A multi-scale approach to hyperbolic evolution equations with limited smoothness *

Fredrik Andersson[†], Maarten V. de Hoop[‡], Hart F. Smith[§] and Gunther Uhlmann[¶]

January 26, 2007

Abstract

We discuss how techniques from multiresolution analysis and phase space transforms can be exploited in solving a general class of evolution equations with limited smoothness. We have wave propagation in media of limited smoothness in mind. The frame that appears naturally in this context is closely related to the one formed by curvelets. The construction considered here implies a full-wave description on the one hand but reveals the geometrical properties derived from the propagation of singularities on the other hand. The approach and analysis we present (i) aids in the understanding of the notion of scale in the wavefield and how field scale is coupled to configurational, i.e. medium, scales, (ii) admits media of limited smoothness, viz. with Hölder regularity $s \geq 2$, and (iii) suggests a novel computational algorithm that requires solving for the mentioned geometry on the one hand and solving a matrix Volterra integral equation of the second kind on the other hand. The Volterra equation can be solved by recursion – as in the computation of certain multiple scattering series – revealing a curvelet-curvelet interaction. We give precise estimates expressing the degree of concentration of curvelets following the propagation of singularities.

Keywords: pseudodifferential evolution equations; paradifferential decomposition; dyadic parabolic decomposition; curvelets.

1 Introduction

We consider evolution equations of the type

$$[\partial_z - \mathrm{i}P(z, x, D_x)]u = 0, \tag{1}$$

subjected to the initial condition $u(z_0, .) = u_0$, where z is an evolution parameter restricted to an interval $[z_0, Z]$, and $x \in X \subset \mathbb{R}^n$. If P is a pseudodifferential operator, with real symbol of order 1, the solution operator to this equation is a Fourier integral operator (FIO). The symbols of pseudodifferential operators are smooth. With a Fourier integral operator is associated the notion of propagation of singularities, namely, the canonical relation of the solution operator prescribes how the wavefront set of $u(z, .)$ is related to the wavefront of $u(z_0, .)$ if $Z > z > z_0$.

In this paper, we are concerned with the solution operators and their construction, if the smoothness of the symbol of pseudodifferential operator P is limited. Our analysis applies to the cases $P \in C^s S^1_{1,0}$ with $s \geq 2$. We follow essentially a multi-scale approach to solving such evolution equations, derived from the approach of Smith [29], and make use of solution representations based on wavepackets or curvelets [8, 6, 7]. The solution operator generalizes the notion of Fourier integral operators the canonical relations of which are generated by canonical transformations. The operator will be described in terms of actions on, and, explicitly, scattering between wavepackets or curvelets. It appears that the multi-scale approach inherits an imprint of the classical propagation of singularities (per scale). A

*This research was supported in part under NSF CMG grant EAR-0417891 and NSF grant EAR-0409816.

[†]Centre for Mathematical Sciences, Lund Institute of Technology / Lund University, SE-221 00 Lund, Sweden

[‡]Center for Computational and Applied Mathematics, Purdue University, 150 N. University Street, West Lafayette IN 47907, USA

[§]Department of Mathematics, University of Washington, Seattle WA 98195-4350, USA

[¶]Department of Mathematics, University of Washington, Seattle WA 98195-4350, USA

key property that emerges out of the operator construction is the phenomenon of concentration of packets or curvelets. This concentration follows, basically, from a careful analysis of the operator kernel, and associated estimates, that generates the scattering between wavepackets. The analysis we present is adapted to computation.

The scattering between wavepackets is described in terms of a Volterra equation of the second kind which admits a series expansion. From a computational perspective, one needs to solve eikonal equations (that is, the associated Hamilton systems) for each scale on the one hand, and the Volterra equation on the other hand. The computation of, and the use of, the geometry from the Hamilton systems is reminiscent of the 'high-frequency' solution to the evolution equation. The Volterra equation can be solved with a step-by-step method; this method has its counterpart in the Trotter product representation (Kumano-go and Taniguchi [24]) of the Fourier integral operators appearing in the smooth pseudodifferential evolution equation case. In each step, the Volterra equation can be solved by recursion – as in the computation of certain multiple scattering series (De Hoop [14]), where wave constituents are replaced by wave packets.

The approach followed here has its roots in the theory of coherent wave packets and Fourier integral operators (Cordoba and Fefferman [12]) and combines elements of the dyadic parabolic decomposition of Fourier integral operators (Stein [30]), of the technique of parabolic cutoffs in the treatment of Fourier integral operators (the class $I^{p,l}$) with certain singular symbols (Greenleaf and Uhlmann [20]), of the parametrix construction for the wave equation with $C^{1,1}$ (essentially Hölder class C^2) coefficients (Smith [28, 29]), and of paradifferential calculus (Bony [3], and Coifman and Meyer [11]; see also Taylor [33]). The concept of the mentioned parabolic cutoffs goes back to Boutet de Monvel [4]. Furthermore, the frame of curvelets and the associated curvelet transform used here can be related to the Fourier-Bros-Iagolnitzer (FBI) transform (Bros and Iagolnitzer [5]) as well as with Gaussian beams and the over-complete frame they form (for example, Shlivinski, Heyman, Boag and Letrou [27]). Geba and Tataru [19] adapted the Bargmann transform to the wave equation in a manner related to curvelets to characterize the associated classes of Fourier integral operators. Curvelets have been used in analyzing wave propagators by Candès and Demanet [9].

The outline of the paper is as follows. In Section 2 we introduce the class of evolution equations considered, and the relevant estimates derived from paradifferential techniques. In Section 3 we summarize the curvelet transform in the context of the dyadic parabolic decomposition of phase space and its generalization in terms of frames and co-frames. Using these frames and co-frames, in Section 4 we initiate the multi-scale approach by constructing approximate solutions to the evolution equations. The full, weak solutions to the evolution equations are computed by solving a system of coupled Volterra equations of the second kind, which is the subject of Section 5. In this section, we furthermore develop a step-by-step approach, prove convergence of the solution of the system of Volterra equations by recursion, and analyze the degree of concentration of curvelets near the flow associated with the canonical transformation generated by the evolution equation, which depends on the Hölder class of the symbol, that is, the medium. We discuss the sparsity of the matrix representation with respect to the frame of curvelets of the full solution operator. In Section 6, we discuss how the frame developed in Section 3 can be discretized while preserving its essential properties. Moreover, we design a numerical forward and inverse transform pair. We further demonstrate how the approximate solution of Section 4, that is, the leading-order recursive solution to the Volterra equation, can be numerically computed, while preserving estimates for the Volterra kernel in Section 5.

A key application of the approach developed here is a solution to the wave-equation imaging problem. We briefly mention an example of the formulation based on 'seismic data downward continuation'; for the detailed mathematical framework of the downward continuation formulation, see Stolk and De Hoop [31, 32]. Then $z \in \mathbb{R}_+$ stands for depth pointing towards the earth's interior, $z_0 = 0$ represents the earth's surface, x stands for source, receiver coordinates and time $(s, r, t) \in \mathbb{R}^{m-1} \times \mathbb{R}^{m-1} \times \mathbb{R}_+$ ($n = 2m - 1$ in the above, with $m = 2, 3$), u_0 stands for seismic reflection data, while

$$P(z, s, r, D_s, D_r, D_t) = B(z, s, D_s, D_t) + B(z, r, D_r, D_t),$$

$$B(z, y, D_y, D_t) = \tfrac{1}{2} D_t^{-1} \sum_{j=1}^{m-1} D_{y_j} c(z, y) D_{y_j}, \quad y = s, r,$$

known as the paraxial approximation, implying that the wavespeed $c(z, .) \in C^s$ (Lipschitz in z is sufficient). The image at depth $z > z_0$ and position $y \in \mathbb{R}^{m-1}$ is extracted from the solution as $u(z, s = y, r = y, t = 0)$. (Here, u is identified as the wavefield in a 'comoving' frame of reference, that is, the seismic wavefield w, say, is written as $w(z, s, r, t) = (2\pi)^{-1} \int \exp[-\mathrm{i}\tau(t' - t + T(z, s, r))] u(z, s, r, t') \, \mathrm{d}t' \mathrm{d}\tau$, with $T(z, s, r) = \int_{z_0}^{z} \mathrm{d}z' [c(z', s)^{-1} +$

$c(z', r)^{-1}$].) From a geophysical perspective, the approach and analysis presented here aids in the fundamental understanding of the notion of scale in the (initial) data (waves), how scale in the data is coupled to scales in the medium (wavespeed), and how scales in the data map to scales in the (solution) image (the above mentioned concentration). The spatial scale of variation in wavespeed is tied to geodynamical processes in the earth's interior. As an additional benefit, seismic data and seismic images are candidates for sparse representations with respect to frames of curvelets. Processes such as data regularization (correcting for missing data traces), correction for illumination of images, in combination with denoising, can well be formulated and performed based on such sparse representations (see Herrmann, Moghaddam and Kirlin [22]). In the quasi-constant coefficient case, the notion of approximate solutions (Section 4) has been tied to FIOs generated by canonical transformations and applied to imaging of the type mentioned here in Douma and De Hoop [16].

2 Hyperbolic evolution equations with limited smoothness

2.1 A paradifferential decomposition

Let C^s denote the Hölder class of order s. The Hölder classes $\{C^s\}$ form a scale of spaces, that is $C^r \subset C^s$ if $s < r$. For integer s the class C^s is defined by continuity of derivatives of order up to s, while for $s \in \mathbb{R}^+ \backslash \mathbb{Z}^+$ the Hölder classes C^s coincide with the Zygmund classes C^s_*, equipped with the norm

$$\|f\|_s := \sup_k 2^{k\,s} \|\psi_k(D) f\|_{L^\infty}.$$

Here, $\psi_k(\xi)$ is supported about $\langle \xi \rangle \sim 2^k$ with $\langle \xi \rangle = (1 + \|\xi\|^2)^{1/2}$, while $\sum_{k=0}^\infty \psi_k(\xi) = 1$ forming a partition of unity.

We consider pseudodifferential operators P with real symbol $p \in C^s S^1_{1,0}$ [33, Section 1.3], and we assume that the symbol p of P is homogeneous of degree 1 in the phase variable ξ for $\|\xi\| \geq 1$. Our operators are pseudodifferential operators in x, parameterized by a variable z; we assume that $z \in [z_0, Z]$. We write $P = P(z, x, D_x)$. Symbols $p \in C^s S^1_{1,0}$ satisfy the estimates:

$$\|\partial_\xi^\alpha p(\cdot, \xi)\|_s \leq C_\alpha (1 + \|\xi\|)^{1-|\alpha|}.$$

Following [33, Section 1.3] we consider a paradifferential decomposition $P = \bar{P}^\sharp + \bar{P}^\flat$ corresponding to the decomposition of p, with frequency localization parameter $\delta = \frac{1}{2}$,

$$\bar{p}^\sharp(z, x, \xi) = \sum_{k=0}^\infty \bar{p}_k(z, x, \xi)\, \psi_k(\xi), \quad \bar{p}_k(z, x, \xi) := \varphi(2^{-k/2} D_x)\, p(z, x, \xi). \tag{2}$$

Here, $\widehat{\varphi}(\xi) = 1$ for $\|\xi\| \leq 1$ and $= 0$ for $\|\xi\| > 2$; thus, $\bar{p}_k$ can be obtained by low-pass filtering p in x for $\|\xi\| < 2^{k/2}$ for each z. The summation for $\bar{p}^\sharp$ follows the Littlewood-Paley decomposition where each annulus in the ξ-space is associated with a dyadic scale 2^k.

With $p \in C^s S^1_{1,0}$ we have $\bar{p}^\sharp \in S^1_{1,\frac{1}{2}}$, that is, a symbol of order 1 and type $(1, \frac{1}{2})$. Furthermore,

$$\partial_x^\beta \bar{p}^\sharp \in \begin{cases} S^1_{1,\frac{1}{2}} & , \quad |\beta| \leq s \\ S^{1+\frac{1}{2}(|\beta|-s)}_{1,\frac{1}{2}} & , \quad |\beta| \geq s \end{cases}.$$

On the other hand, $\bar{p}^\flat \in C^s S^{1-\frac{s}{2}}_{1,\frac{1}{2}}$, [33, Prop. 1.3.E]. This yields that $\bar{P}^\flat$ is of order 0 provided that $s \geq 2$, which we will assume in all of our analysis. We collect some estimates that follow from [33, Lemma 1.3.C],

$$\|\partial_x^\beta \bar{p}_k(z, \cdot, \xi)\|_{L^\infty} \leq C_\beta\, 2^{k(|\beta|-s)/2} \langle \xi \rangle\,, \qquad |\beta| \geq s\,, \tag{3}$$

$$\|p(z, \cdot, \xi) - \bar{p}_k(z, \cdot, \xi)\|_{L^\infty} \leq C\, 2^{-ks/2} \langle \xi \rangle\,. \tag{4}$$

By [33, Prop.2.1.E], it follows that $\bar{P}^\flat \ : \ H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ is bounded for all z and $-\frac{1}{2}s < r < s$; in case $s = 2$, $\bar{P}^\flat$ is bounded for $-1 \le r \le 2$ by [29, Theorem 4.5]. Thus, $\bar{P}^\flat$ acts as an operator of order 0 on $H^r(\mathbb{R}^n)$ for a suitable range of r.

The dependence of the symbol p on z will be assumed to be Lipschitz (Hölder regularity 1). The smoothing of the symbol with respect to z is carried out by convolution against dilates by 2^{-k} of a compactly supported test function. Applying this smoothing to $\bar{p}_k$ we obtain the pseudodifferential operator symbol p_k. With $p^\sharp(z,x,\xi) = \sum_{k=0}^\infty p_k(z,x,\xi)\,\psi_k(\xi)$ we obtain the decomposition $P = P^\sharp + P^\flat$. With $s \ge 2$, the estimate (3) generalizes to

$$\|\partial_z^m \partial_x^\beta p_k\|_{L^\infty([z_0,Z]\times\mathbb{R}^n)} \le C_{m,\beta}\, 2^{k(2m+|\beta|-2)/2}\langle\xi\rangle\,, \qquad 2m+|\beta| \ge 2\,, \tag{5}$$

while estimate (4) generalizes to

$$\|p(\cdot,\cdot,\xi) - p_k(\cdot,\cdot,\xi)\|_{L^\infty([z_0,Z]\times\mathbb{R}^n)} \le C\, 2^{-k}\langle\xi\rangle\,. \tag{6}$$

Estimates (5)-(6) are used to prove Theorem 5.1.

In this paper, we study the hyperbolic evolution equation and the associated Cauchy initial value problem,

$$[\partial_z - \mathrm{i}P(z,x,D_x)]u = 0\,, \quad u(z_0,x) = u_0(x)\,. \tag{7}$$

We denote its solution operator by $F(z,z_0)$, so that $u(z,\cdot) = F(z,z_0)u_0$. Here, $z_0 \le z \le Z$ as before.

2.2 The smooth case: The Hamiltonian flow

If the symbol p of the pseudodifferential operator P is smooth and of order 1, the evolution operator $F(z,z_0)$ for (7) propagates singularities along bicharacteristics. The principal symbol of P (which we also denote by p) defines a Hamiltonian. The bicharacteristics then follow the Hamiltonian flow,

$$\frac{\mathrm{d}x}{\mathrm{d}z} = -\frac{\partial p}{\partial \xi}(z,x,\xi)\,, \qquad \frac{\mathrm{d}\xi}{\mathrm{d}z} = \frac{\partial p}{\partial x}(z,x,\xi)\,, \tag{8}$$

subject to the initial conditions $x = x_0$, $\xi = \xi_0$ at $z = z_0$. Solutions to (8) are denoted by $x = x(z,z_0,x_0,\xi_0)$ and $\xi = \xi(z,z_0,x_0,\xi_0)$; we will also use the shorthand notation $x(z,z_0)$ and $\xi(z,z_0)$.

If P in (7) is a smooth pseudodifferential operator, the solution operator $F(z,z_0)$ of the evolution equation is a family of Fourier integral operators, depending smoothly on parameters (z,z_0), with their canonical relation Λ given by the twisted graph

$$\Lambda = \{(x(z,z_0),\xi(z,z_0);x_0,-\xi_0)\}\,.$$

In the later analysis, we work with the Hamiltonian flows associated to p_k. These flows do not correspond to physical rays of P, but are approximations defined using an appropriate geometry.

To accommodate the implied scale decomposition, we view the norms of the covectors separately from their directions. Hence, we also consider the flow introduced above projected onto the cosphere bundle, $S^*(\mathbb{R}^n)$. Normalizing the cotangent vector, introducing $\nu = \xi/\|\xi\|$, the Hamilton equations (8) result in the system

$$\frac{\mathrm{d}x}{\mathrm{d}z} = -\frac{\partial p}{\partial \xi}, \qquad \frac{\mathrm{d}\nu}{\mathrm{d}z} = \frac{\partial p}{\partial x} - \left\langle \nu, \frac{\partial p}{\partial x}\right\rangle \nu, \tag{9}$$

To describe the rotation of the covector in the flow, we introduce the z-dependent family of orthogonal matrices $\Theta \in O(n)$ that satisfy the equation

$$\frac{\mathrm{d}\Theta}{\mathrm{d}z} = \Theta\left[\nu\otimes\frac{\partial p}{\partial x} - \frac{\partial p}{\partial x}\otimes\nu\right] \tag{10}$$

subject to the initial condition $\Theta = \mathrm{I}$ at $z = z_0$. For the solutions $\Theta(z,z_0)$ and $\nu(z,z_0) = \xi(z,z_0)/\|\xi(z,z_0)\|$ of (10) and (9), we have $\frac{\mathrm{d}}{\mathrm{d}z}[\Theta(z,z_0)\,\nu(z,z_0)] = 0$ so that

$$\nu(z,z_0) = \Theta(z,z_0)^{-1}\,\nu_0\,,$$

with $\nu_0 = \nu(z_0,z_0) = \xi_0/\|\xi_0\|$.

To initiate the approach developed here, we construct approximate solutions to the evolution equation (Section 4). These solutions are derived from the parametrix of the evolution equation in which the symbol of P has been replaced by the smoothed symbol $p^\sharp$.

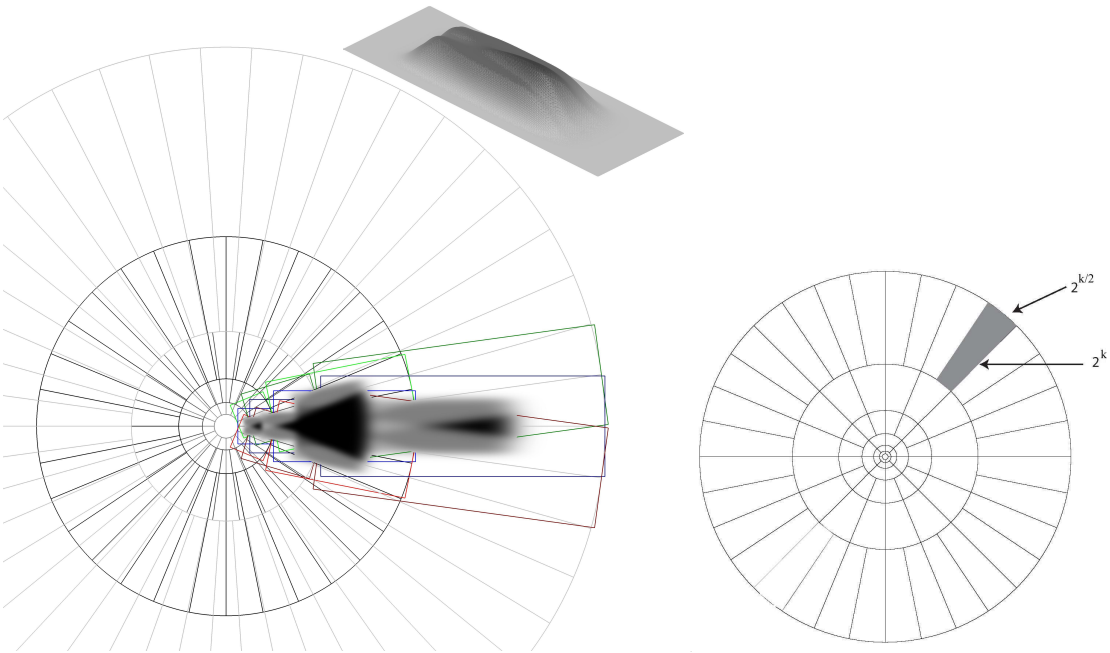

Figure 1: The partition of unity used in Section 6 (left), a window function $\widehat{\chi}_{\nu,k}$ (top), and the common tiling [30] (right) consistent with dyadic parabolic decomposition.

3 Dyadic parabolic decomposition of phase space

We recall the second dyadic decomposition ([18], also [30, IX.4.4]) upon which the curvelet decomposition is based. We introduce the frame of wave packets/curvelets which will be used to construct and analyze the propagator $F(z, z_0)$. Our methods work equally well for tight frames, or for a more general frame/co-frame pair.

3.1 Frames and representations

We begin with an overlapping covering of the positive ξ_1 axis by rectangles of the form

$$B_k = \left[\xi'_k - \frac{L'_k}{2}, \xi'_k + \frac{L'_k}{2}\right] \times \left[-\frac{L''_k}{2}, \frac{L''_k}{2}\right]^{n-1},$$

where the centers ξ'_k, as well as the side lengths L'_k and L''_k, satisfy the parabolic scaling condition

$$\xi'_k \sim 2^k, \quad L'_k \sim 2^k, \quad L''_k \sim 2^{k/2}, \quad \text{as } k \to \infty\,.$$

For $k = 0$, B_0 is a cube centered at $\xi_0 = 0$, with $L'_k = L''_k$. See figure 1 [1]. In this figure (right) we show the usual dyadic parabolic decomposition; figure 1 (left) illustrates a more flexible decomposition with essentially the same scaling properties.

Next, for each $k \geq 1$, let ν vary over a set of approximately $2^{k(n-1)/2}$ uniformly distributed unit vectors. (We can index ν by $\ell = 0, \ldots, N_k - 1$, $N_k = \lfloor 2^{k(n-1)/2} \rfloor$: $\nu = \nu(\ell)$ while we adhere to the convention that $\nu(0) = e_1$ aligns with the ξ_1-axis.) Let $\Theta_{\nu,k}$ denote a choice of rotation matrix which maps ν to e_1, and

$$B_{\nu,k} = \Theta_{\nu,k}^{-1} B_k\,.$$

[1] It is computationally advantageous to depart from a pure dilation representation of B_k for relatively small sized k.

The parameters ξ_k', L_k', L_k'', and ν are chosen so that the $B_{\nu,k}$ amply cover $\mathbb{R}^n$, in the sense that with L_k' and L_k'' multiplied by some fixed $r < 1$, the interiors would still cover.

The final ingredient in the frame construction are two sequences of smooth functions $\widehat{\chi}_{\nu,k}$ and $\widehat{\beta}_{\nu,k}$ on $\mathbb{R}^n$, each supported in $B_{\nu,k}$, so that

$$\widehat{\chi}_0(\xi)\,\widehat{\beta}_0(\xi) + \sum_{k\geq 1}\sum_{\nu} \widehat{\chi}_{\nu,k}(\xi)\,\widehat{\beta}_{\nu,k}(\xi) = 1,$$

yielding a co-partition of unity, and such that

$$|\langle \nu, \partial_\xi\rangle^j\, \partial_\xi^\alpha \widehat{\chi}_{\nu,k}(\xi)| + |\langle \nu, \partial_\xi\rangle^j\, \partial_\xi^\alpha \widehat{\beta}_{\nu,k}(\xi)| \leq C_{j,\alpha}\, 2^{-k(j+|\alpha|/2)}, \tag{11}$$

in which the constants are independent of ν, k. In figure 1 we show $\widehat{\chi}_{\nu,k}(\xi)\,\widehat{\beta}_{\nu,k}(\xi)$ for a typical choice of the above mentioned sequences.

We define

$$\widehat{\psi}_{\nu,k}(\xi) = \rho_k^{-1/2}\,\widehat{\beta}_{\nu,k}(\xi)\,, \qquad \widehat{\varphi}_{\nu,k}(\xi) = \rho_k^{-1/2}\,\widehat{\chi}_{\nu,k}(\xi)\,, \tag{12}$$

with $\rho_k = (2\pi)^{-n}|B_k| = (2\pi)^{-n} L_k'(L_k'')^{n-1}$. Both functions satisfy estimates of the type

$$|\varphi_{\nu,k}(x)| \leq C_N 2^{k(n+1)/4}\,(\,2^k|\langle \nu, x\rangle| + 2^{k/2}|x|\,)^{-N}. \tag{13}$$

We obtain a frame / co-frame pair by subjecting $\varphi_{\nu,k}$ and $\psi_{\nu,k}$ to translations over x_j, resulting in $\varphi_{\nu,k}(x - x_j)$ and likewise for $\psi_{\nu,k}$. Let $\{x_j\}$ denote a set of points in $\mathbb{R}^n$, depending on (ν, k). Introducing triplets $\gamma = (x_j, \nu, k)$, we get $\varphi_\gamma(x) = \varphi_{\nu,k}(x - x_j)$, or [2]

$$\widehat{\varphi}_\gamma(\xi) = \rho_k^{-1/2}\,\widehat{\chi}_{\nu,k}(\xi)\,\exp[-\mathrm{i}\langle x_j, \xi\rangle], \quad k \geq 1. \tag{14}$$

In the further analysis we will consistently filter out coarse-scale ($k \leq 0$) contributions.

The translation factor $\exp[-\mathrm{i}\langle x_j, \xi\rangle]$ is representative of a Fourier basis (with frequencies x_j) for functions of ξ. The (compact) support of $\widehat{\chi}_{\nu,k}$ admits an orthonormal basis defining a Fourier series. Thus, we introduce the lattice

$$X_j := (j_1, \ldots, j_n) \in \mathbb{Z}^n,$$

and capture the scaling of B_k in the dilation matrix

$$D_k = \frac{1}{2\pi}\begin{pmatrix} L_k' & 0_{1\times n-1} \\ 0_{n-1\times 1} & L_k'' I_{n-1} \end{pmatrix}.$$

Choosing $x_j = \Theta_{\nu,k}^{-1} D_k^{-1} X_j$ yields an orthogonal basis, $\exp[-\mathrm{i}\langle x_j, \xi\rangle]$, for functions supported in $B_{\nu,k}$.

LEMMA 3.1. *Let $\gamma = (x_j, \nu, k)$ with $x_j = \Theta_{\nu,k}^{-1} D_k^{-1} X_j$. Then the functions φ_γ and ψ_γ form a frame / co-frame pair in $L^2(\mathbb{R}^n)$, i.e., if*

$$u_\gamma = \langle u|\psi_\gamma\rangle = \int u(x)\overline{\psi_\gamma(x)}\,\mathrm{d}x, \tag{15}$$

then

$$u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x). \tag{16}$$

Furthermore, for each fixed wedge indexed by ν, k, it holds true that

$$\sum_{\gamma'\colon\, k'=k,\ \nu'=\nu} u_{\gamma'}\widehat{\varphi}_{\gamma'}(\xi) = \widehat{u}(\xi)\,\widehat{\beta}_{\nu,k}(\xi)\,\widehat{\chi}_{\nu,k}(\xi)\,. \tag{17}$$

[2] Our Fourier transform convention is $\widehat{u}(\xi) = \int u(x)\,\exp[-\mathrm{i}\langle x, \xi\rangle]\,\mathrm{d}x$, $u(x) = (2\pi)^{-n}\int \widehat{u}(\xi)\,\exp[\mathrm{i}\langle x, \xi\rangle]\,\mathrm{d}\xi$.

The left-hand side of (17) is a sum over j (x_j) for given ν, k. It should be noted that the frame is not orthogonal. In case $\beta_{\nu,k} = \chi_{\nu,k}$ the frame is tight, and corresponds to the frame elements introduced by Smith [29], and also curvelets [8, 6, 10]. In this case, we have a Plancherel formula

$$\|u\|_{L^2}^2 = \sum_\gamma |u_\gamma|^2 .$$

Proof. The functions $\rho_k^{-1/2} \exp[-\mathrm{i}\langle x_j, \xi\rangle]$, $j \in \mathbb{Z}^n$ form an orthonormal basis for $L^2(B_{\nu,k})$. It is natural to express (15) as a convolution, that is

$$u_\gamma = \int u(x)\overline{\psi(x)}\mathrm{d}x = \rho_k^{-1/2} \int u(x)\overline{\beta_{\nu,k}(x - x_j)}\mathrm{d}x = \rho_k^{-1/2}(u * \beta_{\nu,k})(x_j), \tag{18}$$

where we have used (12) and that $\beta_{\nu,k}(x) = \overline{\beta_{\nu,k}(-x)}$, noting that $\widehat{\beta}_{\nu,k}$ is real valued. We express (18) in terms of an inverse Fourier transform,

$$u_\gamma = \langle u|\psi_\gamma\rangle = (2\pi)^{-n}\rho_k^{-1/2} \int \widehat{u}(\xi)\widehat{\beta}_{\nu,k}(\xi) \exp[\mathrm{i}\langle x_j, \xi\rangle]\, \mathrm{d}\xi. \tag{19}$$

Then

$$\sum_{\gamma'\colon k'=k,\ \nu'=\nu} u_{\gamma'}\widehat{\varphi}_{\gamma'}(\xi) = \rho_k^{-1/2} \sum_{j'} u_{j',\nu,k} \exp\left[-\mathrm{i}\langle x_{j'}, \xi\rangle\right] \widehat{\chi}_{\nu,k}(\xi) = \widehat{\beta}_{\nu,k}(\xi)\, \widehat{\chi}_{\nu,k}(\xi)\, \widehat{u}(\xi)\,,$$

writing $u_{x_j,\nu,k} = u_{j,\nu,k}$. This establishes (17). Summing over (ν, k) yields $\widehat{u}(\xi)$, proving (16). □

Equation (15) defines a mapping $U\ :\ u(x) \to (u_\gamma)$, while equation (16) defines a mapping $V\ :\ (u_\gamma) \to u(x)$; V is the left inverse of U. Furthermore, the mapping

$$UV\ :\ u_\gamma \to \int \overline{\psi_{\gamma'}(x)} \sum_\gamma u_\gamma \varphi_\gamma(x)\, \mathrm{d}x$$

is the orthogonal projection onto $\operatorname{ran} U$. We have

LEMMA 3.2. *Let $\gamma = (x_j, \nu, k)$ and $\{x_j\}$ denote the lattice derived from $\{X_j\}$ for given ν, k, and likewise for γ'. Let $c_{\gamma\gamma'} = \langle\varphi_\gamma|\psi_{\gamma'}\rangle$. The following estimate holds: For each $N = 1, 2, \ldots$ there exists a constant C_N such that*

$$|c_{\gamma\gamma'}| \le C_N 1_{\{(\nu',k')\in\mathcal{N}(\nu,k)\}} \langle D_k(x_j - x_{j'})\rangle^{-N}, \tag{20}$$

where $(\nu', k') \in \mathcal{N}(\nu, k)$ if the supports of $\widehat{\varphi}_\gamma$ and $\widehat{\psi}_{\gamma'}$ overlap.

Proof. The factor $1_{\{(\nu',k')\in\mathcal{N}(\nu,k)\}}$ in the estimate follows immediately from the identity $\langle f(.-a)|g(.-b)\rangle = \mathcal{F}^{-1}(\widehat{f}\,\overline{\widehat{g}})(b-a)$. The decay estimate follows from decay estimates of Fourier transforms of smooth functions subjected to parabolic scaling (it should be noticed that $\{x_j\}$ and $\{x_{j'}\}$ are different lattices), and can be found, for example, in [7, Section 5.2]. □

In the case of the tight frame of curvelets, c represents the Gram matrix. The co-frame is related to the frame according to $\psi_\gamma = (VV^*)^{-1}\varphi_\gamma$; $VV^* = I$ in the case of the tight frame of curvelets.

There is a relation between the curvelet transform, the FBI transform and coherent wave packets. In the wave packet approach the analyzing elements can be viewed as Gabor functions where the frequency and window size are connected by the quadratic relation (window size)^2 = spatial frequency. In this context, we mention the almost diagonalization of pseudodifferential operators of type $S^0_{0,0}$ by making use of Gabor frames [21], the underlying theory of which differs from the decomposition followed here.

4 Construction of approximate solutions

We discuss a construction of approximate solutions to (7). The construction is based on a decomposition of the action of solution operator F into wave packets or curvelets. It involves a decomposition into scales, and is reminiscent of high-frequency solutions to the evolution equation.

For each P_k, that is for each scale, we identify its principal symbol p_k with a Hamiltonion as in (8). The associated flows, however, do not correspond with physical rays, and are merely defining an appropriate geometry. Normalizing the cotangent vectors, we obtain (9) with p replaced by p_k.

4.1 'Rigid' motion of wave packets

We approximate the action of $F(z, z_0)$ on frame elements φ_γ by considering rigid motions; the approximation follows a description in terms of particles. We subject φ_γ to a rigid motion in accordance with the Hamiltonian flow defined by p_k for a given scale k. Let $x_\gamma(z, z_0)$ stand for $x(z, z_0)$ evaluated with $p = p_k$ if $\gamma = (x_0, \nu, k)$; we introduce $\Theta_\gamma(z, z_0)$ in a similar manner, satisfying equation (10) with p replaced by p_k. We let $\varphi_\gamma(z, z_0, y)$ denote the function obtained by rigid motion of $\varphi_\gamma(y)$ with the flow out of (x_0, ν) at scale k,

$$\varphi_\gamma(z, z_0, y) = \varphi_\gamma(\Theta_\gamma(z, z_0)(y - x_\gamma(z, z_0)) + x_\gamma) \ , \quad x_\gamma = x_0; \tag{21}$$

we assume $x_0 \in \{x_j\}$ for the given ν, k. The motion is readily evaluated in the Fourier domain (cf. (14)), viz.

$$\widehat{\varphi}_\gamma(z, z_0, \eta) = \rho_k^{-1/2} \widehat{\chi}_{\nu,k}(\Theta_\gamma(z, z_0)\eta) \ \exp[-\mathrm{i}\langle \eta, x_\gamma(z, z_0)\rangle]. \tag{22}$$

Equation (21) approximately solves (7) with initial condition $u_0 = \varphi_\gamma$. The approximation can be motivated, from the infinitesimal generator point of view, as follows. First, we observe that

$$\partial_z \varphi_\gamma(z, z', y) = \mathcal{L}_k(z, x_\gamma(z, z'), \nu_\gamma(z, z'), y, \partial_y)\varphi_\gamma(z, z', y) \, ,$$

with $\mathcal{L}_k(z, x, \nu, y, \partial_y)$ given by

$$\langle \partial_\xi p_k(z, x, \nu), \partial_y\rangle + \langle y - x, \partial_x p_k(z, x, \nu)\rangle\langle \nu, \partial_y\rangle - \langle \nu, y - x\rangle\langle \partial_x p_k(z, x, \nu), \partial_y\rangle \tag{23}$$

in which $\partial_\xi p_k$ and $\partial_x p_k$ arise from the Hamilton system that determines x_γ and Θ_γ, and $\Theta_\gamma(z, z')\nu_\gamma(z, z') = \nu$. We can view $\mathcal{L}_k$ as a pseudodifferential operator (in y) with (elementary) symbol $\mathcal{L}_k(z, x_\gamma(z, z'), \nu_\gamma(z, z'), y, \mathrm{i}\eta)$. The question is up to which order the operator $\mathcal{L}_k$ cancels the action of operator $\mathrm{i}P_k$ ([29, (3.5)]).

The action of $P_k(z, y, D_y)$ on $\varphi_\gamma(z, z', \cdot)$ attains the form (cf. (22))

$$\begin{aligned} P_k(z, y, D_y)\, \varphi_\gamma(z, z', \cdot) &= (2\pi)^{-n} \int p_k(z, y, \eta) \ \exp[\mathrm{i}\langle \eta, y\rangle] \, \widehat{\varphi}_{\gamma'}(z, z', \eta) \, \mathrm{d}\eta \\ &= (2\pi)^{-n} \rho_k^{-1/2} \int p_k(z, y, \eta) \, \widehat{\chi}_{\nu,k}(\Theta_\gamma(z, z')\eta) \ \exp[\mathrm{i}\langle \eta, y - x_\gamma(z, z')\rangle] \, \mathrm{d}\eta \, . \end{aligned} \tag{24}$$

We expand p_k in (y, η) about $(x_1, \xi_1) \equiv (x_\gamma(z, z'), \Theta_\gamma(z, z')\xi_0)$,

$$\begin{aligned} p_k(z, y, \eta) = p_k(z, x_1, \xi_1) &+ \langle \eta - \xi_1, \partial_\eta p_k(z, x_1, \xi_1)\rangle \\ &+ \langle y - x_1, \partial_y p_k(z, x_1, \xi_1)\rangle + \langle y - x_1, \partial_y\rangle\langle \eta - \xi_1), \partial_\eta\rangle p_k(z, x_1, \xi_1)\rangle \\ &\qquad + \tfrac{1}{2}(y - x_1)^2 \partial_y^2 p_k(z, x_1, \xi_1) + \tfrac{1}{2}(\eta - \xi_1)^2 \partial_\eta^2 p_k(z, x_1, \xi_1) + \text{l.o.t.} \end{aligned}$$

where l.o.t. denotes symbols that will lead to terms of order 0 or lower.

Let ν_1 denote the direction of ξ_1. Then $\langle \nu_1, \partial_\eta\rangle \partial_{\eta_j} p_k(z, x_1, \xi_1) = 0$, by homogeneity of $\partial_{\eta_j} p_k$ of order 0. Consequently, the symbol $(\eta - \xi_1)^2 \partial_\eta^2 p_k$ only involves the component of $(\eta - \xi_1)$ perpendicular to ν_1, which is bounded by $2^{k/2}$. Since $\partial_\eta^2 p_k \approx 2^{-k}$, this symbol leads to a bounded term.

Similarly, when applied to $\varphi_\gamma(z, z', \cdot)$, the factor $(y - x_1)$ is of size $2^{-k/2}$, and since $\partial_y^2 p_k \approx 2^k$, the symbol $(y - x_1)^2 \partial_y^2 p_k$ also leads to bounded terms.

By homogeneity, again, the other terms simplify to

$$\langle \partial_\eta p_k(z, x_1, \nu_1), \eta\rangle + \langle y - x_1, \partial_y\rangle\langle \partial_\eta p_k(z, x_1, \nu_1), \eta\rangle\,.$$

In the second term we may replace η by $\langle \nu_1, \eta\rangle\nu_1$, since the difference is $\approx 2^{k/2}$ which with $(y-x_1)$ leads to bounded terms, as $\partial_y\partial_\eta p_k \approx 1$. With this replacement we have the symbol

$$\langle \partial_\eta p_k(z, x_1, \nu_1), \eta\rangle + \langle y - x_1, \partial_y p_k(z, x_1, \nu_1)\rangle\langle \nu_1, \eta\rangle\,.$$

This differs from the symbol of $\mathrm{i}\mathcal{L}_k$ by the term

$$\langle \nu_1, y - x_1\rangle\langle \partial_y p_k(z, x_1, \nu_1), \eta\rangle$$

which leads to bounded terms since $\langle \nu_1, y - x_1\rangle \approx 2^{-k}$ when acting on $\varphi_\gamma(z, z', \cdot)$.

REMARK 4.1. In the development of a numerical approach, while solving the system (9)-(10) for $x_{\gamma'}(z, z')$ and $\Theta_{\gamma'}(z, z')$, the quantities $\varphi_{\gamma'}(z, z', \cdot)$ and $\partial_z\varphi_{\gamma'}(z, z', y) - \mathrm{i}P(z, y, D_y)\,\varphi_{\gamma'}(z, z', \cdot)$ are to be computed in parallel using (23) and (24), see Section 5.

4.2 Superposition of scales

We reconsider the parametrix $F(z, z_0)$ (cf. (7)). Here, we are concerned with developing its action up to leading order, described by decomposing u_0 into wave packets and subjecting, per scale, these packet constituents to the rigid motion elaborated in the previous subsection,

DEFINITION 4.2. *If $u_{\gamma'} = \int \overline{\psi_{\gamma'}(x)}\, u(x)\,\mathrm{d}x$, we define $\mathbf{T}_k(z, z')\, u$ as*

$$(\mathbf{T}_k(z, z')u)(y) = \sum_{\gamma':\ k'=k} u_{\gamma'}\varphi_{\gamma'}(z, z', y), \quad z' \in [z_0, z], \tag{25}$$

using the Hamiltonian p_k. Furthermore, $\mathbf{T}(z, z')u = \sum_{k=0}^{\infty} \mathbf{T}_k(z, z')u$.

Note that $\mathbf{T}(z, z) = \mathbf{I}$, the identity operator. We observe that $\mathbf{T}_k(z, z')$ is localized to wavenumbers of size $\|\xi\| \approx 2^k$. The higher order contributions to the parametrix are described by 'packet-packet' interaction developed in the next section.

In preparation of the further analysis, we consider the *matrix representations* of $\mathbf{T}(z, z')$ with respect to the frame of curvelets. We introduce the elements

$$\mathcal{T}(z, z')_{\gamma\gamma'} = \int \overline{\psi_\gamma(y)}\varphi_{\gamma'}(z, z', y)\,\mathrm{d}y\,.$$

To characterize the matrices, we recall the pseudodistance function on $S^*(\mathbb{R}^n)$ introduced in [28, Definition 2.1], which is given by

$$d(x, \nu; x', \nu') = |\langle \nu, x - x'\rangle| + |\langle \nu', x - x'\rangle| + \min\{\|x - x'\|, \|x - x'\|^2\} + \|\nu - \nu'\|^2.$$

If $\gamma = (x, \nu, k)$ and $\gamma' = (x', \nu', k')$, let $\mathcal{D}(\gamma, \gamma') = \left(1 + \dfrac{d(x, \nu; x', \nu')}{2^{-k} + 2^{-k'}}\right)$. A weight function $\mu_\delta(\gamma, \gamma')$ is then introduced as

$$\mu_\delta(\gamma, \gamma') = (1 + |k' - k|^2)^{-1} 2^{-(\delta + \frac{1}{2}n)|k' - k|}\mathcal{D}(\gamma, \gamma')^{-n-\delta}. \tag{26}$$

We use the following matrix spaces [29, Definitions 2.6-2.8]. Let χ be a mapping on $S^*(\mathbb{R}^n)$, and $\chi(\gamma') = (\chi(x', \nu'), k')$. The matrix $A_{\gamma\gamma'}$ belongs to the class $\mathcal{M}^r_\delta(\chi)$ if

$$|A_{\gamma\gamma'}| \le C_A 2^{rk'}\mu_\delta(\gamma, \chi(\gamma')). \tag{27}$$

Furthermore, $\mathcal{M}^r(\chi) = \cap_{\delta>0}\mathcal{M}^r_\delta(\chi)$. An operator A belongs to the class $\mathcal{I}^r(\chi)$, if its matrix $A_{\gamma\gamma'} = \int \overline{\psi_\gamma(y)}(A\varphi_{\gamma'})(y)\,\mathrm{d}y$ belongs to $\mathcal{M}^r(\chi)$.

There is a natural assignation of an operator to each matrix, but different matrices can lead to the same operator due to the redundancy of the curvelet frame. In particular, the matrix of the identity operator is not the identity matrix, but is easily seen (e.g. [29, Lemma 2.9]) to belong to the class $\mathcal{M}^0(I)$. As a result, any matrix in $\mathcal{M}^r(\chi)$ determines an operator in $\mathcal{I}^r(\chi)$. In this context, we note that according to Definition 4.2, $\mathbf{T}(z,z') = V\,\mathcal{T}(z,z')U$. On the other hand, the elements $\mathbf{T}(z,z')_{\gamma\gamma'} = \langle \mathbf{T}(z,z')\varphi_{\gamma'}|\psi_\gamma\rangle$, are obtained by

$$U\,\mathbf{T}(z,z')V = UV\,\mathcal{T}(z,z')UV;$$

but $UV \in \mathcal{M}^0(I)$ in view of Lemma 3.2.

By [29, Theorem 2.7], $\mathcal{M}^{r_1}(\chi_1)\circ\mathcal{M}^{r_2}(\chi_1) \subseteq \mathcal{M}^{r_1+r_2}(\chi_1\chi_2)$, where $\circ$ denotes matrix composition, and the χ_j are assumed to (approximately) preserve the distance function. Consequently, $\mathcal{I}^{r_1}(\chi_1)\circ\mathcal{I}^{r_2}(\chi_2) \subseteq \mathcal{I}^{r_1+r_2}(\chi_1\chi_2)$. The assumption holds for $\chi = \chi_{z,z'}$, defined by the Hamiltonian flow of a C^2 symbol [29, Lemma 2.2], see (9) with initial conditions at z' derived from γ'. We note that $\chi_{z,z''}\chi_{z'',z'} = \chi_{z,z'}$.

By [29, Theorem 3.2],

$$\mathbf{T}(z,z') \in \mathcal{I}^0(\chi_{z,z'})\,. \tag{28}$$

The results of the previous section yield that also (see [29, Theorem 3.2])

$$\sum_{k=0}^{\infty}[\partial_z - \mathrm{i}P_k(z,x,D_x)]\mathbf{T}_k(z,z') \in \mathcal{I}^0(\chi_{z,z'})\,. \tag{29}$$

Both (28) and (29) are true whether we define $\chi_{z,z'}$ at scale k using the Hamiltonian p or its smooth approximation p_k, since by [29, Lemma 3.6],

$$d(\chi_{z,z'}(x,\nu);(\chi_k)_{z,z'}(x,\nu)) \le C\,2^{-k}\,,$$

uniformly for x, ν, k, and $z_0 \le z, z' \le Z$.

If we set

$$\overline{d}(\gamma;\gamma') = 2^{-\min(k,k')} + d(x,\nu;x',\nu')\,, \tag{30}$$

where $\overline{d}(\gamma;\gamma')$ is obtained by averaging d over the supports of φ_γ and $\varphi_{\gamma'}$, the decay estimates for $\mathbf{T}(z,z')_{\gamma\gamma'}$ attain the form

$$\left|\,\mathbf{T}(z,z')_{\gamma\gamma'}\right| \le C_N\,2^{-N|k-k'|}\left(2^{\min(k,k')}\overline{d}(\gamma;\chi_{z,z'}(\gamma'))\right)^{-N} \tag{31}$$

for all $N > 0$. This implies that the matrix ℓ^p norm of $\mathbf{T}(z,z')_{\gamma\gamma'}$ is bounded for each $p > 0$. Also, (31) implies the concentration of wave packets under the approximate solution operator. Using the microlocal properties of the curvelet transform, one can select coefficients $u_{0,\gamma'}$ with indices γ' close to the wavefront set of u_0; with appropriate thresholding one can obtain a sparse representation of u_0 in terms of curvelets. The sparseness in representation is preserved under $\mathbf{T}(z,z_0)$ in accordance with (31).

5 Multi-scale approach to solving the evolution equation

5.1 A Volterra equation and scattering series

First let us assume that $\mathbf{T}(z,z')$ denotes a family of operators $H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ that are bounded for a given r, such that $[\partial_z - \mathrm{i}P(z,x,D_x)]\mathbf{T}(z,z')$ is a bounded operator, and $\mathbf{T}(z,z) = \mathbf{I}$ for all z. Then the solution u to Cauchy initial value problem (7) can be posed in the form

$$u(z,x) = (\mathbf{T}(z,z_0)u_0)(x) + R(z,x), \tag{32}$$

where

$$R(z,x) = \int_{z_0}^{z}(\mathbf{T}(z,z')G(z',\cdot))(x)\,\mathrm{d}z', \tag{33}$$

with G denoting a 'residual' forcing term (volume source density) yet to be determined. Here, R can be viewed as a scattering contribution to u that satisfies

$$[\partial_z - \mathrm{i}P(z,x,D_x)]R(z,x) \\ = G(z,x) + \int_{z_0}^{z} [\partial_z - \mathrm{i}P(z,x,D_x)](\mathbf{T}(z,z')G(z',\cdot))(x)\,\mathrm{d}z'. \tag{34}$$

The representation for u posed in (32) is a weak solution to the Cauchy initial value problem (7), provided that

$$G(z,x) + \int_{z_0}^{z} [\partial_z - \mathrm{i}P(z,x,D_x)](\mathbf{T}(z,z')G(z',\cdot))(x)\,\mathrm{d}z' = -[\partial_z - \mathrm{i}P(z,x,D_x)](\mathbf{T}(z,z_0)u_0)(x). \tag{35}$$

This equation has the form of a Volterra equation of the second kind, and determines G [3]. Let us set

$$\mathbf{T}(z,z')' = -[\partial_z - \mathrm{i}P(z,x,D_x)]\,\mathbf{T}(z,z')\,.$$

Following the symbol smoothing of the pseudodifferential operator, we decompose $\mathbf{T}(z,z')'$ into a sum of two terms

$$-\sum_{k=0}^{\infty}[\partial_z - \mathrm{i}P_k(z,x,D_x)]\mathbf{T}_k(z,z') + \mathrm{i}\sum_{k=0}^{\infty}[P(z,x,D_x) - P_k(z,x,D_x)]\mathbf{T}_k(z,z')\,. \tag{36}$$

In Subsection 4.2 we observed that $\mathbf{T}(z,z')$ is bounded as an operator $H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ for any r. Also, as noted in (29), the first sum in (36) defines an operator of the class $\mathcal{I}^0(\chi_{z,z'})$, and in particular is a bounded operator $H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ for any r.

As regards the second sum in (36), the operator $\sum_{k=0}^{\infty}[P(z,x,D_x) - P_k(z,x,D_x)]\psi_k(D_x)$ acts as an operator of order 0 on $H^r(\mathbb{R}^n)$ for $-1 \le r \le 2$ provided $s \ge 2$ (see Subsection 2.1). It follows easily that the second sum in (36) is bounded as an operator $H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ for $-1 \le r \le 2$. Thus,

THEOREM 5.1. *The operator $\mathbf{T}(z,z')'$ is a bounded operator $H^r(\mathbb{R}^n) \to H^r(\mathbb{R}^n)$ for $-1 \le r \le 2$.*

The norm of $\mathbf{T}(z,z')'$ is bounded by a constant $C(Z)$ if $z_0 \le z' \le z \le Z$. But then the Volterra equation (35) can be solved by recursion,

$$G(z,x) = \sum_{p=0}^{\infty} G^p(z,x), \tag{37}$$

in which

$$G^p(z^{p+1},x) = \int_{z_0}^{z^{p+1}} (\mathbf{T}(z^{p+1},z^p)'G^{p-1}(z^p,\cdot))(x)\,\mathrm{d}z^p\,, \quad p = 1,2,\dots\,,$$

$$G^0(z^1,x) = (\mathbf{T}(z^1,z_0)'u_0)(x)\,. \tag{38}$$

The series converges in $L^\infty([z_0,Z];H^r(\mathbb{R}^n))$ for $-1 \le r \le 2$, with norm dominated by $C(Z)\exp[ZC(Z)]\,\|u_0\|_{H^r(\mathbb{R}^n)}$.

With $z_0 \le z' \le z \le Z$, the map $(z,z') \mapsto \mathbf{T}(z,z')$ is strongly continuous. Furthermore, the map $(z,z') \mapsto \mathbf{T}(z,z')'v$ is continuous for any particular $v \in L^2(\mathbb{R}^n)$. We notice that $\mathbf{T}(z,z')$ does not satisfy the semi-group property. Still the solution to the Volterra equation admits a *step-by-step* approach, namely

$$G(z+\Delta,x) - \int_{z}^{z+\Delta} (\mathbf{T}(z+\Delta,z')'G(z',\cdot))(x)\,\mathrm{d}z' = (\mathbf{T}(z+\Delta,z)'u(z,\cdot))(x) \tag{39}$$

followed by

$$u(z+\Delta,x) = (\mathbf{T}(z+\Delta,z)u(z,\cdot))(x) + \int_{z}^{z+\Delta} (\mathbf{T}(z+\Delta,z')G(z',\cdot))(x)\,\mathrm{d}z'. \tag{40}$$

[3] With reference to imaging mentioned as an application in the introduction, $\mathbf{T}(z,z_0)$ replaces the notion of 'parsimoneous migration' [2], or local 'plane-wave migration' [26].

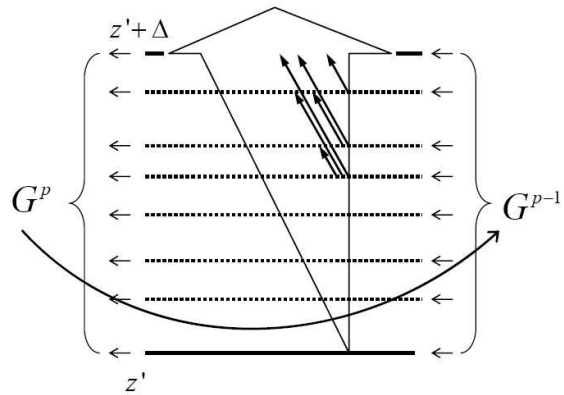

Figure 2: Recursion for the residual force in the step-by-step approach. The dotted lines indicate discretization (sampling) in z.

Having evaluated $u(z,x)$, solving (39) for G in the interval $(z, z+\Delta]$ by recursion (38), we evaluate (40) to obtain $u(z+\Delta, x)$, etc. The recursion (38) applied to (39) is illustrated in figure 2. The step-by-step approach can be exploited in a way similar to the use of Trotter products in solving the evolution equation in the smooth symbol case [15, 25] [4].

Equation (35) can be directly compared with the integral equation defining the generalized Bremmer coupling series [14]: The wave matrix in the Bremmer series becomes G, and the propagator becomes $\mathbf{T}(z,z_0)'$. The Bremmer series describes the scattering between 'up' and 'down' going wave constituents; here, the wave constituents become wave packets or curvelets. We elaborate this aspect in Subsection 6.3.

5.2 Sparsity of Volterra kernel matrix and concentration of wave packets

As mentioned above Theorem 5.1, the term $\sum_{l=0}^{\infty}[\partial_z - \mathrm{i}P_l(z,x,D_x)]\mathbf{T}_l(z,z')$, cf. (36), belongs to the class $\mathcal{I}^0(\chi_{z,z'})$, which yields rapidly decreasing decay estimates on its matrix coefficients,

$$\Big|\Big(\sum_{l=0}^{\infty}[\partial_z - \mathrm{i}P_l(z,x,D_x)]\,\mathbf{T}_l(z,z')\Big)_{\gamma\gamma'}\Big| \le C_N\, 2^{-N|k-k'|}\,\big(2^{\min(k,k')}\overline{d}(\gamma;\gamma')\big)^{-N} \tag{41}$$

for all $N>0$.

The matrix coefficients of the second term, $\mathrm{i}\sum_{l=0}^{\infty}[P(z,x,D_x)-P_l(z,x,D_x)]\mathbf{T}_l(z,z')$, satisfy only a finite rate of decay condition. For s small (including $s=2$), this term-wise decay rate is not sufficiently fast to yield good bounds on the operator. We instead state the decay estimates on this operator in terms of its mapping properties on function spaces defined by a weighted ℓ^2 condition on curvelet coefficients. These will directly yield sparsity conditions on the matrix, i.e. ℓ^p bounds on the columns and rows with $p<2$.

DEFINITION 5.2. *Let $\gamma_0=(x_0,\nu_0,k_0)$ be a triple. We define the space $H^{\sigma,\alpha}_{\gamma_0}$ by the norm*

$$\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} = \sum_{\gamma}\big|\,2^{k\sigma}\,2^{|k-k_0|\alpha}\,\big(2^{\min(k,k_0)}\,\overline{d}(\gamma;\gamma_0)\big)^{\alpha}\,f_\gamma\,\big|^2\,,$$

[4] The exploitation of the step-by-step method is in part motivated by the so-called wavefront construction method applied to solving the Hamilton system for bicharacteristics such as (8).

where $\gamma = (x, \nu, k)$*, and*

$$f_\gamma = \int \overline{\psi_\gamma(y)}\, f(y)\, \mathrm{d}y\,.$$

REMARK 5.3. The choice of a particular curvelet frame in the definition of $H^{\sigma,\alpha}_{\gamma_0}$ changes the norm only by a bounded factor. In particular, $H^{\sigma,\alpha}_{\gamma_0}$ is independent of the particular curvelet frame chosen. This can be seen, for example, using condition (42) below.

LEMMA 5.4. *Suppose that* $f_{\gamma'}$ *vanishes unless* $|k'-k| \le 4$ *and* $\|\nu' - \nu_0\| \approx \theta$*, where* $\theta \ge 2^{-\frac{1}{2}\min(k,k_0)}$*. Then*

$$\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} \approx 2^{2k\sigma + 2\max(k,k_0)\alpha} \int \big(|\langle \nu_0, y - x_0\rangle| + \min(\|y-x_0\|, \|y-x_0\|^2) + \theta^2\big)^{2\alpha} |f(y)|^2\, \mathrm{d}y\,. \tag{42}$$

Under the condition $f_{\gamma'} = 0$ *unless* $|k'-k| \le 4$*, we have*

$$\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} \lesssim 2^{2k\sigma + 2\max(k,k_0)\alpha} \int \big(1 + \|y-x_0\|\big)^{2\alpha} |f(y)|^2\, \mathrm{d}y \tag{43}$$

and

$$\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} \gtrsim 2^{2k\sigma + 2|k-k_0|\alpha} \int \big(1 + \|y-x_0\|\big)^{2\alpha} |f(y)|^2\, \mathrm{d}y\,. \tag{44}$$

Proof. The sum over k' is finite, so we consider terms γ' with $k' = k$. Then

$$\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} \approx 4^{k\sigma + \max(k,k_0)\alpha} \sum_{\nu'} \sum_{x'} \big(|\langle \nu_0, x' - x_0\rangle| + \min(\|x'-x_0\|, \|x'-x_0\|^2) + \theta^2\big)^{2\alpha} |f_{\gamma'}|^2\,.$$

By the spatial localization of curvelets, this in turn is comparable to

$$4^{k\sigma + \max(k,k_0)\alpha} \sum_{\nu'} \int \big(|\langle \nu_0, y - x_0\rangle| + \min(\|y-x_0\|, \|y-x_0\|^2) + \theta^2\big)^{2\alpha} |\widehat{\chi}_{\nu',k}(D_y) f(y)|^2 \mathrm{d}y\,.$$

For $\|\nu' - \nu\| \approx \theta$, the convolution kernel $\chi_{\nu',k}(y)$ is rapidly decreasing at a rate on which the weight function is slowly varying in y, hence this in turn is comparable to

$$4^{k\sigma + \max(k,k_0)\alpha} \int \big(|\langle \nu_0, y - x_0\rangle| + \min(\|y-x_0\|, \|y-x_0\|^2) + \theta^2\big)^{2\alpha} |f(y)|^2 \mathrm{d}y\,.$$

The inequalities (44) and (43) follow by adding the above bound over dyadic values of θ. □

THEOREM 5.5. *Let* $P \in C^s S^1_{1,0}$*, where* $s \ge 2$*. Then for* $0 \le \alpha < \frac{s}{2}$*,* $|\sigma| \le \frac{s}{2} - 1$*, and all* γ_0*, we have*

$$\Big\| \sum_{l=0}^{\infty} [P(z,x,D_x) - P_l(z,x,D_x)] \mathbf{T}_l(z,z') f \Big\|_{H^{\sigma,\alpha}_{\chi_{z,z'}(\gamma_0)}} \le C \|f\|_{H^{\sigma,\alpha}_{\gamma_0}}\,.$$

Proof. The action of $\mathbf{T}_\ell(z,z')$ is essentially a one-to-one map of γ' to $\chi_{z,z'}(\gamma')$, and since $\chi_{z,z'}$ respects the distance function (and preserves k') it suffices to consider the case $z = z'$, in which case we denote $\Delta_l = \mathbf{T}_l(z,z)$, the operator which selects the coefficients $f_{\gamma'}$ at $k' = l$. We are thus considering boundedness in $H^{\sigma,\alpha}_{\gamma_0}$ of the operator

$$P^\flat = \sum_{l=0}^{\infty} \big[P(z,x,D_x) - P_l(z,x,D_x)\big] \Delta_l\,.$$

As in the proof of Lemma 2.1.G of Taylor [33] we may write $P(z,y,\xi)$ as a rapidly converging sum of terms of the form $a(z,y)b(\xi)$, where $a(z,\cdot) \in C^s(\mathbb{R}^n)$ and $b(\xi)$ is a symbol of type S^1_1. We can thus reduce to the case that P

is an elementary symbol of the form $a(z,y)b(D_y)$, and since z is a harmless parameter we ignore it. Against Δ_l, the action of $b(D_y)$ is essentially multiplication by 2^l, and it suffices to consider

$$\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} 2^l \Delta_k\big(a(y)-a_l(y)\big)\Delta_l\,.$$

We consider the various (k,l) terms separately.
Case 1: $l \geq k+4$**.** By frequency separation, we have

$$\Delta_k\big(a(y)-a_l(y)\big)\Delta_l = \Delta_k\big(a(y)-a_{2l}(y)\big)\Delta_l\,.$$

Note that $|a(y)-a_{2l}(y)| \lesssim 2^{-ls}$. By (43) and (44), then

$$\begin{aligned}\|2^l\Delta_k\big(a(y)-a_{2l}(y)\big)\Delta_l f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} &\lesssim 4^{k\sigma+\max(k,k_0)\alpha+l(1-s)}\int\big(1+\|y-x_0\|\big)^{2\alpha}\,|\Delta_l f(y)|^2\,\mathrm{d}y\\ &\lesssim 4^{k\sigma+\max(k,k_0)\alpha+l(1-s)-l\sigma-|l-k_0|\alpha}\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}}\\ &\lesssim 4^{-l(\frac{s}{2}-\alpha)}\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}}\,.\end{aligned}$$

The exponent is strictly negative, so we may sum over (k,l) with $k \leq l$.
Case 2: $k \geq l+4$**.** We similarly derive the bound

$$\begin{aligned}\|2^l\Delta_k\big(a(y)-a_{2k}(y)\big)\Delta_l f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} &\lesssim 4^{k\sigma+\max(k,k_0)\alpha+l-ks-l\sigma-|l-k_0|\alpha}\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}}\\ &\lesssim 4^{-(k-l)-k(\frac{s}{2}-\alpha)}\|f\|^2_{H^{\sigma,\alpha}_{\gamma_0}}\end{aligned}$$

which is summable over (k,l) with $k \geq l$.
Case 3: $|k-l| \leq 4$**.** As representative, we consider $k=l$. Let

$$\theta_j = 2^{j-\frac{1}{2}\min(k,k_0)}\,,\qquad 0 \leq j \leq \frac{1}{2}\min(k,k_0)\,.$$

Let $\Delta_{k,j}$ denote the projection onto coefficients γ' for which $k'=k$ and $\|\nu'-\nu_0\| \in [\theta_j, 2\theta_j)$; in case of $\Delta_{k,0}$ we consider $\|\nu'-\nu_0\| \leq 2\theta_0$.

By orthogonality of the $\Delta_{k,i}$

$$\|2^k\Delta_k\big(a(y)-a_k(y)\big)\Delta_k f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} = \sum_i\Big(\sum_j 2^k\|\Delta_{k,i}\big(a(y)-a_k(y)\big)\Delta_{k,j}f\|_{H^{\sigma,\alpha}_{\gamma_0}}\Big)^2\,. \tag{45}$$

For $i \neq j$, the ranges of $\Delta_{k,i}$ and $\Delta_{k,j}$ are separated in frequency space by distance

$$2^k\max(\theta_i,\theta_j) \geq 2^{|i-j|+\frac{1}{2}k}\,.$$

Consequently,

$$2^k\|\Delta_{k,i}\big(a(y)-a_k(y)\big)\Delta_{k,j}f\|^2_{H^{\sigma,\alpha}_{\gamma_0}} = 2^k\|\Delta_{k,i}\big(a(y)-a_{k+2|i-j|}(y)\big)\Delta_{k,j}f\|^2_{H^{\sigma,\alpha}_{\gamma_0}}\,. \tag{46}$$

Note that

$$|a(y)-a_{k+2|i-j|}(y)| \lesssim 2^{-\frac{s}{2}k-s|i-j|} \leq 2^{-k-s|i-j|}\,.$$

Using this and (42), the term (46) is bounded by

$$2^{k\sigma+\max(k,k_0)\alpha-s|i-j|}\big\|\big(|\langle\nu_0,y-x_0\rangle|+\min(\|y-x_0\|,\|y-x_0\|^2)+\theta_i^2\big)^{\alpha}\Delta_{k,j}f\|_{L^2}$$

which, since $\theta_i \leq 2^{|i-j|}\theta_j$, is bounded by

$$\begin{aligned}&2^{k\sigma+\max(k,k_0)\alpha-(s-\alpha)|i-j|}\big\|\big(|\langle\nu_0,y-x_0\rangle|+\min(\|y-x_0\|,\|y-x_0\|^2)+\theta_j^2\big)^{\alpha}\Delta_{k,j}f\|_{L^2}\\ &\qquad\approx 2^{-(s-\alpha)|i-j|}\|\Delta_{k,j}f\|_{H^{\sigma,\alpha}_{\gamma_0}}\,.\end{aligned}$$

Since $\alpha < s$, we may sum over i and j in (45) and use orthogonality of the $\Delta_{k,j}$ to finish the proof. □

REMARK 5.6. The result of Theorem 5.5 also holds for the transposed operator. Indeed, only the proofs of cases 1 and 2 are affected by transposing the operator.

Combined with the estimates (41), this yields that

$$\|\mathbf{T}(z,z')'f\|_{H^{\sigma,\alpha}_{\chi_{z,z'}(\gamma')}} \leq C(Z)\,\|f\|_{H^{\sigma,\alpha}_{\gamma'}}\,, \tag{47}$$

with constant $C(Z)$ uniform over z and z' in $[z_0, Z]$, and where we abbreviate $\chi_z := \chi_{z,z_0}$. It follows that the recursion formula (37)–(38) converges, uniformly for each z', in $H^{\sigma,\alpha}_{\chi_{z',z_0}(\gamma_0)}$, and that

$$\|G\|_{L^\infty([z_0,Z];H^{\sigma,\alpha}_{\chi_{z,z_0}(\gamma_0)})} \leq C(Z)\exp[ZC(Z)]\,\|u_0\|_{H^{\sigma,\alpha}_{\gamma_0}}\,. \tag{48}$$

This also yields

COROLLARY 5.7. *The exact evolution operator $F(z,z_0)$ is a bounded map,*

$$\|F(z,z_0)u_0\|_{H^{\sigma,\alpha}_{\chi_{z,z_0}(\gamma_0)}} \leq C\,\|u_0\|_{H^{\sigma,\alpha}_{\gamma_0}}\,.$$

Taking $\sigma = 0$, $\gamma_0 = \gamma'$, and $u_0 = \varphi_{\gamma'}$, we obtain that

$$\|F(z,z_0)\varphi_{\gamma'}\|_{H^{0,\alpha}_{\chi_{z,z_0}(\gamma')}} \leq C\,,$$

or equivalently, for all $\alpha < s$,

$$\sup_{\gamma'}\sum_{\gamma} 2^{2|k-k'|\alpha}\left(2^{\min(k,k')}\,\overline{d}(\gamma;\chi_{z,z_0}(\gamma'))\right)^{2\alpha}|F(z,z_0)_{\gamma\gamma'}|^2 \leq C\,. \tag{49}$$

This shows that the curvelet coefficients of the exact evolution of $\varphi_{\gamma'}$ are concentrated near the flow of γ' along the Hamiltonian.

THEOREM 5.8. *The evolution matrix $F(z,z_0)_{\gamma\gamma'}$ satisfies the ℓ^p sparsity condition,*

$$\sup_{\gamma'}\sum_{\gamma}|F(z,z_0)_{\gamma\gamma'}|^p \leq C_p\,, \qquad \sup_{\gamma}\sum_{\gamma'}|F(z,z_0)_{\gamma\gamma'}|^p \leq C_p\,, \qquad p > \frac{2n}{s+n}\,.$$

Proof. We note that, by [29, (2.3)],

$$\sum_{k'=0}^{\infty}\sum_{\nu'}\sum_{x'} 2^{-q|k-k'|}\left(2^{\min(k,k')}\,\overline{d}(\gamma;\gamma')\right)^{-q} \leq C_q\,; \quad q > n\,.$$

Since (49) holds for all $\alpha < \frac{s}{2}$, it follows by Hölder's inequality that

$$\sup_{\gamma'}\sum_{\gamma}|F(z,z_0)_{\gamma\gamma'}|^p \leq C_p\,, \qquad p > \frac{2n}{s+n}\,.$$

The same bound also holds with γ and γ' interchanged. □

6 Discretization

In this section, we address the issue whether the analysis developed, and estimates proved, in the earlier sections can be used and exploited in numerical computation. Indeed, the analysis has an immediate counterpart in computation. We consider $n = 2$ for simplicity of presentation. Furthermore, we restrict ourselves to functions $u = u(z,x)$ which are compactly supported in x (in particular, we assume that $u(z_0,.)$ is compactly supported); again, for simplicity of presentation, we assume that $u(z,.)$ is supported on the disc $\mathcal{D} = \{x \in \mathbb{R}^2 \mid \|x\| < \pi\}$ for all $z \in [z_0, Z]$.

6.1 Frame, co-frame, transforms

We begin by studying the operator $U : u(x) \to (u_\gamma)$ defined by (15). As a point of departure, we take (19). A discretization of the relevant integration is obtained by introducing the approximation

$$\widetilde{u}_\gamma = \frac{1}{(2\pi)^2\sqrt{\rho_k}\sigma'_k\sigma''_k}\sum_l \widehat{u}(\xi_l)\widehat{\beta}_{\nu,k}(\xi_l)\exp[\mathrm{i}\langle x_j,\xi_l\rangle], \tag{50}$$

where the points $\xi_l = \xi_l^{\nu,k}$ are chosen on a lattice. Specifically, we let

$$\Xi^k = \left\{(l_1,l_2)\in\mathbb{Z}^2 \;\middle|\; -\frac{N'_k}{2}\le l_1 < \frac{N'_k}{2}, -\frac{N''_k}{2}\le l_2 < \frac{N''_k}{2}\right\}.$$

Points in this set are denoted by Ξ_l^k (in analogy with the notation X_j). We choose ξ_l (associated with the box $B_{\nu,k}$) as

$$\xi_l = \Theta_{\nu,k}^{-1}S_k^{-1}\Xi_l^k.$$

Here, the parameters $N_k = (N'_k, N''_k)$ are even natural numbers with $N'_k > L'_k$ and $N''_k > L''_k$, while $\sigma'_k = N'_k/L'_k$ and $\sigma''_k = N''_k/L''_k$ are the *oversampling* factors. The matrix S_k is defined as

$$S_k = \begin{pmatrix} \sigma'_k & 0 \\ 0 & \sigma''_k \end{pmatrix} = \begin{pmatrix} \dfrac{N'_k}{L'_k} & 0 \\ 0 & \dfrac{N''_k}{L''_k} \end{pmatrix}.$$

The dot product in the phase of the exponential in (50) then becomes

$$\langle x_j,\xi_l\rangle = (\Xi_l^k)^t S_k^{-1}D_k^{-1}X_j = 2\pi\left(\frac{j_1l_1}{N'_k}+\frac{j_2l_2}{N''_k}\right).$$

This specific choice of points ξ_l allows for a fast evaluation of $\widetilde{u}_\gamma = \widetilde{u}_{j,\nu,k}$ for $j\in\Xi^k$ by means of a two-dimensional fast Fourier transform (FFT).

We discuss the discrete approximation $u_\gamma \approx \widetilde{u}_\gamma$ in more detail. According to Shannon's sampling theorem, we can represent a function with compact support by means of sampling its Fourier transform on a properly chosen equally spaced grid. If $\mathrm{supp}(u)\subset\mathcal{D}$, it is sufficient to sample $\widehat{u}$ at integer points $\xi_j = j\in\mathbb{Z}^2$ to be able to reconstruct $u(x)$. While, in this case, $u(x)$ in (19) has support in $\mathcal{D}$, the convolution $u * \beta_{\nu,k}$ no longer has compact support. However, the fast decay of $\beta_{\nu,k}$ is inherited by $u*\beta_{\nu,k}$. Hence, u_γ will be very small for large x_j's, and in practice it will be sufficient to keep and use only the lattice points located in a neighborhood of $\mathcal{D}$ for representation (16) to be accurate.

The error in the approximation $u_\gamma \approx \widetilde{u}_\gamma$ can be estimated by the numerical integration error associated with the trapezoidal rule [13]. We obtain: Given ν, k, for every p there exists a constant C_p such that

$$|u_\gamma - \widetilde{u}_\gamma| < C_pN^{-p}, \quad N = \min(N'_k, N''_k).$$

Hence, the approximation error made in (50) decays fast subject to the condition that $N'_k > L'_k$ and $N''_k > L''_k$. The oversampling factors will need to be comparatively large for small k but decay as k grows due to the parabolic scaling property of the support of $\widehat{\beta}_{\nu,k}$. Note that the oversampling factors are related to the number of points (x_j) where u_γ is evaluated, but not to their density.

Next, we consider the inverse transform, $V : u_\gamma \to u$, defined by (16). Fourier transforming (16) yields

$$\widehat{u}(\xi) = \sum_\gamma \frac{1}{\sqrt{\rho_k}}u_\gamma\widehat{\chi}_{\nu,k}(\xi)\exp[-\mathrm{i}\langle x_j,\xi\rangle] = \sum_{\nu,k}\frac{1}{\sqrt{\rho_k}}\widehat{\chi}_{\nu,k}(\xi)\sum_j u_{j,\nu,k}\exp[-\mathrm{i}\langle x_j,\xi\rangle]. \tag{51}$$

Thus

$$\begin{aligned} u(x) &= \sum_{\nu,k}\frac{1}{(2\pi)^2\sqrt{\rho_k}}\int\widehat{\chi}_{\nu,k}(\xi)\left(\sum_j u_{j,\nu,k}\exp[-\mathrm{i}\langle x_j,\xi\rangle]\right)\exp[\mathrm{i}\langle x,\xi\rangle]\,\mathrm{d}\xi \\ &\approx \sum_{\nu,k}\frac{1}{(2\pi)^2\sqrt{\rho_k}}\int\widehat{\chi}_{\nu,k}(\xi)\left(\sum_{j\in\Xi^k}\widetilde{u}_{j,\nu,k}\exp[-\mathrm{i}\langle x_j,\xi\rangle]\right)\exp[\mathrm{i}\langle x,\xi\rangle]\,\mathrm{d}\xi. \end{aligned} \tag{52}$$

Here, we discretize the integral using the same arguments as before:

$$u(x) \approx \sum_{\nu,k} \frac{1}{(2\pi)^2\sqrt{\rho_k}\sigma'_k\sigma''_k} \sum_{l\in\Xi^k} \widehat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \left(\sum_{j\in\Xi^k} \widetilde{u}_{j,\nu,k} \exp[-\mathrm{i}\langle x_j, \xi_l^{\nu,k}\rangle] \right) \exp[\mathrm{i}\langle x, \xi_l^{\nu,k}\rangle].$$

Using (50) in combination with the fact that the sum in between the parenthesis above forms a discrete Fourier transform, we obtain

$$\begin{aligned} u(x) &\approx \sum_{\nu,k} \frac{N'_k N''_k}{(2\pi)^4 \rho_k (\sigma'_k\sigma''_k)^2} \sum_{l\in\Xi^k} \widehat{u}(\xi_l^{\nu,k})\, \widehat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \widehat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[\mathrm{i}\langle x, \xi_l^{\nu,k}\rangle] \\ &= \frac{1}{(2\pi)^2} \sum_{\nu,k} \sum_{l\in\Xi^k} \widehat{u}(\xi_l^{\nu,k})\, \frac{\widehat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \widehat{\beta}_{\nu,k}(\xi_l^{\nu,k})}{\sigma'_k\sigma''_k} \exp[\mathrm{i}\langle x, \xi_l^{\nu,k}\rangle], \end{aligned} \tag{53}$$

which can be viewed as a quadrature rule for computing inverse Fourier transforms. As before, for sufficiently regular functions $\widehat{u}$, the approximation can be made arbitrarily accurate.

The quadrature rule referred to above, requires the evaluation of $\widehat{u}$ at unequally spaced points $\xi_l^{\nu,k}$. It is thus natural to invoke algorithms for unequally spaced Fourier transforms (USFFT) [17, 1]. These USFFT algorithms allow for freedom (not limited to, for example, a polar-coordinate grid) in the choice of discretization points $\xi_l^{\nu,k}$.

6.2 Construction of window functions

We discuss how one numerically constructs window functions that define a proper frame/co-frame pair. We make use of some components of the construction of discrete curvelets [10]. We assume that the frame is tight so that $\widehat{\beta}_{\nu,k} = \widehat{\chi}_{\nu,k}$ ($k = 0, 1, 2, \ldots$). We consider $n = 2$, and adopt polar coordinates (r, ϕ), whence $\widehat{\chi}_{\nu,k}(\xi) = w(2^{-k}r)\, v_k(r, \phi - \nu)$, where w and v_k are specified below.

For $w(r)$ we follow the construction of Meyer wavelets:

$$w(r) = \begin{cases} \sin(\frac{\pi}{2} a_n(2r-1)), & \text{if } \frac{1}{2} \le r < 1\,; \\ \cos(\frac{\pi}{2} a_n(\ r-1)), & \text{if } 1 \le r \le 2\,; \\ 0, & \text{otherwise}\,. \end{cases}$$

Here, $a_n(r)$ is defined as

$$a_n(r) = \begin{cases} 0, & r < 0; \\ p_n(r), & 0 \le r \le 1; \\ 1, & r > 1, \end{cases}$$

where p_n is the polynomial of degree $2n+1$ that satisfies

$$\begin{aligned} &p_n(0) = p'_n(0) = \cdots = p_n^{(n)}(0) = 0, \\ &p_n(1) = 1, \quad p'_n(1) = \cdots = p_n^{(n)}(1) = 0. \end{aligned}$$

It is readily verified that $w \in C_0^{n+1}$. For the illustrations in this paper we take $n = 10$, when

$$\begin{aligned} p_{10}(t) = t^{11}\Big(&184756t^{10} - 1939938t^9 + 9189180t^8 - 25865840t^7 \\ &+ 47927880t^6 - 61108047t^5 + 54318264t^4 - 33256080t^3 \\ &+ 13430340t^2 - 3233230t + 352716\Big). \end{aligned}$$

This degree appears to be sufficient to display the appropriate decay properties within computational precision.

Accompanying w is the coarsest scale ($k = 0$) function w_0, given by

$$w_0(r) = \cos\left(\ \frac{\pi}{2} a_n(2r-1)\ \right),$$

which is used to define $\widehat{\chi}_0(\xi) = \widehat{\beta}_0(\xi) = w_0(r)$.

Furthermore, let N_ν denote the number of orientations, ν, at the scale $k = 1$, and let

$$\kappa_k(r,\phi) = \cos\left(\frac{\pi}{2} a_n\left(\frac{r}{2^{k+1}}\frac{N_\nu|\phi|\lfloor 2^{k/2}\rfloor}{2\pi}\right)\right).$$

We define

$$v_k(r,\phi) = \frac{\kappa_k(r,\phi)}{\sqrt{\sum_{\nu'}\kappa_k(r,\phi-\nu')^2}}.$$

We note the particular r dependence in $\kappa_k(r,\phi)$. This dependence is designed to make the supports of $\widehat{\chi}_{\nu,k}$ and $\widehat{\beta}_{\nu,k}$ 'fill out' the boxes $B_{\nu,k}$. As a consequence, the decay in estimate (13) is modified from $(2^k|\langle\nu,x\rangle| + 2^{k/2}|x|)^{-N}$ to $(2^k|\langle\nu,x\rangle| + 4\cdot 2^{k/2}|x|)^{-N}$. This effect, in practice, is significant in as much as it decreases the required oversampling factors σ_k''. However, the angular overlap between the windows (but not the total number of discretization points) increases. It is readily verified that $v_k \in C_0^{n+1}$, and that the resulting $\widehat{\chi}_{\nu,k}$ satisfy (11) for $j, |\alpha| \le n+1$.

The rectangles B_k associated with the construction outlined here are given by

$$B_k = \left[2^{k-1}\cos\left(\frac{8\pi}{N_\nu\lfloor 2^{k/2}\rfloor}\right), 2^{k+1}\right] \times \left[-2^{k+1}\sin\left(\frac{2\pi}{N_\nu\lfloor 2^{k/2}\rfloor}\right), 2^{k+1}\sin\left(\frac{2\pi}{N_\nu\lfloor 2^{k/2}\rfloor}\right)\right],$$

which satisfy the parabolic scaling conditions. Due to its construction in polar coordinates, the support of $v_k(r,\phi)$ in the ξ'-direction will be slightly larger than the interval $[2^{k-1}, 2^{k+1}]$.

In figure 3 we show $\varphi_0(x)$, while in figure 4 we show $\varphi_{\nu,k}(x)$ (cf. (12)) for $k = 2, 4$, as well as part of the associated lattices $\{x_j\}$ (red dots). For the purpose of comparison, we have chosen σ_k' and σ_k'' to be the same for all k considered. Indeed, numerically, we obtain estimates of the type (13). We zoom into the white boxes and follow the coordinate axes: Inserts A and B show (the absolute value of) $\varphi_{\nu,k}(x)$ along the horizontal and vertical axis, respectively. Inserts C and D show the decays (logarithmically) along the horizontal and vertical axis, respectively; the decays are near linear as they should. (The decays are super-algebraic.) We note that the change in slopes between inserts C follows the scaling 2^k, while the change in slopes between inserts D follows the scaling $2^{k/2}$ as it should.

6.3 Approximate solutions and preparation of the kernel of the Volterra equation

To illustrate the approximate solution $\mathbf{T}_k(z,z')$ in (25), we consider the following choice of evolution equation:

$$[\partial_t - \mathrm{i}P(x,D_x)]u = 0, \quad P(x,D_x) = \sqrt{\sum_{i=1}^{2} D_{x_i}\, c(x)^2 D_{x_i}}.$$

Thus z becomes time t and we consider the case $t' = t_0 = 0$. We include the formation of a caustic by constructing a model for $c(x)$ with a low wavespeed lens, see figure 5.

In figure 5 we illustrate solutions to the Hamilton system (8) reduced to (9). We choose $k = 4$. We take 100 values for (x_0, ν_0) corresponding with the decomposition into curvelets of the local plane wave used as initial condition u_0 in figure 6 (top, left); we show 3 values of t to illustrate how the canonical relation can be built. The bicharacteristic indicated by a vertical arrow corresponds with the flow (and location and orientation of the curvelet) used in figure 7.

The initial condition, u_0, is built around one scale only, viz. $k = 4$. The decomposition into curvelets (16) used, only involves this scale. We then use (22) to compute (25). As mentioned before, we set $t' = t_0 = 0$ and plot the outcome (as a function of y) in figure 6, for 4 different values of t (including $t = t_0 = 0$ in the top, left), using the Hamilton flow depicted in figure 5. We ensured that for the final time, the curvelets sufficiently (in the sense of the lattices) sample the wavefield. The computation involved 100 curvelets.

Finally, we verify numerically how well $\mathbf{T}_k(z,z')$ solves the evolution equation, that is, we compute the matrix

$$M_{\gamma\gamma'} = \Big(\big[\partial_t - \mathrm{i}P_l(x,D_x)\big]\mathbf{T}_l(t,t')\Big)_{\gamma\gamma'}$$

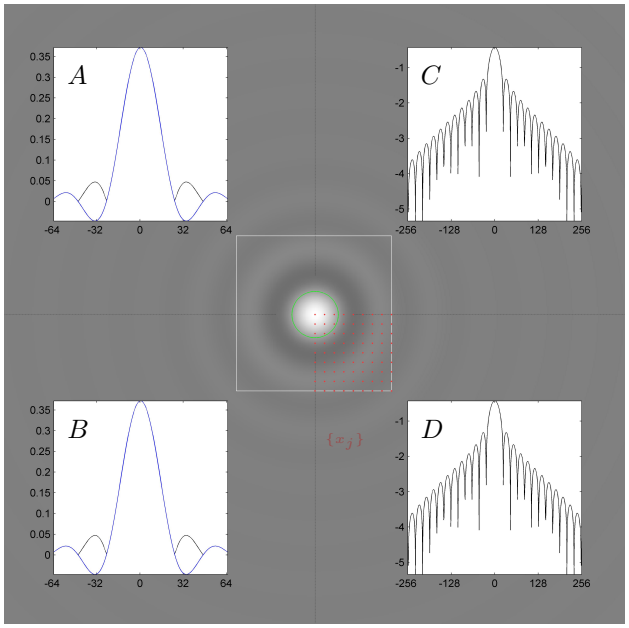

Figure 3: The coarsest scale curvelet, φ_0. Its real part is plotted in gray scale. Panels A and B display φ_0 (its real part, in blue), and $|\varphi_0|$, along the coordinate axes within the white box: A along the 'horizontal' axis, and B along the 'vertical' axis. Panels C and D display $\log_{10}|\varphi_0|$ along the coordinate axes in the entire image: C along the 'horizontal' axis, and D along the 'vertical' axis.

appearing in (41). (This matrix plays an integral part in setting up the kernel of the Volterra equation.) We use (24) to compute the action of P_l and (23) to compute ∂_t, as in a pseudospectral method. (In (41) we only need to account for terms satisfying $|k' - l| \leq 2$.) We restrict ourselves here to the case $l = k'$, k' denoting the scale in γ', that is, $k' = 4$.

We take for γ' the index associated with the curvelet centered at the vertical arrow in figure 5, and consider the case t' being the initial time ($t_0 = 0$) and t being the second time used in figures 5 and 6. We write $\gamma = (x_j, \nu, k)$. We illustrate the absolute values of matrix elements, $M_{\gamma\gamma'}$, in $\log_{10}$-scale, in figure 7: The 'horizontal' planes are spanned by the x_j; ν changes in the 'vertical' direction. We show two values of k: $k = 4$ (top) and $k = 5$ (bottom). We recover, numerically, the decay estimates (near linear in logarithmic scale) in (41).

References

[1] G. Beylkin, On the fast Fourier transform of functions with singularities, Applied and Computational Harmonic Analysis 2 (1995) 363-381.

[2] B. Hua, G.A. McMechan, Parsimonious 2D prestack Kirchhoff depth migration, Geophysics 68 (2003) 1043-1051.

[3] J.N. Bony, Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires, Ann. Scient. E.N.S. 14 (1981) 209-246.

[4] L. Boutet de Monvel, Hypoelliptic operators with double characteristics and related pseudodifferential operators, Comm. Pure Appl. Math. 27 (1974) 585-639.

[5] J. Bros, D. Iagolnitzer, Support essentiel et structure analytique des distributions, Séminaire Goulaouic-Lions-Schwartz, exp. no. 19 (1975-1976).

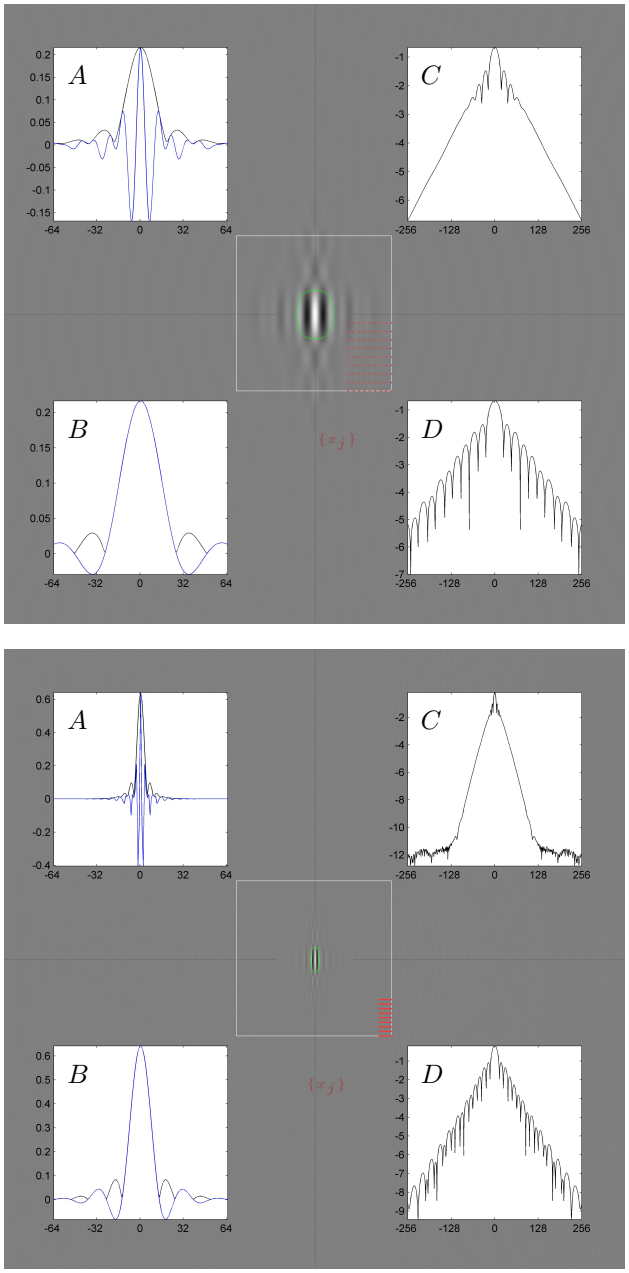

Figure 4: Functions $\varphi_{\nu,k}$ for $k = 2$ (top) and $k = 4$ (bottom). Their real part are plotted in gray scale. Panels A and B display the real parts of $\varphi_{\nu,k}$ (in blue), and $|\varphi_{\nu,k}|$, along the coordinate axes within the white box: A along the 'horizontal' axis, and B along the 'vertical' axis. Panels C and D display $\log_{10} |\varphi_{\nu,k}|$ along the coordinate axes in the entire image: C along the 'horizontal' axis, and D along the 'vertical' axis.

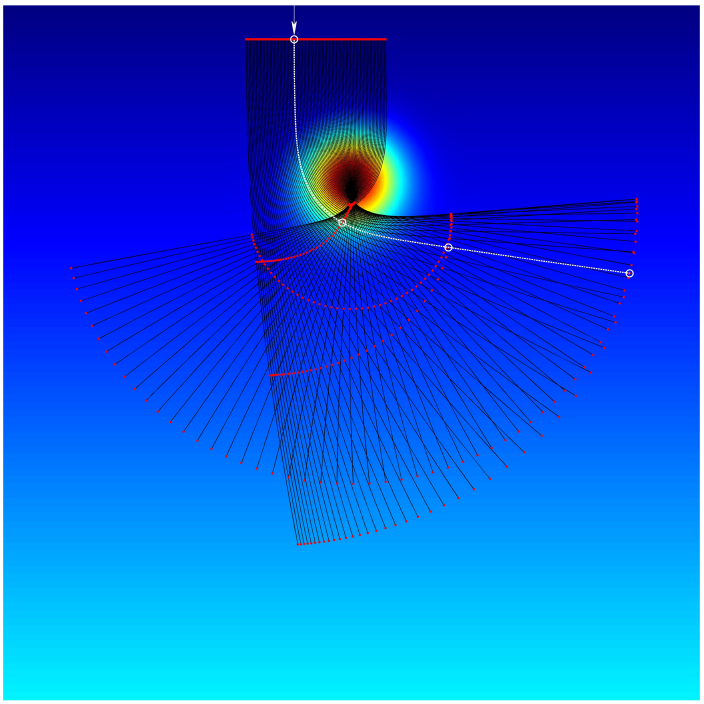

Figure 5: Model $c(x)$ (including a circular lens) and Hamilton flow. The rays are generated by initial conditions corresponding with the curvelets appearing in the decomposition of the initial condition used in the example shown in figure 6. The red dots show positions along the rays at 4 time instances (including the initial time $t_0 = 0$). The white ray corresponds with the flow used to compute figure 7.

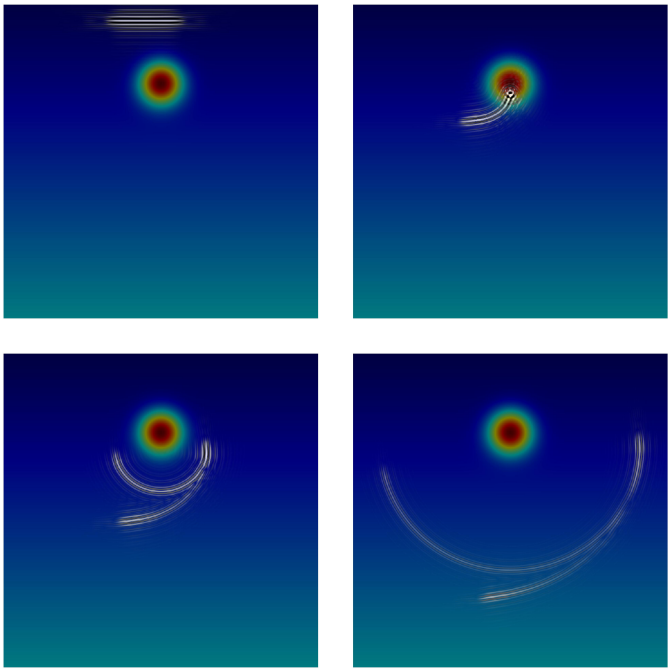

Figure 6: Snapshots of the approximate solution $\mathbf{T}_k(t, t_0 = 0)u_0$. Top, left: Initial condition u_0. The approximate solutions in the 4 panels correspond to the 4 times defining the dotted fronts in figure 5.

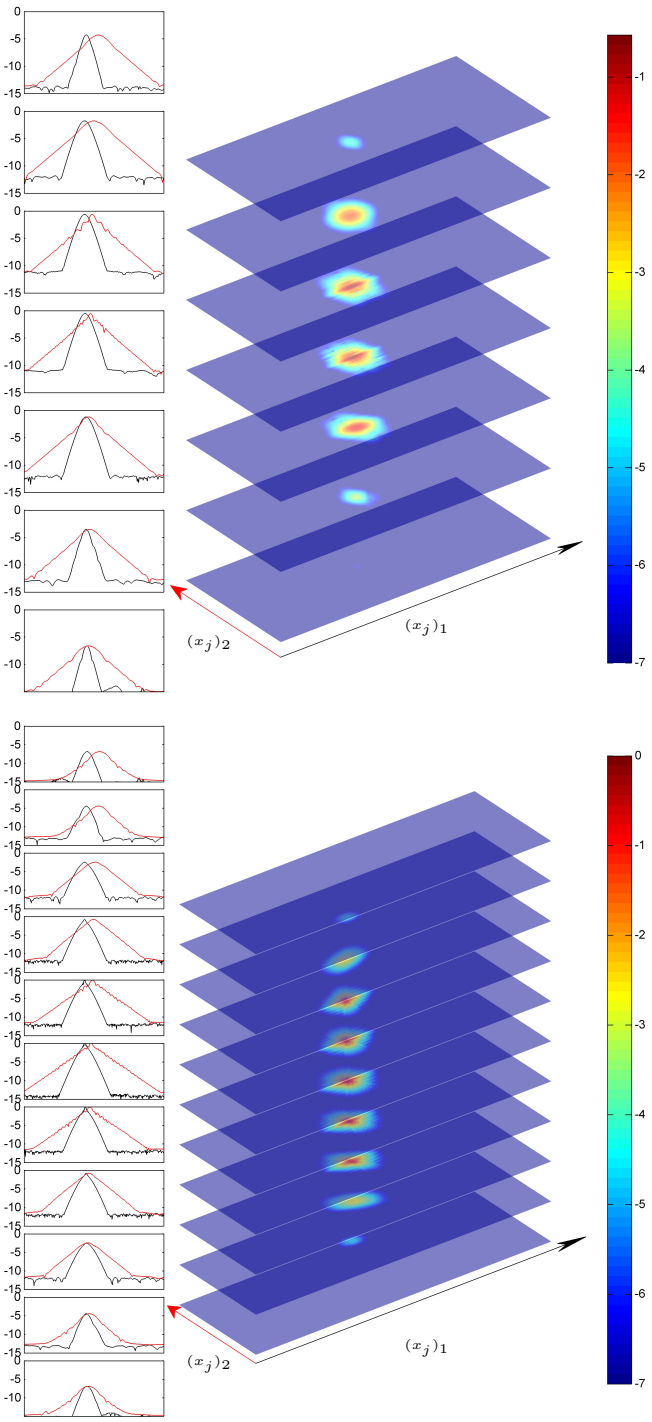

Figure 7: Illustration of a column of the matrix $M_{\gamma\gamma'}$, for $k' = 4$ and $(x_{j'}, \nu')$ corresponding with the initial conditions of the white ray in figure 5. Plotted is $\log_{10} |M_{\gamma\gamma'}|$. Top: $k = 4$; bottom: $k = 5$. The 'horizontal' planes are spanned by the x_j; ν changes in the 'vertical' direction (the ones shown contain an element greater than 10^{-7}). The diagrams on the left are cross sections along the coordinate axes of each plane; red curves are along the $(x_j)_2$-axes and the black curves are along the $(x_j)_1$-axes.

[6] E.J. Candès, D. Donoho, Continuous curvelet transform: I. Resolution of the wavefront set, Applied and Computational Harmonic Analysis 19 (2005) 162-197.

[7] E.J. Candès, D. Donoho, Continuous curvelet transform: II. Discretization and frames, Applied and Computational Harmonic Analysis 19 (2005) 198-222.

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

[9] E.J. Candès, L. Demanet, The curvelet representation of wave propagators is optimally sparse, Comm. Pure Appl. Math. 58 (2005) 1472-1528.

[10] E.J. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, SIAM Multiscale Model. Simul. (2006) in print.

[11] R.R. Coifman, Y. Meyer, Au delà des opérateurs pseudo-differentiels, Astérisque, Soc. Math. France 57 (1978).

[12] A. Córdoba, C. Fefferman, Wave packets and Fourier integral operators, Comm. Partial Differential Equations 3-11 (1978) 979-1005.

[13] P.J. Davis, P. Rabinowitz, Methods of numerical integration, Academic Press Inc., Orlando, (1975) 167-168.

[14] M.V. de Hoop, Generalization of the Bremmer coupling series, J. Math. Phys. 37 (1996) 3246-3282.

[15] M.V. de Hoop, J.H. Le Rousseau, B. Biondi, Symplectic structure of wave equation imaging: A path-integral approach based on the double-square-root equation, Geoph. J. Int. 153 (2003) 52-74.

[16] H. Douma, M.V. de Hoop, Leading-order seismic imaging using curvelets, Geophysics (2006) submitted.

[17] A. Dutt, V. Rokhlin, Fast Fourier transforms for nonequispaced data, SIAM Journal on Scientific Computing, 14 (1993) 1368-1393.

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

[19] D.-A. Geba, D. Tataru, A phase space transform adapted to the wave equation, preprint (2005).

[20] A. Greenleaf, G. Uhlmann, Estimates for singular Radon transforms and pseudodifferential operators with singular symbols, J. Funct. Anal. 89 (1990) 202-232.

[21] K. Gröchenig, Foundations of Time-Frequency Analysis, Birkhäuser, Boston, 2001.

[22] F.J. Herrmann, P.P Moghaddam, R. Kirlin, Optimization strategies for sparseness- and continuity-enhanced imaging: Theory, EAGE 67th Conference & Exhibition Proceedings, 2005.

[23] L. Hörmander, The Analysis of Linear Partial Differential Operators, Volume IV, Springer-Verlag, Berlin, 1985.

[24] H. Kumano-go, K. Taniguchi, Fourier integral operators of multi-phase and the fundamental solution for a hyperbolic system, Funkcialaj Ekvacioj 22 (1979) 161-196.

[25] J.H. Le Rousseau, Fourier-integral-operator approximation of solutions to first-order hyperbolic pseudodifferential equations I: Convergence in Sobolev spaces, Comm. Partial Differential Equations 31 (2006) 867-906.

[26] F. Akbar, M. Sen, P. Stoffa, Prestack plane-wave Kirchhoff migration in laterally varying media, Geophysics 61 (1996) 1068-1079.

[27] A. Shlivinski, E. Heyman, A. Boag, C. Letrou, A phase-space beam summation formulation for ultrawide-band radiation, IEEE Trans. Antennas Propagat. 52 (2004) 2042-2056.

[28] H.F. Smith, A Hardy space for Fourier integral operator, Jour. Geom. Anal. 8 (1998) 629-653.

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

[30] E.M. Stein, Harmonic Analysis: Real Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, 1993.

[31] C.C. Stolk, M.V. de Hoop, Modelling of seismic data in the downward continuation approach, SIAM J. Appl. Math. 65 (2005) 1388-1406.

[32] C.C. Stolk, M.V. de Hoop, Seismic inverse scattering in the downward continuation approach, Wave Motion 43 (2006) 579-598.

[33] M.E. Taylor, Pseudodifferential operators and nonlinear PDE, Birkhäuser, Boston, 1991.