The Curvelet Representation of Wave Propagators is Optimally Sparse

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

June 2004

Abstract

This paper argues that *curvelets* provide a powerful tool for representing very general linear symmetric systems of hyperbolic differential equations. Curvelets are a recently developed multiscale system [10, 7] in which the elements are highly anisotropic at fine scales, with effective support shaped according to the parabolic scaling principle $width \approx length^2$ at fine scales. We prove that for a wide class of linear hyperbolic differential equations, the curvelet representation of the solution operator is both optimally sparse and well organized.

- It is sparse in the sense that the matrix entries decay nearly exponentially fast (i.e. faster than any negative polynomial),

- and well-organized in the sense that the very few nonnegligible entries occur near a few shifted diagonals.

Indeed, we show that the wave group maps each curvelet onto a sum of curvelet-like waveforms whose locations and orientations are obtained by following the different Hamiltonian flows—hence the diagonal shifts in the curvelet representation. A physical interpretation of this result is that curvelets may be viewed as coherent waveforms with enough frequency localization so that they behave like waves but at the same time, with enough spatial localization so that they simultaneously behave like particles.

1 Introduction

This paper is concerned with the representation of symmetric systems of linear hyperbolic differential equations of the form

$$\frac{\partial u}{\partial t} + \sum_k A_k(x)\frac{\partial u}{\partial x_k} + B(x)u = 0, \qquad u(0,x) = u_0(x), \tag{1.1}$$

where u is an m-dimensional vector and $x \in \mathbb{R}^n$. The matrices A_k and B may depend on the spatial variable x, and the A_k's are symmetric. Linear hyperbolic systems are ubiquitous in the sciences and a classical example are the equations for acoustic waves which read

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_0 u) &= 0 \\
\rho_0 \frac{\partial u}{\partial t} + \nabla(c_0^2 \rho) &= 0,
\end{aligned} \tag{1.2}$$

where u and ρ are the velocity and density disturbances respectively. (Here, $\rho_0 = \rho_0(x)$ is the density and $c_0 = c_0(x)$ is the speed of sound.) Other well-known examples include Maxwell's equations of electrodynamics

and the equations of linear elasticity. All the results presented in this paper equally apply to higher-order scalar wave equations, e.g., of the form

$$\frac{\partial^2 u}{\partial t^2} - \sum_{ij} a_{ij}(x) \frac{\partial^2 u}{\partial x_j \partial x_k} = 0, \qquad u(0, x) = u_0(x), \quad \frac{\partial u}{\partial t}(0, x) = u_1(x),$$

(where u is now a scalar and $a_{ij}(x)$ is taken to be symmetric and positive definite) as it is well-known that such single second-order equations can be reduced to a symmetric system of first-order equations (1.1) by appropriate changes of variables.

1.1   About representations

We are interested in representations of the solution operator $E(t)$ to the system (1.1)

$$u(t, \cdot) = E(t)u_0,$$

which may be expressed as an integral involving the so-called Green's function

$$u(t, x) = \int E(t; x, y) u_0(y) \, dy.$$

To introduce the concept of representation, suppose that the coefficient matrices do not depend on x. As is well-known, the Fourier transform is a powerful tool for studying differential equations in this setting. Indeed, in the Fourier domain, (1.1) takes the form

$$(\partial_t + i \sum_k A_k \xi_k + B)\hat{u}(t, \xi) = 0;$$

in short, (1.1) reduces to a system of ordinary differential equations which can be solved analytically. This shows the power of the representation; in the frequency domain, the solution operator is diagonal and the study becomes ridiculously simple.

These desirable properties are very fragile, however. Both mathematicians and computational scientists know that Fourier methods are not really amenable to differential equations with variable coefficients, and that we need to find alternatives. Instead of considering the evolution of Fourier coefficients, we may want to think, instead, of the action of the propagation operator $E(t)$ on other types of basis elements. This connects with the viewpoint of modern harmonic analysis whose goal is to develop representations, e.g. an orthonormal basis (f_n) of $L^2(\mathbb{R}^n)$, say, in which the solution operator

$$E(t; n, n') = \langle f_n, E(t) f_{n'} \rangle \tag{1.3}$$

is as simple as possible; that is, such that $E(t)f_n$ is a *sparse* superposition of those elements $f_{n'}$. Such sparse representations are extremely significant both in mathematical analysis, where sparsity allows for sharper inequalities and in numerical applications where sparsity allows for faster algorithms.

- In the field of mathematical analysis, for example, Calderón introduced what one would nowadays call the Continuous Wavelet Transform (CWT) in which objects are represented as a superpositions of simple elements of the form $\psi((x - b)/a)$, with $a > 0$ and $b \in \mathbb{R}^n$; i.e, objects are represented as a superposition of dilates and translates of a single function ψ. These elements proved to be *almost* eigenfunctions of large classes of operators, the Calderón-Zygmund operators which are special types of singular integrals, some of which arising in connection with elliptic problems. It was later gradually realized that tools like atomic decompositions of Hardy spaces [18, 30] and orthonormal bases of Wavelets [23, 24] provide a setting in which some aspects of the mathematical analysis of these operators is dramatically eased.

- Clever representation of scientific and engineering computations can make previously intractable computations tractable. Here, sparsity may allow the design of fast matrix multiplication and/or fast matrix inversion algorithms. For example, Beylkin, Coifman and Rokhlin [2] exploited the sparsity of those singular integrals mentioned above, and showed how to use wavelet bases to compute such integrals with very low complexity algorithms.

In short, a single representation, namely, the wavelet transform provides sparse decompositions of large classes of operators simultaneously.

1.2 Limitations of Classical Multiscale Ideas

Our goal in this paper is to find a representation which provides sparse representations of the solution operators to fairly general classes of systems of hyperbolic differential equations. Now the last two decades have seen the widespread development of multiscale ideas such as Multigrid, Fast Multipole Methods, Wavelets, Finite Elements with or without adaptive refinement, etc. All these representations propose dictionaries of roughly isotropic elements occurring at all scales and locations; the templates are rescaled treating all directions in essentially the same way. Isotropic scaling may be successful when the object under study does not exhibit any special features along selected orientations. This is the exception rather than the rule.

Tools from traditional multiscale analysis are very powerful for representing certain elliptic problems but unfortunately, they are definitely ill-adapted to hyperbolic problems such as 1.1). Indeed,

1. they fail to sparsify the wave propagation, i.e. the solution operator $E(t)$,

2. and they fail to provide a sparse representation of oscillatory signals which are the solutions of those equations.

To make things concrete, consider the problem of propagating elastic waves as in geophysics. Consider the scalar wave equation in two dimensions

$$\partial_{tt}u = c^2(x)\Delta u, \tag{1.4}$$

where Δ is the Laplacian defined by $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$ (we may take the velocity field to be constant). To describe the action of the wave group, we assume that the initial condition takes the form of a wavelet with vanishing initial velocity, say. Then it is clear that at a later time, the wavefield is composed of large concentric rings (imagine throwing a stone in a lake). Now, it is also clear that many wavelets are needed to represent the wavefield. In other words, the wavefield is a rather dense superposition of wavelets. Note that this may be quantified. Suppose that the velocity field is identically equal to one, say, and that the initial condition is a wavelet at scale 2^{-j}; that is, of the form $2^{j/2}\psi(2^j x)$ so that in frequency, the energy is concentrated near the dyadic annulus $|\xi| \sim 2^j$. Then one would need at least $O(2^j)$ wavelets to reconstruct the wavefield at time $t = 1$ to within reasonable accuracy.

Our simple example above shows wave-like flows do not preserve the geometry and characteristics of classical multiscale systems. To achieve sparsity, we need to rethink the geometry of multiscale representations.

1.3 A New Form of Multiscale Analysis

As we will see in section 2, curvelets are waveforms which are highly anisotropic at fine scales, with effective support obeying the parabolic principle $length \approx width^2$. Just as for wavelets, there is both a continuous and a discrete curvelet transform. A curvelet is indexed by three parameters which—adopting a continuous description of the parameter space—are: a scale a, $0 < a < 1$; an orientation θ, $\theta \in [-\pi/2, \pi/2)$ and a location b, $b \in \mathbb{R}^2$. At scale a, the family of curvelets is generated by translation and rotation of a basic element φ_a

$$\varphi_{a,b,\theta}(x) = \varphi_a(R_\theta(x - b)).$$

Here, $\varphi_a(x)$ is some kind of directional wavelet with spatial width $\sim a$ and spatial length $\sim \sqrt{a}$, and with minor axis pointing in the horizontal direction

$$\varphi_a(x) \approx \varphi(D_a x), \quad D_a = \begin{pmatrix} 1/a & 0 \\ 0 & 1/\sqrt{a} \end{pmatrix};$$

D_a is a parabolic scaling matrix, R_θ is a rotation by θ radians.

An important property is that curvelets obey the principle of harmonic of analysis which says that it is possible to analyze and reconstruct an arbitrary function $f(x_1, x_2)$ as a superposition of such templates. It is possible to construct tight-frames of curvelets and one can, indeed, easily expand an arbitrary function $f(x_1, x_2)$ as a series of curvelets, much like in an orthonormal basis. Continuing at an informal level of exposition, there is a sampling of the plane (a, b, θ)

$$a_j = 2^{-j}, \quad \theta_{j,\ell} = 2\pi\ell \cdot 2^{-\lfloor j/2 \rfloor}, \quad R_{\theta_{j,\ell}} b_k^{(j,\ell)} = (k_1 2^{-j}, k_2 2^{-j/2}),$$

such that with μ indexing the triples $(a_j, \theta_{j,\ell}, b_k^{(j,\ell)})$ the collection φ_μ is a tight-frame:

$$f = \sum_\mu \langle f, \varphi_\mu \rangle \varphi_\mu, \quad \|f\|_2^2 = \sum_\mu |\langle f, \varphi_\mu \rangle|^2. \tag{1.5}$$

(Note that these formulae allow us to analyze and synthesize arbitrary functions in $L^2(\mathbb{R}^2)$ as a superposition of curvelets in a stable and concrete way.)

As we have seen, a curvelet is well-localized in space but it is also well-localized in frequency. Recall that a given scale, curvelets φ_μ are obtained by applying shifts and rotations to a 'mother' curvelet $\varphi_{j,0,0}$. In the frequency domain then

$$\hat{\varphi}_{j,0,0}(\xi) = 2^{-3j/4} W(2^{-j}|\xi|) V(2^{\lfloor j/2 \rfloor}\theta).$$

Here, W, V are smooth windows compactly supported near the intervals $[1, 2]$ and $[-1/2, 1/2]$ respectively. Whereas in the spatial domain curvelets live near an oriented rectangle R of length $2^{-j/2}$ and width 2^{-j}, in the frequency domain, they are located in a parabolic wedge of length 2^j and width $2^{j/2}$ and whose orientation is orthogonal to that of R. The joint localization in both space and frequency allows us to think about curvelets as occupying a 'Heisenberg cell' in phase-space with parabolic scaling in both domains. Figure 1 offers a schematic representation of this joint localization. As we shall see, this microlocal behavior is key to understanding the properties of curvelet-propagation. Additional details are given in section 2.

1.4 Curvelets and Geometrical Optics

A hyperbolic system can typically be considered in the approximation of high-frequency waves, also known as *geometrical optics*. In order to best describe our main result, it is perhaps suitable first to exhibit the connections between curvelets and geometrical optics. In that setting it is not necessary to describe the dynamics in terms of the wavefield $u(t, x)$. Only its prominent features are studied: wave fronts, or equivalently rays. The latter are trajectories $(x(t), \xi(t))$ in phase-space $\mathbb{R}^2 \times \mathbb{R}^2$, and are the solutions to the m Hamiltonian flows (indexed by ν)

$$\begin{cases} \dot{x}(t) &= \nabla_\xi \lambda_\nu^0(x, \xi), & x(0) &= x_0, \\ \dot{\xi}(t) &= -\nabla_x \lambda_\nu^0(x, \xi), & \xi(0) &= \xi_0. \end{cases} \tag{1.6}$$

The system (1.6) is also called the *bicharacteristic flow* and the rays $(x(t), \xi(t))$ the *bicharacteristics*. To see how this system arises, consider the classical high-frequency wave-propagation approximation to a wavefield $u(x, t)$ of the form

$$u(x, t) = \sigma(x, t) e^{i\omega \Phi(x, t)}.$$

where ω is a large parameter; it then follows (after substituting the approximation in the wave equation (1.1)) that the phase function Φ must obey one of the Hamilton-Jacobi equations (indexed by ν)

$$\partial_t \Phi_\nu + \lambda_\nu^0(x, \nabla_x \Phi_\nu) = 0, \tag{1.7}$$
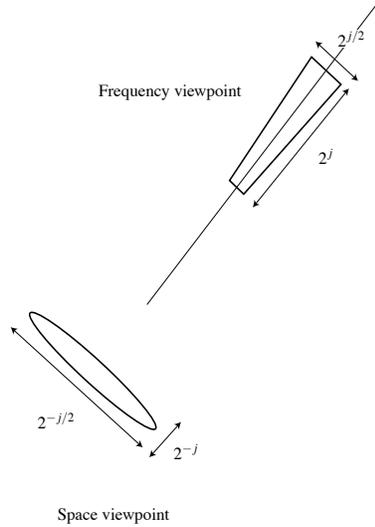
4

Figure 1: Schematic representation of the support of a curvelet in both space and frequency. In the spatial domain, a curvelet has an envelope strongly aligned along a specified 'ridge' while in the frequency domain, it is supported near a box whose orientation is aligned with the co-direction of the ridge.

and that the amplitude must obey a transport equation we shall not detail here (see however section 4). In the above expression (and in the Hamiltonian equations (1.6)), the $\lambda_\nu^0(x, \xi)$ are the eigenvalues of the dispersion matrix

$$a^0(x, \xi) = \sum_k A_k(x)\xi_j. \tag{1.8}$$

It is well-known that the Hamiltonian equations describe the evolution of the wavefront set of the solution as $\Phi_\nu(t, x(t))$ is actually constant along the νth Hamiltonian flow (1.6).

With this background, we are now in a position to qualitatively describe the behavior of the wave-propagation operator $E(t)$ acting on a curvelet φ_μ. However, we first need to introduce a notion of vector-valued curvelet since $E(t)$ is acting on vector fields. Let $r_\nu^0(x, \xi)$ be the eigenvector of the dispersion matrix associated with the eigenvalue $\lambda_\nu^0(x, \xi)$. We then define hyper-curvelets by

$$\varphi_{\mu\nu}^{(0)}(x) = \frac{1}{(2\pi)^2} \int e^{ix \cdot \xi} r_\nu^0(x, \xi) \hat\varphi_\mu(\xi) \, d\xi. \tag{1.9}$$

Later in this section, we will motivate this special choice but for now simply observe that $\varphi_{\mu\nu}^{(0)}$ is a vector-valued waveform.

Consider then the solution to the wave equation $\varphi_{\mu\nu}^{(0)}(t, x)$ with initial value $\varphi_{\mu\nu}^{(0)}(x)$. Our claim is as follows:

> *the wave group maps each hyper-curvelet onto another curvelet-like waveform whose location and orientation are obtained from the corresponding Hamiltonian flow.*

To examine this claim, let (x_μ, ξ_μ) be the center of $\varphi_{\mu\nu}^{(0)}$ in phase-space and define the rotation matrix $U(t)$ by

$$U(t)\frac{\xi(t)}{|\xi(t)|} = \frac{\xi_\mu}{|\xi_\mu|}$$

where $(x(t), \xi(t))$ is the solution to (1.6) with initial condition (x_μ, ξ_μ). Our claim says that the solution to the wave equation nearly follows the dynamics of the reduced Hamiltonian flow, i.e.,

$$\varphi_{\mu\nu}^{(0)}(t, x) = \tilde\varphi_{\mu\nu}^{(0)}(U_\mu(t)(x - x_\mu(t)) + x_\mu). \tag{1.10}$$
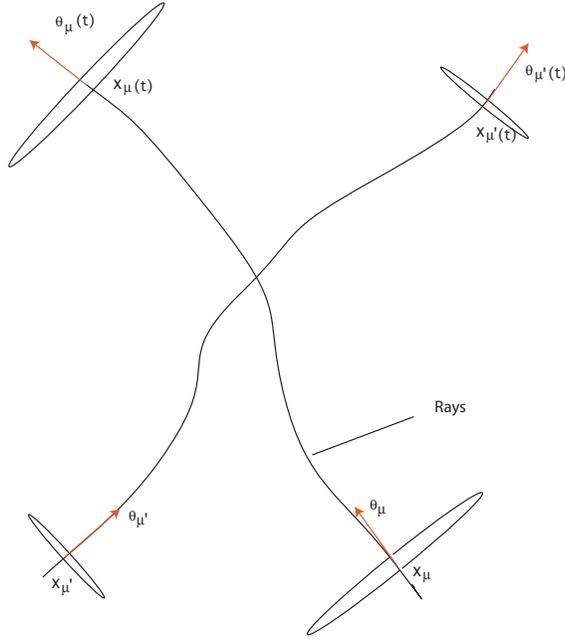
5

Figure 2: Schematic representation of the action of the wave group on a hyper-curvelet. The new positions and orientations are given by the Hamiltonian flow. The two waveforms at time 0 and t are not quite the same although they have very similar profiles.

The point of this paper is that the waveform $\tilde{\boldsymbol{\varphi}}_{\mu\nu}^{(0)}$ has the same strong spatial and frequency localization properties as the initial curvelet $\boldsymbol{\varphi}_{\mu\nu}^{(0)}$ itself. For an illustration, see Figure 2.

We now return to the interpretation of a hyper-curvelet. Suppose that r_ν^0 only depends on ξ as in the case of the acoustic system (1.2)

$$r_0^0(\xi) = \begin{pmatrix} \xi^\perp/|\xi| \\ 0 \end{pmatrix}, \qquad r_\pm^0(\xi) = \frac{1}{\sqrt{2}} \begin{pmatrix} \pm\xi/|\xi| \\ 1 \end{pmatrix}.$$

(Here and below, $\xi^\perp$ denotes the vector obtained from ξ after applying a rotation by 90 degrees). In this special case, we see that the hyper-curvelet is obtained by multiplying—in the frequency domain—a scalar-valued curvelet with the eigenvectors of the dispersion matrix

$$\hat{\boldsymbol{\varphi}}_{\mu\nu}^{(0)}(\xi) = r_\nu^0(\xi)\hat{\varphi}_\mu(\xi), \quad \nu \in \{+, -, 0\}.$$

This is useful for the curvelet $\hat{\boldsymbol{\varphi}}_{\mu\nu}^{(0)}$ will essentially follow only one flow, namely, the νth flow. Suppose we had started, instead, with an initial value of the form $\boldsymbol{\varphi}_{\mu\nu} = \varphi_\mu \mathbf{e}_\nu$, where $\mathbf{e}_\nu$ is the canonical basis of $\mathbb{R}^3$,

say. Then our curvelet would have interacted with the three eigenvectors of the dispersion matrix, and would have 'split' and followed the three distinct flows. By forcing $\hat{\boldsymbol{\varphi}}_{\mu\nu}^{(0)}(\xi)$ to be aligned with $r_\nu^0(\xi)$, we essentially removed the components associated with the other flows. In the general case (1.9), we build hyper-curvelets by applying R_ν^0, which is now a pseudo-differential operator with symbol $r_\nu^0(x,\xi)$, mapping scalars to m-dimensional vectors, and independent of time. The effect is of course the same.

Note that when r_ν^0 is independent of x, hyper-curvelets build-up a (vector-valued) tight-frame; letting $[F,G]$ be the usual inner product over three-dimensional vector fields in $L^2(\mathbb{R}^2)$, the family $(\boldsymbol{\varphi}_{\mu\nu}^{(0)})_{\mu\nu}$ obeys the reconstruction formula

$$u = \sum_{\mu,\nu} [u, \boldsymbol{\varphi}_{\mu\nu}^{(0)}] \boldsymbol{\varphi}_{\mu\nu}^{(0)} \tag{1.11}$$

and the Parseval relation

$$\|u\|_{L^2}^2 = \sum_{\mu,\nu} |[u, \boldsymbol{\varphi}_{\mu\nu}^{(0)}]|^2. \tag{1.12}$$

Just as one can decompose a scalar field as a superposition of scalar curvelets, one can analyze and synthesize any wavefield as a superposition of hyper-curvelets in a stable and concrete way. For arbitrary $r_\nu^0(x,\xi)$, this is, however, in general not true.

We would like to emphasize that although the Hamilton-Jacobi equations only have solutions for small times, the approximation (1.10) and, more generally, all of our results are valid for all times since the rays (1.6) are always well-defined, see section 1.6 below for a more detailed discussion.

1.5 Curvelets and Hyperbolic Systems

The previous section gave a qualitative description of the action of the wave group on a curvelet and we we shall now quantify this fact. The evolution operator $E(t)$ acting on a curvelet $\boldsymbol{\varphi}_{\mu_0\nu_0}^{(0)}$ is of course not exactly another curvelet $\boldsymbol{\varphi}_{\mu_0(t)\nu_0}^{(0)}$ which occurs at a displaced location and orientation. Instead, it is a superposition of curvelets $\sum_{\mu,\mu} \alpha_{\mu\nu} \boldsymbol{\varphi}_{\mu\nu}^{(0)}$ such that

1. the coefficients $(\alpha_{\mu\nu})$ decay nearly exponentially,

2. and the significant coefficients of this expansion are all located at indices (μ,ν) 'near' $(\mu_0(t),\nu_0)$. By near, we mean nearby scales, orientations and locations.

To state the key result of this paper, we need a notion of distance ω between curvelet indices which will be formally introduced in section 2. Crudely, $\omega(\mu,\mu')$ is small if and only if both curvelets are at roughly the same scale, have similar orientation and are at nearby spatial locations. In the same spirit, the distance $\omega(\mu,\mu')$ increases as the distance between the scale, angular, and location parameters increases.

For each $\mu = (j,k,\ell)$ and $\nu = 1,\ldots,m$, define the vector-valued curvelets

$$\boldsymbol{\varphi}_{\mu\nu} = \mathbf{e}_\nu \varphi_\mu, \tag{1.13}$$

where $\mathbf{e}_\nu$ is the νth canonical basis vector in $\mathbb{R}^m$. The $\boldsymbol{\varphi}_{\mu\nu}$'s inherit the tight-frame property (1.11)–(1.12). We would like to again remind the reader that these vector-valued curvelets are simpler and different from the hyper-curvelets $\boldsymbol{\varphi}_{\mu\nu}^{(0)}$ defined in the previous section. Consider now the representing the operator $E(t)$ in a tight-frame of vector-valued curvelets, namely,

$$E(t;\mu,\nu;\mu',\nu') = \langle \boldsymbol{\varphi}_{\mu\nu}, E(t)\boldsymbol{\varphi}_{\mu'\nu'} \rangle. \tag{1.14}$$

We will refer to $E(t;\mu,\nu;\mu',\nu')$ or simply E as the curvelet matrix of $E(t)$, with row index μ,ν and column index μ',ν'. Decompose the initial wavefield $u_0 = \sum_{\mu,\nu} c_{\mu\nu} \boldsymbol{\varphi}_{\mu\nu}$. Then one can express the action of $E(t)$ on u_0 in the curvelet domain as

$$E(t)u_0 = \sum_{\mu\nu} c_{\mu\nu}(t)\boldsymbol{\varphi}_{\mu\nu}, \quad c_{\mu\nu}(t) = \sum_{\mu',\nu'} E(t;\mu',\nu';\mu,\nu)c_{\mu'\nu'}$$

with convergence in $L^2(\mathbb{R}^2, \mathbb{C}^m)$. In short, the curvelet matrix maps the curvelet coefficients of the initial wavefield $u_0(\cdot)$ into those of the solution $u(t, \cdot)$ at time t.

Theorem 1.1. *Suppose that the coefficients $A_k(x)$ and $B(x)$ of the hyperbolic system are C^∞, with uniform smoothness constants, and that the multiplicity of the eigenvalues of the dispersion matrix $\sum_k A_k(x)\xi_k$ is constant in x and ξ. Then*

- *The matrix E is sparse. Suppose a is either a row or a column of E, and let $|a|_{(n)}$ be the n-largest entry of the sequence $|a|$, then for each $M > 0$, $|a|_{(n)}$ obeys*

$$|a|_{(n)} \leq C_{tM} \cdot n^{-M}. \tag{1.15}$$

- *The matrix E is well-organized. For each $N > 0$, the coefficients obey*

$$|E(t; \mu, \nu; \mu', \nu')| \leq C_{tN} \cdot \sum_{\nu''=1}^{m} \omega(\mu, \mu'_{\nu''}(t))^{-N}. \tag{1.16}$$

Here $\mu_\nu(t)$ is the curvelet index μ flown along the νth Hamiltonian system.

Both constants C_{tM} and C_{tN} grow in time at most like $C_1 e^{C_2 t}$ for some $C_1, C_2 > 0$ depending on M, resp. N.

In effect, the curvelet matrix of the solution operator resembles a sum of m permutation matrices where m is the order of the hyperbolic system; first, there are significant coefficients along m shifted diagonal and second, coefficients away from these diagonals decay nearly exponentially; i.e. faster than any negative polynomial. Now just as wavelets provide sparse representations to the solution operators to certain elliptic differential equations, our theorem shows that curvelets provide an optimally sparse representation of solution operators to systems of symmetric hyperbolic equations.

We can also resort to hyper-curvelets as defined in the previous section and formulate a related result where the curvelet matrix is sparse around a *single* shifted diagonal. This refinement approximately decouples the evolution into polarized components and will be made precise later.

To grasp the implications of Theorem 1.1, consider the following corollary:

Corollary 1.2. *Consider the truncated operator A_B obtained by keeping $m \cdot B$ elements per row—the B closest to each shifted diagonal in the sense of the pseudo-distance ω. Then the truncated matrix obeys*

$$\|A - A_B\|_{L^2 \to L^2} \leq C_M \cdot B^{-M}, \tag{1.17}$$

for each $M > 0$.

The proof follows from that of Theorem 1.1 by an application of Schur's lemma and is omitted. Hence, whereas the Fourier or wavelet representations are dense, curvelets faithfully model the geometry of wave propagation as only a few terms are needed to represent the action of the wave group accurately.

1.6 Strategy

In his seminal paper [19], Lax constructed approximate solution operators to linear and symmetric hyperbolic systems, also known as *parametrices*. He showed that these parametrices are oscillatory integrals in the frequency domain which are commonly referred to as Fourier integral operators (FIO) (the development and study of FIOs is motivated by the connection). An operator T is said to be an FIO if it is of the form

$$Tf(x) = \int e^{i\Phi(x,\xi)}\sigma(x,\xi)\hat{f}(\xi)\,d\xi. \tag{1.18}$$

We suppose the phase function Φ and the amplitude σ obey the following standard assumptions [30]:

8

- the phase $\Phi(x, \xi)$ is C^∞, with uniform smoothness constants in x, 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

$$|\det \Phi_{x\xi}(x, \xi)| > c > 0, \tag{1.19}$$

  uniformly in x and ξ;

- the amplitude σ is a symbol of order m, which means that σ is C^∞, and obeys

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x, \xi)| \leq C_{\alpha\beta}(1 + |\xi|)^{m-|\alpha|}. \tag{1.20}$$

Lax's insight is that the solution of the initial value problem for a variable coefficient hyperbolic system can be well approximated by a superposition of integrals of the form (1.18) with matrix-valued amplitudes of order 0. The phases of these FIO's are, of course, those solving the Eikonal equations (1.7). Hence, a substantial part of our argument will be about proving that curvelets sparsify FIO's. Now an important aspect of this construction is that this approximation is only valid for small times whereas our theorem is valid for *all* times. The reason is that the solutions to the Eikonal equations (1.7) are not expected to be global in time, because Φ_ν would become multi-valued when rays originating from the same point x_0 cross at a later time. This typically happens at cusp points, when caustics start developing. We refer the reader to [15, 34]. Because, we are interested in a statement valid for all times, we need to bootstrap the construction of the FIO parametrix by composing the small time FIO parametrix with itself. Now this creates an additional difficulty. Each parametrix convects a curvelet along m flows, and we see that after each composition, the number of curvelets would be multiplied by m, see section 4.1 for a proper discussion. This would lead to matrices with poor concentration properties. Therefore, the other part of the argument consists in decoupling the equations so that this phenomenon does not occur. In summary, the general architecture of the proof of Theorem 1.1 is as follows:

- We first decompose the wave-field into m-one way components, i.e. components which essentially travel along only one flow. We show that this decomposition is sparse in tight-frames of curvelets.

- Second, we show that curvelet representations of FIO's are optimally sparse in tight-frame of curvelets, a result of independent interest.

As a side remark, we would like to point out that the result about optimally sparse representations of FIO's was announced without a proof in the companion paper [4]. This paper, however, gives the first proof of this optimality result.

As the title of this paper suggests, we claim that curvelets provide optimally sparse representations of wave propagators just as wavelets are often said to provide the sparsest representation of large classes of pseudo-differential operators. Admittedly, this is an abuse since a rigorous justification would need to argue that there are no bases or tight frames which would provide faster decay; that is, in which the matrix would decay faster than any negative polynomial *simultaneously over all propagators* with sufficiently smooth coefficients. We have not pursued this issue here although there is considerable evidence supporting our claim; for example, it is likely that this is linked to the impossibility of approximating C^∞ functions at arbitrary high rates.

1.7 Inspiration and Relation to Other Work

Underlying our results is a mathematical insight concerning the central role for the analysis of hyperbolic differential equations, played by the *parabolic scaling*, in which analysis elements are supported in elongated regions obeying the relation *width* $\approx$ *length*^{2}. In fact, curvelets imply the same tiling of the frequency plane as the Second Dyadic Decomposition (SDD), a construction introduced in the seventies by Stein and Fefferman [16, 30], originally for the purpose of understanding boundedness of Riesz spherical means, and later widely adapted to the study of various Fourier integral operators. More specifically, we would like to

single out the work of Hart Smith with which we became familiar while working on this project. Smith [27] used parabolic scaling to define function spaces preserved by Fourier integral operators [27], and to analyze the behavior of wave equations with low-regularity coefficients [28]. The latter reference actually develops curvelet-like systems which provide a powerful tool to derive so-called sharp Strichartz estimates for solutions to such equations in space dimensions $d = 2, 3$ (a Strichartz estimate is a bound on the norm of the solution in some appropriate functional space, e.g. L^p). We find the connection with the work of Smith especially stimulating. From a broader viewpoint, the literature on the subject indicates that curvelets are in some sense compatible with a long tradition in harmonic analysis.

The fact that the action of a FIO should be seen as a 'rearrangement of wave packets' was discovered by Córdoba and Fefferman in their visionary paper [11]. They show how simple proofs of L^2 boundedness and the Garding inequality follow in a straightforward way from a decomposition into Gaussian wave packets.

Next, there is of course the inspiration of modern computational harmonic analysis (CHA) whose agenda is the development of orthobases, tight-frames, which are 'optimal' for representing objects (operators, functions) of scientific interest together with rapid algorithms to compute such representations. The point of view here is to develop new mathematical ideas and and turn these ideas into effective algorithms and these effective algorithms into effective and targeted applications. At the beginning of this introduction, we mentioned an instance of this scientific vision: (1) wavelets provide sparse representations of objects with punctuated smoothness and of large classes of singular integrals and other pseudo-differential operators [23, 24]; (2) there are fast discrete wavelet transforms operating in $O(N)$ for a signal of length N [22]; (3) this creates an opportunity for targeted applications in signal processing where wavelets allow for better compression [12], scientific computing where wavelets allow for faster algorithms [2], and for statistical estimation where wavelets allow for sharper reconstructions [13]. This vision was perhaps championed in [2] where (1)–(3) were combined to demonstrate how one can 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.

1.8 Significance

We would like to mention how we see our work fit with the vision described above.

- *Curvelets and wavefronts.* Curvelets are ideal for representing wavefront phenomena [8], or objects which display *curve-punctuated smoothness* —smoothness except for discontinuity along a general curve with bounded curvature [6, 7]. This fact originally motivated their construction [6, 7]. For example, [7] established that curvelets provide the sparsest representations of functions which are C^2 away from piecewise C^2 edges. Such representations are nearly as sparse as if the object were not singular and turn out to be far more sparse than the wavelet decomposition of the object.

  Hence, we see that curvelets provide the unique opportunity for having a representation giving enhanced sparsity of wave groups, and *simultaneously* of the solution space. We believe that this will eventually be of great practical significance for applications in fields which are great consumers of these mathematical models, e.g., seismic imaging.

- *New ideas for new numerical solvers.* Clearly, Theorem 1.1 may serve as a basis for faster geometric multiscale PDE solvers. In fact, this paper is the first of a projected series showing how one can exploit the structure of the curvelet transform and the enhanced sparsity of wave groups to derive new numerical low-complexity algorithms for accurately computing the solution to large classes of differential equations, see the concluding section for a discussion.

- *Digital curvelet transforms.* In order to deploy curvelet-like ideas in practical applications, one would need a digital notion of curvelet transform which (1) would be rapidly computable and (2) would be geometrically faithful in the sense that one would want an accurate digital analog of the corresponding geometric ideas defined at the level of the continuum. There actually is progress on this front. The

authors along with Donoho and Ying recently proposed two architectures for a Digital Curvelet Transform, one via Unequispaced Fast Fourier Transforms, and one using a Wedge Wrapping technique [5]. Both are fast algorithms which allow analysis and synthesis of Cartesian arrays as superpositions of discrete curvelets; for practical purposes, the algorithms run in $O(N \log N)$ operations for input array of size N. Digital curvelets obey sharp frequency and spatial localization.

In short, this paper is an essential piece of a much larger body of work.

1.9 Contents

Section 2 below reviews the construction of Curvelets. Section 3 below examines second-order scalar hyperbolic equations and gives a heuristic indicating why the sparsity may be expected to hold. Section 4 links our main result with properties of FIOs. Section 5 proves that FIO's are optimally sparse in scalar curvelet tight-frames. Section 6 discusses implications of this work, namely, in the area of scientific computing. Finally, proofs of key estimates supporting our main result are given in Section 7.

2 Curvelets

This section briefly introduces tight frames of curvelets, see [7] for more details.

2.1 Definition

We work throughout in $\mathbb{R}^2$, with spatial variable x, with ξ a frequency-domain variable, and with r and θ polar coordinates in the frequency-domain. We start with a pair of windows $W(r)$ and $V(t)$, which we will call the 'radial window' and 'angular window', respectively. These are both smooth, nonnegative and real-valued, with W taking positive real arguments and supported on $r \in [1/2, 2]$ and V taking real arguments and supported on $t \in [-1, 1]$. These windows will always obey the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \qquad r > 0; \tag{2.1}$$

$$\sum_{\ell=-\infty}^{\infty} V^2(t - \ell) = 1, \qquad t \in \mathbb{R}. \tag{2.2}$$

Now, for each $j \geq j_0$, we introduce $\varphi_j(x_1, x_2)$ defined in the Fourier domain by

$$\hat{\varphi}_j(\xi) = 2^{-3j/4} W(2^{-j}|\xi|) \cdot V(2^{\lfloor j/2 \rfloor} \theta) \tag{2.3}$$

Thus the support of $\hat{\varphi}_j$ is a polar 'wedge' defined by the support of W and V, the radial and angular windows, applied with scale-dependent window widths in each direction.

We may think of φ_j as a "mother" curvelet at scale 2^{-j} in the sense that all curvelets at that scaled are obtained by rotations and translations of φ_j. Introduce

- the equispaced sequence of *rotation angles* $\theta_{j,\ell} = 2\pi \cdot 2^{-\lfloor j/2 \rfloor} \cdot \ell$, $0 \leq \ell < L_j = 2^{\lfloor j/2 \rfloor}$,

- and the sequence of *translation parameters* $k = (k_1, k_2) \in \mathbb{Z}^2$.

With these notations, we define curvelets (as function of $x = (x_1, x_2)$) at scale 2^{-j}, orientation $\theta_{j,\ell}$ and position $b_k^{(j,\ell)} = R_{\theta_{j,\ell}}(k_1 \cdot 2^{-j}/\delta_1, k_2 \cdot 2^{-j/2}/\delta_2)$ for some adequate constants δ_1, δ_2 by

$$\varphi_{j,k,\ell}(x) = \varphi_j \left(R_{-\theta_{j,\ell}}(x - b_k^{(j,\ell)}) \right).$$

As in wavelet theory, we also have coarse scale elements. We introduce the low-pass window W_0 obeying

$$|W_0(r)|^2 + \sum_{j \geq 0} |W(2^{-j}r)|^2 = 1,$$

and for $k_1, k_2 \in \mathbb{Z}$, define coarse scale curvelets as

$$\Phi_{j_0,k}(x) = \Phi_{j_0}(x - 2^{-j_0}k), \quad \hat{\Phi}_{j_0}(\xi) = 2^{-j_0}W_0(2^{-j_0}|\xi|).$$

Hence, coarse scale curvelets are nondirectional. The 'full' curvelet transform consists of the fine-scale directional elements $(\varphi_{j,\ell,k})_{j \geq j_0, \ell, k}$ and of the coarse-scale isotropic father wavelets $(\Phi_{j_0,k})_k$. For our purposes, it is the behavior of the fine-scale directional elements that matters.

In the remainder of the paper, we will use the generic notation $(\varphi_\mu)_{\mu \in M}$ to index the elements of the curvelet tight-frame. The dyadic-parabolic subscript μ stands for the triplet (j, k, ℓ). We will also make use of the convenient notations

- $x_\mu = b_k^{(j,\ell)}$ is the center of φ_μ in space.

- $\theta_\mu = \theta_{j,\ell}$ is the orientation of φ_μ with respect to the vertical axis in x.

- $\xi_\mu = (2^j \cos\theta_\mu, 2^j \sin\theta_\mu)$ is the center of $\hat{\varphi}_\mu$ in frequency.

- $e_\mu = \xi_\mu/|\xi_\mu|$ indicates the codirection of φ_μ.

2.2   Properties

We now list a few properties of the curvelet transform which will play an important role throughout the remainder of this paper.

1. **Tight-frame**. Much like in an orthonormal basis, we can easily expand an arbitrary function $f(x_1, x_2) \in L^2(\mathbb{R}^2)$ as a series of curvelets: we have a reconstruction formula

$$f = \sum_\mu \langle f, \varphi_\mu \rangle \varphi_\mu,$$

with equality holding in an L^2 sense; and a Parseval relation

$$\sum_\mu |\langle f, \varphi_\mu \rangle|^2 = \|f\|_{L^2(\mathbb{R}^2)}^2, \quad \forall f \in L^2(\mathbb{R}^2).$$

2. **Parabolic scaling**. The frequency localization of φ_j implies the following spatial structure: $\varphi_j(x)$ is of rapid decay away from an 2^{-j} by $2^{-j/2}$ rectangle with minor axis pointing in the horizontal direction. In short, the effective length and width obey the anisotropy scaling relation

$$length \approx 2^{-j/2}, \quad width \approx 2^{-j} \quad \Rightarrow \quad width \approx length^2. \tag{2.4}$$

3. **Oscillatory behavior**. As is apparent from its definition, $\hat{\varphi}_j$ is actually supported away from the vertical axis $\xi_1 = 0$ but near the horizontal $\xi_2 = 0$ axis. In a nutshell, this says that $\varphi_j(x)$ is oscillatory in the x_1-direction and lowpass in the x_2-direction. Hence, 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'.
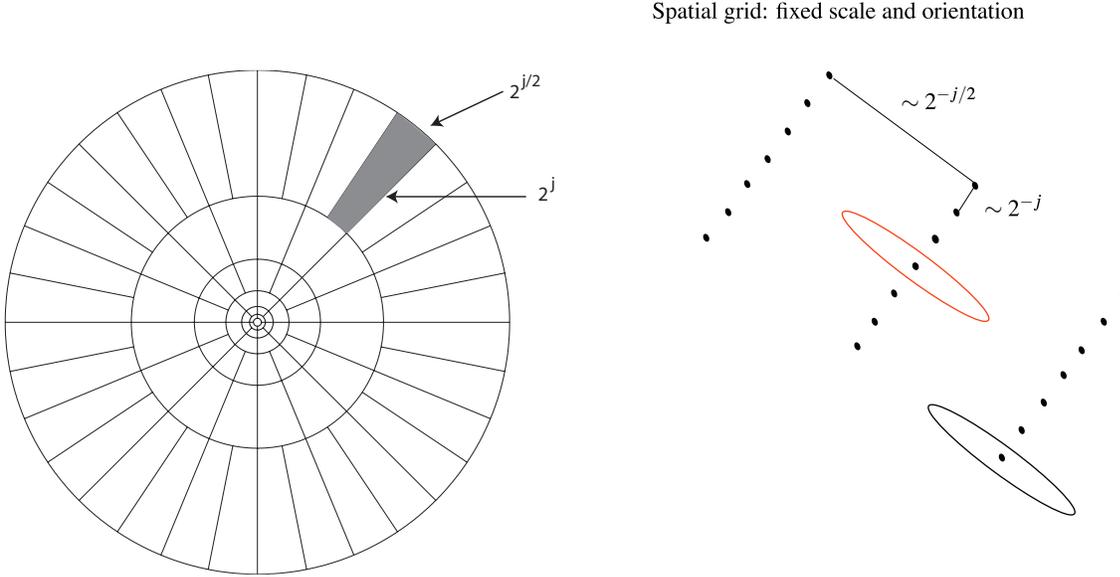
Spatial grid: fixed scale and orientation

Figure 3: Curvelet tiling of Phase-Space. The figure on the left represents the sampling in the frequency plane. In the frequency domain, curvelets are supported near a 'parabolic' wedge. The shaded area represents such a generic wedge. The figure on the right schematically represents the spatial Cartesian grid associated with a given scale and orientation.

4. **Phase-Space Tiling/Sampling**. We can really think about curvelets as Heisenberg tiles of minimum volume in phase-space. In x, the essential support of φ_μ has size $O(2^{-j} \times 2^{-j/2})$. In frequency, the support of $\hat{\varphi}_\mu$ has size $O(2^{j/2} \times 2^j)$. The net volume in phase-space is therefore

$$O(2^{-j} \times 2^{-j/2}) \cdot O(2^{j/2} \times 2^j) = O(1),$$

which is in accordance with the uncertainty principle. The parameters (j, k, ℓ) of the curvelet transform induce a new non-trivial sampling of phase-space, Cartesian in x, polar in ξ, and based on the parabolic scaling.

5. **Complex-valuedness**. Since curvelets do not obey the symmetry $\hat{\varphi}_\mu(-\xi) = \overline{\hat{\varphi}_\mu(\xi)}$, φ_μ is complex-valued. There exists a related construction for real-valued curvelets by simply symmetrizing the construction, see [7]. The complex-valued transform is better adapted to the purpose of this paper.

Figure 3 summarizes the key components of the construction.

2.3   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 we have just introduced. Our formulation is inspired by the notion of 'vaguelettes' in wavelet analysis [24]. Our motivation for introducing this concept is the fact that operators of interest do not map curvelets into curvelets, but rather into these molecules. Note that the terminology 'molecule' is somewhat standard in the literature of harmonic analysis [18].

Definition 2.1. *A family of functions $(m_\mu)_\mu$ is said to be a family of curvelet molecules with regularity R if (for $j > 0$) they may be expressed as*

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

13

where $k' = (\frac{k_1}{\delta_1}, \frac{k_2}{\delta_2})$ and where for all μ, the $a^{(\mu)}$'s verify the following properties:

- Smoothness and spatial localization: for each $|\beta| \leq R$, and each $M = 0, 1, 2 \ldots$, there is a constant $C_M > 0$ such that

$$|\partial_x^\beta a^{(\mu)}(x)| \leq C_M \cdot (1 + |x|)^{-M}. \tag{2.5}$$

- Nearly vanishing moments: for each $N = 0, 1, \ldots, R$, there is a constant $C_N > 0$ such that

$$|\hat{a}^{(\mu)}(\xi)| \leq C_N \cdot \min(1, 2^{-j} + |\xi_1| + 2^{-j/2}|\xi_2|)^N. \tag{2.6}$$

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 (2.5).

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

$$|m_\mu(x)| \leq C_M \cdot 2^{3j/4} \cdot \left(1 + |2^j x_1 - \frac{k_1}{\delta_1}| + |2^{j/2} x_2 - \frac{k_2}{\delta_2}|\right)^{-M} \tag{2.7}$$

for each $M > 0$ and $|\beta| \leq R$, and similarly for its derivatives

$$|\partial_x^\beta m_\mu(x)| \leq C_M \cdot 2^{3j/4} \cdot 2^{(\beta_1 + \beta_2/2)j} \cdot \left(1 + |2^j x_1 - \frac{k_1}{\delta_1}| + |2^{j/2} x_2 - \frac{k_2}{\delta_2}|\right)^{-M}. \tag{2.8}$$

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

$$|\hat{m}_\mu(\xi)| \leq C_N \cdot 2^{-3j/4} \cdot \min(1, 2^{-j}(1 + |\xi|))^N, \tag{2.9}$$

which is valid for every $N \leq R$, which gives the frequency localization

$$|\hat{m}_\mu(\xi)| \leq C_N \cdot 2^{-3j/4} \cdot |S_\mu(\xi)|^N \tag{2.10}$$

where for $\mu_0 = (j, 0, 0)$,

$$S_{\mu_0}(\xi) = \min(1, 2^{-j}(1 + |\xi|)) \cdot (1 + |2^{-j}\xi_1| + |2^{-j/2}\xi_2|)^{-1}. \tag{2.11}$$

For arbitrary μ, S_μ is obtained from S_{μ_0} by a simple rotation of angle θ_μ, i.e. $S_{\mu_0}(R_{\theta_\mu}\xi)$. Similar estimates are available for the derivatives of $\hat{\varphi}_\mu$.

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. It is easy to show that curvelets as introduced in the previous section are indeed curvelet molecules for arbitrary degrees R of regularity.

2.4   Near Orthogonality of Curvelet Molecules

Curvelets are not necessarily orthogonal to each other[1], but in some sense they are almost orthogonal. As we show below, the inner product between two molecules m_μ and $p_{\mu'}$ decays nearly exponentially as a function of the 'distance' between the subscripts μ and μ'.

This notion of distance in phase-space, tailored to curvelet analysis, is to be understood as follows. Given a pair of indices $\mu = (j, k, \ell), \mu' = (j', k', \ell')$, define the dyadic-parabolic pseudo-distance

$$\omega(\mu, \mu') = 2^{|j - j'|} \cdot \left(1 + \min(2^j, 2^{j'}) d(\mu, \mu')\right), \tag{2.12}$$

---

[1]It is an open problem whether orthobases of curvelets exist or not.

where
$$d(\mu, \mu') = |\theta_\mu - \theta_{\mu'}|^2 + |x_\mu - x_{\mu'}|^2 + |\langle e_\mu, x_\mu - x_{\mu'}\rangle|.$$

Angle differences like $\theta_\mu - \theta_{\mu'}$ are understood modulo π. As introduced earlier, e_μ is the codirection of the first molecule, i.e., $e_\mu = (\cos\theta_\mu, \sin\theta_\mu)$.

The pseudo-distance (2.12) is a slight variation on that introduced by Smith [28]. We see that ω increases by at most a constant factor every time the distance between the scale, angular, and location parameters increases. The extension of the definition of ω to arbitrary points (x, ξ) and (x', ξ') is straightforward. Observe that the extra term $|\langle e_\mu, x_\mu - x_{\mu'}\rangle|$ induces a non-Euclidean notion of distance between x_μ and $x_{\mu'}$. The following properties of ω are proved in section A.1. (The notation $A \asymp B$ means that $C_1 \leq A/B \leq C_2$ for some constants $C_1, C_2 > 0$.)

Proposition 2.2.   *1. Symmetry: $\omega(\mu, \mu') \asymp \omega(\mu', \mu)$.*

2. Triangle inequality: $d(\mu, \mu') \leq C \cdot (d(\mu, \mu'') + d(\mu'', \mu'))$ for some constant $C > 0$.

3. Composition: for every integer $N > 0$, and some positive constant C_N

$$\sum_{\mu''} \omega(\mu, \mu'')^{-N} \cdot \omega(\mu'', \mu')^{-N} \leq C_N \cdot \omega(\mu, \mu')^{-(N-1)}.$$

4. Invariance under Hamiltonian flows: $\omega(\mu, \mu') \asymp \omega(\mu_\nu(t), \mu'_\nu(t))$.

We can now state the almost orthogonality result

Lemma 2.3. *Let $(m_\mu)_\mu$ and $(p_{\mu'})_{\mu'}$ be two families of curvelet molecules with regularity R. Then for $j, j' \geq 0$,*
$$|\langle m_\mu, p_{\mu'}\rangle| \leq C_N \cdot \omega(\mu, \mu')^{-N}. \tag{2.13}$$
for every $N \leq f(R)$ where $f(R)$ goes to infinity as R goes to infinity.

Proof. Throughout the proof of (2.13), 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 $M \leq M'$. Here and throughout, the constants C may vary from expression to expression.

For notational convenience put $\Delta\theta = \theta_\mu - \theta_{\mu'}$ and $\Delta x = x_\mu - x_{\mu'}$. We abuse notation by letting m_{μ_0} be the molecule $a^{(\mu)}(D_{2^{-j}} R_{\theta_\mu} x)$, i.e., m_{μ_0} 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

$$I_{\mu\mu'} = \frac{1}{(2\pi)^2} \int \hat{m}_{\mu_0}(\xi) \overline{\hat{p}_{\mu'_0}(\xi)}\, e^{-i(\Delta x)\cdot\xi}\, d\xi.$$

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

$$\int |S_{\mu_0}(\xi) S_{\mu'_0}(\xi)|^N\, d\xi \leq C \cdot 2^{3j/4 + 3j'/4} \cdot 2^{-|j-j'|N} \cdot \left(1 + 2^{j_0}|\Delta\theta|^2\right)^{-N}, \tag{2.14}$$

where S_{μ_0} is defined in equation (2.11). Therefore, the frequency localization of the curvelet molecules (2.10) gives

$$\begin{aligned}
\int |\hat{m}_{\mu_0}(\xi)|\, |\hat{p}_{\mu'_0}(\xi)|\, d\xi &\leq C \cdot 2^{-3j/4 - 3j'/4} \cdot \int |S_{\mu_0}(\xi) S_{\mu'_0}(\xi)|^N\, d\xi \\
&\leq C \cdot 2^{-|j-j'|N} \cdot \left(1 + 2^{j_0}|\Delta\theta|^2\right)^{-N}.
\end{aligned} \tag{2.15}$$

This inequality explains the angular decay. A series of integrations by parts will introduce the spatial decay, as we now show.

The partial derivatives of $\hat{m}_\mu$ obey

$$|\partial_\xi^\alpha \hat{m}_\mu(\xi)| \le C \cdot 2^{-3j/4} \cdot 2^{-j(\alpha_1 + \frac{\alpha_2}{2})} \cdot |S_\mu(\xi)|^N.$$

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

$$|\Delta_\xi \hat{p}_{\mu'}(\xi)| \le C \cdot 2^{-3j'/4} \cdot 2^{-j'} \cdot |S_{\mu'}(\xi)|^N,$$

$$|\frac{\partial^2}{\partial \xi_1^2} \hat{p}_{\mu'}(\xi)| \le C \cdot 2^{-3j'/4} \cdot (2^{-2j'} + 2^{-j'} |\sin(\Delta\theta)|^2) \cdot |S_{\mu'}(\xi)|^N.$$

Recall that for $t \in [-\pi/2, \pi/2]$, $2/\pi \cdot |t| \le |\sin t| \le |t|$, 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, $L^k(\hat{m}_\mu \overline{\hat{p}_{\mu'}})$ obeys

$$|L^k(\hat{m}_\mu \overline{\hat{p}_{\mu'}})(\xi)| \le C \cdot 2^{-3j/4 - 3j'/4} \cdot |S_\mu(\xi)|^N \cdot |S_{\mu'}(\xi)|^N.$$

On the other hand

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

Therefore, a few integrations by parts give

$$|I_{\mu\mu'}| \le C \cdot 2^{-|j-j'|N} \cdot \left(1 + 2^{j_0} |\theta_\mu - \theta_{\mu'}|^2\right)^{-N}$$

$$\cdot \left(1 + 2^{j_0} |\Delta x|^2 + \frac{2^{2j_0}}{1 + 2^{j_0} |\Delta\theta|^2} |\langle e_\mu, \Delta x\rangle|^2\right)^{-N},$$

and then

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

One can simplify this expression by noticing that

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

This yields equation (2.13) as required. $\qquad\square$

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

$$|\langle m_\mu, p_{\mu'}\rangle| \le C \cdot 2^{-jN} \cdot \left(1 + |x_\mu - x_{\mu'}|^2 + |\langle e_\mu, x_\mu - x_{\mu'}\rangle|\right)^{-N}.$$

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

$$|\langle m_\mu, p_{\mu'}\rangle| \le C \cdot (1 + |x_\mu - x_{\mu'}|)^{-N}.$$

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 2.4. *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

$$\sum_{\mu \in M_{j'}} \left(1 + 2^{j_0}(d(\mu, \mu'))\right)^{-Np} \leq C \cdot 2^{2|j-j'|} \tag{2.16}$$

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. $\qquad\square$

3   Heuristics

This section explains the organization of the argument underlying the proof of the main result, namely, Theorem 1.1, and gives the main reasons why curvelets are special.

3.1   Architecture of the proof of the main result

- *Decoupling into polarized components.* The first step is to decouple the wavefield $u(t, x)$ into m one-way components $f_\nu(t, x)$

$$u(t, x) = \sum_{\nu=1}^m R_\nu f_\nu(t, x),$$

where the R_ν are operators mapping scalars to m-dimensional vectors, and independent of time. The f_ν will also be called 'polarized' components. This allows a separate study of the m flows corresponding to the m eigenvalues of the matrix $\sum_{k=1}^m A_k(x)\xi_k$. In the event these eigenvalues are simple, the evolution operator $E(t)$ can be decomposed as

$$E(t) = \sum_{\nu=1}^m R_\nu e^{-it\Lambda_\nu} L_\nu + \text{ negligible}, \tag{3.1}$$

where the L_ν's are operators mapping m-dimensional vectors to scalars and the Λ_ν's are one-way wave operators acting on scalar functions. In effect, each operator $E_\nu(t) = e^{-it\Lambda_\nu}$ convects wave-fronts and other singularities along a separate flow. The 'negligible' contribution is a smoothing operator—not necessarily small. The composition operators R_ν and decomposition operators L_ν are provably pseudo-differential operators, see section 4.2.

17

- *Fourier integral operator parametrix.* We then approximate for small times $t > 0$ each $e^{-it\Lambda_\nu}$, $\nu = 1, \ldots, m$, by an oscillatory integral or Fourier integral operator (FIO) $F_\nu(t)$. Such operators take the form

$$F_\nu(t)f(x) = \int e^{i\Phi_\nu(t,x,\xi)} \sigma_\nu(t,x,\xi)\hat{f}(\xi)\,d\xi,$$

under suitable conditions on the phase function $\Phi_\nu(t,x,\xi)$ and the amplitude $\sigma_\nu(t,x,\xi)$. Again, the identification of the evolution operator $E_\nu(t) = e^{-it\Lambda_\nu}$ with F_ν is valid up to a smoothing and localized additive remainder. The construction of the so-called *parametrix* $F_\nu(t)$ and its properties are detailed in section 4.3.

  Historically [19], the construction of an oscillatory integral parametrix did not involve the decoupling into polarized components as a preliminary step. When applied directly to the system (4.1), the construction of the parametrix gives rise to a matrix-valued amplitude $\sigma(t,x,\xi)$ where all the couplings are present. This somewhat simpler setting, however, is not adequate for our purpose. The reason is that we want to bootstrap the construction of a parametrix to *large* times by composing the small time FIO parametrix with itself, $F(nt) = [F(t)]^n$. Without decoupling of the propagation modes, each $E(t)$ or $F(t)$ involves convection of singularities along m families of characteristics or flows. Applying $F(t)$ again, each flow would artificially split into m flows again, yielding m^2 fronts to keep track of. At time $T = nt$, that would be at most m^n fronts. This flow-splitting situation is not physical and can be avoided by isolating one-way components *before* constructing the parametrix. The correct large-time argument is to consider $E_\nu(nt)$ for small $t > 0$ and large integer n as $[E_\nu(t)]^n$. This expression involves one single flow, indexed by ν.

- *Sparsity of Fourier integral operators.* The core of the proof is found in section 5 and consists in showing that very general FIO's $F(t)$, including the parametrices $F_\nu(t)$, are sparse and well-structured when represented in tight frames of (scalar) curvelets φ_μ. The scalar analog of Theorem 1.1 for FIO's is Theorem 5.1—a statement of independent interest. Observe that pseudo-differential operators are a special class of FIO's and, therefore, are equally sparse in a curvelet frame.

Section 4.5 assembles key intermediate results and proves Theorem 1.1.

3.2   The parabolic scaling is special

Why is the curvelet parabolic scaling the only correct way to scale a family of wave packets to sparsely represent wave groups? Assume for a moment that the curvelet scaling $width \approx length^2$ is replaced with the more general power-law

$$width \approx length^\alpha, \quad 1 \le \alpha \le \infty,$$

and that one has available a tight frame φ_μ of "α-wave-packets." For example, $\alpha = 1$ would correspond to wavelets and $\alpha = \infty$ to ridgelets [3].

  Consider a wave packet $\varphi_\mu(x)$ centered around x_μ in space and ξ_μ in frequency. The action of a Fourier integral operator on this wave packet can be viewed as the composition of two transformations, (1) non-rigid convection along the Hamiltonian flow due to the phase factor $\Phi(x,t,\xi)$ (or more precisely its linearization $\xi \cdot \nabla_\xi \Phi(t,x,\xi_\mu)$ around ξ_μ) and (2) microlocal dispersion due to the remainder after linearization and the amplitude $\sigma(t,x,\xi)$. Depending upon the size of the essential support in phase-space (controlled by the value of α), these two transformations may leave the shape of the waveform nearly invariant, or not. We now argue that the curvelet parabolic scaling, $\alpha = 2$, offers the correct compromise.

1. **Spatial localization**. For simplicity, suppose that one can model the convective effect by a smooth diffeomorphism $g(x)$, so that a wave-packet $\varphi_\mu(x)$ is effectively mapped into $\varphi_\mu(g(x))$. If we Taylor expand $g(x)$ around $y_\mu = g^{-1}(x_\mu)$, where x_μ is the center of $\phi_\mu(x)$, we obtain

$$g(x) = x_\mu + (x - x_\mu)g'(y_\mu) + O((x - x_\mu)^2).$$

The first two terms induce an essentially rigid motion, while the remainder is responsible for deforming the waveform. The requirement for optimal sparsity, as it turns out, is that the extent of the deformation should not exceed the width of the wave packet. In the case of curvelets, this imposes the correct condition for $\varphi_\mu(g(x))$ to remain a 'molecule' in the sense defined in section 2, see how equation (5.25) combines with the molecule estimate (2.7).

If the spatial width is of the order of $a = 2^{-j}$, then the wave packet should essentially be supported in a region obeying $(x - x_\mu) \sim 2^{-j/2}$. This is satisfied if and only if $1 \leq \alpha \leq 2$. In short, *any scaling more isotropic than the parabolic scaling works.*

2. **Frequency localization**. Dispersive effects are already present in the wave equation with constant velocity $c = 1$,

$$\frac{\partial^2 u}{\partial t^2} = \Delta u,$$

with initial conditions $u(x, 0) = u_0(x)$, $\frac{\partial u}{\partial t}(x, 0) = u_1(x)$. In the Fourier domain, the solution is given by

$$\hat{u}(t, \xi) = \cos(|\xi|t)\, \hat{u}_0(\xi) + \frac{\sin(|\xi|t)}{|\xi|}\, \hat{u}_1(\xi).$$

These multipliers are of course associated with the phases $\Phi_\pm(t, x, \xi) = x \cdot \xi \pm t|\xi|$ (express sine and cosine in terms of complex exponentials). Linearize $\Phi_\pm$ around ξ_μ, the center of $\hat{\varphi}_\mu$, and obtain

$$\Phi_\pm(t, x, \xi) = (x \pm t\mathbf{e}_\mu) \cdot \xi \pm t(|\xi| - \xi \cdot \mathbf{e}_\mu),$$

where $\mathbf{e}_\mu = \frac{\xi_\mu}{|\xi_\mu|}$. The first term is responsible for convection as before while the second is responsible for dispersion (transverse to the oscillations of the wave packet). Again, we must invoke more sophisticated arguments to see that to achieve sparsity, one needs $\delta(\xi) = |\xi| - \xi \cdot \mathbf{e}_\mu$ to be uniformly bounded over the frequency support of the wave packet $\hat{\varphi}_\mu$ as to make the remainder $e^{it\delta(\xi)}$ non-oscillatory. This would effectively transform each wave-packet into a proper 'molecule.' For curvelets, see how equation (5.6) depends on the crucial estimate (A.8) about the phase, and how this implies the molecule inequality (2.10).

It is easy to see that $\delta(\xi)$ is zero on the line $\xi = \text{const} \times \mathbf{e}_\mu$, and proportional to $\frac{(\xi \cdot \mathbf{e}_\mu^\perp)^2}{|\xi \cdot \mathbf{e}_\mu|}$ away from it. If $\hat{\varphi}_\mu$ is supported around ξ_μ so that $|\xi_\mu| \sim 2^j$, and the support lies well away from the origin, then $\delta(\xi) \leq \text{const}$ implies that $\xi \cdot \mathbf{e}_\mu^\perp$ be bounded by constant times $2^{j/2}$. This is saying that the width of the support should be at most the square root of the length (in frequency), i.e. $2 \leq \alpha \leq \infty$. In short, *any scaling more anisotropic than the parabolic scaling works.*

In conclusion, only the parabolic scaling, $\alpha = 2$, allows to formulate a sparsity result like Theorem 1.1 because it meets both requirements of small warping and small dispersion effects.

4 Representation of Linear Hyperbolic Systems

We now return to the main theme of this paper and consider linear initial-value problems of the form

$$\frac{\partial u}{\partial t} + \sum_{k=1}^m A_k(x) \frac{\partial u}{\partial x_k} + B(x)u = 0, \qquad u(0, x) = u_0(x), \tag{4.1}$$

where in addition to the properties listed in the introduction, A_k and B together with all their partial derivatives are uniformly bounded for $x \in \mathbb{R}^n$. As explained in section 4.2, we need to make the technical assumption that for every set of real parameters ξ_k, the (real) eigenvalues of the matrix $\sum_k A_k(x)\xi_k$ have constant multiplicity in x and ξ.

Our goal is to construct a concrete 'basis' of $L^2(\mathbb{R}^n, \mathbb{C}^m)$ in which the evolution is as simple/sparse as possible. We present a solution based on the newly developed curvelets—which were introduced in section 2—and choose to specialize our discussion to $n = 2$ spatial dimensions. The reason is twofold: first, this setting is indeed that in which the exposition of curvelets is the most convenient; and second, this is not a restriction as similar results would hold in arbitrary dimensions.

4.1 Main result

We need to prove

$$|E(t; \mu, \nu; \mu', \nu')| \leq C_{t,N} \cdot \sum_{\nu''} \omega(\mu, \mu'_{\nu''}(t))^{-N}, \tag{4.2}$$

for some constant $C_{t,N} > 0$ growing at most like $C_N e^{K_N t}$ for some $C_N, K_N > 0$. The sum over ν'' indexes the different flows and takes on as many values as there are distinct eigenvalues $\lambda^0_{\nu''}$.

It is instructive to notice that the estimate (4.2) for $t = 0$ is already the strongest of its sort on the off-diagonal decay of the Gram matrix elements for a tight frame of curvelets. For $t > 0$, equation (4.2) states that the strong phase-space localization of every curvelet is preserved by the hyperbolic system, thus yielding a sparse and well-organized structure for the curvelet matrix. These warped and displaced curvelets are 'curvelet molecules' as introduced in section 2.3 because, as we will show, they obey the estimates (2.5) and (2.6).

The choice of the curvelet family being complex-valued in the above theorem is not essential. $E(t)$ acting on real-valued curvelets would yield *two* molecules per flow (upstream and downstream). 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 the curvelet matrix can be recovered by expressing each real curvelet as a superposition of two complex curvelets.

The following two sections present results which are for the most part established knowledge in the theory of hyperbolic equations. For example, we borrow some methods and results from geometric optics [19] and most notably from Taylor [32] and Stolk and de Hoop [31]. The goal here is to keep the exposition self-contained and at a reasonable level, and to recast prior results in the framework adopted here, which is sometimes significantly different than that used by the original contributors.

4.2 Decoupling into polarized components

How to disentangle the vector wavefield into m independent components is perhaps best understood in the special case of constant coefficients, $A_k(x) = A_k$, and with $B(x) = 0$. In this case, applying the 2-dimensional Fourier transform on both sides of (4.1) gives a system of ordinary differential equations

$$\frac{d\hat{u}}{dt}(t, \xi) + ia(\xi)\hat{u}(t, \xi) = 0, \quad a(\xi) = \sum_k A_k \xi_k.$$

(Note that $a(\xi)$ is a symmetric matrix with real entries.) It follows from our assumptions that one can find m real eigenvalues $\lambda_\nu(\xi)$ and orthonormal eigenvectors $r_\nu(\xi)$, so that

$$a(\xi)r_\nu(\xi) = \lambda_\nu(\xi)r_\nu(\xi).$$

Put $f_\nu(t, \xi) = r_\nu(\xi) \cdot \hat{u}(t, \xi)$. Then our system of equations is of course equivalent to the system of independent scalar equations

$$\frac{df_\nu}{dt}(t, \xi) + i\lambda_\nu(\xi)f_\nu(\xi) = 0,$$

which can then be solved for explicitly;

$$f_\nu(t, \xi) = e^{-it\lambda_\nu(\xi)} f_\nu(0, \xi).$$

Hence, the diagonalization of $a(\xi)$ decouples the original equation (4.1) into m *polarized* components; these can be interpreted as waves going in definite directions, for example 'up and down' or 'outgoing and incoming' depending on the geometry of the problem. This is the reason why f_ν is also referred to as being a 'one-way' wavefield.

The situation is more complicated when $A_k(x)$ is non-uniform since Fourier techniques break down. A useful tool in the variable coefficient setting is the calculus of pseudo-differential operators. An operator T is said to be *pseudo-differential* with symbol σ if it can be represented as

$$Tf(x) = \sigma(x, D)f = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{ix\cdot\xi}\sigma(x,\xi)\hat{f}(\xi)\,d\xi, \qquad (4.3)$$

with the convention that $D = -i\nabla$. It is of type $(1,0)$ and order m if σ obeys the estimate

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x,\xi)| \le C_{\alpha,\beta} \cdot (1 + |\xi|)^{m-|\alpha|}$$

for every multi-indices α and β. Unless otherwise stated, all pseudo-differential operators in this paper are of type $(1,0)$. An operator is said to be smoothing of order $-\infty$, or simply smoothing if its symbol satisfies the above inequality for every $m < 0$. Observe that this is equivalent to the property that T maps boundedly distributions in the Sobolev space H^{-s} to functions in H^s for every $s > 0$, in addition to a strong localization property of its kernel $G(x,y)$ which says that for each $N > 0$, there is a constant $C_N > 0$ such that G obeys

$$|G(x,y)| \le C_N \cdot (1 + |x - y|)^{-N} \qquad (4.4)$$

as in [30][Chapter 6].

Now set

$$a(x, D) = \sum_{k=1}^{m} A_k(x)D_k - iB(x),$$

and its principal part

$$a^0(x, D) = \sum_{k=1}^{m} A_k(x)D_k,$$

so that equation (4.1) becomes $\partial_t u + ia(x, D)u = 0$. The matrices $a(x,\xi)$ (resp. $a^0(x,\xi)$) are called the symbol of the operator $a(x, D)$ (resp. $a^0(x, D)$). Note that $a^0(x,\xi)$ is homogeneous of degree one in ξ; a^0 also goes by the name of *dispersion matrix*.

It follows from the symmetry of A_k and B that for every set of real parameters $\xi_1, \ldots, \xi_m$, the matrix $a^0(x,\xi) = \sum_k A_k(x)\xi_k$ is also symmetric and thus admits real eigenvalues $\lambda_\nu^0(x,\xi)$ and an orthonormal basis of eigenvectors $r_\nu^0(x,\xi)$,

$$a^0(x,\xi)r_\nu^0(x,\xi) = \lambda_\nu^0(x,\xi)r_\nu^0(x,\xi). \qquad (4.5)$$

The eigenvalues being real and the set of eigenvectors complete is a *hyperbolicity* condition and ensures that equation (4.1) will admit wave-like solutions. We assume throughout this paper that the multiplicity of each $\lambda_\nu^0(x,\xi)$ is constant in x and ξ.

By analogy with the special case of constant coefficients, a first impulse may be to introduce the components $r_\nu^0(x, D) \cdot u$, where $r_\nu^0(x, D)$ is the operator associated to the eigenvector $r_\nu^0(x,\xi)$ by the standard rule (4.3). In particular this is how we defined hyper-curvelets from curvelets in section 1.4. Unfortunately, this does not perfectly decouple the system into m polarized modes—it only approximately decouples. Instead, we would achieve perfect decoupling if we could solve the eigenvalue problem

$$a(x, D)r_\nu(x, D) = r_\nu(x, D)\lambda_\nu(x, D). \qquad (4.6)$$

Here, each $\Lambda_\nu = \lambda_\nu(x, D)$ is a scalar operator and $R_\nu = r_\nu(x, D)$ is an m-by-1 vector of operators. Equation (4.6) must be understood in the sense of composition of operators. Now let f_ν be the polarized components obeying the scalar equation

$$\frac{\partial f_\nu}{\partial t} + i\Lambda_\nu f_\nu = 0, \qquad (4.7)$$

with initial condition $f_\nu(0, x)$ and consider the superposition

$$u = \sum_\nu u_\nu, \quad u_\nu = R_\nu f_\nu.$$

Then u is a solution to our initial-value problem (4.1). (We will make this rigorous later, and detail the dependence between the initial values u_0 and the $f_\nu(0, \cdot)$'s.)

The following result shows how in some cases, (4.6) can be solved up to a smoothing remainder of order $-\infty$. When all the eigenvalues $\lambda_\nu^0(x, \xi)$ are simple, the exact diagonalization is, in fact, possible. The situation is more complicated when some of the eigenvalues are degenerate; further decoupling within the eigenspaces is in general not possible. This complication does not compromise, however, any of our results.

The theorem is due to Taylor [32], Stolk and de Hoop [31].

Lemma 4.1. *Suppose our hyperbolic system satisfies all the assumptions stated below (4.1). Then there exists an m-by-m block-diagonal matrix of operators Λ and two m-by-m matrices of operators R and S such that*

$$a(x, D)R = R\Lambda + S,$$

where Λ, R and S are componentwise pseudo-differential with Λ of order one, R of order zero, and S of order $-\infty$. Each block of Λ corresponds to a distinct eigenvalue λ_ν^0 whose size equals the multiplicity of that eigenvalue. The principal symbol of Λ is diagonal with the eigenvalues $\lambda_\nu^0(x, \xi)$ as entries.

Sketch of proof. Let us only give the ideas of what might be an alternative easy proof of the Taylor-Stolk-de Hoop lemma.

It is common practice to define the *twisted product* of two symbols σ and τ as

$$(\sigma \sharp \tau)(x, D) = \sigma(x, D)\tau(x, D),$$

so that (4.6) becomes the symbol equation

$$a \sharp r_\nu = r_\nu \sharp \lambda_\nu. \tag{4.8}$$

It turns out that $\sigma \sharp \tau$ is the product $\sigma\tau$ up to terms that are at least one order lower in ξ. For more information on twisted products, see [17].

The idea is now to introduce polyhomogeneous expansions

$$r_\nu \sim r_\nu^0 + r_\nu^1 + r_\nu^2 + \dots, \qquad \lambda_\nu \sim \lambda_\nu^0 + \lambda_\nu^1 + \lambda_\nu^2 + \dots$$

so that r_ν^n is of order $-n$ in ξ and λ_ν^n of order $-n+1$, and find the terms iteratively by identification of the integer powers of $|\xi|$ in equation (4.8). It can be seen that the contribution at the leading order is, of course, $a^0 r_\nu^0 = \lambda_\nu^0 r_\nu^0$. All the higher-order equations can be solved for. Issues about convergence to valid symbols are typically treated by using suitable cutoffs, see for example [29, 33]. If an eigenvalue λ_ν^0 has multiplicity p, each λ_ν^j stands for a p-by-p matrix and the reasoning is very similar. $\square$

The above construction indeed provides efficient decoupling of the original problem (4.1) into polarized modes. The following lemma is a straightforward consequence of Lemma 4.1 although we have not been able to find it in the literature. See [31] for related results.

Lemma 4.2. *In the setting of Lemma 4.1, the solution operator $E(t)$ for (4.1) may be decomposed for all times $t > 0$ as*

$$E(t) = Re^{-it\Lambda}L + \tilde{S}(t),$$

where the matrices of operators Λ and R are defined in Lemma 4.1 and $\tilde{S}(t)$ is (another) matrix of smoothing operators of order $-\infty$. In addition,

1. *L is an approximate inverse of R, i.e. $RL = I$ and $LR = I$ (mod smoothing).*

2. L is a pseudo-differential of order zero (componentwise).

Observe that $e^{-it\Lambda}$ inherits the block structure from Λ, and is diagonal in the case where all the eigenvalues λ_ν^0 are simple.

Proof. Begin by observing that $R = r(x, D)$—as an operator acting on $L^2(\mathbb{R}^2, \mathbb{C}^m)$—is invertible modulo a smoothing additive term. This means that one can construct a parametrix L so that $LR = I$ and $RL = I$ with both equations holding modulo a smoothing operator. To see why this is true, note that the matrix $r(x, \xi)$ is a lower-order perturbation from the unitary matrix $r^0(x, \xi)$ of eigenvectors of the principal symbol $a^0(x, \xi)$. The inverse of $r^0(x, \xi)$ is explicitly given by $\ell^0(x, \xi) = r^0(x, \xi)^*$. The symbol of L can now be built as an expansion $\ell^0 + \ell^1 + \ldots$, where each $\ell^n(x, \xi)$ is homogeneous of degree $-n$ in ξ and chosen to suppress the $O(|\xi|^{-j})$ contribution in $RL - I$ as well as in $LR - I$. This construction implies that L is pseudo-differential of order zero (componentwise). All of this is routine and detailed in [17][page 117].

In the sequel, S, S_1 and S_2 will denote a generic smoothing operator whose value may change from line to line. The composition of a pseudo-differential operator and a smoothing operator is obviously still smoothing. Set $f = Lu$ and let $A = a(x, D)$, so that $\partial_t u = -iAu$. On the one hand, $u = Rf - Su$ and

$$\partial_t u = R\partial_t f - S\partial_t u = R\partial_t f - SAu. \tag{4.9}$$

On the other hand, Lemma 4.1 gives

$$Au = ARf - ASu = R\Lambda f + S_1 f + S_2 u = R\Lambda f + Su \tag{4.10}$$

Comparing (4.9) and (4.10), and applying L gives

$$\partial_t f = -i\Lambda f + Su. \tag{4.11}$$

This can be solved by Duhamel's formula,

$$f(t) = e^{-it\Lambda} f(0) + \int_0^t e^{-i(t-\tau)\Lambda} Su(\tau) \, d\tau. \tag{4.12}$$

We now argue that the integral term is, indeed, a smoothing operator applied to the initial value u_0.

- First, the evolution operator $E(t) = e^{-itA}$ has a kernel $K(t, x, y)$ supported inside a neighborhood of the diagonal $y = x$ and for each $s \geq 0$, is well-known to map $H^s(\mathbb{R}^2, \mathbb{C}^m)$ boundedly onto itself [19]. Therefore, $SE(\tau)$ maps H^{-s} to H^s boundedly for every $s > 0$ and has a well-localized kernel in the sense of (4.4). This implies that $SE(\tau)$ is a smoothing operator.

- Second, section 4.3 shows that $e^{-it\Lambda}$ is, for small t, a FIO of type $(1, 0)$ and order zero, modulo a smoothing remainder. The composition of a FIO and a smoothing operator is smoothing. For larger t, think about $e^{-it\Lambda}$ as the product $(e^{-i\frac{t}{n}\Lambda})^n$ for appropriately large n.

- And third, the integral extends over a finite interval $[0, t]$ and may be thought as an average of smoothing operator—hence smoothing.

In short, $f(t) = e^{-it\Lambda} f(0) + Su_0$. Applying R on both sides of (4.11) finally gives

$$u = Re^{-it\Lambda} Lu_0 + S_1 u_0 + S_2 u = (Re^{-it\Lambda} L + S)u_0$$

which is what we set to establish.

It remains to see that the evolution operator $e^{-it\Lambda}$ for the polarized components has the same block-diagonal structure as Λ itself. This is gleaned from equation (4.11): evolution equations for two components f_{ν_1}, f_{ν_2} (corresponding to distinct eigenvalues) are completely decoupled. $\qquad \square$

4.3 The Fourier integral operator parametrix

Lemma 4.2 explained how to turn the evolution operator $E(t)$ into the block-diagonal representation $e^{-it\Lambda}$. In this section, we describe how each of these blocks can be approximated by a Fourier integral operator. The ideas here are standard and our exposition is essentially taken from [15] and [29]. The original construction is due to Lax [19].

Let us first assume that all eigenvalues of the principal symbol $a^0(x, \xi)$ are simple. This is the situation where the matrix of operators Λ (Lemma 4.1) is diagonal with elements Λ_ν. Put $E_\nu(t) = e^{-it\Lambda_\nu}$, the (scalar) evolution operator relative to the νth polarized mode. We seek a parametrix $F_\nu(t)$ such that $S_\nu(t) = E_\nu(t) - F_\nu(t)$ is smoothing of order $-\infty$.

Formally,

$$f(t, x) = \int E_\nu(t)(e^{ix\cdot\xi})\, \widehat{f_0}(\xi)\, d\xi.$$

Our objective is to build a high-frequency asymptotic expansion for $E_\nu(t)(e^{ix\cdot\xi})$ of the form

$$e^{i\Phi_\nu(t,x,\xi)}\sigma_\nu(t, x, \xi), \tag{4.13}$$

where $\sigma_\nu \sim \sigma_\nu^0 + \sigma_\nu^1 + \ldots$ with σ_ν^n homogeneous of degree $-n$ in ξ, and Φ homogeneous of degree one in ξ.

As is classical in asymptotic analysis, we proceed by applying $M_\nu = \partial_t + i\Lambda_\nu$ to the expansion (4.13) and successively equate all the coefficients of the negative powers of $|\xi|$ to zero, hence mimicking the relation $M_\nu E_\nu(t)(e^{ix\cdot\xi}) = 0$ which holds by definition. For obvious reasons, we also impose that (4.13) evaluated at $t = 0$ be $e^{ix\cdot\xi}$. Note that, in accordance to Lemma 4.1, Λ_ν is taken as a polyhomogeneous expansion $\sum_{j\geq 0} \lambda_\nu^j(x, D)$, where each symbol $\lambda_\nu^j(x, \xi)$ is homogeneous of degree $-j + 1$ in ξ.

After elementary manipulations, one finds that the phases must satisfy the standard Hamilton-Jacobi equations

$$\frac{\partial \Phi_\nu}{\partial t} + \lambda_\nu^0(x, \nabla_x \Phi_\nu) = 0, \tag{4.14}$$

with $\Phi_\nu(0, x, \xi) = x\cdot\xi$. The amplitudes σ_ν^n are successively determined as the solutions of transport equations along each Hamiltonian vector field,

$$\frac{\partial \sigma_\nu^n}{\partial t} + \nabla_\xi \lambda_\nu^0(x, \nabla_x \Phi_\nu) \cdot \nabla_x \sigma_\nu^n = P_n(\sigma_\nu^0, \ldots, \sigma_\nu^n), \tag{4.15}$$

where P_n is a known differential operator applied to $\sigma_\nu^0, \ldots, \sigma_\nu^n$.

In the case where some eigenvalue λ_ν^0 has multiplicity $p > 1$, the construction of a FIO parametrix goes the same way, except that Λ_ν denotes the p-by-p block corresponding to λ_ν^0 in the matrix Λ of Lemma 4.1. Also each σ_ν^n is now a p-by-p matrix of amplitudes.

It is important to notice that Φ_ν may be defined only for small times, because it would become multi-valued when rays originating from the same point x_0 cross again later. This typically happens at cusp points, when caustics start developing. We refer the interested reader to [15, 34].

We skipped a lot of justifications in the above exposition, in particular on convergence issues, but these technicalities are standard and detailed in some very good monographs. The following result summarizes all that we shall need.

Lemma 4.3. *Define t^* as half the infimum time for which a solution to (4.14) ceases to exist, uniformly in ν and ξ. In the setting of Lemma 4.1, denote by Λ_ν a block of Λ and $E_\nu(t) = e^{-it\Lambda_\nu}$. Then for every $0 < t \leq t^*$, there exists a parametrix $F_\nu(t)$ for the evolution problem $\partial_t f + i\Lambda_\nu f = 0$ which takes the form of a Fourier integral operator,*

$$F_\nu(t)f_0(x) = \int e^{i\Phi(t,x,\xi)}\sigma_\nu(t, x, \xi)\hat{f_0}(\xi)\, d\xi,$$

For each $t \leq t^*$, the phase function Φ_ν is positive-homogeneous of degree one in ξ and smooth in x and ξ; the amplitude σ_ν is a symbol of type $(1,0)$ and order zero. The remainder $S_\nu(t) = E_\nu(t) - F_\nu(t)$ is a smoothing operator of order $-\infty$.

Proof. The proof is for the most part presented in [29][Pages 120 and below]. See also [14, 15, 33]. $\qquad\square$

4.4 Sparsity of smoothing terms

The specialist will immediately recognize that a smoothing operator of order $-\infty$ is very sparse in a curvelet frame. This is the content of the following lemma.

Lemma 4.4. *The curvelet entries of a smoothing operator S obey the following estimate: for each $N > 0$, there is a constant C_N such that*

$$|\langle \varphi_\mu, S\varphi_{\mu'} \rangle| \leq C_N \cdot 2^{-|j+j'|N}(1 + |x_\mu - x_{\mu'}|)^{-N}. \tag{4.16}$$

Note that (4.16) is a stronger estimate than that of Theorem 1.1. Indeed, our lemma implies that

$$|\langle \varphi_\mu, S\varphi_{\mu'} \rangle| \leq C_N \cdot \omega(\mu, h_\nu(t, \mu'))^{-N}$$

which is valid for each $N > 0$ and regardless of the value of ν.

Proof. We know that S maps H^{-s} to H^s for arbitrary large s, so does its adjoint S^*. As a result,

$$
\begin{aligned}
|\langle \varphi_\mu, S\varphi_{\mu'} \rangle| &\leq |\langle S^*\varphi_\mu, \varphi_{\mu'} \rangle|^{1/2}|\langle \varphi_\mu, S\varphi_{\mu'} \rangle|^{1/2} \\
&\leq \|S^*\varphi_\mu\|_{H^s}^{1/2}\|\varphi_{\mu'}\|_{H^{-s}}^{1/2}\|\varphi_\mu\|_{H^{-s}}^{1/2}\|S\varphi_{\mu'}\|_{H^s}^{1/2} \\
&\leq C \cdot \|\varphi_{\mu'}\|_{H^{-s}}\|\varphi_\mu\|_{H^{-s}} \leq C \cdot 2^{-(j+j')s}.
\end{aligned}
$$

Next, recall that curvelets have an essential spatial support of size at most $O(1) \times O(1)$. (Coarse scale curvelets have support size about $O(1) \times O(1)$ and the size decreases at increasingly finer scales.) The action of S is local on this range of distances, so that

$$|\langle \varphi_\mu, S\varphi_{\mu'} \rangle| \leq C_N \cdot (1 + |x_\mu - x_{\mu'}|)^{-N}.$$

for arbitrary large $N > 0$. These two bounds can be combined to conclude that the matrix elements of S are negligible in the sense defined above. $\qquad\square$

4.5 Proof of Theorem 1.1

Let us first show how the first assertion on the near-exponential decay of the curvelet matrix elements follows immediately from the second one, equation (1.16). Let a be either a row or a column of the curvelet matrix and let $|a|_{(n)}$ be the n-largest entry of the sequence $|a|$. We have

$$n^{1/p} \cdot |a|_{(n)} \leq \|a\|_{\ell_p}^{1/p}$$

and, therefore, it is sufficient to prove that the matrix E has rows and columns bounded in ℓ_p for every $p > 0$. Consider the columns. We need to establish

$$\sup_{\mu',\nu'} \sum_{\mu,\nu} |E(t; \mu, \nu; \mu', \nu')|^p \leq C_{t,p}, \tag{4.17}$$

for some constant $C_{t,p} > 0$ growing at most like $C_p e^{K_p t}$ for some $C_p, K_p > 0$.

The sum over ν and the sup over ν' do not come in the way since these subscripts take on a finite number of values. The fine decoupling between the m one-way components, crucial for equation (1.16), does not play any role here.

Let us now show that there exists N so that

$$\sum_\mu \omega(\mu, \mu')^{-Np} \leq C_{N,p},$$

uniformly in μ'. We can use the bound (A.2) with Np in place of N for the sum over k and ℓ. This gives

$$\sum_{\mu} \omega(\mu, \mu')^{-Np} \leq C_{N,p} \cdot \sum_{j \geq 0} 2^{-|j-j'|Np} \cdot 2^{2|j-j'|},$$

which is bounded by a constant depending on N and p provided again that $Np \geq 2$.

Hence we proved the property for the columns. The same holds for the rows because the same conclusion is true for the adjoint $E(t)^*$; indeed, the adjoint solves the backward initial-value problem for the adjoint equation $u_t = A^* u$, and A^* satisfies the same hyperbolicity conditions as A. We can therefore interchange the role of the two curvelets and obtain

$$\sup_{\mu',\nu'} \sum_{\mu,\nu} |E(t;\mu',\nu';\mu,\nu)|^p \leq C_{t,p}.$$

Note that the classical interpolation inequality shows that $E(t)$ is a bounded operator from ℓ_p to ℓ_p for every $0 < p \leq \infty$.

We now turn to (1.16). Let us assume first that all eigenvalues λ_ν^0 of the principal symbol a^0 are simple. According to Lemmas 4.2 and 4.3, each matrix element $E(t;\nu;\nu') = \mathbf{e}_\nu \cdot E(t)\mathbf{e}_{\nu'}$ of $E(t)$ can for fixed (possibly large) time $t > 0$ be written as

$$E(t;\nu;\nu') = \sum_{\nu''=1}^{m} R_{\nu,\nu''}(e^{-i\frac{t}{n}\Lambda_{\nu''}})^n L_{\nu'',\nu'} + S_{\nu,\nu'}(t). \tag{4.18}$$

We have taken n large enough—proportional to t—so that $e^{-i\frac{t}{n}\Lambda_\nu}$ is a Fourier integral operator (mod smoothing) for every ν. Each $R_{\nu,\nu'}$ and $L_{\nu,\nu'}$ is pseudo-differential of order zero and $S_{\nu,\nu'}(t)$ is smoothing.

Thanks to Lemma 4.4, we only need to prove the claim for the first term of (4.18) which follows from Theorem 5.1 in section 5 about the sparsity of FIO's in a curvelet tight-frame.

As is well-known, the ray dynamics is equivalently expressed in terms of Hamiltonian flows

$$\begin{cases} \dot{x}(t) & = \nabla_\xi \lambda^0(x(t), \xi(t)), & x(0) = x_0, \\ \dot{\xi}(t) & = -\nabla_x \lambda^0(x(t), \xi(t)), & \xi(0) = \xi_0, \end{cases} \tag{4.19}$$

or in terms of canonical transformations generated by the phase functions Φ_ν,

$$\begin{cases} x_0 & = \nabla_\xi \Phi(t, x(t), \xi_0), \\ \xi(t) & = \nabla_x \Phi(t, x(t), \xi_0), \end{cases} \tag{4.20}$$

provided $\Phi(t, x, \xi)$ satisfies the Hamilton-Jacobi equation $\frac{\partial \Phi}{\partial t} + \lambda^0(x, \nabla_x \Phi) = 0$ with initial condition $\Phi(0, x, \xi_0) = x \cdot \xi_0$. We obviously need this property to ensure that the geometry of FIO's is the same as that of hyperbolic equations.

Pseudo-differential operators are a special instance of Fourier integral operators so the theorem equally applies to them. For $E(t;\mu,\nu;\mu',\nu') = \langle \varphi_\mu, E(t;\nu;\nu')\varphi'_\mu \rangle$ we get

$$|E(t;\mu,\nu;\mu',\nu')| \leq C_N \sum_{\nu''=1}^{m} \sum_{\mu_0} \cdots \sum_{\mu_n} \omega(\mu, \mu_0)^{-N} \omega(\mu_0, \mu_{1\nu''}(\tfrac{t}{n}))^{-N} \cdots$$
$$\omega(\mu_{n-1}, \mu_{n\nu''}(\tfrac{t}{n}))^{-N} \omega(\mu_n, \mu')^{-N},$$

for all $N > 0$. Inequality (4.2) then follows from repeated applications of properties 3 and 4 of the distance ω, see proposition 2.2. The power growth in t of the overall multiplicative constant comes from the number of intermediate sums over $\mu_0, \ldots, \mu_n$. There are $n + 1 \sim t$ such sums and they each introduce the same multiplicative constant C_N.

The reasoning is the same when at least some eigenvalues λ_ν^0 are degenerate. The subscript ν'' now denotes the flows i.e., the eigenvalues $\lambda_{\nu''}^0$ *not* counting their multiplicity. Each $R_{\nu,\nu''}$ is a row vector, $e^{-i\frac{t}{n}\Lambda_{\nu''}}$ a matrix and $L_{\nu'',\nu'}$ a column vector. The FIO parametrix for $e^{-i\frac{t}{n}\Lambda_{\nu''}}$ was constructed in such a way that only one flow $h_{\nu''}$ appears in the majoration of its curvelet elements (componentwise). There is no intermediate sum over $\nu_0, \ldots, \nu_n$ and this is the whole point of decoupling the polarized components *before* constructing the FIO parametrix.

4.6   Relation to hyper-curvelets

In section 1.4 we introduced hyper-curvelets as 'polarized' curvelets which would not split into m molecules along the m different flows. In light of section 4.2, it is interesting to reformulate our main result (4.2) in terms of hyper-curvelets. We recall that

$$\boldsymbol{\varphi}_{\mu\nu}^{(0)}(x) = r_\nu^0(x,D)\varphi_\mu(x) = \frac{1}{(2\pi)^2} \int e^{ix\cdot\xi} r_\nu^0(x,\xi)\hat{\varphi}_\mu(\xi)\, d\xi.$$

Corollary 4.5. *Define* $E^{(0)}(t;\mu,\nu;\mu',\nu') = \langle \boldsymbol{\varphi}_{\mu\nu}^{(0)}, E(t)\boldsymbol{\varphi}_{\mu'\nu'}^{(0)}\rangle$. *Then under the same assumptions as those of Theorem 1.1 we have for all* $N > 0$

$$|E^{(0)}(t;\mu,\nu;\mu',\nu')| \le C_{tN} \cdot [\, \omega(\mu,\mu'_{\nu'}(t))^{-N} + 2^{-j'} \sum_{\nu''\neq\nu'} \omega(\mu,\mu'_{\nu''}(t))^{-N}\,]. \tag{4.21}$$

The main contribution to the right-hand side is due to the ν'th flow. All other flows are weighted by the small factor $2^{-j'}$ (which is about equal to $|\xi|^{-1}$ on the support of $\hat{\varphi}_{\mu'}$). In other words, there might be some 'cross talk' between the various components corresponding to the different flows but it is at most smoothing of order -1, hence small at small scales.

Proof. Equation (4.21) follows from Theorem 1.1 and the fact that the adjoint of the matrix operator R^0 whose columns are the $R_\nu^0 = r_\nu^0(x,D)$ is an approximate left inverse for R^0—up to an error smoothing of order -1. Indeed, by the standard rules for composition and computation of the adjoint of pseudo-differential operators,

$$\begin{aligned}(R_\nu^0)^* R_{\nu'}^0 &= ((r_\nu^0)^* \,\sharp\, r_{\nu'}^0)(x,D) \\ &= ((r_\nu^0)^* r_{\nu'}^0)(x,D) + \text{ order } -1 \\ &= \delta_{\nu\nu'} I + \text{ order } -1.\end{aligned}$$

We have used the fact that the dispersion matrix $a^0(x,\xi)$ is assumed to be symmetric, hence admits an orthobasis of eigenvectors $r_\nu^0(x,\xi)$. We then conclude from Theorem 5.1 applied to pseudo-differential operators of order -1. $\qquad\square$

Alternatively, we could have defined hyper-curvelets as

$$\boldsymbol{\varphi}_{\mu\nu}^{(\infty)} = r_\nu(x,D)\varphi_\mu(x) = \frac{1}{(2\pi)^2} \int e^{ix\cdot\xi} r_\nu(x,\xi)\hat{\varphi}_\mu(\xi)\, d\xi.$$

This would have given the same result.[2] The reason why we did not use hyper-curvelets in the preceding sections is that they do not necessarily constitute a suitable practical basis to decompose wavefields onto. We do not even know if they always constitute a frame. Digital implementation would also seem less obvious.

---

[2]We can only conjecture that the decoupling should be better if we use the improved $\boldsymbol{\varphi}_{\mu\nu}^{(\infty)}$.

5 Representation of FIOs

The purpose of this section is to show that Fourier integral operators admit a sparse and well-organized structure in a curvelet frame. The main result, Theorem 5.1, is a key step in completing the discussion of the previous section. (Observe that by construction, the FIO's encountered in the previous section satisfy all the assumptions stated in section 1.6 right below (1.18).) As in the previous section, we will restrict the discussion to $x \in \mathbb{R}^2$ which is no loss of generality, see section 6.

5.1 Main results

In the introduction section, we detailed a notion of Hamiltonian correspondence for hyperbolic equations. This correspondence also exists for FIO's and is 'encoded' in the phase function Φ of the FIO. It is called the canonical transformation associated to Φ, and is defined as the mapping $(x, \xi) \to (y, \eta)$ of phase-space

$$x = \nabla_\xi \Phi(y, \xi), \qquad \eta = \nabla_x \Phi(y, \xi). \tag{5.1}$$

As suggested in section 4.5, this formulation is equivalent to that involving trajectories along the bicharacteristic flow as in equation (1.6), provided the phase function solves an appropriate Hamilton-Jacobi equation.

This canonical transformation induces a mapping of curvelet subscripts, denoted by $\mu' = h(\mu)$. It is defined via the closest point $(x_{\mu'}, \xi_{\mu'})$, on the curvelet lattice, to the image of (x_μ, ξ_μ) by the canonical transformation. We can already remark that mistaking a point $(x_{\mu'}, \xi_{\mu'})$ for one of its neighbors will not compromise the following result, only increase the value of the constant C_N in front of the estimate.

The main result for this section reads as follows.

Theorem 5.1. *Let T be a Fourier integral operator of order m acting on functions of $\mathbb{R}^2$, with the assumptions stated above, and $T(\mu; \mu')$ denote its matrix elements in the complex curvelet tight frame. Then with h the curvelet index mapping and ω the distance defined in (2.12), the elements $T(\mu; \mu')$ obey for each $N > 0$*

$$|T(\mu; \mu')| \leq C_N \cdot 2^{mj'} \omega(\mu, h(\mu'))^{-N},$$

for some $C_N > 0$. Moreover, for every $0 < p \leq \infty$, $(T(\mu, \mu'))$ is bounded from ℓ^p to ℓ^p.

The interpretation of Theorem 5.1 is in strong analogy with that of Theorem 1.1. Namely, a FIO has the property of transporting and warping a curvelet into another curvelet-like molecule. (Again, the choice of using complex-valued curvelets is not essential, as a real curvelet would be mapped onto two molecules.)

The proof of Theorem 5.1 relies on the factorization of T on the space-frequency support of φ_μ as a nice pseudo-local operator $T_{1,\mu}$ followed by a smooth change of variables, or warping $T_{2,\mu}$. This decomposition goes as follows.

Let φ_μ be a fixed curvelet centered around the lattice point (x_μ, ξ_μ) in phase-space. The phase of our FIO can be decomposed as

$$\Phi(x, \xi) = \Phi_\xi(x, \xi_\mu) \cdot \xi + \delta(x, \xi), \qquad \phi_\mu(x) = \Phi_\xi(x, \xi_\mu). \tag{5.2}$$

In effect, the above decomposition 'linearizes' the frequency variable and is classical, see [25, 30]. With these notations, we may rewrite the action of T on a curvelet φ_μ as

$$(T\varphi_\mu)(x) = \int e^{i\phi_\mu(x) \cdot \xi} e^{i\delta(x, \xi)} \sigma(x, \xi) \hat{\varphi}_\mu(\xi) \, d\xi. \tag{5.3}$$

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

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

where

$$(T_{1,\mu} f)(x) = \int e^{ix \cdot \xi} b_\mu(x, \xi) \hat{f}(\xi) \, d\xi, \quad (T_{2,\mu} f)(x) = f(\phi_\mu(x)), \tag{5.4}$$

with $b_\mu(x, \xi) = e^{i\delta(\phi_\mu^{-1}(x), \xi)}\sigma(\phi_\mu^{-1}(x), \xi))$. 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 5.2. *Let $(\varphi_\mu)_\mu$ be a tight frame of curvelets compactly supported in frequency. For each μ, $T_{1,\mu}$ maps φ_μ into a curvelet molecule m_μ with arbitrary regularity R, uniformly over μ in the sense that the constants in estimates (2.5) and (2.6) do not depend on μ.*

As we shall see, the proof of Theorem 5.2, presented in section 5.2, relies on the property of compact support in frequency of the φ_μ. In contrast the corresponding result for the operators $T_{2,\mu}$ which we present next, is extraordinarily simplified if one uses curvelets compactly supported in space. Although well localized in space, the tight frame introduced in section 2 does not meet this requirement. In order to circumvent this technical difficulty, we introduce compactly supported *curvelet atoms* in section 5.3. They are built on the model of atomic decompositions, standard in approximation theory [18].

Theorem 5.3. *Let $(\rho_\mu)_\mu$ be a family of complex-valued curvelet atoms, compactly supported in space, with regularity R. Denote by h the canonical index correspondence associated to Φ, as defined above. For each μ, $T_{2,\mu}$ maps ρ_μ into a molecule $m_{h(\mu)}$ of the same regularity R, uniformly over μ.*

The latter theorem says that the 'warped' atom $\rho_\mu \circ \phi_\mu$ is still an atom, only its scale, orientation, and location may have been changed. That a smooth warping preserves the sparsity of curvelet expansions is a result of independent interest.

The remaining 3 sections are devoted to the proofs of Theorems 5.2, 5.3 and 5.1. The dependence of ϕ_μ upon μ is not essential in proving Theorems 5.2, 5.3 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 the next sections we will drop the explicit dependence on μ and work with a generic warping ϕ.

5.2   Proof of Theorem 5.2

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

Set $m_\mu = T_1\varphi_\mu$. We first show that m_μ obeys the smoothness and spatial localization estimate of a molecule (2.5). With the same notations as before, recall that m_μ is given by

$$m_\mu(x) = \int e^{ix\cdot\xi} b_\mu(x, \xi)\hat{\varphi}_\mu(\xi)\, d\xi, \quad b_\mu(x, \xi) = e^{i\delta(\phi^{-1}(x), \xi)}\sigma(\phi^{-1}(x), \xi). \tag{5.5}$$

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 (5.5) 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 $N \geq 0$,

$$|L_\xi^N[b_\mu(x, \xi)\hat{\varphi}_\mu(\xi)]| \leq C \cdot 2^{-3j/4}. \tag{5.6}$$

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

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

Since $|L_\xi^N[b_\mu(x,\xi)\hat\varphi_\mu(\xi)]| \le C \cdot 2^{-3j/4}$ and is supported on a dyadic rectangle R_μ, 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

$$\partial_x^\alpha(e^{ix\cdot\xi} b_\mu(x,\xi)) = \sum_{\beta+\varphi\le\alpha} \partial^\beta(e^{ix\cdot\xi}) \partial^\varphi(b_\mu(x,\xi))$$

$$= \sum_{\beta+\varphi\le\alpha} \partial^\varphi(b_\mu(x,\xi))\, \xi^\beta e^{ix\cdot\xi}.$$

Therefore, the partial derivatives of m_μ are given by

$$(\partial_x^\alpha m_\mu)(x) = \sum_{\beta+\varphi\le\alpha} I_{\beta,\varphi}(x), \tag{5.7}$$

where

$$I_{\beta,\varphi}(x) = \int e^{ix\cdot\xi} \partial_x^\varphi(b_\mu(x,\xi)) \xi^\beta \hat\varphi_\mu(\xi)\, d\xi. \tag{5.8}$$

First, observe that on the support of $\hat\varphi_\mu$, $|\xi|^\beta$ obeys $|\xi|^\beta \le C \cdot 2^{j\beta_1} \cdot 2^{j\beta_2/2}$. Second, the term $\partial_x^\varphi b(x,\xi)$ 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 $N \ge 0$,

$$|L_\xi^N[\partial_x^\varphi b_\mu(x,\xi)\xi^\beta\hat\varphi_\mu(\xi)]| \le C \cdot 2^{-3j/4} \cdot 2^{j\beta_1} \cdot 2^{j\beta_2/2}. \tag{5.9}$$

Hence, the same argument as before gives

$$|I_{\beta,\varphi}(x)| \le C \cdot \frac{2^{3j/4} \cdot 2^{j\beta_1} \cdot 2^{j\beta_2/2}}{\left(1 + |2^j x_1|^2 + |2^{j/2} x_2|^2\right)^N}.$$

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

$$|(\partial_x^\alpha m_\mu)(x)| \le C \cdot \frac{2^{3j/4} \cdot 2^{j\alpha_1} \cdot 2^{j\alpha_2/2}}{\left(1 + |2^j x_1|^2 + |2^{j/2} x_2|^2\right)^N}.$$

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 (2.6)–(2.9).

Formally, the Fourier transform of m_μ is given by

$$\hat m_\mu(\xi) = \iint e^{ix\cdot(\eta-\xi)} b_\mu(x,\eta)\hat\varphi_\mu(\eta)\, dx d\eta. \tag{5.10}$$

We should point out that because the amplitude b is not of compact support in x, the sense in which (5.10) 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 $D_1 = -i\frac{\partial}{\partial x_1}$. To develop bounds on $|\hat m_\mu(\xi)|$, observe that

$$D_1^N e^{ix\cdot\eta} = (\eta_1)^N.$$

An integration by parts then gives

$$\hat{m}_\mu(\xi) = \iint e^{ix\cdot\eta} D_1^N \left(e^{-ix\cdot\xi} b_\mu(x,\xi)\right) \eta_1^{-N} \hat{\varphi}_\mu(\eta)\, dx d\eta.$$

Hence,

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

where

$$F_m(x) = \int e^{ix\cdot\eta} (\partial_{x_1}^{n-m} b(x,\xi))\, \eta_1^{-n} \hat{\varphi}_\mu(\eta)\, d\eta.$$

Note that F_m is exactly of the same form as (5.8)—but with η_1^{-n} instead of η^β—and therefore, the exact same argument as before gives

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

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 5.2.

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.

5.3   Atomic decompositions

As we will see later, to prove our main result and especially Theorem 5.3, 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 a rotation

$$f_{a,\theta}(x) = a^{-3/4} f\left(D_a R_\theta x\right), \quad D_a = \begin{pmatrix} 1/a & 0 \\ 0 & 1/\sqrt{a} \end{pmatrix},$$

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

$$q(\xi) \int_{a\le 1} \hat{\tilde{\psi}}_{a,\theta}(\xi) \hat{\psi}_{a,\theta}(\xi)\, a\, dad\theta = r(\xi); \tag{5.11}$$

here r is a smooth cut-off function obeying

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

and q is a standard Fourier multiplier of order zero; that is, for each multiindex α, there exists a constant C_α such that

$$|\partial_\xi^\alpha q(\xi)| \le C_\alpha (1 + |\xi|)^{-|\alpha|}.$$

This formula is useful because it allows us to express any object whose Fourier transform vanishes on $\{|\xi| \le 2\}$ 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 $\tilde{\psi}(x) = \psi(-x)$ and the function ψ of the form

$$\psi(x_1, x_2) = \psi^D(x_1)\,\varphi(x_2), \tag{5.12}$$

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

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

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

$$\int \psi^D(x_1) x_1^k \, dx_1 = 0, \quad k = 0, 1, \dots, D. \tag{5.13}$$

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

$$\psi_{a,\theta,b}(x) := \psi_{a,\theta}(x - b) = a^{-3/4}\psi\left(D_a R_\theta(x - b)\right); \tag{5.14}$$

and given an object f, define coefficients by

$$CCT(f)(a, b, \theta) = \int \overline{\psi_{a,\theta,b}(x)} f(x) dx. \tag{5.15}$$

Now, suppose for instance that $\hat{f}$ vanishes over $|\xi| \le 2$, then (5.11) gives the exact reconstruction formula

$$f(x) = \int_{a \ge 1} CCT(q(D)f))(a, b, \theta)\psi_{a,\theta,b}(x)\mu(dad\theta db), \tag{5.16}$$

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 $j \ge 0$, $\ell = 0, 1, \dots 2^{\lfloor j/2 \rfloor} - 1$ and $k = (k_1, k_2) \in \mathbb{Z}^2$, the cell Q_μ is the collections of triples (a, θ, b) for which

$$2^{-(j+1)} \le a < 2^{-j}, \quad |\theta - \theta_\mu| \le \frac{\pi}{2} 2^{-\lfloor j/2 \rfloor}$$

and

$$D_{2^{-j}} R_{\theta_\mu} b \in [k_1, k_1 + 1) \times [k_2, k_2 + 1).$$

Note that $\int_{Q_\mu} d\mu = 3\pi/2$ for j even, and 3π for j odd. We may then break the integral (5.16) into a sum of terms arising from different cells, namely,

$$f(x) = \sum_\mu \alpha_\mu \rho_\mu(x) \tag{5.17}$$

where

$$\alpha_\mu = \|CCT(q(D)f)\|_{L_2(Q_\mu)}, \tag{5.18}$$

$$\rho_\mu(x) = \frac{1}{\alpha_\mu} \int_{Q_\mu} CCT(q(D)f))(a, b, \theta)\psi_{a,\theta,b}(x)\, d\mu.$$

Of course, the decomposition (5.17) greatly resembles the tight frame expansion, compare (1.5). In particular, the atoms ρ_μ are curvelet-like in the sense that they share all the properties of the tight frame $(\varphi_\mu)_\mu$ – 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 5.4. *Rewrite the atoms ρ_μ as $\rho_\mu(x) = 2^{3j/4}a^{(\mu)}\left(D_{2^{-j}}R_{\theta_\mu}x - k\right)$. In other words, ρ_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;*

$$Supp\ a^{(\mu)} \subset cQ. \tag{5.19}$$

- *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*

$$\int a^{(\mu)}(x_1, x_2)x_1^k\,dx_1 \le C_m \cdot 2^{-j(m+1)}. \tag{5.20}$$

- *Regularity; for every multiindex α*

$$|\partial_x^\alpha a^{(\mu)}(x)| \le C_\alpha. \tag{5.21}$$

In (5.20) and (5.21), the constants may be chosen independently of μ and f.

Proof. See appendix A.2 □

Needless to say that curvelet atoms are molecules with spatial compact support, compare lemma 5.4 with the definition of a molecule. Finally, observe (and this is important) that 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 2.3

$$|\alpha_{\mu\mu'}| \le C_N \cdot \omega(\mu, \mu')^{-N}, \tag{5.22}$$

and in particular, for each $p > 0$,

$$\sup_\mu \sum_{\mu'} |\alpha_{\mu\mu'}|^p < C_p.$$

This is briefly justified in appendix A.2.

5.4   Proof of Theorem 5.3

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 5.3, 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 5.3, we suppose our curvelet atom is of the form

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

where $a^{(\mu)}$ obeys the conditions of Lemma 5.4. (Here, the location x_μ may be formally defined by $x_\mu = (D_{2^{-j}}R_{\theta_\mu})^{-1}k_\delta$.) Define y_μ and A_μ by

$$y_\mu = \phi^{-1}(x_\mu), \quad \text{and} \quad A_\mu = (\nabla\phi)(y_\mu) \tag{5.23}$$

so that

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

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

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

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 $\theta_\mu = 0$ and $y_\mu = 0$ (statements for arbitrary orientations and locations are obtained in an obvious fashion) so that

$$\rho_\mu(\phi(y)) = 2^{3j/4} a^{(\mu)}\left(D_{2^{-j}}(y + g(y))\right) = 2^{3j/4} b^{(\mu)}(D_{2^{-j}} y), \tag{5.24}$$

with

$$b^{(\mu)}(y) = a^{(\mu)}\left(y + D_{2^{-j}} g(D_{2^j} y)\right).$$

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

$$|\partial^\alpha b^{(\mu)}(y)| \leq C_\alpha, \quad |\alpha| \leq R. \tag{5.25}$$

Over the support of $\rho_\mu \circ \phi$, $g = (g_1, g_2)$ deviates little from zero and for each $k = 1, 2$, g_k obeys

$$|g_k(y)| \leq C \cdot 2^{-j}, \quad |\partial^\alpha g_k(y)| \leq C \cdot 2^{-j/2}, \ |\alpha| = 1.$$

Similarly, for each α, $|\alpha| > 1$,

$$|\partial^\alpha g_k(y)| \leq C_\alpha. \tag{5.26}$$

These estimates hold uniformly over μ. It follows that for $|y_1|, |y_2| \leq C$ and each α, the perturbation g obeys

$$2^j \cdot |\partial^\alpha g_1(2^{-j} y_1, 2^{-j/2} y_2)| \leq C_\alpha, \quad 2^{j/2} \cdot |\partial^\alpha g_2(2^{-j} y_1, 2^{-j/2} y_2)| \leq C_\alpha. \tag{5.27}$$

The bound (5.25) is then a simple consequence of (5.27) 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.

$$\widehat{\rho_\mu \circ \phi}(\xi) = \int e^{-ix\cdot\xi} \rho_\mu(\phi(x))\, dx$$

$$= \int e^{-i\phi^{-1}(x)\cdot\xi} \rho_\mu(x) \frac{dx}{|\det \nabla\phi|(\phi^{-1}(x))}.$$

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 $N \leq R$, we have available the following upper bound on the partial derivative of S_ξ

$$|\partial_1^N S_\xi(x)| \leq C_N \cdot (1 + |\xi|)^N.$$

Classical arguments give

$$\widehat{\rho_\mu \circ \phi}(\xi) = \sum_{k=0}^{n-1} \int \frac{\partial_1^k S_\xi(0, x_2)}{k!}\, dx_2 \int \rho_\mu(x_1, x_2) x_1^k\, dx_1 dx_2 + E, \tag{5.28}$$

where E is a remainder term obeying

$$|E| \leq C_n \cdot 2^{-3j/4} \cdot 2^{-jn} \cdot \sup |\partial_1^n S_\xi(x)| \leq C_n \cdot 2^{-3j/4} \cdot 2^{-jn}(1 + |\xi|^n). \tag{5.29}$$

The near-vanishing moment property gives that each term in the right-hand side of (5.28) obeys the estimate in (5.29). This proves that the Fourier transform of $\rho_\mu \circ \phi$ obeys

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

as required.

We now discuss the case where the matrix A_μ is not the identity. In this case, (5.24) becomes

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

with

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

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.

5.5 Proof of Theorem 5.1

Let φ_{μ_0} be a fixed curvelet and decompose T as $T_{2,\mu_0} \circ T_{1,\mu_0}$. First, Theorem 5.2 proved that $T_{1,\mu_0}\varphi_{\mu_0}$ is a curvelet molecule m_{μ_0} which we will express as a superposition of curvelet atoms ρ_{μ_1}

$$T_{1,\mu_0}\varphi_{\mu_0} = m_{\mu_0} = \sum_{\mu_1} \beta_0(\mu_1, \mu_0)\rho_{\mu_1}.$$

Second, for each μ_1, Theorem 5.3 shows that $T_{2,\mu_1}\rho_{\mu_1}$ is a molecule $m_{h(\mu_1)}$ at the location $h(\mu_1)$. We are not exactly in that setting since in $T_{2,\mu_0}\rho_{\mu_1}$, the subscripts do not in general match. This does not pose any difficulty since Theorem 5.3 can be understood as a statement concerning general warpings ϕ. We can define the map h_{μ_0} as induced by the transformation $(x, \xi) \to (y, \eta)$ given by

$$x = \nabla_x \Phi(y, \xi_{\mu_0}), \qquad \eta = \nabla_\xi(y, \xi_{\mu_0})$$

(compare this with equation (5.1)). Then, according to Theorem 5.3, $T_{2,\mu_0}\rho_{\mu_1}$ is a molecule $m_{h_{\mu_0}(\mu_1)}$ at the location $h_{\mu_0}(\mu_1)$. So

$$\langle \varphi_{\mu_2}, T_{2,\mu_0}\rho_{\mu_1} \rangle = \beta_1(\mu_2, h_{\mu_0}(\mu_1)).$$

Hence,

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

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

- By Theorem 5.2 and Lemma 2.3, $|\beta_0(\mu_1, \mu_0)| \le C_n \cdot \omega(\mu_1, \mu_0)^{-N}$ for arbitrarily large $N > 0$, provided that the selected atoms are regular enough.

- By Theorem 5.3 and Lemma 2.3, $|\beta_1(\mu_2, h_{\mu_0}(\mu_1))| \le C_N \cdot \omega(\mu_2, h_{\mu_0}(\mu_1))^{-N}$ for arbitrarily large $N > 0$, provided that the selected atoms are regular enough.

Theorem 5.1 now follows from the observation that

$$\sum_{\mu_1} \omega(\mu_2, h_{\mu_0}(\mu_1))^{-N} \cdot \omega(\mu_1, \mu_0)^{-N} \le C_n \cdot \omega(\mu_2, h_{\mu_0}(\mu_0))^{-(N-1)}, \tag{5.30}$$

This is an immediate consequence of properties 3 and 4 of the pseudo-distance ω, see proposition 2.2.

Cases involving coarse scale elements are treated similarly and we omit the proof. The boundedness from ℓ^p to ℓ^p for every $p > 0$ follows from the same argument as in the proof of Theorem 1.1.

6   Discussion

All along we specialized our discussion to the special case where the dimension of the spatial variable is $n = 2$. It is clear that nothing in our arguments depends upon this specific assumption. Indeed, we could just as well construct tight-frames of curvelets in arbitrary dimensions by smoothly partitioning the frequency plane into dyadic coronae, which would then be angularly localized near regions of sidelength length 2^j in the radial direction and $2^{j/2}$ in all the other directions; in order to this, we would use smooth partitions of the unit sphere of $\mathbb{R}^n$ into spherical caps of radius about $2^{-j/2}$. All of our analysis would apply as is, and would prove versions of Theorem 1.1 in arbitrary dimensions.

Our main result assumes that the coefficients of the equation (1.1) be smooth. In many applications of interest, however, the coefficients may be smooth away from singular smooth surfaces. In geophysics for example, we typically have different layers with very different physical properties. A very important question would be to know how our analysis would adapt to this situation. In fact, it seems natural to believe that sparsity would continue to hold in this more general setting. Intuitively, the wave group would still be approximated by rigid motion along the Hamiltonian flow. Only, one would need to account for possible reflections/refractions. A curvelet hitting a singularity at a small angle of incidence would typically produce two curvelets, a reflected and a refracted curvelet. This is merely an intuition which one would need to justify by a careful analysis quantifying the behavior of a curvelet near the interface (here, the singular surface). We regard this type of question as an important extension to this work.

The estimates proven in this paper are motivated by efforts towards applications as sparser expansions theoretically lead to better and faster algorithms. Our goal is to transform our theoretical insights into effective algorithms, and derive fast accurate solvers to certain classes of partial differential equations for which our methods have a comparative advantage. In this direction, suppose that a discretized version $E_N(t)$ of the curvelet matrix $E(t)$ is known in the sense that it has been precomputed once for all. Assuming that discretized version inherits the sparsity from its continuous analog, then for each initial value problem, we would have available a fast algorithm for calculating the solution of the full wave equation. Indeed, Corollary 1.2 asserts that the truncated matrix $\text{trunc}(E_N(t))$ obtained by keeping only $O(\epsilon^{1/m})$ terms per row obeys

$$\|E_N(t) - \text{trunc}(E_N(t))\| \leq \epsilon.$$

Therefore, ignoring the cost of the digital curvelet transform, the total cost of an algorithm calculating the solution of a full wave equation to within accuracy ϵ would be linear in the number of unknowns (number of voxels) and scale at most like $C(\epsilon) \cdot N$. This is not the most practically relevant situation, however, as in general, one would not know $E_N(t)$. Work in progress shows that one can still design effective algorithms to 'build' the matrix $E_N(t)$, and for all practical purposes compute the solution of the full wave equation for a given accuracy ϵ in $O(N(\log N)^2)$.

Finally, we would like to conclude by pointing out that there is now considerable evidence that the curvelet transform is a very useful mathematical architecture; curvelets can do things that other classical systems simply cannot do. This paper proved that tight-frames of curvelets provide optimally sparse representations of large classes of linear systems of hyperbolic differential equations. But they also allow for optimally sparse representations of wavefront phenomena [7]. Further, they can also be useful in many different settings. For example, they have useful microlocal features which make them especially suitable for deployment in many inverse problems and especially limited-angle tomography [9]. In short, curvelets addresses a new range of problems, going beyond what traditional multiscale systems offer.

A   Appendix

A.1   Additional proofs for section 2

Proof of proposition 2.2. These four properties were already formulated in [28], although with a slightly weaker definition of pseudo-distance. Properties 1 and 2 are not proved in that reference, and property 3 is not extensively documented. We give the justification for these three results for completeness.
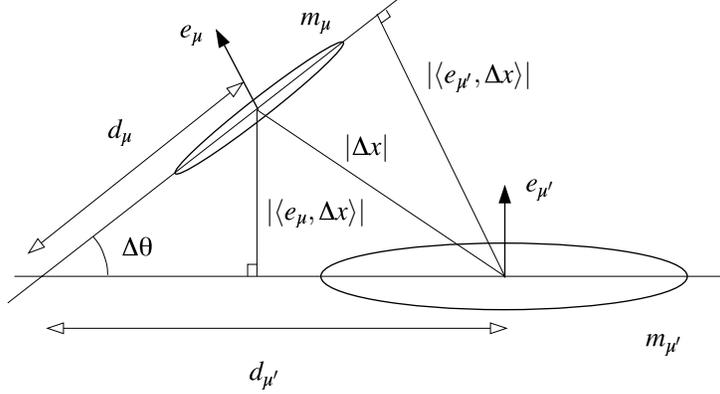
Figure 4: Relative position and orientation of two curvelet molecules in x-space. The ellipses indicate their essential support.

Claim (1). We are to show that $d(\mu,\mu') \asymp d(\mu',\mu)$. With $e_\mu = \xi_\mu/|\xi_\mu|$, this is

$$|\langle e_\mu, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2 \asymp |\langle e_{\mu'}, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2.$$

It is sufficient to notice that

$$|\langle e_\mu, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2 \asymp |\langle e_\mu, \Delta x\rangle| + |\langle e_{\mu'}, \Delta x\rangle| + |\Delta x|^2 + |\Delta\theta|^2.$$

In order to justify the nontrivial inequality, use the law of cosines illustrated in Figure 4:

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

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

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

Claim (2). Recall that $\omega(\mu,\mu') = 2^{|j-j'|}(1 + 2^{\min(j,j')}d(\mu,\mu'))$. Let us show that $d(\mu,\mu') \le C\cdot(d(\mu,\mu'') + d(\mu'',\mu'))$. To simplify notations, set in the coordinates defined by $\{e_\mu, e_\mu^\perp\}$,

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

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

$$\begin{aligned}
\epsilon|x_1| \le |y_1| &+ |\cos\alpha(x_1 - y_1) + \sin\beta(x_2 - y_2)| \\
&+ (|\beta| + |\alpha - \beta|)(|y_1| + |x_1 - y_1| + |y_2| + |x_2 - y_2|),
\end{aligned}$$

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 $|\beta| + |\alpha - \beta| \leq 2\epsilon^2$. By induction, $\alpha = \beta = 0$ and $|y_1| + |x_1 - y_1| \geq |x_1|$ yields a contradiction.

Claim (3). We need to establish that $\sum_{\mu_1} \omega(\mu_0, \mu_1)^{-N} \cdot \omega(\mu_1, \mu_2)^{-N} \leq C_N \cdot \omega(\mu_0, \mu_2)^{-(N-1)}$. We closely follow and expand the argument in [28]. We will need to use $d(\mu_0, \mu_1) \asymp d(\mu_1, \mu_0)$, as we have just showed. Define I_{μ_1} by

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

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,

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

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

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

with

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

Note that

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

and, therefore,

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

In the sequel we will repeatedly make use of the bound

$$
\sum_{k, \ell}(1 + 2^q d(\mu, \mu'))^{-N} \leq C \cdot 2^{2(j-q)_+}, \tag{A.2}
$$

38

valid for $N \geq 2$, any real q and where the subscript $+$ denotes the positive part. This is justified as follows. Without loss of generality, assume that $\mu' = (j', 0, 0)$ so that the curvelet $\gamma_{\mu'}$ 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 $x_\mu = R_{\theta_\mu} D_j^{-1} k$, say. Then the left-hand side is

$$\sum_{\ell=0}^{2^{\lfloor j/2 \rfloor} - 1} \sum_{k \in \mathbb{Z}^2} \left(1 + 2^q(|2^{-j/2}\ell|^2 + |2^{-j/2}k_2|^2 + |2^{-j}k_1|)\right)^{-N}. \tag{A.3}$$

For $j \geq q$ this can be seen as a Riemann sum and bounded—up to a numerical multiplicative constant—by the corresponding integral

$$\int_{\mathbb{R}^2} \frac{dx}{2^{-3j/2}} \int_{\mathbb{R}} \frac{dy}{2^{-j/2}} [1 + 2^q(y^2 + x_2^2 + |x_1|)]^{-N}$$

which in turn is less than $C \cdot 2^{2(j-q)}$ provided $N \geq 2$. For $j \leq q$, the sum (A.3) essentially consists of a few terms, giving a $O(1)$ contribution. This gives the bound $C \cdot 2^{2(j-q)_+}$.

By symmetry, we can now assume $j_0 \leq j_2$. Let us consider three cases.

- **$0 \leq j_2 \leq j_1$.** In that case we have the bound

$$(L_1)^{-N} \leq C \cdot [(1 + 2^{j_2}d_{01})^{-N} + (1 + 2^{j_2}d_{12})^{-N}].$$

Summing this quantity over k_1 and ℓ_1 i.e., over all μ_1 that correspond to a given j_1, and using (A.2), we obtain for $j_1 \geq j_2$

$$\sum_{\mu_1} I_{\mu_1} \leq C \cdot (1 + 2^{j_0}d_{02})^{-N} \sum_{j_1 \geq j_2} 2^{-(2j_1 - j_0 - j_2)N} \cdot 2^{2(j_1 - j_2)}$$

$$\leq C \cdot 2^{-(j_2 - j_0)N}(1 + 2^{j_0}d_{02})^{-N} = C \cdot \omega(\mu_0, \mu_2)^{-N}.$$

- **$0 \leq j_1 \leq j_0$.** We now have

$$(L_1)^{-N} \leq C \cdot [(1 + 2^{j_1}d_{01})^{-N} + (1 + 2^{j_1}d_{12})^{-N}].$$

According to (A.2), the sum over k_1 and ℓ_1 of $(L_1)^{-N}$ is bounded by a constant independent of j_1. The remaining sum is

$$\sum_{\mu_1} I_{\mu_1} \leq C \cdot 2^{-(j_0 + j_2)N} \sum_{j_1 \leq j_0} 2^{2j_1 N} \cdot (1 + 2^{j_1}d_{02})^{-N}.$$

Observe that $2^{j_1 N}(1 + 2^{j_1}d_{02})^{-N} \leq 2^{j_0 N}(1 + 2^{j_0}d_{02})^{-N}$, therefore

$$\sum_{\mu_1} I_{\mu_1} \leq C \cdot 2^{-(j_2 - j_0)N}(1 + 2^{j_0}d_{02})^{-N} = C \cdot \omega(\mu_0, \mu_2)^{-N}.$$

- **$j_0 \leq j_1 \leq j_2$.** In that case we still have

$$(L_1)^{-N} \leq C \cdot [(1 + 2^{j_1}d_{01})^{-N} + (1 + 2^{j_1}d_{12})^{-N}].$$

summed over k_1 and ℓ_1 into a $O(1)$ contribution. What remains is

$$\sum_{\mu_1} I_{\mu_1} \leq C \cdot 2^{-(j_2 - j_0)N}(1 + 2^{j_0}d_{02})^{-N} \sum_{j_0 \leq j_1 \leq j_2} 1$$

$$\leq C \cdot \omega(\mu_0, \mu_2)^{-(N-1)}.$$

We conclude by collecting the estimates corresponding to the three different cases. Remark that the loss of one (fractional) power of ω in the third case is unavoidable unless one modifies its definition in the spirit of [28]. This would however make notations unnecessarily heavy.

Claim (4). See [28] p. 804. □

Proof of the inequality (2.14). Assume without loss of generality that $\mu = \mu_0$. We may express $S_{\mu'}(\xi)$ as $S_{\mu_0'}(R_{\Delta\theta}\xi)$, with $\Delta\theta = \theta_\mu - \theta_{\mu'}$. We begin by expressing the integral in polar coordinates,

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

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

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

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 [24] p.56,

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

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 (2.14). □

A.2 Additional proofs for section 5

Proof of lemma 5.4. By definition $a^{(\mu)}(x) = 2^{-3j/4}m_\mu\left(D_{2^{-j}}R_{\theta_\mu}x - k\right)$ and, therefore,

$$a^{(\mu)}(x) = \frac{1}{\alpha_\mu}\int(Rf)(a,\theta,b)a^{3/4}2^{-3j/4}\psi(D_aR_\theta(R_{\theta_\mu}^{-1}D_{2^j}(x+k) - b)\,d\mu$$
$$= \frac{1}{\alpha_\mu}\int(Rf)(a,\theta,b)|A|^{1/2}\psi\left(A(x - (\beta - k))\,d\mu,\right. \tag{A.4}$$

where $A = D_aR_\delta D_{2^j}$ with $\delta = \theta - \theta_\mu$ and $\beta = D_{2^{-j}}R_{\theta_\mu}b$.

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

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

Next $\text{Supp}\,\psi(Ax) \subset A^{-1}[0,1]^2$ with $A^{-1} = D_{2^{-j}}R_{-\delta}D_a^{-1}$. 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 (5.19).

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

$$\left|\int\psi(A(x - \beta))x_1^k\,dx_1\right| \le C_m\cdot 2^{-j(m+1)}. \tag{A.5}$$

uniformly over the $(a, \theta, b) \in Q_\mu$. The property (5.20) 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

$$\left| \int a^{(\mu)}(x_1, x_2)\, x_1^k \, dx_1 \right| \leq \frac{1}{\alpha_\mu} \|Rf\|_{L_2(Q_\mu)} \left(\int_{Q_\mu} \left| \int |A|^{1/2} \psi(A(x - \beta)) x_1^k \, dx_1 \right|^2 d\mu \right)^{1/2}$$

$$= \left(\int_{Q_\mu} \left| \int |A|^{1/2} \psi(A(x - \beta)) x_1^k \, dx_1 \right|^2 d\mu \right)^{1/2}.$$

The uniform bound (A.5) together with the fact that $\int_{Q_\mu} d\mu$ is either 3π or $3\pi/2$ gives (5.20).

We then need to establish (A.5). Let ∂_2 be $\partial/\partial x_2$, recall that by assumption (5.12)–(5.13), we have that for all $x_2 \in \mathbb{R}$,

$$\int \partial_2^n \psi(x_1, x_2) x_1^k \, dx_1 = 0, \quad k = 0, 1, \ldots, R,$$

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

$$\int \partial_2^n \psi(\alpha x_1 + \beta, x_2) x_1^k \, dx_1 = 0, \quad k = 0, 1, \ldots, R. \tag{A.6}$$

We shall use (A.6) to prove (A.5). 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 $a_{21} = -\frac{2^{-j} \sin \delta}{\sqrt{a}}$. As $a \geq 2^{-(j+1)}$ and $|\delta| \leq \pi/2 \cdot 2^{-\lfloor j/2 \rfloor}$, we have

$$|a_{21}| \leq c \cdot 2^{-j}. \tag{A.7}$$

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 $k \leq D$ and pick $N = D - k + 1$ so that for $n = 0, 1, \ldots, N - 1$, $n + k \leq D$. By virtue of (A.6) all the integrals in the sum vanish and the only remaining term is $O(a_{21}^N)$ which because of (A.7) is $O(2^{-jN})$. As a consequence, setting $m = D/2$, we conclude that

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

this is the content of (A.5).

The careful reader will notice that inequality (A.5) or equivalently (5.20) is a weaker statement than inequality (2.6) for the definition of nearly vanishing moments. There is no doubt that the stronger estimate (2.6) also holds for curvelet atoms. The proof of this fact uses standard arguments and we choose not to reproduce it here.

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

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

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

Proof of (5.22). 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 2.3 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. $\qquad\square$

Proof of inequality (5.6). Derivatives of $\hat{\gamma}_\mu$ and σ are treated using the following estimates.

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

We now develop size estimates for the phase perturbation δ. Following closely the discussion in [30], p.407, we claim that on W_μ,

$$
|\partial_\xi^\alpha \partial_x^\beta \delta(x, \xi)| \leq C_{\alpha\beta} \cdot 2^{-\alpha_1 j} 2^{-\alpha_2 j/2}. \tag{A.8}
$$

The derivations in x add no complications. Hence, assume that $\beta = 0$. As the above result (A.8) 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 $|\xi_1| \leq C \cdot 2^j$ and $|\xi_2| \leq C \cdot 2^{j/2}$. Taylor

series expansions about $\xi_2 = 0$ together with homogeneity assumptions give

$$\frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}} \delta(x,\xi) = O(|\xi_2|^2 |\xi|^{-1-\alpha_1}) = O(2^{-\alpha_1 j}),$$

$$\frac{\partial}{\partial \xi_2} \frac{\partial^{\alpha_1}}{\partial \xi_1^{\alpha_1}} \delta(x,\xi) = O(|\xi_2||\xi|^{-1-\alpha_1}) = O(2^{-j/2} 2^{-\alpha_1 j}),$$

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

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

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

and therefore (5.6). $\qquad\qquad\square$

References

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

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

[3] Candès, E. J. Harmonic analysis of neural networks. *Appl. Comput. Harmon. Anal.* **6** (1999), 197–218.

[4] Candès, E. J.; Demanet, L. Curvelets and Fourier integral operators. *C. R. Acad. Sci. Paris, Ser. I* **336** (2003), 395–398.

[5] Candès, E.J.; Demanet, L.; Donoho, D.L.; Ying L. Fast Discrete Curvelet Transforms. Technical Report, California Institute of Technology, 2005.

[6] Candès, E. J.; Donoho, D. L. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. *Curves and Surfaces*, 105–120. Edited by C. Rabut A. Cohen and L. L. Schumaker, Vanderbilt University Press, Nashville, TN, 2000.

[7] Candès, E. J.; Donoho, D. L. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. on Pure and Appl. Math.* **57** (2004), 219–266.

[8] Candès, E. J.; Donoho, D. L. Continuous curvelet transform: I. resolution of the wavefront set. To appear in *Appl. Comput. Harmon. Anal.*

[9] Candès, E. J.; Donoho, D. L. Curvelets: new tools for limited-angle tomography. Technical Report, California Institute of Technology, 2004.

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

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

[12] Donoho, D. L.; Vetterli, M.; DeVore, R. A.; Daubechies, I. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory* **44** (1998), 2435–2476.

[13] Donoho, D. L. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory* **41** (1995), 613–627.

[14] Duistermaat, J. *Fourier integral operators*. Birkhauser, Boston, 1996.

[15] Garding, L. *Singularities in linear wave propagation.* Lecture notes in Mathematics, 1241, Springer, 1987.

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

[17] Folland, G. *Harmonic analysis in phase space*, Princeton University Press, 1989.

[18] Frazier, M.; Jawerth, B.; Weiss, G. *Littlewood-Paley theory and the study of function Spaces.* CBMS, 79, American Mathematical Society, Providence, 1991.

[19] Lax, P. Asymptotic solutions of oscillatory initial value problems. *Duke Math J.* **24** (1957), 627–646.

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

[21] Meyer. Y. *Wavelets: algorithms and applications.* SIAM, Philadelphia, 1993.

[22] Mallat, S. *A wavelet tour of signal processing.* Second edition. Academic Press, Orlando-San Diego, 1999.

[23] Meyer, Y. *Ondelettes et opérateurs.* Hermann, Paris, 1990.

[24] Meyer, Y.; Coifman, R. R. *Wavelets, Calderón-Zygmund and multilinear operators.* Cambridge University Press, Cambridge, 1997.

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

[26] Seeley, R. Complex powers of an elliptic operator. *Proc. Symp. Pure Math.* **10** (1968), 288–307.

[27] Smith, H. A Hardy space for Fourier integral operators. *J. Geom. Anal.* **8** (1998), 629–653.

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

[29] Sogge, C. *Fourier integrals in classical analysis.* Cambridge University Press, 1993.

[30] Stein, E. *Harmonic analysis.* Princeton University Press, Princeton, N.J., 1993.

[31] Stolk, C.; de Hoop, M. Microlocal analysis of seismic inverse scattering in anisotropic elastic media. *Comm. Pure and Appl. Math.* **55** (2002), 261–301.

[32] Taylor, M. Reflection of singularities of solutions to systems of differential equations. *Comm. Pure and Appl. Math.* **28** (1975), 457–478.

[33] Trèves, F. *Introduction to pseudo-differential and Fourier integral operators.* Plenum press, 1982, 2 volumes.

[34] Whitham, G. *Linear and nonlinear waves.* Wiley Interscience, 1999.