) = 0$. Recall that the support conditions are $|\xi_1| \le C \cdot 2^j$ and $|\xi_2| \le C \cdot 2^{j/2}$. Taylor series expansions about $\xi_2 = 0$ together with homogeneity assumptions give

$$\begin{aligned}
\frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}} \delta(x,\xi) &= O(|\xi_2|^2 |\xi|^{-1-\alpha_1}) = O(2^{-\alpha_1 j}), \\
\frac{\partial}{\partial \xi_2} \frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}} \delta(x,\xi) &= O(|\xi_2||\xi|^{-1-\alpha_1}) = O(2^{-j/2} 2^{-\alpha_1 j}), \\
\frac{\partial^{\alpha_2}}{\partial \xi_2^{\alpha_2}} \frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}} \delta(x,\xi) &= O(|\xi|^{1-\alpha_1-\alpha_2}) = O(2^{-\alpha_1 j} 2^{-\alpha_2 j/2}) \qquad \text{when } \alpha_2 \ge 2,
\end{aligned}$$

as claimed. The point about these estimates is that they exhibit exactly the parabolic scaling of curvelets. We conclude

$$|\partial_\xi^\alpha e^{i\delta(\phi^{-1}(x),\xi)}| \le C_\alpha \cdot 2^{-\alpha_1 j} 2^{-\alpha_2 j/2} \qquad \text{on } W_\mu$$

and therefore (6.2).

Proof of Lemma 8.1. To simplify notations, set in the coordinates defined by $\{e_1, e_2\}$,

$$\begin{array}{lll}
x_\mu = (0,0) & x_{\mu'} = (x_1, x_2) & x_{\mu''} = (y_1, y_2) \\
e_1 = (1,0) & e_1' = (\cos\alpha, \sin\alpha) & e_1'' = (\cos\beta, \sin\beta) \\
|\theta_l - \theta_{l''}| = |\beta| & |\theta_{l'} - \theta_{l''}| = |\alpha - \beta| &
\end{array}$$

It is enough to show that there exists $\epsilon > 0$ such that

$$\epsilon|x_1| \le |y_1| + |\cos\alpha(x_1 - y_1) + \sin\beta(x_2 - y_2)| + (|\beta| + |\alpha - \beta|)(|y_1| + |x_1 - y_1| + |y_2| + |x_2 - y_2|),$$

because then $(|\beta| + |\alpha - \beta|)(|y_1| + |x_1 - y_1| + |y_2| + |x_2 - y_2|) \le C \cdot (|\beta|^2 + |\alpha - \beta|^2 + |y_1|^2 + |x_1 - y_1|^2 + |y_2|^2 + |x_2 - y_2|^2)$. By contradiction let us assume that the inequality fails. Then we must have $|y_1| < \epsilon|x_1|$. It is always true that $|x_1 - y_1| + |y_1| \ge |x_1|$ so it is necessary that $|\beta| + |\alpha - \beta| < \epsilon$. But then $|\alpha| < 2\epsilon$ thus $\cos\alpha > 1 - 4\epsilon^2$ and $|\sin\alpha| < 2\epsilon$. The term $|\cos\alpha(x_1 - y_1) + \sin\beta(x_2 - y_2)|$ is therefore always greater than $(1 - 4\epsilon^2)|x_1 - y_1| - \epsilon|x_2 - y_2|$. But this quantity must also be less than $\epsilon|x_1 - y_1|$, otherwise its sum with $|y_1|$ would exceed $\epsilon|x_1|$. So we must have $|x_2 - y_2| > \frac{1-\epsilon-4\epsilon^2}{\epsilon}|x_1 - y_1|$. But then the sum $|y_1| + |x_1 - y_1| + |x_2 - y_2|$ must dominate $\frac{|x_1|}{2\epsilon}$, which implies $|\beta| + |\alpha - \beta| \le 2\epsilon^2$. By induction, $\alpha = \beta = 0$ and $|y_1| + |x_1 - y_1| \ge |x_1|$ yields a contradiction.

References

[1] J.P. Antoine, R. Murenzi, Two-dimensional directional wavelets and the scale-angle representation, *Sig. Process.*, **52** (1996) 259-281

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

[3] E. Candès, Near-Optimality of Curvelet Expansions, manuscript, 1999.

[4] E. Candès, L. Demanet, Curvelets and Fourier Integral Operators, submitted to Comptes-Rendus de l'Académie des Sciences Paris, 2002.

[5] 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

[6] E. Candès, D. Donoho, New Tight Frames of Curvelets and Optimal Representations of Objects with C^2 Singularities, (2002).

[7] E. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis : Application to edge-preserving image reconstruction, *Sig. Process.*, special issue on Image and Video Coding Beyond Standards, **82** (2002), 1519–1543.

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

[9] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992

[10] J. Duistermaat, *Fourier Integral Operators*, Birkhauser, Boston, 1996

[11] M. Do, M. Vetterli, Contourlets, in *Beyond Wavelets*, J. Stoeckler and G. V. Welland (eds.), Academic Press, 2001, 1-26

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

[13] M. Frazier, B. Jawerth, G. Weiss, *Littlewood-Paley Theory and the Study of Function Spaces*, CBMS 79, AMS, Providence, 1991

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

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

[16] Y. Meyer, *Ondelettes et Opérateurs*, Hermann, Paris, 1990

[17] Y. Meyer, R. Coifman, *Wavelets, Calderón-Zygmund and Multilinear Operators*, Cambridge Univresity Press, Cambridge, 1997

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

[19] H. Smith, A Hardy space for Fourier integral operators, *J. Geom. Anal.* **7** (1997)

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

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

[22] B. Torrésani, *Analyse Continue par Ondelettes*, InterEditions/CNRS Editions, Paris, 1995

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