

MULTISCALE DISCRETE APPROXIMATION OF FOURIER INTEGRAL OPERATORS*

FREDRIK ANDERSSON[†], MAARTEN V. DE HOOP[‡], AND HERWIG WENDT[‡]

Abstract. We develop a discretization and computational procedures for approximation of the action of Fourier integral operators the canonical relations of which are graphs. Such operators appear, for instance, in the formulation of imaging and inverse scattering of seismic reflection data. Our discretization and algorithms are based on a multiscale low-rank expansion of the action of Fourier integral operators using the dyadic parabolic decomposition of phase space and on explicit constructions of low-rank separated representations using prolate spheroidal wave functions, which directly reflect the geometry of such operators. The discretization and computational procedures connect to the discrete almost symmetric wave packet transform. Numerical wave propagation and imaging examples illustrate our computational procedures.

Key words. Fourier integral operators, multiscale computations, wave packets, dyadic parabolic decomposition, separated representation, operator compression, reflection seismology

AMS subject classifications. 35S30, 65T99, 86A15

DOI. 10.1137/100808174

1. Introduction. Fourier integral operators (FIOs), and their calculus, have played an important role in analyzing many problems involving linear hyperbolic partial differential equations. We mention parametrix constructions and developments in scattering and inverse scattering theories. In these developments, typically, the FIOs correspond with canonical relations, describing the propagation of singularities by these operators, which are the graphs of canonical transformations. In the present work, we focus on discretizing the action of FIOs in this class and on developing computational algorithms for their numerical evaluation.

The action of an FIO F in the mentioned class on a function $u(x)$ in L^2 is given by

$$(1.1) \quad (Fu)(y) = \int a(y, \xi) \exp(iS(y, \xi)) \hat{u}(\xi) d\xi,$$

where \hat{u} denotes the Fourier transform $\mathcal{F}_{x \rightarrow \xi}$ of u , $a(y, \xi)$ is the amplitude function, and the phase $S(y, \xi)$ is the generating function. Without restriction we assume that a is homogeneous of order zero in ξ . Furthermore, we assume that $\frac{\partial^2 S}{\partial y \partial \xi}$ is nonsingular. The propagation of singularities by F , $(x, \xi) \rightarrow (y, \eta)$, follows from S and is described by the transformation

$$(1.2) \quad \chi : \left(\frac{\partial S}{\partial \xi}, \xi \right) \rightarrow \left(y, \frac{\partial S}{\partial y} \right).$$

*Received by the editors September 10, 2010; accepted for publication (in revised form) November 30, 2011; published electronically March 8, 2012. This research was supported in part by NSF CMG grant DMS-0724644 and in part by the members of the Geo-Mathematical Imaging Group: BP, ConocoPhillips, ExxonMobil, Statoil, and Total.

<http://www.siam.org/journals/mms/10-1/80817.html>

[†]Mathematics LTH, Centre for Mathematical Sciences, Lund Institute of Technology, Lund University, Sweden (fa@maths.lth.se).

[‡]CCAM, Department of Mathematics, Purdue University, West Lafayette, IN (mdehoop@math.purdue.edu, hwendt@math.purdue.edu).

The operator F has a sparse matrix representation with respect to the frame of curvelets [10, 53], which originates from the dyadic parabolic decomposition of phase space and which will be briefly discussed below. We will refer to curvelets ([11] and references therein) by their collective name “wave packets.”

To arrive, through discretization, at an efficient algorithm for the action of an FIO it is natural to seek expansions of the amplitude function and complex exponential in terms of tensor products in phase space. (This strategy has been followed to develop algorithms for propagators since the advent of paraxial approximations of the wave equation, their higher-order extensions, and phase-screen methods and their generalizations. See Beylkin and Mohlenkamp [6] for a comprehensive analysis.) In the case of pseudodifferential operators, which are included in the class of operators considered here, χ is the identity and the generating function $S(y, \xi) = \langle y, \xi \rangle$ is linear in ξ and naturally separated. Typically, one introduces a radial partition of unity in ξ -space, the functions of which scale dyadically. On each annulus of this partition, the amplitude function or symbol can then be expanded in spherical harmonics. This results in a tensor product expansion [57]; each term in this expansion is also referred to as an elementary symbol. Bao and Symes [2] developed a computational method for pseudodifferential operators based on such a type of expansions. They considered a Fourier series expansion of the symbol in the angular variables $\arg \xi$ and a polyhomogeneous expansion in $|\xi|$. More recently other, fastly converging separated symbol expansions were introduced by Demanet and Ying [23] in adequate systems of rational Chebyshev functions or hierarchical splines with control points placed in a multiscale way in ξ -space. Alternative expansions of the action of Calderón–Zygmund operators, using bases of wavelets, were introduced and analyzed by Beylkin, Coifman, and Rokhlin [5].

Here, we consider FIOs in the class mentioned above and focus on expansions of the complex exponential in (1.1) separated in base and cotangent coordinates. A natural way to initiate the discretization and associated approximation is via the dyadic parabolic decomposition of phase space, enabling a natural connection with the geometry of the operators. Recently, De Hoop et al. [21] proposed an explicit *multiscale* expansion of low phase space separation rank of the action of FIOs associated with canonical graphs using the dyadic parabolic decomposition of phase space. The second-order term in the expansion provides an accuracy $\mathcal{O}(2^{-k/2})$ at frequency scale 2^k . For each frequency scale, the separation rank depends on k but is otherwise independent of the problem size. The present work elaborates on this result and develops a discretization, numerical approximation, and procedure for computing the action (1.1). We obtain an algorithm of complexity $\mathcal{O}(N^{\frac{3d-1}{2}} \log(N))$, or $\mathcal{O}(D N^d \log(N))$ if D is the number of significant tiles in the dyadic parabolic decomposition of u , valid in arbitrary finite dimension d . Our separated representation is expressed in terms of *geometric attributes* of the canonical relation of the FIO. We make use of *prolate spheroidal wave functions* (PSWFs) in connection with the dyadic parabolic decomposition, while the propagation of singularities or canonical transformation is accounted for via an unequally spaced FFT (USFFT). The use of PSWFs was motivated by the work of Beylkin and Sandberg [7] and the proposition of an efficient algorithm for their numerical evaluation by Xiao, Rokhlin, and Yarvin [62]. We note that it is also possible to obtain low-rank separated representations of the complex exponential in (1.1) purely numerically at the cost of losing the explicit relationship with the geometry. The algorithm presented here can be applied to computing parametrices of hyperbolic evolution and wave equations; we show that then our approximation corresponds to the solution of the paraxial wave equation in curvilinear coordinates,

i.e., directionally developed relative to the central wave vector. However, it also forms the basis of a computational procedure following the construction of *weak* solutions of Cauchy initial value problems for the wave equation if the medium is $C^{2,1}$, in which, in addition, a Volterra equation needs to be solved (de Hoop et al. [18]).

We derive our discretization from the (inverse) transform based on discrete almost symmetric wave packets [25]. The connection of our algorithm to discrete almost symmetric wave packets is important in imaging and inverse scattering applications, where the FIOs act on data (u in the above). The wave packets can aid in regularizing the data from a finite set of samples through sparse optimization (instead of standard interpolation, for example) [14, 15, 17, 58]. Moreover, the mentioned connection enables multiscale imaging and, in the context of directional pointwise regularity analysis [1, 30, 32, 33, 34, 35, 42], the numerical estimation and study of propagation of scaling exponents by the FIO, extending the corresponding results for Calderón–Zygmund operators using wavelets [43].

Imaging and inverse scattering of seismic reflection data can be generally formulated in terms of FIOs in the class considered here. In the presence of caustics, the construction of such FIOs requires an extension of standard (single) scattering operators; see Stolk and De Hoop [56, 54, 55]. First-order evolution equations and associated propagators play a role in implementations of wave-equation imaging and inverse scattering; we mention time and depth extrapolation (or downward continuation), and velocity continuation [27]. Furthermore, extended imaging can be described in terms of solving a Cauchy initial value problem for an evolution equation (Duchkov and De Hoop [27]), that is, an associated parametrix. We provide an explicit estimate of the paraxial approximation of the evolution operator. In connection with paraxial approximations, we also mention “beam” migration [9]. In the present contribution, we account for caustics only in parametrices of evolution equations using the semi-group property. The general case of caustics is the subject of a forthcoming paper.

We hasten to mention the work by Candès, Demanet, and Ying [12], who recently considered the fast computation of FIOs (in dimension $d = 2$). In their work, the ξ -space is decomposed into angular wedges which satisfy a parabolic relationship reminiscent of the dyadic parabolic decomposition for the finest available scale. The separated expansion of the complex exponential makes use of the Taylor series for the exponential function—as in the generalized-phase-screen expansions introduced by De Hoop, Le Rousseau, and Wu [20]—and a polar coordinates Taylor (or Maclaurin) expansion of the phase function in ξ ; the wedges can be chosen sufficiently narrow (which corresponds with large k in our analysis) so that only the first term in the latter expansion needs to be accounted for. In [13], a butterfly algorithm was obtained through a balanced tiling of the space and frequency domain which also admits low-rank separated representations of the complex exponential. An alternative approach is based on compressing operators by decomposing them in properly chosen bases of L^2 . Once a sparse representation has been obtained, the action of the operator is carried out by applying a sparse matrix in the transform domain. In dimension 1, such an approach was developed by Bradie, Coifman, and Grossman [8] for the computation of oscillatory integrals related to acoustic wave scattering. Here, we present an algorithm with a controlled error of $\mathcal{O}(2^{-k/2})$, essentially structured around the geometry (canonical graphs) of the FIOs. Our algorithm differs in structure from the methods introduced in [8, 12, 13]; those methods are accurate to arbitrary precision. In principle, in our approach, the phase function can be expanded to higher order, reducing the error accordingly; however, this would yield a significant loss of efficiency.

The outline of this paper is as follows. Below we give a brief introduction to the dyadic parabolic decomposition of phase space, the copartition of unity, and wave packets. In section 2, we summarize the multiscale operator expansion proposed in [21], and we construct the separated expansion of the complex exponential in (1.1) in explicit form using PSWFs and provide an analysis of its rank properties. In section 3, we establish the discretization of the operator expansion from the discrete almost symmetric wave packet transform, which we briefly summarize for convenience. We discuss the deformation of the phase space discretization under the operator action, suggesting strategies for choosing the oversampling factors and for the evaluation of the canonical transformation by USFFT. We obtain a box (frequency tile), individual packet, and hybrid packet-box based algorithm for the evaluation of (1.1) and investigate and compare their computational properties. In section 4, we detail their application to parametrices of evolution equations. We establish the explicit relationship with paraxial ray theory, the expansion terms of the phase in (1.1) being obtained from the propagator matrix of the associated Hamilton–Jacobi system along paraxial rays. As a special case, we consider solution operators of evolution equations represented as Trotter products, highlighting the connection with phase-space localized paraxial approximations. In section 5, the proposed algorithms are compared and illustrated in numerical examples including wave propagation in a heterogeneous isotropic medium and evolution-equation based (common-offset) imaging involving a homogeneous anisotropic Hamiltonian. In section 6, we draw conclusions on the present work and discuss future perspectives.

Wave packets. We briefly discuss the (co)frame of curvelets and wave packets [11, 25, 53]. Let $u \in L^2(\mathbb{R}^d)$ represent a (seismic) velocity field. We consider the Fourier transform $\hat{u}(\xi) = \int u(x) \exp[-i\langle x, \xi \rangle] dx$.

One begins with an overlapping covering of the positive ξ_1 axis ($\xi' = \xi_1$) by boxes of the form

$$(1.3) \quad B_k = \left[\xi'_k - \frac{L'_k}{2}, \xi'_k + \frac{L'_k}{2} \right] \times \left[-\frac{L''_k}{2}, \frac{L''_k}{2} \right]^{d-1},$$

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

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

Next, for each $k \geq 1$, let ν vary over a set of approximately $2^{k(d-1)/2}$ uniformly distributed unit vectors. (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

$$(1.4) \quad B_{\nu,k} = \Theta_{\nu,k}^{-1} B_k.$$

The $B_{\nu,k}$ are illustrated in Figure 1 (left). We denote for later use by $\mathbf{1}_{\nu,k}(\xi)$ the indicator function of $B_{\nu,k}$. In the (co)frame construction, one encounters two sequences of smooth functions, $\hat{\chi}_{\nu,k}$ and $\hat{\beta}_{\nu,k}$, on \mathbb{R}^d , each supported in $B_{\nu,k}$, so that they form a copartition of unity

$$(1.5) \quad \hat{\chi}_0(\xi) \hat{\beta}_0(\xi) + \sum_{k \geq 1} \sum_{\nu} \hat{\chi}_{\nu,k}(\xi) \hat{\beta}_{\nu,k}(\xi) = 1$$

and satisfy the estimates

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

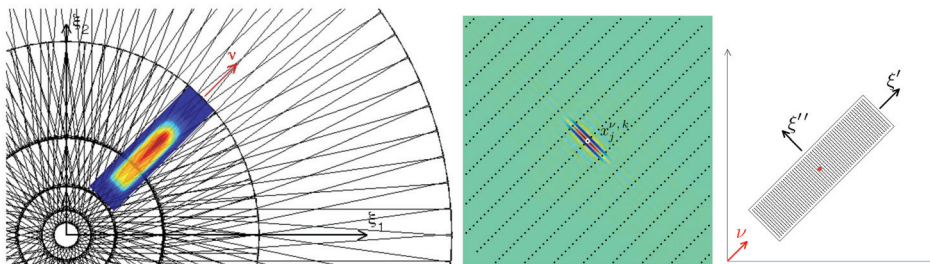


FIG. 1. Geometry for two-dimensional wave packets. Frequency domain boxes $B_{\nu,k}$ and window function $\hat{\chi}_{\nu,k}(\xi)$ for one particular box for scale $k = 3$ with orientation ν (left). One wave packet corresponding to the box highlighted in the subfigure on the left and central locations of wave packets in this box (center). Orientations of ξ' and ξ'' in the Taylor expansion of $S(y, \xi)$ (right).

A function $\hat{\chi}_{\nu,k}$ is plotted in color in Figure 1 (left). One then forms

$$(1.6) \quad \hat{\psi}_{\nu,k}(\xi) = \rho_k^{-1/2} \hat{\beta}_{\nu,k}(\xi), \quad \hat{\varphi}_{\nu,k}(\xi) = \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi)$$

with ρ_k the volume of B_k . These functions satisfy the estimates

$$(1.7) \quad \left. \begin{aligned} |\varphi_{\nu,k}(x)| \\ |\psi_{\nu,k}(x)| \end{aligned} \right\} \leq C_N 2^{k(d+1)/4} (2^k |\langle \nu, x \rangle| + 2^{k/2} \|x\|)^{-N}$$

for all N . To obtain a (co)frame, one introduces the integer lattice: $X_j := (j_1, \dots, j_n) \in \mathbb{Z}^d$, the dilation matrix

$$D_k = \frac{1}{2\pi} \begin{pmatrix} L'_k & 0_{1 \times d-1} \\ 0_{d-1 \times 1} & L''_k I_{d-1} \end{pmatrix}, \quad \det D_k = (2\pi)^{-d} \rho_k,$$

and points $x_j^{\nu,k} = \Theta_{\nu,k}^{-1} D_k^{-1} X_j$, which change with (ν, k) . The frame elements ($k \geq 1$) are then defined in the Fourier domain as

$$(1.8) \quad \hat{\varphi}_\gamma(\xi) = \hat{\varphi}_{\nu,k}(\xi) \exp[-i \langle x_j^{\nu,k}, \xi \rangle], \quad \gamma = (j, \nu, k),$$

and similarly for $\hat{\psi}_\gamma(\xi)$. A function $\varphi_{\nu,k}$ —referred to as a wave packet—as well as the corresponding lattice with points $x_j^{\nu,k}$ are plotted in Figure 1 (middle). One obtains the transform pair

$$(1.9) \quad u_\gamma = \int u(x) \overline{\psi_\gamma(x)} dx, \quad u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x)$$

with the property that $\sum_{\gamma': k'=k, \nu'=\nu} u_{\gamma'} \hat{\varphi}_{\gamma'}(\xi) = \hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi)$ for each ν, k .

2. Expansion of FIOs.

2.1. Dyadic parabolic decomposition and separated representation. Let $\varphi_\gamma(x)$, $\gamma = (j, \nu, k)$, denote a single wave packet with central position $x_j^{\nu,k}$, orientation ν , and scale k . The action of the operator F on $\varphi_\gamma(x)$ is

$$(2.1) \quad (F\varphi_\gamma)(y) = \rho_k^{-1/2} \int a(y, \xi) \exp[i(S(y, \xi) - \langle \xi, x_j^{\nu,k} \rangle)] \hat{\chi}_{\nu,k}(\xi) d\xi,$$

where $\hat{\varphi}_\gamma(\xi) = \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi) \exp[-i\langle \xi, x_j^{\nu,k} \rangle]$ is the Fourier transform of φ_γ . The action (1.1) on a function u is then recovered as

$$(2.2) \quad (Fu)(y) = \sum_\gamma u_\gamma(F\varphi_\gamma)(y).$$

In [21], three approximations of $(F\varphi_\gamma)(y)$ to order $\mathcal{O}(2^{-k/2})$ are obtained. They will underly our discretization and algorithms and are briefly summarized here. The strategy of [21] consists in replacing $S(y, \xi)$ and $a(y, \xi)$ by Taylor expansions near the microlocal support of φ_γ . The amplitudes $a(y, \xi)$ can be replaced by $a(y, \nu)$ without giving rise to errors larger than $\mathcal{O}(2^{-k/2})$ [21, Lemma 3.1]. By homogeneity in ξ of $S(y, \xi)$, the first-order Taylor expansion of S yields

$$S(y, \xi) - \langle \xi, x_j^{\nu,k} \rangle = \left\langle \xi, \frac{\partial S}{\partial \xi}(y, \nu) - x_j^{\nu,k} \right\rangle + h_2(y, \xi)$$

along the ν axis, where the error term $h_2(y, \xi)$ is homogeneous of order 1 and of class $S_{\frac{1}{2}, \text{rad}}^0$ on $\mathbf{1}_{\nu,k}(\xi)$ (cf. [21, (22)]). We introduce the *coordinate transform*,

$$(2.3) \quad y \rightarrow T_{\nu,k}(y) = \frac{\partial S}{\partial \xi}(y, \nu),$$

which describes the propagation of the wave packet φ_γ along rays according to geometrical optics (cf. (1.2)). Replacing $S(y, \xi) - \langle \xi, x_j^{\nu,k} \rangle$ by $\langle \xi, T_{\nu,k}(y) - x_j^{\nu,k} \rangle$ in (2.1) results in the approximation

$$(2.4) \quad (F\varphi_\gamma)(y) = a(y, \nu)\varphi_\gamma(T_{\nu,k}(y)) + \mathcal{O}(2^0).$$

We will use this approximation for comparison in our numerical examples and refer to it as the *zero order approximation*.

To refine the approximation to $\mathcal{O}(2^{-k/2})$, we need to include the second-order terms in the ξ'' directions perpendicular to the radial $\nu = \xi'$ direction in the Taylor expansion of $S(y, \xi)$. (The expansion directions are illustrated in Figure 1 (right).) Making use again of the homogeneity of S in ξ , we obtain the expansion

$$S(y, \xi) = \left\langle \xi, \frac{\partial S}{\partial \xi}(y, \nu) \right\rangle + \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y, \nu) \xi'' \right\rangle + h_3(y, \xi),$$

where $h_3(y, \xi)$ is $S_{\frac{1}{2}, \text{rad}}^{-\frac{1}{2}}$ on $\mathbf{1}_{\nu,k}(\xi)$ (cf. [21, (22)]). In view of the dyadic parabolic scaling, the argument of the complex exponential $\exp[i\frac{1}{2\xi'} \langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y, \nu) \xi'' \rangle]$ is bounded by a constant, c , say. The expansion leads to a tensor-product representation, separating the y and ξ variables, and yields the following result [21, Theorem 4.1].

THEOREM 2.1. *With functions $T_{\nu,k}(y)$ defined by (2.3) and functions $\alpha_{\nu,k}^{(r)}(y)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ such that*

$$(2.5) \quad \exp \left[i \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y, \nu) \xi'' \right\rangle \right] \mathbf{1}_{\nu,k}(\xi) \approx \sum_{r=1}^R \alpha_{\nu,k}^{(r)}(y) \hat{\vartheta}_{\nu,k}^{(r)}(\xi),$$

one may express

$$(2.6) \quad (F\varphi_\gamma)(y) = a(y, \nu) \sum_{r=1}^R \alpha_{\nu,k}^{(r)}(y) (\hat{\vartheta}_{\nu,k}^{(r)} * \varphi_\gamma)(T_{\nu,k}(y)) + 2^{-k/2} f_\gamma$$

with $R \sim k/\log(k)$, where f_γ is a curvelet-like function (cf. [21, (23)]) centered at $\chi(\gamma)$.

Theorem 2.1 hence approximates (1.1) to order $\mathcal{O}(2^{-k/2})$ as the sum of R modified wave packets $\tilde{\phi}_{r;\gamma}(x) = (\vartheta_{\nu,k}^{(r)} * \varphi_\gamma)(x)$ with amplitude corrections $a(y, \nu) \alpha_{\nu,k}^{(r)}(y)$, followed by a coordinate transform $T_{\nu,k}(y)$ accounting for the canonical transformation. This expansion can be extended to any order.

Further approximations. Let $y_j^{\nu,k} = T_{\nu,k}^{-1}(x_j^{\nu,k})$. It is possible to replace the functions $a(y, \nu)$, $\frac{\partial S}{\partial \xi}(y, \nu)$, and $\frac{\partial^2 S}{\partial \xi'^2}(y, \nu)$ with $a(y_j^{\nu,k}, \nu)$, $\frac{\partial S}{\partial \xi}(y_j^{\nu,k}, \nu)$, and $\frac{\partial^2 S}{\partial \xi'^2}(y_j^{\nu,k}, \nu)$ with error remaining of order $\mathcal{O}(2^{-k/2})$. This yields the following alternative result [21, Theorem 4.2]. With

$$(2.7) \quad \hat{\vartheta}_\gamma(\xi) = \exp \left[i \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 S}{\partial \xi'^2}(y_j^{\nu,k}, \nu) \xi'' \right\rangle \right] \mathbf{1}_{\nu,k}(\xi),$$

one may express

$$(2.8) \quad (F\varphi_\gamma)(y) = a(y_j^{\nu,k}, \nu) (\vartheta_\gamma * \varphi_\gamma)(T_{\nu,k}(y)) + 2^{-k/2} f_\gamma,$$

where f_γ is a curvelet-like function centered at $\chi(\gamma)$ (cf. [21, (23)]).

Furthermore, the change of coordinates $T_{\nu,k}$ can be approximated by Taylor expansion of $S(y, \nu)$ about $(y_j^{\nu,k}, \nu)$ [21, Theorem 4.3]: One may express

$$(2.9) \quad (F\varphi_\gamma)(y) = a(y_j^{\nu,k}, \nu) (\vartheta_\gamma * \varphi_\gamma) \left(DT_\gamma(y - y_j^{\nu,k}) + M_\gamma \cdot (y - y_j^{\nu,k})^2 \right) + 2^{-k/2} f_\gamma,$$

where $DT_\gamma = \frac{\partial T_{\nu,k}}{\partial y}(y_j^{\nu,k}) = \frac{\partial^2 S}{\partial \xi \partial y}(y_j^{\nu,k}, \nu)$, $M_\gamma = \frac{1}{2} \frac{\partial^2 S}{\partial y^2}(y_j^{\nu,k}, \nu)$, and f_γ is a curvelet-like function centered at $\chi(\gamma)$ (cf. [21, (23)]). In this approximation, M_γ captures the curvature of a localized plane wave attached to φ_γ under the underlying canonical transformation, and DT_γ contains rigid motion, shear along the wave front and dilations along and perpendicular to the wave front. It is important to note that the further approximations are tied to particular wave packets, unlike the expansion given in Theorem 2.1.

2.2. PSWFs and tensor product. Here we revisit (2.5). The argument of the exponential on the left-hand side consists of terms each of which reveals a separation of variables in phase space and is reminiscent of the kernel of specific operators whose eigenfunctions are the PSWFs. Motivated by the fast decay of the corresponding eigenspectrum, we aim at obtaining an explicit low-rank realization of (2.5) by constructing the functions $\alpha_{\nu,k}^{(r)}(y)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ from PSWFs.

2.2.1. PSWFs. We give a brief summary on PSWFs and refer to, e.g., [38, 39, 40, 41, 49, 50, 51] for details and to [44, 47, 61, 62] for recent methods for their numerical evaluation. The (generalized) prolate spheroidal wave functions ψ are the eigenfunctions of the integral operator

$$(F^c \psi)(x) = \int_{\mathcal{R}} \exp[ic\langle x, z \rangle] \psi(z) dz, \quad c \in \mathbb{R}^+, \quad \|x\| \leq 1,$$

on the unit ball \mathcal{R} in $\mathcal{D} \geq 1$ dimensions (for $\mathcal{D} = 1$, \mathcal{R} is the interval $[-1, 1]$). For each $c \in \mathbb{R}^+$, there exists a countable set of numbers λ_κ^c , which are either real or imaginary, such that the equation

$$(2.10) \quad \lambda_\kappa^c \psi_\kappa^c(x) = \int_{\mathcal{R}} \exp[ic\langle x, z \rangle] \psi_\kappa^c(z) dz, \quad \|x\| \leq 1,$$

has a continuous solution on \mathcal{R} , where κ is a multi-index. The functions ψ_κ^c are bounded, purely real, orthonormal, and complete in $L^2(\mathcal{R})$. The eigenvalue spectrum consists of few eigenvalues λ_κ^c with significant magnitude, the precise number depending on the bandwidth parameter c , and then decays exponentially fast to values close to zero [38, 40, 41]. (For example, for $\mathcal{D} = 1$, the spectrum contains roughly $2c/\pi$ eigenvalues with magnitude close to $\sqrt{2\pi/c}$ and decays exponentially beyond.)

For $\mathcal{D} = 1$, $\kappa = n$ is a simple index, and the functions $\psi_n(x)$ are also the eigenfunctions of the self-adjoint differential operator $(L^c \cdot)(x) = (1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx} - c^2x^2$. For their numerical construction, expansions in Legendre polynomials are used, where the expansion coefficients are obtained by recurrence relations derived from L^c . In $\mathcal{D} \geq 2$, PSWFs are constructed in polar coordinates (ρ, Ω) , in which their radial parts separate from their angular parts:

$$\psi_\kappa^c = \psi_{(N,n,l)}^c(\rho, \Omega) = \Psi_{(N,n)}^c(\rho) S_l(\Omega).$$

Let $p = \mathcal{D} - 2$. The angular functions $S_l(\Omega)$ are given by complete sets of orthonormal surface harmonics of degree $N + p$. (In the practically most interesting case $\mathcal{D} = 3$, these are the spherical harmonics.) The radial functions are given by $\Psi_{N,n}^c(\rho) = \rho^{-(p+1)/2} \varphi_{N,n}^c(\rho)$, where $\varphi^c(\rho)$ are the bounded solutions to the eigen equation (J_N are the Bessel functions):

$$(2.11) \quad \gamma_{N,n}^c \varphi_{N,n}^c(\rho) = \int_0^1 J_{N+\frac{p}{2}}(c\rho\rho') \sqrt{c\rho\rho'} \varphi_{N,n}^c(\rho') d\rho'.$$

The functions $\varphi_{N,n}^c(\rho)$ are also the eigenfunctions of the self-adjoint differential operator $(L^c \cdot)(\rho) = (1-\rho^2)\frac{d^2}{d\rho^2} - 2\rho\frac{d}{d\rho} + (\frac{1/4-(N+p/2)^2}{\rho^2} - c^2\rho^2)$. Similar to the case $\mathcal{D} = 1$, the (numerical) construction of $\Psi_{N,n}^c(\rho)$ is based on recurrence relations, derived from the differential operator L^c , for the coefficients of expansions in Jacobi polynomials. Recent numerical procedures allow the construction of PSWFs for most values of c encountered in practice [47, 62]. (See also, e.g., [44, 61] for asymptotic results and approximations.) The corresponding eigenvalues λ_κ^c are obtained by numerical integration of (2.10) ($\mathcal{D} = 1$) and (2.11) ($\mathcal{D} \geq 2$, here $\lambda_{N,n}^c = i^N (2\pi)^{1+p/2} c^{-(P+1)/2} \gamma_{N,n}^c$); cf. [47, 49]. Both the expansion coefficients in the numerical construction and the eigenvalues λ_κ^c can be precomputed and tabulated for given bandwidth parameters c .

2.2.2. Tensor product. We proceed with the construction of the tensor product functions $\alpha_{\nu,k}^{(r)}(y)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ from PSWFs. The kernel of operator (2.10) admits the representation

$$(2.12) \quad \exp[i\langle x, z \rangle] = \sum_{\kappa} \lambda_{\kappa}^c \psi_{\kappa}^c(x) \psi_{\kappa}^c(z), \quad \|x\|, \|z\| \leq 1.$$

Our strategy is to manipulate the left-hand side of (2.5) to match this expression. We begin with extracting from the matrices $\frac{\partial^2 S}{\partial \xi_j'' \partial \xi_l''}(y, \nu)$ and $\frac{\xi_j'' \xi_l''}{\xi'}$ the vector-valued functions $\tilde{f}_\nu : \mathbb{R}^d \rightarrow \mathbb{R}^{\mathcal{D}(d)}$ $\tilde{g} : \mathbb{R}^d \rightarrow \mathbb{R}^{\mathcal{D}(d)}$,

$$\begin{aligned} \tilde{f}_{m(j,l)}(y) &= \left[(2 - \delta_{jl}) \frac{\partial^2 S}{\partial \xi_j'' \partial \xi_l''}(y, \nu) \right], \\ \tilde{g}_{m(j,l)}(\xi', \xi'') &= \left[(2 - \delta_{jl}) \frac{\xi_j'' \xi_l''}{\xi'} \right], \quad m(j, l) = 1, \dots, \mathcal{D}, \end{aligned}$$

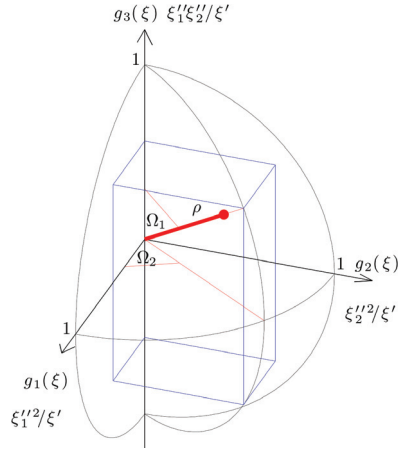


FIG. 2. Illustration of PSWF coordinates for $g(\xi)$ and $\mathcal{D} = 3 (d = 3)$. The Cartesian boxes $f(y)$ and $g(\xi)$ are included in the unit ball \mathcal{R} on which $\psi_\kappa^c(\rho, \Omega)$ form an orthonormal basis.

where, due to symmetry in partial derivatives and in $\xi_j'' \xi_l''$,

$$\mathcal{D}(d) = (d-1)d/2.$$

Proper normalization confines the (transformed) Cartesian boxes $\tilde{f}(y)$ and $\tilde{g}(\xi', \xi'')$ to the unit ball \mathcal{R} (cf. illustration in Figure 2):

$$(2.13) \quad f(y) = \frac{\tilde{f}(y)}{\sup_y |\tilde{f}(y)|}, \quad g(\xi', \xi'') = \frac{\tilde{g}(\xi', \xi'')}{\sup_{\mathbf{1}_{\nu,k}(\xi)} |\tilde{g}(\xi', \xi'')|}.$$

We absorb the normalization constants in the bandwidth parameter

$$(2.14) \quad c = c(\nu) = \frac{1}{2} \sup_{\mathbf{1}_{\nu,k}(\xi)} |\tilde{g}(\xi', \xi'')| \sup_y |\tilde{f}(y)|.$$

With these definitions, we obtain by elementary manipulations of the left-hand side of (2.5)

$$(2.15) \quad \begin{aligned} \exp \left[i \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 S}{\partial \xi'^2}(y, \nu) \xi'' \right\rangle \right] \mathbf{1}_{\nu,k}(\xi) &= \exp \left[i \frac{1}{2} \sum_{j,l=2}^d \frac{\xi_j'' \xi_l''}{\xi'} \frac{\partial^2 S}{\partial \xi_j'' \partial \xi_l''}(y, \nu) \right] \mathbf{1}_{\nu,k}(\xi) \\ &= \exp \left[i \frac{1}{2} \sum_{m=1}^{\mathcal{D}(d)} \tilde{g}_m(\xi', \xi'') \tilde{f}_m(y) \right] \mathbf{1}_{\nu,k}(\xi) \\ &= \exp [ic \langle f(y), g(\xi', \xi'') \rangle] \mathbf{1}_{\nu,k}(\xi) \\ &= \sum_{\kappa} \lambda_{\kappa}^c \psi_{\kappa}^c(f(y)) \psi_{\kappa}^c(g(\xi', \xi'')) \mathbf{1}_{\nu,k}(\xi). \end{aligned}$$

Now let the sequence of multi-indices $\kappa_1, \kappa_2, \dots$ correspond to the sorted sequence of eigenvalues $|\lambda_{\kappa_1}^c| \geq |\lambda_{\kappa_2}^c| \geq \dots$, and truncate the infinite sum over the multi-index κ

at the R th term to within precision $\varepsilon(k)$:

$$\begin{aligned}
 & \exp \left[i \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 S}{\partial \xi'^2}(y, \nu) \xi'' \right\rangle \right] \mathbf{1}_{\nu,k}(\xi) \\
 &= \sum_{r=1}^{R_{\nu,k}} \lambda_{\kappa_r}^c \psi_{\kappa_r}^c(f(y)) \psi_{\kappa_r}^c(g(\xi', \xi'')) \mathbf{1}_{\nu,k}(\xi) + \varepsilon(k) \\
 (2.16) \quad &= \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y) \hat{\vartheta}_{\nu,k}^{(r)}(\xi) + \varepsilon(k).
 \end{aligned}$$

Here, in view of Theorem 2.1, $\varepsilon(k) \sim 2^{-k/2}$ in order to achieve accuracy $\mathcal{O}(2^{-k/2})$ at frequency scale k . We identify the functions

$$(2.17) \quad \alpha_{\nu,k}^{(r)}(y) = \psi_{\kappa_r}^c(f(y)),$$

$$(2.18) \quad \hat{\vartheta}_{\nu,k}^{(r)}(\xi) = \lambda_{\kappa_r}^c \psi_{\kappa_r}^c(g(\xi', \xi'')),$$

which completes the construction of the tensor-product (2.5), given by (2.16)–(2.18). The eigenvalues λ_{κ}^c can alternatively be absorbed in either of the functions¹ $\alpha_{\nu,k}^{(r)}(y)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$.

Rank properties. The rank R of the separated expansion (2.16) is controlled by the desired precision ε and by the bandwidth parameter c defined in (2.14). The bandwidth is in turn determined by the largest value that the function $|\frac{\partial^2 S}{\partial \xi'^2}(y, \nu)|$ attains over y on the calculation domain and by the size of the boxes B_k in the frequency tiling (cf. (1.3)) through the values that $\frac{\xi''^2}{\xi'}$ can attain on them. Under our assumption that there are no caustics, the former is always bounded on finite domains over y , and the latter also is by virtue of the dyadic parabolic decomposition; hence c is bounded. The exponentially fast decay of the eigenvalue spectrum and the orthonormality of the functions ψ_j^c then guarantee the fast convergence of (2.16) and finite rank R for finite precision ε . The choice of frequency tiling can be seen as a trade-off between the number of frequency boxes $B_{\nu,k}$ to be computed in (2.2) and the number of tensor product terms to be included in (2.16). We note that in view of the parabolic scaling, the bandwidth parameter (2.13) is (asymptotically) independent of scale. Indeed, $\sup_{j,l,1_{\nu,k}(\xi)} \xi_j'' \xi_l'' / \xi' = \sup_{j,l,1_{\nu,1}(\xi)} (\xi_j'' 2^{k/2}) (2^{k/2} \xi_l'' 2^{k/2}) / \xi' / (\xi' 2^k) = \sup_{j,l,1_{\nu,1}(\xi)} \xi_j'' \xi_l'' / \xi'$ and $\frac{\partial^2 S}{\partial \xi'^2}(y, \nu)$ are scale independent. In the following, we revisit bounds on the precision ε for given rank R for $\mathcal{D} = 1$ ($d = 2$). From [48], we have the following estimates:

$$(2.19) \quad |\lambda_r^c| = \frac{\sqrt{\pi} c^r (r!)^2}{(2r)! \Gamma(r + \frac{3}{2})} \exp \left[\int_0^c \left(\frac{2(\psi_r^b(1))^2 - 1}{2b} - \frac{r}{b} \right) db \right] \leq \frac{\sqrt{\pi} c^r (r!)^2}{(2r)! \Gamma(r + \frac{3}{2})}$$

and $|\psi_r^c(1)| < \sqrt{r + 1/2}$, hence

$$(2.20) \quad |\lambda_r^c| \leq \frac{\sqrt{\pi} c^r (r!)^2}{(2r)! \Gamma(r + \frac{3}{2})} \leq \frac{\sqrt{\pi} c^r r!}{(2r)!} \leq \sqrt{\pi} c^r 2^{-r \log_2(r)} = \sqrt{\pi} 2^{-r[\log_2(r) - \log_2(c)]},$$

¹With the exception of the next paragraph, we will omit explicit reference to the bandwidth parameter c hereafter for convenience of notation.

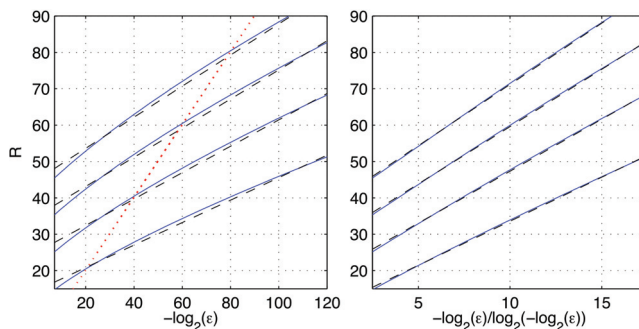


FIG. 3. Plots of numerical evaluation of (2.24) for $c = \{10, 20, 30, 50\}$ (blue solid line) in $-\log(\varepsilon)$ (left) and in $-\log(\varepsilon)/\log(-\log(\varepsilon))$ (right) vs. R coordinates. Black dashed lines correspond to linear fits in the respective coordinates. Plot of bound (2.22) (left, red dotted line).

and for $r \geq 2c$,

$$(2.21) \quad |\lambda_r^c| \leq 2^{-r+1},$$

which together with $M_r^c = \max_{s \leq r} \max_{-1 \leq x \leq 1} |\psi_s^c(x)| \leq 2\sqrt{r}$ gives the L^∞ bound

$$\varepsilon^\infty(R) = \left| F(x, y) - \sum_{r=1}^R \lambda_r^c \psi_r^c(x) \psi_r^c(y) \right|_\infty \leq \sum_{r=R+1}^\infty |\lambda_r^c| (M_r^c)^2 \leq \frac{8(R+2)}{2^R},$$

valid for $R \geq 2c$. By orthonormality on the unit ball of the functions ψ_r , we obtain the L^2 bound

$$(2.22) \quad \varepsilon(R) = \left\| \sum_{r=R+1}^\infty \lambda_r \psi_r^c(x) \psi_r^c(z) \right\|_{L^2(-1,1)} = \sqrt{\sum_{r=R+1}^\infty |\lambda_r|^2} \leq \frac{4}{\sqrt{3}} 2^{-(R+1)},$$

valid for $R \geq 2c$, and the corresponding rank estimate

$$(2.23) \quad R(\varepsilon) \geq -\log_2(\varepsilon) + \log_2(4/\sqrt{3}) - 1.$$

The bounds (2.22) and (2.23) are based on (2.21), which enables us to obtain closed form expressions but is a very conservative estimate. A refined estimate on the order of $R(\varepsilon)$ is obtained from the rightmost inequality in (2.20). Results for the numerical evaluation of

$$(2.24) \quad \varepsilon(R) = \sqrt{\sum_{r=R+1}^\infty |\lambda_r|^2} \leq \sqrt{\pi} \sqrt{\sum_{r=R+1}^\infty 2^{-2r[\log_2(r) - \log_2(c)]}}$$

are plotted in Figure 3 for different bandwidths c , together with (2.23), clearly indicating that

$$(2.25) \quad R(\varepsilon) = \mathcal{O}(-\log(\varepsilon)/\log(-\log(\varepsilon))).$$

For accuracy $\varepsilon(k) = \mathcal{O}(2^{-k/2})$ we therefore have, in agreement with Theorem 2.1,

$$(2.26) \quad R(k) = \mathcal{O}(k/\log(k)).$$

3. Discretization. We develop an algorithm, based on the operator expansion Theorem 2.1 and on the separated representation (2.16)–(2.18), for the evaluation of the approximate action of F on a function u for discrete space and frequency points y_n and ξ_l , respectively. Our discretization is chosen to match the structure of the discrete wave packet transform [25]. This enables us to switch from the coefficients of the wave packet transform to data in the frequency domain—the input to (1.1)—efficiently through standard FFTs. We assume here that the partial derivatives $\frac{\partial^2 S}{\partial \xi^{m2}}(y, \nu)$ and the functions $T_{\nu,k}(y)$ and $T_{\nu,k}^{-1}(x)$ are known.

3.1. Discrete almost symmetric wave packets and operator action. We initiate the discretization of Theorem 2.1 from the adjoint discrete almost symmetric wave packet transform. We begin with writing the convolutions $(\vartheta_{\nu,k}^{(r)} * \varphi_\gamma)(T_{\nu,k}(y))$ in (2.6) in the Fourier domain,

$$(3.1) \quad \tilde{\phi}_\gamma(y) = (F\varphi_\gamma)(y) \approx a(y, \nu) \rho_k^{-1/2} \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y) \sum_{\xi \in \mathbf{1}_{\nu,k}} e^{i\langle T_{\nu,k}(y), \xi \rangle} \hat{\vartheta}_{\nu,k}^{(r)}(\xi) \hat{\chi}_{\nu,k}(\xi),$$

and obtain the action (2.2) on an input function $u(x)$:

$$(3.2) \quad \begin{aligned} (Fu)(y) &\approx \sum_\gamma u_\gamma \tilde{\phi}_\gamma(y) \\ &= \sum_{\nu,k} a(y, \nu) \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y) \sum_{\xi \in \mathbf{1}_{\nu,k}} e^{i\langle T_{\nu,k}(y), \xi \rangle} \hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi) \hat{\vartheta}_{\nu,k}^{(r)}(\xi). \end{aligned}$$

Below, the amplitudes $a(y, \nu)$ are, with slight abuse of notation, absorbed in the functions $\alpha_{\nu,k}^{(r)}(y)$. The structure of (3.2) is reminiscent of the (adjoint) wave packet transform (1.9),

$$(3.3) \quad u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x) = \sum_\xi \sum_{\nu,k} e^{i\langle x, \xi \rangle} \hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi),$$

and we will indeed use the same discretization, which we briefly summarize for convenience (see [25] for details). We assume that the data $u(x_i)$ are given in discrete form at sampling points² $x_i = N^{-1}2\pi i$, $i \in \mathbb{R}^d$, $-\frac{N}{2} \leq i_n < \frac{N}{2}$. Following the discretization of the “inner” forward transform

$$(3.4) \quad \tilde{u}_{j,\nu,k} = \frac{1}{\rho_k^{1/2}} \frac{1}{(2\pi)^d} \frac{1}{\sigma'_k(\sigma''_k)^{d-1}} \sum_l \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[i\langle x_j^{\nu,k}, \xi_l^{\nu,k} \rangle] \approx u_\gamma,$$

the discretization of the “inner” adjoint transform $\hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi) = \sum_{\gamma': \nu'=\nu, k'=k} u_{\gamma'} \hat{\varphi}_{\gamma'}(\xi)$ is obtained as

$$(3.5) \quad \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) = \rho_k^{-1/2} \left(\sum_j \tilde{u}_{j,\nu,k} \exp \left[-i\langle x_j^{\nu,k}, \xi_l^{\nu,k} \rangle \right] \right) \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}).$$

²When the data $u(x_i)$ are sampled at sampling intervals Δ_n^x in direction n , then $x_n^{phys} = N\Delta_n^x x_{l_n}$ and $\xi_n^{phys} = \xi_{l_n}/(N\Delta_n^x)$. Below, the normalization constants are assumed to be absorbed in the functions $\alpha_{\nu,k}^{(r)}(y)$, $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ and $T_{\nu,k}$.

The points $\xi_l^{\nu,k}$ are chosen on a (regular) rotated grid. Specifically, we let

$$(3.6) \quad \Xi^k = \left\{ l \in \mathbb{Z}^d \mid -\frac{N'_k}{2} \leq l_1 < \frac{N'_k}{2}, \dots, -\frac{N''_k}{2} \leq l_d < \frac{N''_k}{2} \right\}.$$

The points in this set are denoted by Ξ^k_l . The parameters (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, determining the accuracy of approximation (3.4) to the inverse Fourier transform. The set Ξ^k contains $N^k_\xi \sim \sigma'_k(\sigma''_k)^{d-1}N^{\frac{d+1}{2}}$ points. We choose the $\xi_l^{\nu,k}$ (covering the box $B_{\nu,k}$) as

$$(3.7) \quad \xi_l^{\nu,k} = \Theta_{\nu,k}^{-1} (D_k S_k^{-1} \Xi^k_l + \xi'_k e_1),$$

where the matrix S_k is defined as $S_k = \frac{1}{2\pi} \begin{pmatrix} N'_k & 0_{1 \times d-1} \\ 0_{d-1 \times 1} & N''_k I_{d-1} \end{pmatrix}$. The dot product in the phase of the exponential in (3.5) then becomes

$$(3.8) \quad \langle x_j^{\nu,k}, \xi_l^{\nu,k} \rangle = (D_k S_k^{-1} \Xi^k_l + \xi'_k e_1)^t D_k^{-1} X_j = \frac{2\pi j_1 \xi'_k}{L'_k} + 2\pi \left(\frac{j_1 l_1}{N'_k} + \frac{j_2 l_2 + \dots + j_d l_d}{N''_k} \right).$$

Thus, the specific choice of points $\xi_l^{\nu,k}$ allows for a fast evaluation of $\hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$ from the data wave packet coefficients $\tilde{u}_{j,\nu,k}$ (cf. (3.4), (3.5)) for $l \in \Xi^k$,

$$(3.9) \quad \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp(2\pi i j_1 \xi'_k / L'_k) = \rho_k^{-1/2} N'_k (N''_k)^{d-1} \sum_j \tilde{u}_{j,\nu,k} \exp[-i \langle x_j, \xi_l \rangle],$$

where $\xi_l = l$ and $x_j = S_k^{-1} j$ with $j \in \Xi^k$, while $(N'_k (N''_k)^{d-1}) = (2\pi)^d \det S_k$. One can use a d -dimensional FFT for the evaluation of $\hat{u}(\xi_l^{\nu,k})$ and $\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$ in (3.9) when the values for $\tilde{u}_{j,\nu,k}$ are given. The discrete “outer” adjoint transform completes the discretization of (3.3):

$$(3.10) \quad u(x_i) \approx \sum_{\nu,k} \sum_{l \in \Xi^k} e^{i \langle x_i, \xi_l^{\nu,k} \rangle} \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}).$$

It is evaluated by USFFT [3, 28, 29] from the irregularly spaced set of points $\xi_l^{\nu,k}$ to x_i .

Now let $y_i = T_{\nu,k}^{-1}(x_i)$. Then the dot product in the phase of the complex exponential in (3.2) becomes

$$\langle T_{\nu,k}(y_i), \xi_l \rangle = \langle x_i, \xi_l \rangle,$$

and we obtain the discretization of (3.2):

$$(3.11) \quad (Fu)(y_i) \approx \sum_{\nu,k} \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y_i) \sum_{l \in \Xi^k} e^{2\pi i \langle x_i, \xi_l^{\nu,k} \rangle} \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \hat{\vartheta}_{\nu,k}^{(r)}(\xi_l^{\nu,k}).$$

As above, d -dimensional FFT is used for the fast evaluation of $\hat{u}(\xi_l^{\nu,k})$ and $\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$ from the wave packet transform of the data. Unlike (3.10), the “outer” transform USFFT $\xi_l^{\nu,k} \rightarrow x_i$ now has to be evaluated for each box (ν, k) separately, since the

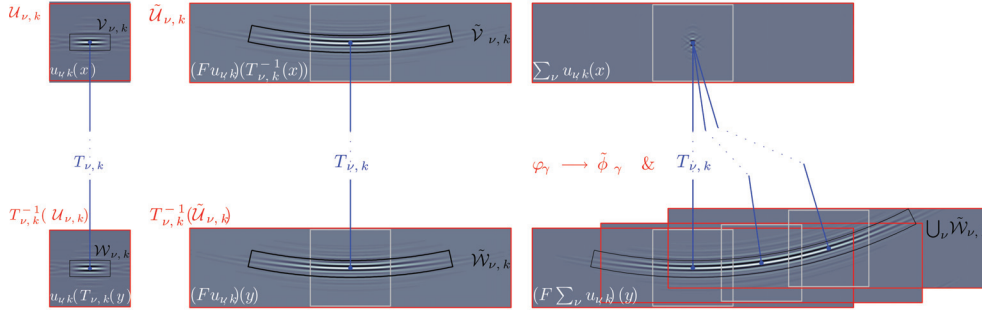


FIG. 4. Illustration of oversampling and "compression" of computational domain for one single wave packet (zero order approximation (left column), approximations to $\mathcal{O}(2^{-k/2})$ (center column)), and for three wave packets with common central position and frequency scale and neighboring orientations (approximation to $\mathcal{O}(2^{-k/2})$ (right column)). The domains $\mathcal{V}_{\nu,k} = \bigcup_j b_{j,\nu,k}$ and $\mathcal{W}_{\nu,k} = T_{\nu,k}(\mathcal{V}_{\nu,k})$ with effective nonzero data components are schematically indicated with black borders; the upper and lower rows are related through the coordinate transform $T_{\nu,k}$.

functions $T_{\nu,k}(y)$, $\alpha_{\nu,k}^{(r)}(y)$ differ for each box. Denoting by $u_{\nu,k}$ the data component corresponding to the box (ν, k) ,

$$(3.12) \quad u_{\nu,k}(x_i) = \sum_{\gamma': k'=k, \nu'=\nu} u_{\gamma'} \varphi_{\gamma'}(x_i),$$

reveals the organization by boxes of (3.11), $(Fu)(y_i) \approx \sum_{\nu,k} (Fu_{\nu,k})(y_i)$.

3.2. Deformation, compression, and oversampling. Here, we consider the deformation of phase space induced by the operator and account for it in the discretization (3.11) of (3.2). The action of F on the data components (3.12) is twofold: modification of their spatial support and deformation under the coordinate transformation $y \rightarrow T_{\nu,k}(y)$. We account for both by the introduction of additional oversampling factors, while keeping the structure of the discrete wave packet transform.

Oversampling. We first consider the operator action $(Fu_{\nu,k})$ for one single box (3.12) as a function of x within the frame of reference

$$(3.13) \quad \mathcal{E}(x) = T_{\nu,k}^{-1}(x).$$

The data component $u_{\nu,k}(x)$ has spatial support $\mathcal{U}_{\nu,k} = \text{supp } u_{\nu,k}(x) \subset (-\frac{\pi}{2}, \frac{\pi}{2}]^d$. As a result of the application of the frequency domain windows $\hat{\vartheta}_{\nu,k}^{(r)}$, the functions $\tilde{\varphi}_{\gamma}(\mathcal{E}(x)) = (F\varphi_{\gamma})(\mathcal{E}(x))$ which constitute $(Fu_{\nu,k})(\mathcal{E}(x))$ spread out in the ξ'' directions and have enlarged spatial support w.r.t. $\varphi_{\gamma}(x)$: $\tilde{\mathcal{U}}_{\nu,k} = \text{supp } (Fu_{\nu,k})(\mathcal{E}(x)) \subset (-\zeta\frac{\pi}{2}, \zeta\frac{\pi}{2}]^d$ with $\zeta \geq 1$ and $\mathcal{U}_{\nu,k} \subseteq \tilde{\mathcal{U}}_{\nu,k}$. Consequently, the sampling density in ξ has to be increased by a factor $\zeta \geq 1$ w.r.t. the original discretization $\xi_l^{\nu,k}$. We account for this by initiating the above discretization for zero-padded data $u^{zp}(x_i)$, consisting of the data $u(x_i)$ augmented in each direction with $\lceil (\zeta - 1)N \rceil$ zeros (cf. Figure 4). We denote the corresponding box data components by $u_{\nu,k}^{zp}$.

We can relate the amount of spreading of $\tilde{\varphi}_{\gamma}(\mathcal{E}(x))$ (and hence the oversampling factor ζ) to the partial derivatives $\frac{\partial^2 S}{\partial \xi''^2}(\mathcal{E}(x), \nu)$ and to the size of the boxes $B_{\nu,k}$ by geometrically imposing connectivity, under the action of F , of wave packets sharing scale and position at neighboring orientations. For instance, if l_k'' and \tilde{l}_k'' are measures

for the width of the effective numerical support of $\varphi_\gamma(x)$ and $\tilde{\phi}_\gamma(\mathcal{E}(x))$, respectively, in $d = 2$ dimensions, $\tilde{l}_k'' \approx \max(l_k'', \frac{1}{2} \frac{\partial^2 S}{\partial \xi'^2} \tan(\mathcal{C}\pi/2/N_\nu(k)))$, where the constant \mathcal{C} depends on the overlap of two neighboring boxes, and $N_\nu(k)$ the number of boxes at frequency scale k . We note that $\tilde{\phi}_\gamma(\mathcal{E}(x))$ (and consequently $(Fu_{\nu,k})(\mathcal{E}(x))$) as functions of x have compact support $\mathbf{1}_{\nu,k}(\xi)$, as is clear from (3.11) and the fact that $\hat{\chi}_{\nu,k}(\xi)$ and $\hat{\beta}_{\nu,k}(\xi)$ have compact support $\mathbf{1}_{\nu,k}(\xi)$.

Deformation and spatial grid resolution. Now we apply the coordinate transform $y \rightarrow x = T_{\nu,k}(y)$ and map the frame of reference $\mathcal{E}(x)$ onto y . We obtain the functions $\tilde{\phi}_{j,\nu,k}(y)$ in (y, η) phase space, which are translated, rotated, and deformed versions of the (x, ξ) phase space functions $\tilde{\phi}_\gamma(\mathcal{E}(x))$. The map $x \rightarrow T_{\nu,k}^{-1}(x)$ contracts and expands locally, inducing a local change in frequency; indeed, for two points \tilde{x} and \tilde{y} connected by $\tilde{y} = T_{\nu,k}^{-1}(\tilde{x})$, it follows from (2.9) that

$$(3.14) \quad DT(\tilde{y}, \nu) = \left. \frac{\partial x}{\partial y}(y, \nu) \right|_{\tilde{y}} = \frac{\partial^2 S(\tilde{y}, \nu)}{\partial \xi \partial y},$$

and the sampling density in y has to be chosen accordingly. Furthermore, the map $y_i = T_{\nu,k}^{-1}(x_i)$ yields irregularly spaced samples y_i from regularly spaced samples x_i , placed differently for each box (ν, k) . We point out that the evaluation of the sum over boxes $\sum_{\nu,k}$ in (3.11) requires $(Fu_{\nu,k})(y)$ to be evaluated on discrete points y_n that are common for all boxes. We therefore compute $(Fu_{\nu,k})(y_n)$ for points y_n on a rectangular grid defined by an (arbitrarily chosen) common reference point $y_{n,0}$, and global sampling density $\Delta y = (1/N) \inf_{i,\nu} \text{ev}(DT^{-1}(x_i, \nu))$. Alternatively, we can adapt the grid resolution locally through a hierarchical set of resolution levels $\{\Delta_l y\}$, reflecting (3.14) and constructed, for instance, in a multiresolution manner as $\{\Delta_l y = 2^l \Delta y\}$, $l = 0, 1, \dots$. The USFFTs in (3.11) are now evaluated from discrete frequencies $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to irregularly spaced discrete samples $x_n = T_{\nu,k}(y_n)$ and realize the coordinate transform onto the grid y_n . This completes our discretization (3.11) of (3.2).

Computational domain. In general, only a fraction of the wave packets $\varphi_{\gamma'}, \gamma' : k' = k, \nu' = \nu$ yield numerically significant contributions to $u_{\nu,k}$ and $(Fu_{\nu,k})$, resulting in effective compression in the wave packet domain [10, 53]. This reduces the computational domain on which $(Fu_{\nu,k})(y_n)$ actually needs to be evaluated (cf. schematic illustration in Figure 4). Indeed, the wave packets $\varphi_\gamma(x)$ and $\tilde{\phi}_{j,\nu,k}(\mathcal{E}(x))$ have, to precision ε , support in a box $b_{j,\nu,k} = l_k' \times (l_k'')^{d-1}$ and $\tilde{b}_{j,\nu,k} = \tilde{l}_k' \times (\tilde{l}_k'')^{d-1}$, respectively, with $l_k', \tilde{l}_k' \sim 2^{-k}$, $l_k'', \tilde{l}_k'' \sim 2^{-k/2}$, and their volumes decay as $\mathcal{O}(2^{-k} 2^{-k \frac{d-1}{2}})$ with scale k (cf. (1.7)).

The rate of compression and the resulting reduction in computational domain, the output sampling density Δy , and the oversampling factor ζ are data and problem dependent. Below, we consider them as being absorbed in one common oversampling factor ζ .

3.3. “Box” algorithm. We can now summarize and analyze the sequence of operations for the evaluation of (3.11). We first consider a single box and evaluate $(Fu_{\nu,k})(y_n)$. Assuming that the “inner adjoint” discrete transform (3.9) for zero-padded data $u^{zp}(x_i)$, $\hat{u}^{zp}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$ is given, we perform the following operations.

BOX ALGORITHM (FOR SINGLE BOX (ν, k)).

1. for each tensor product term, $r = 1, \dots, R_{\nu,k}$:
 - (a) evaluate tensor product functions $\alpha_{\nu,k}^{(r)}(y_n)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi_l^{\nu,k})$, $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}$
 - (b) multiply $\hat{u}^{zp}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$ with $\hat{\vartheta}_{\nu,k}^{(r)}(\xi_l^{\nu,k})$
 - (c) compute adjoint USFFT of (b) from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = T_{\nu,k}(y_n)$:

$$\Phi_{\nu,k}^{(r)}(x_n) = \sum_{\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)} e^{i\langle x_n, \xi_l^{\nu,k} \rangle} \hat{u}^{zp}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})\hat{\vartheta}_{\nu,k}^{(r)}(\xi_l^{\nu,k})$$
 - (d) multiply $\Phi_{\nu,k}^{(r)}(x_n)$ with amplitudes $\alpha_{\nu,k}^{(r)}(y_n)$
2. sum $R_{\nu,k}$ tensor-product contributions:

$$(Fu_{\nu,k})(y_n) \approx \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y_n)\Phi_{\nu,k}^{(r)}(x_n)$$

The number of operations, including explicitly the constants involved, is

- $\mathcal{O}(cR_{\nu,k}(\zeta N)^d)$ for evaluation of tensor product functions,³
- $\mathcal{O}(R_{\nu,k}(\zeta N)^d)$ for multiplications and additions,
- $\mathcal{O}(dR_{\nu,k}(\sigma_u \zeta N)^d \log(N))$ for USFFTs, where σ_u is the oversampling factor of the USFFT,

and the complexity of the box algorithm is therefore

$$(3.15) \quad \sim \mathcal{O}(dN^d \log(N)).$$

We can slightly modify the algorithm and reduce the number of USFFTs by substitution with standard FFTs, decreasing the constants in (3.15) and hence computation time. We assume here that the “inner adjoint” discrete transform (3.9) of the *original* data $u(x_i)$, $\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$ is given. We first obtain the box contribution $u_{\nu,k}(x_i)$ via USFFT, zero-pad it, and compute its FFT, inducing regularly spaced frequencies $\tilde{\xi}_j$. Now computations are performed on x_i and $\tilde{\xi}_j$, and standard FFTs replace the USFFTs in 1(c). Eventually, the change of coordinates to $x_n = T_{\nu,k}(y_n)$ is evaluated by a single USFFT, as follows.

MODIFIED BOX ALGORITHM (FOR SINGLE BOX (ν, k)).

1. adjoint USFFT of $\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$ from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to x_i
2. zero-pad and compute FFT
3. for each tensor product term, $r = 1, \dots, R_{\nu,k}$:
 - (a) evaluate tensor product functions $\alpha_{\nu,k}^{(r)}(y_i)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\tilde{\xi}_l)$, $\tilde{\xi}_j \in \mathbf{1}_{\nu,k}$
 - (b) multiply $\hat{u}^{zp}(\tilde{\xi}_j)\hat{\beta}_{\nu,k}(\tilde{\xi}_j)\hat{\chi}_{\nu,k}(\tilde{\xi}_j)$ with $\hat{\vartheta}_{\nu,k}^{(r)}(\tilde{\xi}_j)$
 - (c) compute inverse FFT of (b):

$$\Phi_{\nu,k}^{(r)}(x_i) = \sum_j e^{i\langle x_i, \tilde{\xi}_j \rangle} \hat{u}^{zp}(\tilde{\xi}_j)\hat{\beta}_{\nu,k}(\tilde{\xi}_j)\hat{\chi}_{\nu,k}(\tilde{\xi}_j)\hat{\vartheta}_{\nu,k}^{(r)}(\tilde{\xi}_j)$$
 - (d) multiply $\Phi_{\nu,k}^{(r)}(x_i)$ with amplitudes $\alpha_{\nu,k}^{(r)}(y_i)$
4. sum $R_{\nu,k}$ tensor-product contributions $\alpha_{\nu,k}^{(r)}(y_i)\Phi_{\nu,k}^{(r)}(x_i)$ and compute FFT of sum
5. compute adjoint USFFT of (4) from $\tilde{\xi}_j \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = T_{\nu,k}(y_n)$:

$$(Fu_{\nu,k})(y_n) \approx \sum_{r=1}^{R_{\nu,k}} \alpha_{\nu,k}^{(r)}(y_n)\Phi_{\nu,k}^{(r)}(x_n)$$

This modified algorithm requires $R_{\nu,k} + 2$ FFTs and only two USFFTs. The computational complexity remains the same as for the original box algorithm and is given by (3.15).

The action of F on u is now given by $\sum_{\nu,k} (Fu_{\nu,k})(y_n)$, the sum of the contributions of all significant boxes (ν, k) . Assuming that all $D \sim N^{\frac{d-1}{2}}$ boxes contribute,

³The evaluation of a PSWF at one point is $\mathcal{O}(c)$ [62].

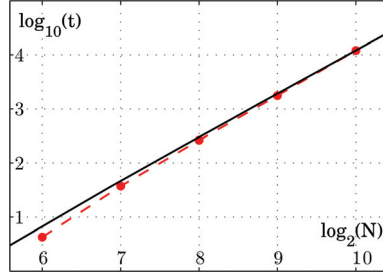


FIG. 5. Computation time as a function of sample size N (red dots and broken line) and complexity estimate (3.16) (black solid line) for parametriz of half-wave equation (cf. section 5.1) in $d = 2$ dimensions in homogeneous medium ($v = 2\text{km/s}$; evolution time is $T = 5\text{s}$).

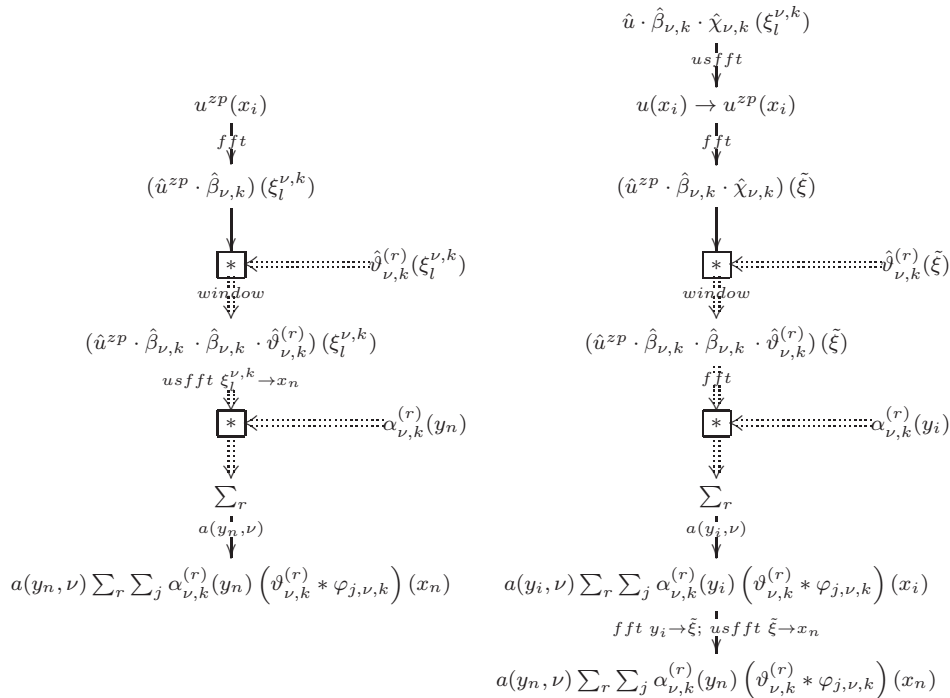


FIG. 6. Box algorithm (left) with FFTs replacing USFFTs (right) for one box (ν, k) . Double arrows indicate operations performed for each individual tensor-product term, $r = 1, \dots, R_{\nu,k}$.

the complexity of the above algorithms for the evaluation of (3.11) is

$$(3.16) \quad \sim \mathcal{O}\left(dN^{\frac{3d-1}{2}} \log(N)\right).$$

Actual computation time as a function of problem size N for $d = 2$ ($\mathcal{D} = 1$) is plotted in Figure 5 and compared to the complexity estimate (3.16). The diagrams in Figure 6 schematically summarize the box algorithm and the modified box algorithm.

3.4. Further approximations: “Packet” algorithms. We proceed with the further approximations (2.8) and (2.9) and describe algorithms for their evaluation on the discrete set of points y_n . Both approximations are tied to individual wave

packets, since the functions $\hat{\vartheta}_\gamma(\xi)$ are identified with a single wave packet $\varphi_\gamma(x)$. Consequently, the modified packets $\tilde{\phi}_{j,\nu,k}(\mathcal{E}(x_n)) = (\vartheta_\gamma * \varphi_{j,\nu,k})(\mathcal{E}(x_n))$ have to be constructed one at a time. Note that under approximation (2.9), the expansion of the coordinate transform also ties the change of coordinates to individual wave packets, yielding a pure “packet” algorithm, whereas under approximation (2.8) the change of coordinates can still be evaluated for all packets of a box (ν, k) at once since the coordinate transform $T_{\nu,k}$ is independent of index j , and we obtain a “hybrid packet-box” algorithm. The input to both algorithms are the data wave packet coefficients u_γ (cf. (3.4)), assumed to be obtained for zero-padded data $u^{zp}(x_i)$.

HYBRID PACKET-BOX ALGORITHM FOR APPROXIMATION (2.8).

- for each box (ν, k) :
 1. for each coefficient, $\gamma' : k' = k, \nu' = \nu$:
 - (a) set $u_{\gamma'}|_{j \neq j'} = 0$, FFT $u_{\gamma'}$ to $\xi_l^{\nu',k} \in \mathbf{1}_{\nu,k}(\xi)$
 - (b) evaluate window function $\hat{\vartheta}_{\gamma'}(\xi_l^{\nu',k})$
 - (c) multiply $\hat{u}_{\gamma'}(\xi_l^{\nu',k})\hat{\beta}_{\nu,k}(\xi_l^{\nu',k})\hat{\chi}_{\nu,k}(\xi_l^{\nu',k})$ with $a(y_j^{\nu,k}, \nu)\hat{\vartheta}_{\gamma'}(\xi_l^{\nu',k})$
 2. sum $\hat{\Phi}_{\nu,k}(\xi_l^{\nu,k}) = \sum_{\gamma'} \hat{u}_{\gamma'}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})a(y_j^{\nu,k}, \nu)\hat{\vartheta}_{\gamma'}(\xi_l^{\nu,k})$
 3. compute adjoint USFFT of $\hat{\Phi}_{\nu,k}(\xi_l^{\nu,k})$ from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = T_{\nu,k}(y_n)$: $(Fu_{\nu,k})(y_n) \approx \sum_{\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)} e^{i\langle x_n, \xi_l^{\nu,k} \rangle} \hat{\Phi}_{\nu,k}(\xi_l^{\nu,k}) = \sum_j a(y_j^{\nu,k}, \nu) (\vartheta_\gamma * \varphi_\gamma)(x_n)$
- sum the contributions of the individual boxes (ν, k) .

PACKET ALGORITHM FOR APPROXIMATION (2.8).

- for each coefficient γ :
 1. set $u_\gamma|_{j \neq j'} = 0$, FFT u_γ to $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$
 2. evaluate window function $\hat{\vartheta}_\gamma(\xi_l^{\nu,k})$
 3. multiply (a) and (b):
 $\hat{\Phi}_\gamma(\xi_l^{\nu,k}) = \hat{u}_\gamma(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})a(y_j^{\nu,k}, \nu)\hat{\vartheta}_\gamma(\xi_l^{\nu,k})$
 4. compute adjoint USFFT of $\hat{\Phi}_\gamma(\xi_l^{\nu,k})$ from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = (DT_\gamma(y_n - y_j^{\nu,k}) + M_\gamma \cdot (y_n - y_j^{\nu,k})^2)$:
 $(F\varphi_\gamma)(y_n) \approx \sum_{\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)} e^{i\langle x_n, \xi_l^{\nu,k} \rangle} \hat{\Phi}_\gamma(\xi_l^{\nu,k})$
- sum the contributions $(F\varphi_\gamma)(y_n)$ of the individual packets.

Evaluated for a single wave packet $\varphi_\gamma(x)$, both algorithms have complexity:

$$(3.17) \quad \sim \mathcal{O}(dN^d \log(N)).$$

Assuming that all $\mathcal{O}(N^d)$ data wave packets are significant,⁴ the evaluation of $(Fu)(y_n)$ requires $\mathcal{O}(dN^{2d} \log(N))$ operations with the packet algorithm and $\mathcal{O}(dN^{\frac{3d+1}{2}} \log(N))$ operations with the hybrid packet-box algorithm. The hybrid packet-box algorithm hence has complexity above the box algorithm but below the packet algorithm, since we can perform the coordinate transform via a USFFT per box (ν, k) . The diagrams in Figure 7 schematically summarize the hybrid packet-box algorithm and the packet algorithm.

⁴Note that this assumption is unrealistic in many applications, where typically the number of data wave packets with practically nonzero coefficients amounts to a small fraction.

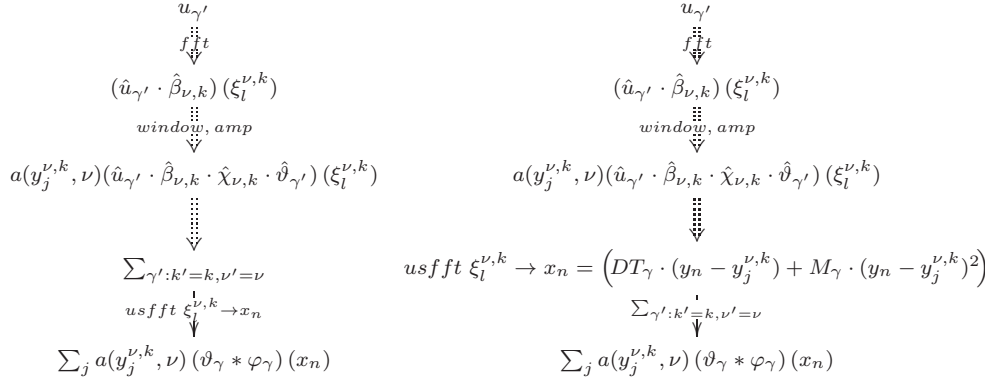


FIG. 7. Hybrid box-packet algorithm (left) and packet algorithm (right) for one box (ν, k) . Double arrows indicate operations performed for each individual wave packet.

4. Parametrix. Evaluation of approximations (2.6), (2.8), and (2.9) with the proposed algorithms requires knowledge of the values of the first- and second-order derivatives of the generating function $S(y, \xi)$. Here, we detail how these derivatives can be computed numerically for parametrices of evolution equations. Evolution equations play an important role in inverse scattering applications and general extended imaging [26, 27]. We obtain the derivatives of S from the Hamilton system describing the propagation of singularities and from the fundamental matrix of the Hamilton–Jacobi system for perturbations of the bicharacteristics.

Effectively, the numerical procedures described in the previous section yield (approximate) solvers for Cauchy initial value problems for evolution equations from the initial time to arbitrarily large later time. As a special case, we revisit in subsection 4.2 “thin-slab” propagation, in which straight rays and closed form expressions approximate the first- and second-order terms of the phase expansion for small time steps and obtain a directionally developed paraxial approximation. The result is closely related to so-called beam migration [52].

4.1. Hamiltonian system and perturbed system. We consider evolution equations of type

$$(4.1) \quad [\partial_t + ip(t, x, D_x)]u(t, x) = 0, \quad u(t_0, x) = u_0(x),$$

on a domain $X \subset \mathbb{R}^d$ and on the interval $t \in [t_0, T]$, where p is a pseudodifferential operator with symbol P in $S_{1,0}^1$ (in the case of the half-wave equation, $P = P(x, \xi) = \sqrt{c(x)^2 ||\xi||^2}$), and denote the associated parametrix by F , $u(t, y) = (F(t, t_0)u_0)(y)$, $F(t_0, t_0) = \text{Id}$. We introduce the Hamiltonian system that gives the propagation of singularities, cf. (1.2), for (4.1). For every $(x, \xi) \in \mathbb{R}^d \times \mathbb{R}^d \setminus \{0\}$, the integral curves $(y(x, \xi; t, t_0), \eta(x, \xi; t, t_0))$ of

$$(4.2) \quad \frac{dy}{dt} = \frac{\partial P(t, y, \eta)}{\partial \eta}, \quad \frac{d\eta}{dt} = -\frac{\partial P(t, y, \eta)}{\partial y}$$

with initial conditions $y(x, \xi; t_0, t_0) = x$ and $\eta(x, \xi; t_0, t_0) = \nu$ at time $t = t_0$ define the mapping from $(x, \xi; t, t_0)$ to (y, η) , which is the canonical relation of the solution operator of (4.1). Integrating the system (4.2) from t_0 to T hence yields the map

$$(4.3) \quad y(x, \nu; T, t_0) = T_{\nu,k}^{-1}(x).$$

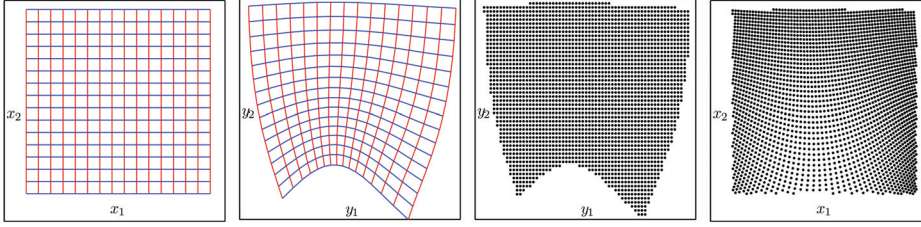


FIG. 8. Schematic illustration of discrete evaluation of the coordinate transform $T_{\nu,k}$: $y_m = T_{\nu,k}^{-1}(x_m)$ from regularly spaced x_m (left); interpolation of $x_m = T_{\nu,k}(y_m)$ at regularly spaced y_n gives $x_n = T_{\nu,k}(y_n)$ (right).

Under the assumption of absence of caustics, ξ and y determine η and x . We note that for T sufficiently close to t_0 the assumption is always satisfied. In approximations (2.6) and (2.8), the numerical evaluation of $T_{\nu,k}$ for the predefined (regular) grid y_n , $x_n = T_{\nu,k}(y_n)$, is performed by backward ray tracing from y_n subject to $\xi_n/||\xi_n|| = \nu$. Alternatively, we first integrate (4.2) for initial conditions (x_m, ν) with x_m a discrete set of points on $\tilde{\mathcal{V}}_{\nu,k}$ and obtain the map $x_m = T_{\nu,k}(y_m^T)$; interpolation on the grid y_n then yields the desired map $x_n = T_{\nu,k}(y_n)$ (cf. Figure 8).

Now consider the perturbations of (y, η) w.r.t. initial conditions (x, ξ) :

$$(4.4) \quad W(x, \xi; t, t_0) = \begin{pmatrix} W_1 & W_2 \\ W_3 & W_4 \end{pmatrix} = \begin{pmatrix} \partial_x y & \partial_\xi y \\ \partial_x \eta & \partial_\xi \eta \end{pmatrix}.$$

The system for the $2d \times 2d$ matrix W is given by the Hamilton–Jacobi equations,

$$(4.5) \quad \frac{dW}{dt}(x, \xi; t, t_0) = \begin{pmatrix} \partial_{\eta y} P(t, y, \eta) & \partial_{\eta \eta} P(t, y, \eta) \\ -\partial_{yy} P(t, y, \eta) & -\partial_{y\eta} P(t, y, \eta) \end{pmatrix} W(x, \xi; t, t_0),$$

which are integrated for initial conditions $W|_{t=t_0} = \mathbb{I}_{2d}$. Note that under our assumptions, the $d \times d$ submatrix W_1 is always invertible. Since $x = \frac{\partial S}{\partial \xi}$ and $\eta = \frac{\partial S}{\partial y}$ (cf. (1.2)), integration of (4.5) along (y, η) for t_0 to T yields

$$(4.6) \quad \frac{\partial^2 S}{\partial y \partial \xi}(y, \xi; T, t_0) = \frac{\partial x}{\partial y} = W_1^{-1},$$

$$(4.7) \quad \frac{\partial^2 S}{\partial \xi^2}(y, \xi; T, t_0) = \frac{\partial x}{\partial \xi} = \frac{\partial x}{\partial y} \frac{\partial y}{\partial \xi} = -W_1^{-1} W_2,$$

$$(4.8) \quad \frac{\partial^2 S}{\partial y^2}(y, \xi; T, t_0) = \frac{\partial \eta}{\partial y} = \frac{\partial \eta}{\partial x} \frac{\partial x}{\partial y} = W_3 W_1^{-1},$$

which we evaluate at points (y_n, ν) .

The leading-order amplitude follows to be

$$(4.9) \quad a(y, \nu; T, t_0) = \sqrt{1/\det W_1(x^{t_0}(y, \nu; T, t_0), \nu; T, t_0)},$$

where $x^{t_0}(y, \nu; T, t_0)$ is the backward solution to (4.2) with initial time T , evaluated at t_0 . The system (4.5) is given in Cartesian coordinates and can be reduced to a paraxial system evaluated in Fermi- or ray-centered coordinates; see, e.g., [36] and [59]. This reduced paraxial system and the expressions for the matrices $\frac{\partial^2 S}{\partial \xi^2}$, DT_γ ,

and M_γ in terms of its fundamental matrix are given in Appendix A. We finally detail the expression for the propagation of a wave packet φ_γ :

$$(F(t, t_0)\varphi_\gamma)(y) = \int \sqrt{\frac{1}{\det W_1(y, \nu)}} e^{i\langle \xi, x^{t_0}(y, \nu) - x_j^{\nu, k} \rangle} \hat{\chi}_{\nu, k}(\xi) e^{-i\frac{1}{2\xi'} \langle \xi'', [W_1^{-1}(y, \nu)W_2(y, \nu)]'' \xi'' \rangle} d\xi.$$

Here, $(\cdot)''$ indicates the square submatrix with entries corresponding to the coordinates of ξ'' .

4.2. Example: Trotter product. We analyze approximations (2.6), (2.8), and (2.9) for the evolution equation (4.1) for the specific case of discretization of evolution time into a sequence of small time steps. The solution operator $F(t, t_0)$ can be written in the form of a Trotter product, resulting in a computational scheme driven by marching-on-in- t . If $t \geq t_N > t_{N-1} > \dots > t_0$, we let the operator $\mathcal{W}_N(t, t_0)$ be defined as $\mathcal{W}_N(t, t_0) = \bar{F}(t, t_N) \Pi_{i=N}^1 \bar{F}(t_i, t_{i-1})$, assuming that $T \geq t_{N+1} \geq t \geq t_N$. We have $\Delta_i = t_i - t_{i-1}$, $\Delta_i \leq \Delta = \mathcal{O}(N^{-1})$ as $N \rightarrow \infty$. We consider a single component operator $\bar{F}(t_{i-1} + \Delta_i, t_{i-1})$ and set $t' = t_{i-1}$ and $\Delta = \Delta_i$. It can be approximated by the “short-time” propagator, given by

$$(4.10) \quad \bar{F}(t' + \Delta, t')u(t', \cdot)(y) = (2\pi)^{-n} \int \exp[i(P(t', y, \xi)\Delta - \langle \xi, y \rangle)] \hat{u}(t', \xi) d\xi.$$

This is an FIO of order 0 in the class considered in this paper with the simple substitution

$$(4.11) \quad a(y, \xi) = 1, \quad S(y, \xi) = P(t', y, \xi)\Delta - \langle \xi, y \rangle.$$

The associated canonical transformation is given by

$$\chi : (-\partial_\xi P(t', y, \xi)\Delta + y, \xi) \rightarrow (y, -\partial_y P(t', y, \xi)\Delta + \xi);$$

with the Hamilton system,

$$(4.12) \quad \frac{dx}{dt} = \frac{\partial P}{\partial \xi}(t, x, \xi), \quad \frac{d\xi}{dt} = -\frac{\partial P}{\partial x}(t, x, \xi),$$

it follows that

$$\chi : \left(y - \frac{dx}{dt}(t', y, \xi)\Delta, \xi \right) \rightarrow \left(y, \xi + \frac{d\xi}{dt}(t', y, \xi)\Delta \right),$$

which describes straight rays in the interval $[t', t' + \Delta]$. The canonical transformation χ reflects a numerical integration scheme for the Hamilton system, viz., the Euler method.

The first-order term in the expansion of the phase yields $T_{\nu, k} = \partial_\xi P(t', y, \nu)$. Under the map $T_{\nu, k}$, y follows from solving $x + \partial_\xi P(t', y, \nu)\Delta = y$, which involves backtracking a straight ray that connects $(t' + \Delta, y)$ with (t', x) . The second-order term in the expansion, $(\partial_{\xi''^2} P)(t', y, \nu)$, is directly related to solving the Hamilton–Jacobi system for paraxial rays (in ray-centered coordinates) using Euler’s method and discretization step Δ , as discussed in detail in the previous section.

In the case of so-called depth extrapolation [16], t is replaced by the depth z and x is replaced by the transverse coordinates and time, $(x, t) \in \mathbb{R}^n$. The principal symbol of P becomes

$$(4.13) \quad P(z, (x, t), (\xi, \tau)) = -\tau \sqrt{c(z, x)^{-2} - \tau^{-2} |\xi|^2}$$

and

$$(4.14) \quad S((y, t), (\xi, \tau)) = P(z', (y, t), (\xi, \tau)) \Delta - \langle \xi, y \rangle - \tau t.$$

We introduce (ξ_ν, τ_ν) using projective coordinates $(\tau_\nu^{-1} \xi_\nu, 1) / \sqrt{\tau_\nu^{-2} |\xi_\nu|^2 + 1} = \nu$, $\tau_\nu \neq 0$; ν determines $\tau_\nu^{-1} \xi_\nu$, and the propagation direction at depth z' , $c(z', y)(\tau_\nu^{-1} \xi_\nu, \sqrt{c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2})$. The expansion of S yields the (principal) symbol of the paraxial wave equation, directionally developed relative to ν ,

$$(4.15) \quad \begin{aligned} \frac{\partial P}{\partial \xi}(z', (y, t), \nu) &= \frac{\tau_\nu^{-1} \xi_\nu}{\sqrt{c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2}}, \\ \frac{\partial P}{\partial \tau}(z', (y, t), \nu) &= -\frac{c(z', y)^{-2}}{\sqrt{c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2}}, \end{aligned}$$

(in the classical paraxial expansion, $\xi_\nu = 0$), and

$$(4.16) \quad \begin{aligned} \tau_\nu \frac{\partial^2 P}{\partial \xi^2}(z', (y, t), \nu) &= \frac{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2] I - \tau_\nu^{-2} \xi_\nu \otimes \xi_\nu}{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2]^{3/2}}, \\ \tau_\nu \frac{\partial^2 P}{\partial \tau^2}(z', (y, t), \nu) &= -\frac{c(z', y)^{-2} \tau_\nu^{-2} |\xi_\nu|^2}{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2]^{3/2}}, \\ \tau_\nu \frac{\partial^2 P}{\partial \xi \partial \tau}(z', (y, t), \nu) &= -\frac{c(z', y)^{-2} \tau_\nu^{-1} \xi_\nu}{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2]^{3/2}}. \end{aligned}$$

Hence, with $(\xi', \xi'') = R_\nu^{-1}(\xi, \tau)$ and $\xi'' = \tilde{R}_\nu^{-1}(\xi, \tau)$,⁵ and with

$$\begin{aligned} \partial_{(\xi', \xi'')} P(\cdot, (\cdot, \cdot), R_\nu(\xi', \xi'')) &= R_\nu^{-1}(\partial_{(\xi, \tau)} P)(\cdot, (\cdot, \cdot), R_\nu(\xi', \xi'')), \\ \partial_{\xi''^2} P(\cdot, (\cdot, \cdot), R_\nu(\xi', \xi'')) &= \tilde{R}_\nu^{-1} \left[\tilde{R}_\nu^{-1}(\partial_{(\xi, \tau)^2} P)(\cdot, (\cdot, \cdot), R_\nu(\xi', \xi'')) \right]^T, \end{aligned}$$

the expression for the phase expansion of the operator is

$$(4.17) \quad \begin{aligned} &\left\langle \xi, \frac{\partial P}{\partial \xi}(z', (y, t), \nu) \right\rangle + \tau \frac{\partial P}{\partial \tau}(z', (y, t), \nu) + \frac{1}{2\xi'} \left\langle \xi'', \frac{\partial^2 P}{\partial \xi''^2}(z', (y, t), \nu) \xi'' \right\rangle \\ &= \frac{\langle \xi, \tau_\nu^{-1} \xi_\nu \rangle - \tau c(z', y)^{-2}}{\sqrt{c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2}} \\ &+ \frac{1}{2\xi'} \left\langle \xi'', \left(\tilde{R}_\nu^{-1} \left[\tau_\nu^{-1} \tilde{R}_\nu^{-1} \left(\begin{aligned} &\frac{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2] I - \tau_\nu^{-2} \xi_\nu \otimes \xi_\nu}{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2]^{3/2}} - \frac{c(z', y)^{-2} \tau_\nu^{-1} \xi_\nu}{[c(z', y)^{-2} - \tau_\nu^{-2} |\xi_\nu|^2]^{3/2}} \end{aligned} \right)^T \right]^T \right) \xi'' \right\rangle. \end{aligned}$$

⁵That is, \tilde{R}_ν^{-1} is R_ν^{-1} without the first row.

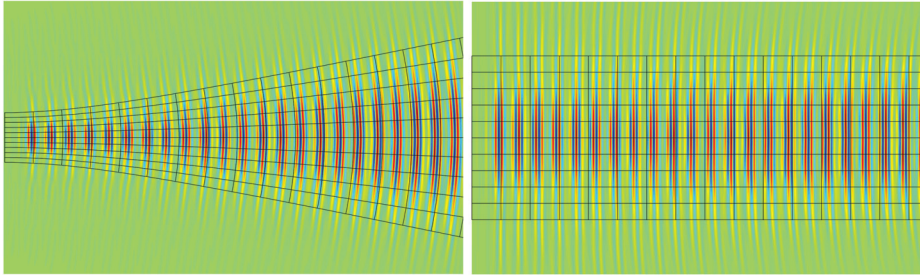


FIG. 9. A beam of wave packets in homogeneous background under approximation (2.6) for the half-wave equation in Cartesian coordinates (x, z) (left; z horizontal) and elliptic coordinates $x = a \cosh(\mu) \cos(\varsigma)$, $z = a \sinh(\mu) \sin(\varsigma)$ (right; μ horizontal); elliptic coordinate system (black grids). The horizontal elliptic coordinate axis on the right has been transformed according to $\tilde{\mu} = \sinh(\mu)$ in order to achieve regular horizontal spacing. Propagation is confined to a tube in curvilinear coordinates.

Indeed, for $\xi_\nu = 0$ (that is, $\xi' = \tau$ and $\xi'' = \xi$), this expression reduces to the standard paraxial (15°) approximation $-\tau c(z', y)^{-1} + \frac{1}{2} \frac{|\xi|^2}{\tau} c(z', y)$; then $T_{\nu, k}$ defines the so-called comoving frame of reference. We refer to the corresponding “short-time” propagator as the “thin-slab” propagator.

The operator $\mathcal{W}_N(z, z_0)$ is reminiscent of the Trotter product representation of the parametrix:⁶ it converges in the Sobolev operator norm to $F(t, t_0)$ as $\Delta^{s/2}$ with s depending on the Hölder regularity α of P w.r.t. z : For $\frac{1}{2} \leq \alpha$, $s = 1$, and the balance of accuracies $\mathcal{O}(\Delta^{1/2})$ and $\mathcal{O}(2^{-k/2})$ requires $\Delta \sim 2^{-k}$ [19, 45]. The underlying multiproduct of FIOs can be estimated using the Kumano-go–Taniguchi theorem [37].

We can now construct a process similar to (back) propagation in “beam migration.” We decompose the data into its wave packet components. Each wave packet initializes a solution to the (half-)wave equation, which, through the Trotter product representation, reveals a *phase-space localized* paraxial approximation. The standard paraxial approximation is commonly exploited in beam migration, for example, expressed in terms of geodesic coordinates. In Figure 9 (left), we show curvilinear coordinates particular to wave packets, which enable us to define tubes to which the propagation is confined⁷ (see, e.g., [9, Figures 1 and 2]).

5. Applications and numerical examples. Here, we illustrate and compare the proposed approximations for $(Fu)(y)$ in numerical examples for propagators corresponding to evolution equations in $d = 2$ dimensions. In the first examples, we illustrate Cauchy initial value problems for the half-wave equation in isotropic homogeneous medium and in isotropic heterogeneous medium. The second example demonstrates evolution equation based imaging and involves an anisotropic but homogeneous Hamiltonian.

5.1. Wave propagation: Isotropic, heterogeneous case. We consider the initial value problem (4.1) for the half-wave equation, i.e., with symbol

$$(5.1) \quad P(x, \xi) = \sqrt{c(x)^2 \|\xi\|^2},$$

⁶Geometrically, $\mathcal{W}_N(z, z_0)$ has some similarities with the wave front construction method for computing the propagation of singularities.

⁷Here, we use elliptic coordinates $x = a \cosh(\mu) \cos(\varsigma)$, $z = a \sinh(\mu) \sin(\varsigma)$. In $d = 3$ dimensions, the corresponding curvilinear coordinates are the oblate spheroidal coordinates, $x = a \cosh(\mu) \cos(\varsigma) \cos(\phi)$, $y = a \cosh(\mu) \cos(\varsigma) \sin(\phi)$, $z = a \sinh(\mu) \sin(\varsigma)$, with tubes in the z direction.

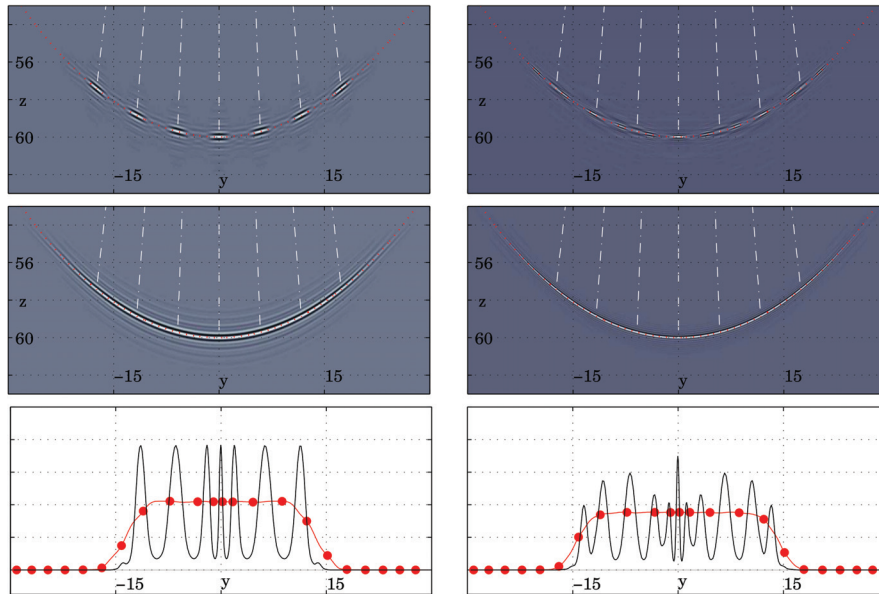


FIG. 10. Wave propagation in isotropic, homogeneous medium for initial conditions (5.2) with $k' = 2$ (left column) and $k' = 1 - 4$ (right column): zero order approximation (top row), approximations (2.6), (2.8), and (2.9) (center row), and corresponding amplitudes along wave front (bottom row; solid black line corresponds to zero order approximation). The white dot-dashed lines indicate rays of seven wave packets at scale $k = 3$. Note that the aspect ratio is not equal to one.

where $c(x)$ is the medium velocity. We compare the accuracy of the box algorithm for approximation (2.6), the hybrid packet-box algorithm for approximation (2.8), and the packet algorithm for approximation (2.9) to zero order approximation (2.4) for large evolution time T . The initial data $u_0(x)$ are a band-limited Dirac, defined in the ξ domain as

$$(5.2) \quad \hat{u}_0(\xi) = \sum_{k'} \sum_{|\nu - \nu_c| \leq \Delta} \hat{\chi}_{\nu, k'}(\xi),$$

i.e., \hat{u}_0 defines a wedge with half-opening angle Δ and smooth cutoff. We set $\nu_c = (0, 1)$ (vertical downward), $\Delta = 21$ degrees, and let the initial data domain extend over $x \in [-5km, 5km] \times [-5km, 5km]$, and insert $u_0(x)$ in its center. The initial data consist of $N \times N = 256 \times 256$ samples, resulting in maximum scale $k_{\max} = 4$. We consider two background velocities: homogeneous with $c(x) = c_0$, and heterogeneous with a low-velocity lens

$$(5.3) \quad c(x) = c_0 + \mu \exp(-|x - x_0|^2 / \sigma^2)$$

with $c_0 = 2km/s$, $\mu = -0.3km/s$, $\sigma = 5km$, and $x_0 = (0, 35)km$. The output spatial sampling density Δy is set equal to the initial sampling density Δx . We consider evolution time $T = 30s$ for the homogeneous case and $T = 20s$ for the heterogeneous case⁸ ($t_0 = 0$ throughout this section).

Figure 10 (homogeneous case) and Figure 11 (heterogeneous case) compare the different approximations of $(Fu)(y_n)$: zero order approximation (2.4) (top row), box

⁸With this parameter setup, the calculation domains containing $(Fu)(y)$ are rectangles of roughly $N_1 \times N_2 = 1900 \times 300$ and 1100×300 samples for the homogeneous and for the heterogeneous case, respectively.

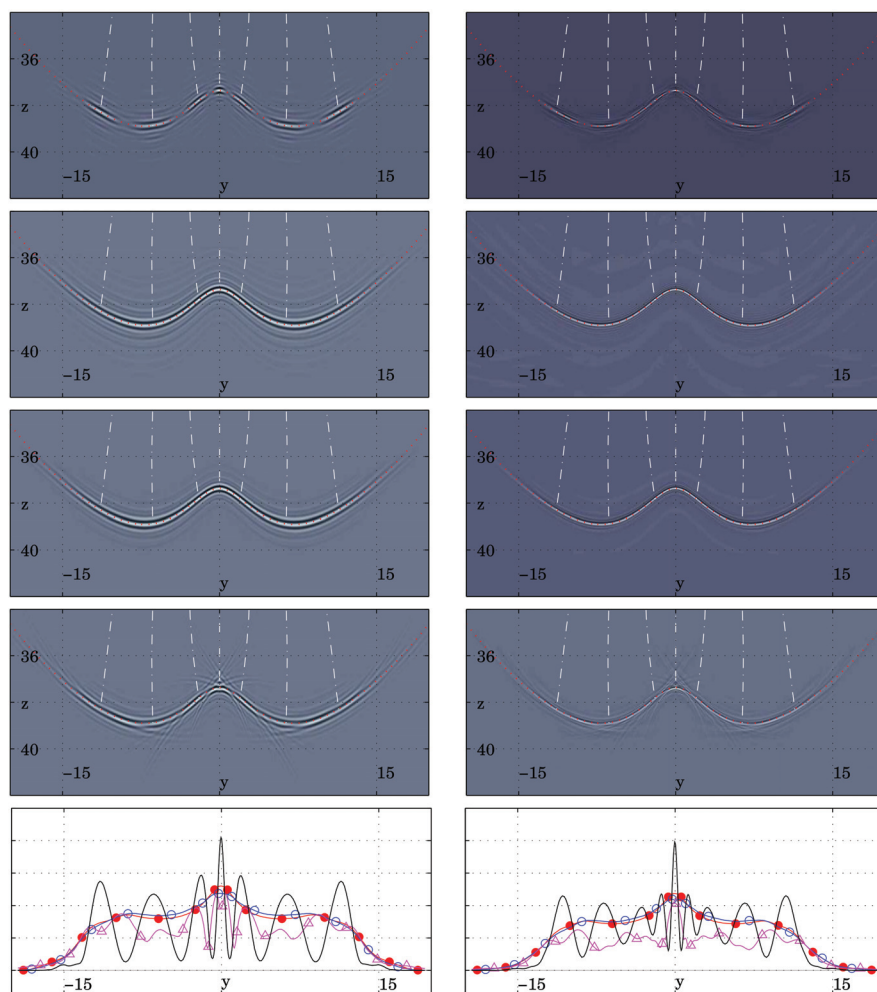


FIG. 11. Wave propagation in isotropic, heterogeneous medium for initial conditions (5.2) with $k' = 2$ (left column) and $k' = 1 - 4$ (right column) including physical amplitudes $a(y, \nu)$: zero order approximation (top row), approximation (2.6) (second row), (2.8) (third row), and (2.9) (fourth row); corresponding amplitudes along wave front (bottom row): zero order approximation (solid black), approximation (2.6) (red dot), approximation (2.8) (black circle), approximation (2.9) (triangle). The white dot-dashed lines indicate rays of seven wave packets at scale $k = 3$. Note that the aspect ratio is not equal to one.

algorithm approximation (2.6), hybrid packet-box algorithm approximation (2.8), and box algorithm approximation (2.9) (second row for homogeneous case, second to fourth rows for heterogeneous case). The bottom row compares the amplitudes along the wave front. The left column corresponds to initial condition (5.2) with frequency scale limited to $k' = 3$ only; the column on the right includes all frequency scales $k' = 1 - 4$.

We start with investigating the homogeneous case (cf. Figure 10). Note that in this case, approximations (2.6), (2.8), and (2.9) are equivalent, since $\frac{\partial^2 S}{\partial \xi^2}(y, \nu) = c_0 T$ is independent of y , $T_{\nu, k}$ describes, for fixed ν , parallel straight rays of path length $c_0 T$, $DT_\gamma = I_{d \times d}$, and $M_\gamma = 0_{d \times d}$. Also, the zero order approximation is

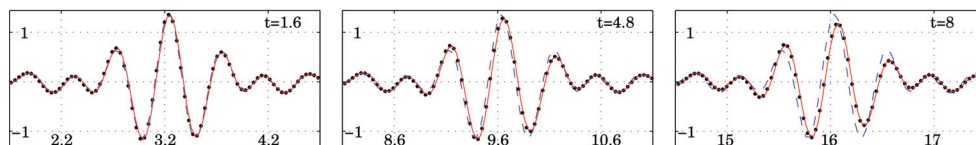


FIG. 12. Central cross sections of wave packets at frequency scale $k = 2$ propagating in isotropic, homogeneous medium for evolution times $t = [1.6s, 4.8s, 8s]$, respectively: zero order approximation (blue dashed line), approximation (2.6) (red solid line), time domain finite difference computation (black dots).

equal to rigid motion. As observed in [24], the wave front breaks apart under the zero order approximation with constituting wave packets at a given scale ending up disconnected (top row). The wave packets do not receive any deformation and are merely displaced data wave packets, resulting in large gaps in the wave front due to the geometry of propagation when $c_0 T$ is large w.r.t. initial data domain x . As a further consequence, only the center points of the wave packets sit exactly on the wave front. The error of the zero order approximation does not decrease with increasing scale. Indeed, including all scales $k' = 2 - 4$ does not fill up the gaps. In contrast, under the approximations to order $\mathcal{O}(2^{-k/2})$, the wave packets spread out and bend to perfectly align and overlap along the wave front, without any visible artifacts. These differences between zero order approximation and the proposed algorithms are also reflected by amplitudes along the wave front (bottom row). Unlike zero order approximation, which results in strong fluctuations regardless of scale k , amplitudes under approximations (2.6), (2.8), and (2.9) are essentially constant.

Furthermore, we point out the difference in phase between the zero order approximation and approximations (2.6), (2.8), and (2.9), which we illustrate in Figure 12. A wave packet evolving under the action of a propagator is subject to a phase rotation, which resides in the second-order term of the phase expansion on the frequency support of the wave packet entering the approximations to order $\mathcal{O}(2^{-k/2})$. These precisely match a finite difference reference computation both in phase and amplitude. It cannot be reproduced by the zero order approximation, which relies purely on the travel time and ray geometry for the central orientation ν .

We now turn our attention to the heterogeneous case; cf. Figure 11. As above, the wave front breaks apart under the zero order approximation (top row) with error not decreasing when scale k is increased. Only center points $y_l^{\nu,k}$ sit precisely on the singularity. In contrast, under approximation (2.6) (second row), the data wave packets bend, spread out, and connect along the singularity and form a visually perfect wave front. We note the dilations in the vicinity of the vertical symmetry axis at $x = 0$ caused by the low-velocity lens, resulting in packets being “squeezed” in their direction of propagation. Results obtained under approximation (2.8) (third row) are very similar, since in this example the dependence of $\frac{\partial^2 S}{\partial \xi^{1/2}}(y, \nu)$ on y is weak within the support of the individual wave packets. Note that under approximation (2.9) (fourth row), significant artifacts result from the additional approximate (second-order) expansion of the coordinate transform. The spatial extent of the background perturbation is too small w.r.t. the spatial extent of the modified wave packets for the expansion to be accurate on the entire support of the packets. In particular, we observe artifacts from wave packets that “stick out” of the wave front into regions toward the vertical symmetry axis, close to which the coordinate transform gradually contracts more and more violently due to the low velocity lens at position $(0, 35)km$.

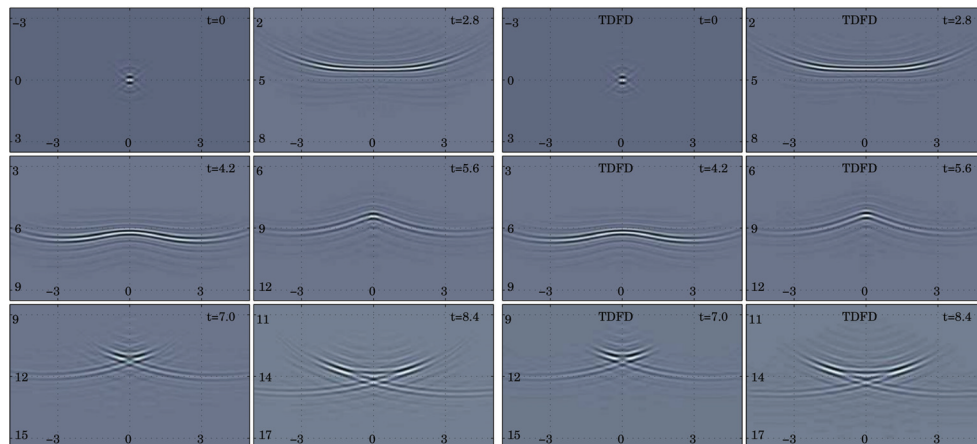


FIG. 13. Propagation through caustic using the semigroup property (columns 1–2): initial wave front (top left) and wave fronts at $t = [2.8s, 4.2s, 5.6s, 7.0s, 8.4s]$, obtained by stepwise continuation in time of the wave fronts obtained at $t = [0s, 2.8s, 2.8s, 5.6s, 5.6s]$, using approximation (2.6). Time domain finite difference computation (columns 3–4).

Nevertheless, approximation (2.9) appears to produce a more accurate image of the wave front than the zero order approximation. The above statements are further confirmed by investigation of the amplitudes along the wave fronts (bottom row): zero order approximation produces large gaps, while amplitudes under approximation (2.6) and (2.8) are nearly fluctuation free. Note that unlike zero order approximation, amplitude fluctuation under approximation (2.9) decreases for finer scales.

Remark 5.1. We note that by using the semigroup property, we can apply the procedures developed in this work also in the presence of caustics. We illustrate this in Figure 13, where we stepwise continue in time a wave front initiated by (5.2) with $\nu_c = (0, 1)$, $\Delta = 40$, $k' = 3$, through the low-velocity lens (5.3) with parameters $c_0 = 2km/s$, $\mu = -0.4km/s$, $\sigma = 3km$, and $x_0 = (0, 5)km$.

Limited aperture array retrofocussing via phase space localization. We apply the box algorithm for approximation (2.6) in a retrofocus experiment for one single wave packet φ_γ ,

$$(5.4) \quad F(0, T) (F(0, T)^* \varphi_\gamma)(x),$$

where F is the solution operator to (4.1) with symbol (5.1). As the background c we use a high-velocity lens, given by (5.3) with $c_0 = 2km/s$, $\mu = +0.3km/s$, $\sigma = 6km$, and $x_0 = (5, 16)km$. The initial conditions $u(x, t_0)$ consist of one single wave packet $\varphi_\gamma(x)$ at scale $k = k_0 = 3$ with $\nu = \nu_0 = (0, 1)$ in the vertical direction, depicted in Figure 14 (top left). The initial data are discretized at $N \times N = 512 \times 512$ sample points. Spatial sampling density Δy is set to equal the initial sampling density Δx , and the evolution time is $T = 8s$.

We begin with evaluating $\tilde{\phi}_\gamma(y) = (F(0, T)^* \varphi_\gamma)(y)$, plotted in Figure 14 (second row, left). Then, we compress $\tilde{\phi}_\gamma(y)$ by simple hard thresholding of wave packet coefficients with magnitude below 10% the magnitude of the largest coefficient. Note that significant boxes are concentrated in a narrow cone about the central wave vector of $\tilde{\phi}_\gamma(y)$. Finally, we evaluate $\tilde{\psi}_\gamma(x) = (F(0, T)\tilde{\phi}_\gamma)(x)$ on the limited aperture array detected by $\tilde{\phi}_\gamma(x)$ and obtain the retrofocused wave packet (Figure 14,

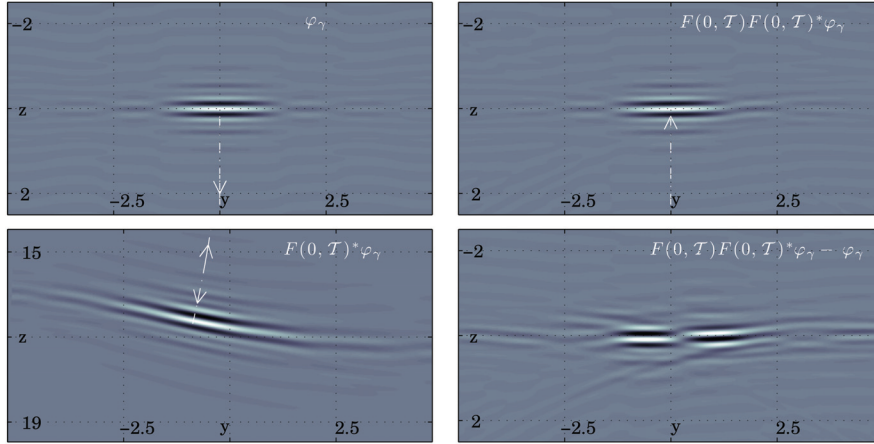


FIG. 14. Retrofocus experiment. Top row: initial single wave packet $\varphi_\gamma(x)$ at scale $k = k_0 = 3$ (left) and retrofocused wave packet $\tilde{\psi}_\gamma(x) = (FF^*\varphi_\gamma)(x)$ (right). Bottom row: downward propagated wave packet $\tilde{\phi}_\gamma(y) = (F^*\varphi_\gamma)(y)$ (left) and difference $((FF^* - I)\varphi_\gamma)(x)$ (right, magnified by a factor of 8).

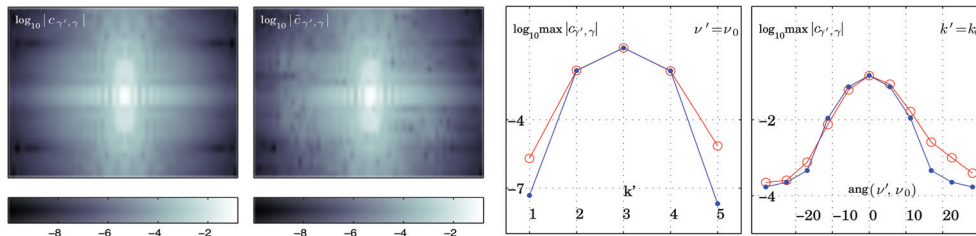


FIG. 15. Decay properties of $\tilde{c}_{\gamma',\gamma} = \langle \varphi_{\gamma'}, FF^*\varphi_\gamma \rangle$ versus $c_{\gamma',\gamma} = \langle \varphi_{\gamma'}, \varphi_\gamma \rangle$. Logarithmic magnitude of $c_{\gamma',\gamma}$ (left) and $\tilde{c}_{\gamma',\gamma}$ (center left) for the box $B_{\nu=\nu_0, k=k_0}$. Decay of coefficients $\max_j |c_{\gamma',\gamma}|$ and $\max_j |\tilde{c}_{\gamma',\gamma}|$ for $\nu' = \nu_0$ fixed as a function of $k' = k_0 \pm [0, 1, 2]$ (center right) and for $k' = k_0$ fixed as a function of $\text{ang}(\nu', \nu)$ (right); blue dots correspond to $c_{\gamma',\gamma}$, red circles to $\tilde{c}_{\gamma',\gamma}$.

top right). Figure 14 (second row, right) depicts the difference $\tilde{\psi}_\gamma(x) - \varphi_\gamma(x)$ between the retrofocused and the original wave packet, i.e., $(FF^* - I)\varphi_\gamma(x)$ (magnified by a factor of 8). In Figure 15, we visualize in more detail the decay of $\tilde{c}_{\gamma',\gamma} = \langle \varphi_{\gamma'}, \tilde{\psi}_\gamma \rangle = \langle \varphi_{\gamma'}, FF^*\varphi_\gamma \rangle$ away from the diagonal and compare it to the decay of the original wave packet, $c_{\gamma',\gamma} = \langle \varphi_{\gamma'}, \varphi_\gamma \rangle$: magnitude of $c_{\gamma',\gamma}$ (left) and $\tilde{c}_{\gamma',\gamma}$ (center left) for the box $(k' = k_0, \nu' = \nu_0)$; maxima of $c_{\gamma',\gamma}$ and $\tilde{c}_{\gamma',\gamma}$ as a function of scale k' ($\nu' = \nu_0$, center right) and of orientation ν' ($k' = k_0$, right). Note that this corresponds to the analysis of the decay properties of the kernel of the pseudodifferential operator FF^* . The propagated wave packet $\tilde{\phi}_\gamma(y) = (F^*\varphi_\gamma)(y)$ remains well localized in space. The original and retrofocused wave packets $\varphi_\gamma(x)$ and $\tilde{\psi}_\gamma(x)$ are visually very close, and $\tilde{\psi}_\gamma(x)$ essentially preserves the decay properties of $\varphi_\gamma(x)$ while detecting $\tilde{\phi}_\gamma(y)$ on a limited aperture array only. These properties can be exploited in illumination analysis [60], interferometry [46], and partial reconstruction [21].

5.2. Common-offset imaging: Anisotropic, homogeneous case. Many processes in seismic data analysis and imaging can be identified with solution operators of evolution equations. In [26], isochrons defined by imaging operators are

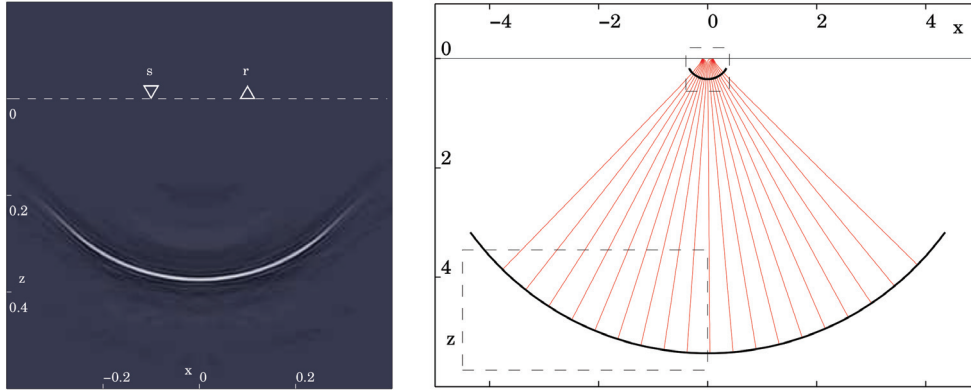


FIG. 16. Source-receiver geometry and initial band-limited isochron front (left). Geometry of evolution of initial isochron front under the flow of Hamiltonian (5.5) (right): initial isochron and front after evolution for $T = 5s$ (small and large black solid curves, respectively) and isochron rays (red solid). The small dashed rectangle corresponds to the region of the initial data depicted in the left figure, the larger dashed rectangle to the calculation domain.

identified with wave fronts of solutions of evolution equations. The bicharacteristics of the Hamiltonian associated with such evolution equations provide a natural way for implementing prestack map migration by evolution in the prestack imaging volume. We illustrate the principle of imaging with common-offset isochrons for homogeneous medium in $d = 2$ dimensions. The Hamiltonian governing the evolution of the common-offset isochron fronts is given by [26]

$$\begin{aligned} \mathcal{H}(y, z, \omega, k_y, k_z) &= \omega - \frac{c}{k_y z \sqrt{2}} \left(\frac{\sqrt{Q_- Q_+}}{\sqrt{Q_-} + \sqrt{Q_+}} \right), \\ Q_{\pm} &= z^2(k_y^2 + k_z^2)^2 + (2hk_y k_z \pm z(k_y^2 - k_z^2))q_{\pm}, \\ (5.5) \quad q_{\pm} &= 2hk_y k_z \pm \sqrt{4h^2 k_y^2 k_z^2 + z^2(k_y^2 + k_z^2)^2}. \end{aligned}$$

Note that the formulation of imaging operators in terms of solution operators of evolution equations is in general obtained through embedding in an extended image domain [26, 27] in at least $d = 3$ dimensions. Restriction from the prestack imaging volume formulation to $d = 2$ dimensions implies that the Hamiltonian (5.5) has a singularity at $z = 0$. Hence, evolution must be initialized at $z > 0$: the initial data have to be isochrons at early two-way travel times T_0 [31]. Hamiltonian (5.5) is anisotropic and can create caustics for initial conditions that differ from isochron fronts (e.g., for local plane waves).

We use background $c = c_0 = 2km/s$ and half-offset $h = 100m$. The initial condition is an isochron front for $T_0 = 0.39s$ (resulting in maximum initial depth $z < 400m$), hard-thresholded in the wave packet domain and plotted in Figure 16 (left). We evaluate approximation (2.9) for evolution of the isochron front for $T = 5s$ and compare it to zero order approximation. The geometry of the problem is depicted in Figure 16 (right). Note that the area including the calculation domain is significantly larger than the area including the initial conditions. Results are plotted in Figure 17 for frequency scales $k = 2$ and $k = 3$ and including all frequency scales k . The zero order approximation fails to correctly image the isochron front at the different frequency scales and produces an image with large gaps and amplitude fluctuations

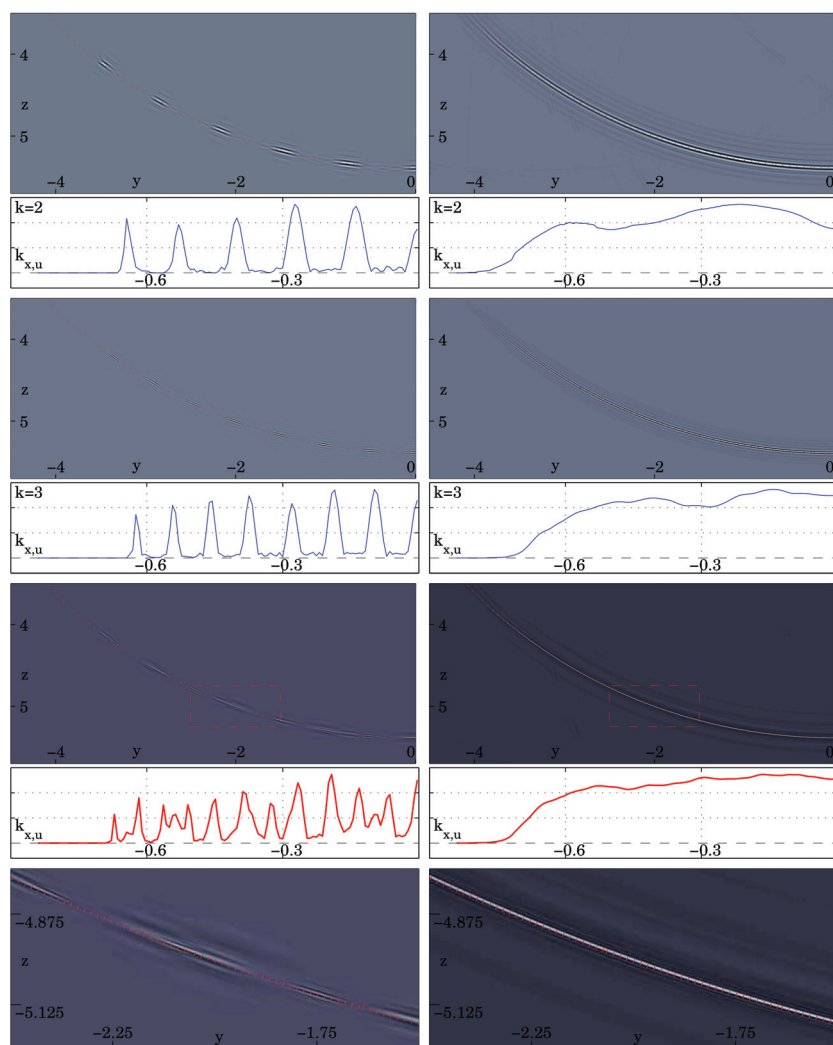


FIG. 17. Evolution of common offset isochron under Hamiltonian (5.5): zero order approximation (left column) and approximation (2.9) (right column). Isochron front band-limited to frequency scales $k = 2$ (top row) and $k = 3$ (third row) and corresponding amplitudes along the fronts (second and fourth row, respectively); isochron fronts and amplitudes including all frequency scales k (fifth and sixth row). Bottom row: magnification of the image regions indicated in the fifth row by red dashed rectangles (theoretical position of the isochron indicated by the red dotted line).

along the isochron. Resorting to finer scales or including all scales does not improve the image. In contrast, approximation (2.9) produces a very satisfactory image isochron that is sharply aligned along the theoretical position of the front without gaps or major amplitude variations.⁹

⁹The smooth amplitude variation results from the initial conditions we have used. The hard-thresholding preprocessing step does not guarantee that the energy from different orientations ν is kept balanced. Indeed, it is clear from inspection of the amplitudes of the zero order approximation in Figure 17 (left) that different box orientations do not contribute equal energy and that amplitude fluctuations under approximation (2.9) depicted in Figure 17 (right) merely reflect these variations.

6. Discussion. We have devised numerical procedures enabling the discrete evaluation of the action of Fourier integral operators on general input functions through approximations, yielding accuracy $\mathcal{O}(2^{-k/2})$ at frequency scale k . While numerical examples have been given in $d = 2$ dimensions, the procedures are valid for arbitrary dimension. The discretization being initiated from the dyadic parabolic decomposition of phase space, the algorithms reflect the geometry of the operators and are tightly interwoven with discrete almost symmetric wave packet transforms. (We mention that it would be possible to construct alternative discretizations, potentially resulting in faster algorithms, yet at the price of losing the explicit connection with discrete almost symmetric wave packets and the geometry of the operators. An alternative method, following the decomposition into wave atoms, has been introduced in [22].) This aspect accounts for the inherent practical flexibility common to all three approximations (2.6), (2.8), and (2.9): they naturally enable, for instance, the embedding of data regularization, emphasis or muting of coherent data structures, and modeling or imaging with subsets of wave packets at selected orientations, frequency scales, and spatial positions, viz. phase-space filtering. The elementary building blocks being directly connected to geometric phase-space attributes, such operations are particularly useful in modeling, illumination analysis, partial reconstruction, and target-oriented imaging applications. Moreover, in the case of parametrices of hyperbolic evolution equations, which also generate extended imaging, we obtain effective one-step procedures for (large) time steps that are insensitive to numerical dispersion.

Any of the described algorithms is embarrassingly parallel, since computations per individual box, tensor product term, or packet can be performed independently. The parallelization of the (global) USFFT in the forward transform of the data has been described in [25]. Also, the procedures offer the possibility of incorporating real-time image visualization: computations can be hierarchically organized such that useful intermediate results—for instance, successively finer scales (as, e.g., in Figure 11 or 17) and updated output points for successively refined grid resolution—can be visualized as they become available during computations, which can be favorably exploited in data intensive and computationally demanding real-world applications.

When all boxes contribute to the output, the computational complexity of our box algorithm is $\mathcal{O}(RN^{(d-1)/2} \log(N))$ above a diagonal approximation. The additional factor R is fundamental in the approximation (2.6) leading to the algorithm, stemming from the necessity for separating a complex exponential and reflecting its (numerical) rank, while the factor $\mathcal{O}(N^{(d-1)/2})$ results from the total number of boxes that have to be evaluated separately. The action of $(Fu_{\nu,k})(y)$ on portions $u_{\nu,k}(x)$ of data corresponding to different boxes does in general result in contributions to different regions in the output domain. Consequently, $(Fu)(y)$ potentially provides information on an output image that can be (significantly) larger than the original N^d data cube $u(x)$. Obviously, the output is practically restricted to the domain on which a model is given (i.e., on which $a(y, \nu)$ and derivatives of $S(y, \nu)$ are accessible).

In the context of imaging, comparison with the generalized radon transform (GRT) is favorable. In $d = 2$ ($d = 3$) dimensions, the box algorithm yields complexity $\mathcal{O}(N^{2.5} \log(N))$ ($\mathcal{O}(N^4 \log(N))$), as compared to the GRT with $\mathcal{O}(N^3)$ ($\mathcal{O}(N^5)$), respectively.

Our hybrid packet-box and packet algorithms for approximations (2.8) and (2.9) appear very attractive at first, since no tensor product representation as in approximation (2.6) needs to be constructed. Indeed, when applied to one single wave packet, they yield low complexity $\mathcal{O}(dN^d \log(N))$. Yet, when used as approximations to the

global operator action for input data with diverse coherent structures, they become less advantageous since the organization of computations by boxes (ν, k) is partially lost, and certain computations need to be performed packet by packet, worsening overall complexity.

We note that the procedures developed in this paper aid in understanding and estimating the accuracy of so-called beam migration. Here, we obtain beams as data wave packets which are propagated using a phase-space localized paraxial approximation in geodesic coordinates. For approximations (2.6) and (2.8), we can also form boxed beams from data wave packets that share the same frequency scale and dip. We finally mention the effective reduction in calculation domain per beam resulting from the spatial decay properties of individual wave packets.

Future work will include extension of the proposed procedures to the presence of caustics. This is currently being investigated. The idea of further separation of variables—within y and within ξ —to unidirectional separated representations has been put forward for computations in high dimensions by Beylkin et al. [4, 7]. Incorporation of such strategies promises computational advantages, in particular for dimensions $d > 3$, while typically resulting in purely numerical algorithms.

Appendix A. System for perturbations in Fermi coordinates. To simplify notation, let (x, ξ) and t_0 be fixed. Let f_I , $I = 2, \dots, d$, be a set of orthonormal vectors in the plane tangent to the wave front at time t_0 , and let $f_1(t_0) = \frac{\partial y}{\partial t}(t_0)$. The subset $f_I(t_0)$ can be chosen with arbitrary orientation in the tangent plane and will be fixed here to coincide with the unit vectors of the ξ'' axes (cf. Figure 1, right). Denote by $f_i(t)$ the coordinate system f_i transported parallel along the integral curve (ray) $(y(t), \eta(t))$. The Fermi- or ray-centered coordinates $y_f(t)$ are coordinates in this system, where $y_{f,1}(t) = t$ is time along ray, and $y_{f,2}(t), \dots, y_{f,d}(t)$ essentially describe the distance from $(y(t), \eta(t))$. The transformation matrices with global Cartesian coordinates read

$$(A.1) \quad H_{ij}(t) = \frac{\partial y_i}{\partial y_{f,j}}(t) = f_j^i(t), \quad \bar{H}_{jl}(t) = \frac{\partial y_{f,j}}{\partial y_l}(t), \quad H\bar{H} = I_{d \times d},$$

where f_j^i is the i th component of f_j with corresponding cotangent vectors denoted by η_f^i . The components f_I satisfy $\frac{df_I(t)}{dt} = -1/\langle \eta(t), \eta(t) \rangle \langle f_I(t), -\frac{\partial P(t, y, \eta)}{\partial y} \eta(t) \rangle$, and $f_1(t)$ is known from integration of (4.2). Transformation of the system (4.5) to Fermi coordinates and reduction to the subsystem f_I , $I = 2, \dots, d$, in the tangent plane yields the system (cf., e.g., [59])

$$(A.2) \quad \left. \frac{d}{dt} \frac{\partial(y_{f,I}(t), \eta_{f,I}(t))}{\partial(y_{f,I}(t_0), \eta_{f,I}(t_0))} (y_{f,I}(t)) \right|_{y_{f,I}(t)=0} = \begin{pmatrix} A_{f_I}(t) & B_{f_I}(t) \\ C_{f_I}(t) & D_{f_I}(t) \end{pmatrix} \cdot \frac{\partial(y_{f,I}(t), \eta_{f,I}(t))}{\partial(y_{f,I}(t_0), \eta_{f,I}(t_0))} (y_{f,I}(t)) \Big|_{y_{f,I}(t)=0},$$

$$(A.3) \quad A_{f_I, MN}(t) = \bar{H}_{Mn}(t) H_{mN}(t) \left(\frac{\partial^2 P(t, y, \eta)}{\partial \eta_n \partial y_m} - \frac{1}{\langle \eta(t), \eta(t) \rangle} \eta_n \frac{\partial P(t, y, \eta)}{\partial y_m} \right),$$

$$(A.4) \quad B_{f_I, MN}(t) = \bar{H}_{Mn}(t) \bar{H}_{Nm}(t) \frac{\partial^2 P(t, y, \eta)}{\partial \eta_n \partial \eta_m},$$

$$(A.5) \quad C_{f_I, MN}(t) = H_{nM}(t) H_{mN}(t) \frac{\partial^2 P(t, y, \eta)}{\partial y_n \partial y_m},$$

$$(A.6) \quad D_{f_I, MN}(t) = H_{nM}(t) \bar{H}_{Nm}(t) \left(\frac{\partial^2 P(t, y, \eta)}{\partial y_n \partial \eta_m} - \frac{\partial P(t, y, \eta)}{\partial y_n} \frac{\partial P(t, y, \eta)}{\partial y_m} \right).$$

From the fundamental matrix $W = \frac{\partial(y_{f,I}(T), \eta_{f,I}(T))}{\partial(y_{f,I}(t_0), \eta_{f,I}(t_0))}$ of (A.3) (cf. (4.4)) we obtain

$$(A.7) \quad \frac{\partial^2 S}{\partial \xi'^2}(y(T), \nu) = -W_1^{-1}W_2.$$

We finally give expressions for the matrices DT_γ and M_γ in approximation (2.9). Let G^t, \bar{G}^t be the transformation matrices between global Cartesian coordinates and the local Cartesian coordinate system $g_i(t) = \{g_1(t), f_2(t), \dots, f_d(t)\}$ defined as in (A.1), where $g_1(t)$ is the unit vector normal to the wave front at its intersection with ray $(y(t), \eta(t))$. The matrix DT_γ consists of dilation terms in directions $g_i(T)$, shear terms in directions $g_i(T)$, and rotation from $g_i(T)$ to $g_i(t_0)$, or, equivalently, dilations in $f_i(T)$ and transformation from $f_i(T)$ to $f_i(t_0)$, and is given by

$$(A.8) \quad DT_\gamma = \bar{H}(t_0)(H(T)Q_{\gamma,T}), \quad \text{where} \quad Q_{\gamma,T} = \begin{bmatrix} 1 & 0_{1 \times d-1} \\ 0_{d-1 \times 1} & W_1 \end{bmatrix}^{-1}.$$

The matrix M_γ consists of quadratic terms in the $f_i(T)$ directions and rotation from $g_i(T)$ to $g_i(t_0)$ and can be obtained as

$$(A.9) \quad M_\gamma \cdot (y - y_j^{\nu,k})^2 = \bar{G}(t_0)[(y - y_j^{\nu,k})^T (G^T(T)P_{\gamma,T}G(T))(y - y_j^{\nu,k})e_1],$$

$$P_{\gamma,T} = \begin{bmatrix} 0 & 0_{1 \times d-1} \\ 0_{d-1 \times 1} & W_3 W_1^{-1} \end{bmatrix}.$$

REFERENCES

- [1] A. ARNEODO, N. DECOSTER, AND S.G. ROUX, *A wavelet-based method for multifractal image analysis. I. Methodology and test applications on isotropic and anisotropic random rough surfaces*, Eur. Phys. J. B, 15 (2000), pp. 567–600.
- [2] G. BAO AND W.W. SYMES, *Computation of pseudo-differential operators*, SIAM J. Sci. Comput., 17 (1996), pp. 416–429.
- [3] G. BEYLKIN, *On the fast Fourier transform of functions with singularities*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 363–381.
- [4] G. BEYLKIN, V. CHERUVU, AND F. PÉREZ, *Fast adaptive algorithms in the non-standard form for multidimensional problems*, Appl. Comput. Harmon. Anal., 24 (2008), pp. 354–377.
- [5] G. BEYLKIN, R. COIFMAN, AND V. ROKHLIN, *Fast wavelet transforms and numerical algorithms I*, Comm. Pure Appl. Math., 44 (1991), pp. 141–183.
- [6] G. BEYLKIN AND M.J. MOHLENKAMP, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.
- [7] G. BEYLKIN AND K. SANDBERG, *Wave propagation using bases for bandlimited functions*, Wave Motion, 41 (2005), pp. 263–291.
- [8] B. BRADIE, R. COIFMAN, AND A. GROSSMANN, *Fast numerical computations of oscillatory integrals related to acoustic scattering, I*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 94–99.
- [9] S. BRANDSBERG-DAHL AND J. ETGEN, *Beam-wave imaging*, in Proceedings of Society of Exploration Geophysicists, 2003.
- [10] E. CANDÈS AND L. DEMANET, *Curvelets and Fourier integral operators*, C. R. Acad. Sci. Paris, I (2003), pp. 395–398.
- [11] E. CANDÈS, L. DEMANET, D. DONOHO, AND L. YING, *Fast discrete curvelet transforms*, SIAM Multiscale Model. Simul., 5 (2006), pp. 861–899.
- [12] E. CANDÈS, L. DEMANET, AND L. YING, *Fast computation of Fourier integral operators*, SIAM J. Sci. Comput., 29 (2007), pp. 2464–2493.
- [13] E. CANDÈS, L. DEMANET, AND L. YING, *A fast butterfly algorithm for the computation of Fourier integral operators*, SIAM Multiscale Model. Simul., 7 (2009), pp. 1727–1750.
- [14] E. CANDÈS AND D. DONOHO, *Continuous curvelet transform: I. Resolution of the wavefront set*, Appl. Comput. Harmon. Anal., 19 (2003), pp. 162–197.
- [15] E. CANDÈS AND D. DONOHO, *Continuous curvelet transform: II. Discretization and frames*, Appl. Comput. Harmon. Anal., 19 (2003), pp. 198–222.

- [16] J.F. CLAERBOUT, *Imaging the Earth's Interior*, Blackwell Science, Oxford, UK, 1985.
- [17] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.
- [18] M.V. DE HOOP, S. HOLMAN, H. SMITH, AND G. UHLMANN, *Regularity in, and multi-scale discretization of the solution construction of hyperbolic evolution equations of limited smoothness*, in Proceedings of Geo-Mathematical Imaging Group, 2011.
- [19] M.V. DE HOOP, J.H. LE ROUSSEAU, AND B.L. BIONDI, *Symplectic structure of wave-equation imaging: A path-integral approach based on the double-square-root equations*, Geophys. J. Int., 153 (2003), pp. 52–74.
- [20] M.V. DE HOOP, J.H. LE ROUSSEAU, AND R.S. WU, *Generalization of the phase-screen approximation for the scattering of acoustic waves*, Wave Motion, 31 (2000), pp. 43–70.
- [21] M.V. DE HOOP, H. SMITH, G. UHLMANN, AND R.D. VAN DER HILST, *Seismic imaging with the generalized radon transform: A curvelet transform perspective*, Inverse Problems, 25 (2009), p. 025005+.
- [22] L. DEMANET AND L. YING, *Wave atoms and time upscaling of wave equations*, Numer. Math., 113 (2009), pp. 1–71.
- [23] L. DEMANET AND L. YING, *Discrete symbol calculus*, SIAM Rev., 53 (2011), pp. 71–104.
- [24] H. DOUMA AND M.V. DE HOOP, *Leading-order seismic imaging using curvelets*, Geophys., 72 (2007), pp. 231–248.
- [25] A. DUCHKOV, F. ANDERSSON, AND M.V. DE HOOP, *Discrete almost symmetric wave packets and multi-scale representation of (seismic) waves*, IEEE Trans. Geosci. Remote Sensing, 48 (2010), pp. 3408–3423.
- [26] A.A. DUCHKOV AND M.V. DE HOOP, *Extended isochron rays in prestack depth (map) migration*, Geophys., 75 (2010), pp. S139–S150.
- [27] A.A. DUCHKOV, M.V. DE HOOP, AND A. SÁ BARETTO, *Evolution-equation approach to seismic image, and data, continuation*, Wave Motion, 45 (2008), pp. 952–969.
- [28] A. DUTT AND V. ROKHLIN, *Fast Fourier transforms for nonequispaced data*, SIAM J. Sci. Comput., 14 (1993), pp. 1368–1393.
- [29] A. DUTT AND V. ROKHLIN, *Fast Fourier transforms for nonequispaced data II*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 85–100.
- [30] M. FEDI, *Global and local multiscale analysis of magnetic susceptibility data*, Pure Appl. Geophys., 160 (2003), pp. 2399–2417.
- [31] E. IVERSEN, *The isochron ray in seismic modeling and imaging*, Geophys., 69 (2004), pp. 1053–1070.
- [32] S. JAFFARD, *Exposants de Hölder en des points données et coefficients dondelettes*, C. R. Acad. Sci. Paris, 308 (1989), pp. 79–81.
- [33] S. JAFFARD, *Multifractal formalism for functions part I: Results valid for all functions*, SIAM J. Math. Anal., 28 (1997), pp. 944–970.
- [34] S. JAFFARD, *Multifractal formalism for functions part II: Self-similar functions*, SIAM J. Math. Anal., 28 (1997), pp. 971–998.
- [35] S. JAFFARD, *Wavelet techniques in multifractal analysis*, in Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus, and M. van Frankenhuysen, eds., Proc. Sympos. Pure Math. 72, 2004, American Mathematical Society, Providence, RI, pp. 91–152.
- [36] W. KLINGENBERG, *Riemannian Geometry*, Walter de Gruyter, Berlin, 1995.
- [37] H. KUMANO-GO, K. TANIGUCHI, AND Y. TOZAKI, *Multi-products of phase functions for Fourier integral operators with an application*, Comm. Partial Differential Equations, 3 (1978), pp. 349–380.
- [38] H.J. LANDAU, *On Szegő's eigenvalue distribution theorem and non-Hermitean kernels*, J. Anal. Math., 28 (1975), pp. 335–357.
- [39] H.J. LANDAU AND H.O. POLLAK, *Prolate spheroidal wave functions, Fourier analysis and uncertainty-II*, Bell Systems Tech. J., Jan. (1961), pp. 65–94.
- [40] H.J. LANDAU AND H.O. POLLAK, *Prolate spheroidal wave functions, Fourier analysis and uncertainty III: The dimension of space of essentially time- and band-limited signals*, Bell Systems Tech. J., July (1962), pp. 1295–1337.
- [41] H.J. LANDAU AND H. WIDOM, *Eigenvalue distribution of time and frequency limiting*, J. Math. Anal. Appl., 77 (1980), pp. 469–481.
- [42] C.F. LI AND C. LINER, *Wavelet-based detection of singularities in acoustic impedance from surface seismic reflection data*, Geophysics, 73 (2008), pp. V1–V9.
- [43] Y. MEYER AND R. COIFMAN, *Wavelets: Calderón-Zygmund and Multilinear Operators*, Cambridge University Press, Cambridge, UK, 1997.

- [44] V. ROKHLIN AND H. XIAO, *Approximate formulae for certain prolate spheroidal wave functions valid for large values of both order and band-limit*, Appl. Comput. Harmon. Anal., 22 (2007), pp. 105–123.
- [45] 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), pp. 867–906.
- [46] G. SCHUSTER, *Seismic Interferometry*, Cambridge University Press, Cambridge, UK, 2009.
- [47] Y. SHKOLNISKY, *Prolate spheroidal wave functions on a disc—integration and approximation of two-dimensional bandlimited functions*, Appl. Comput. Harmon. Anal., 22 (2007), pp. 235–256.
- [48] Y. SHKOLNISKY, M. TYGERT, AND V. ROKHLIN, *Approximation of bandlimited functions*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 413–420.
- [49] D. SLEPIAN, *Prolate spheroidal wave functions, Fourier analysis and uncertainty IV: Extensions to many dimensions, generalized prolate spheroidal wave functions*, Bell Systems Tech. J., Nov. (1964), pp. 3009–3057.
- [50] D. SLEPIAN, *Prolate spheroidal wave functions, Fourier analysis and uncertainty V: The discrete case*, Bell Systems Tech. J., 57 (1978), pp. 1371–1430.
- [51] D. SLEPIAN AND H.O. POLLAK, *Prolate spheroidal wave functions, Fourier analysis and uncertainty I*, Bell Systems Tech. J., Jan. (1961), pp. 43–63.
- [52] H. SMITH, *Lecture Notes*, <http://www.math.washington.edu/~hyliu/SummerSchool/uwss3.pdf>.
- [53] H. SMITH, *A parametrix construction for wave equations with $c^{1,1}$ coefficients*, Ann. Inst. Fourier (Grenoble), 48 (1998), pp. 797–835.
- [54] C.C. STOLK AND M.V. DE HOOP, *Modeling of seismic data in the downward continuation approach*, SIAM J. Appl. Math., 65 (2005), pp. 1388–1406.
- [55] C.C. STOLK AND M.V. DE HOOP, *Seismic inverse scattering in the downward continuation approach*, Wave Motion, 43 (2006), pp. 579–598.
- [56] C.C. STOLK AND M.V. DE HOOP, *Microlocal analysis of seismic inverse scattering in anisotropic, elastic media*, Comm. Pure Appl. Math., 55 (2002), pp. 261–301.
- [57] M.E. TAYLOR, *Partial Differential Equations III: Nonlinear Equations*, Springer, Berlin, 1996.
- [58] E. VAN DEN BERG AND M.P. FRIEDLANDER, *Sparse Optimization with Least-Squares Constraints*, Technical report TR-2010-02, Department of Computer Science, University of British Columbia, 2011.
- [59] V. ČERVENÝ, *Seismic Ray Theory*, Cambridge University Press, Cambridge, UK, 2001.
- [60] S. WANG, M. DE HOOP, AND B. URSIN, *Illumination analysis of wave-equation imaging with curvelets*, Inverse Problems, 26 (2010), p. 115013.
- [61] H. XIAO AND V. ROKHLIN, *High-frequency asymptotic expansions for certain prolate spheroidal wave functions*, J. Fourier Anal. Appl., 9 (2003), pp. 575–596.
- [62] H. XIAO, V. ROKHLIN, AND N. YARVIN, *Prolate spheroidal wave functions, quadrature and interpolation*, Inverse Problems, 17 (2001), pp. 805–838.