Proceedings of the Project Review, Geo-Mathematical Imaging Group (Purdue University, West Lafayette IN), Vol. 1 (2010) pp. 93-134.

MULTI-SCALE DISCRETE APPROXIMATION OF FOURIER INTEGRAL OPERATORS

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

Abstract. We develop a discretization and computational procedures for the approximation of the action of Fourier integral operators whose canonical relations are graphs. Such operators appear in many physical contexts and computational problems, for instance in the formulation of imaging and inverse scattering of seismic reflection data. Our discretization and algorithms are based on a multi-scale 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 that directly reflect the nature of such operators. The discretization and computational procedures explicitly connect and can be seamlessly overlaid to the discrete almost symmetric wave packet transformation. Numerical wave propagation and imaging examples illustrate our algorithms.

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

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

$$(Fu)(y) = \int a(y,\xi)\exp(\mathrm{i}S(y,\xi))\hat{u}(\xi)\mathrm{d}\xi, \tag{1.1}$$

where $\hat{u}$ denotes the Fourier transform of u, $a(y,\xi)$ is the amplitude function and $S(y,\xi)$ denotes the generating function. The propagation of singularities by F, $(x,\xi)\to(y,\eta)$, follows from S and is described by the transformation

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

These FIOs include the pseudodifferential operators, for which $S(y,\xi)=\langle y,\xi\rangle$ and χ is the identity. We will assume that a is of order zero. The operator F has a sparse matrix representation with respect to the frame of curvelets [12, 51], which originates from the dyadic parabolic decomposition of phase space and which will be briefly discussed below. We will refer to curvelets ([13] 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 [9] for a general analysis.)

In the case of pseudodifferential operators, the generating function is linear in ξ and is 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 [55]; each term in this expansion is also referred to as an elementary symbol. Bao and Symes [3] developed a

*CCAM, Department of Mathematics, Purdue University, West Lafayette, IN, USA (hwendt@math.purdue.edu)
†CCAM, Department of Mathematics, Purdue University, West Lafayette, IN, USA (mdehoop@math.purdue.edu)
‡Mathematics LTH, Centre for Mathematical Sciences, Lund Institute of Technology, Lund University, SE (fa@maths.lth.se)

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 [19] 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 [6].

Here, we consider FIOs and focus on separated expansions of the complex exponential in (1.1). Recently, De Hoop *et al.* ([18]) proposed an explicit *multi-scale* expansion of low phase space separation rank of the action of FIOs in the class mentioned above using the dyadic parabolic decomposition of phase space. The first-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 problem size. In the work presented here, we elaborate on this result and develop a discretization, approximation and numerical procedure for computing the action of this class of operators. 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 signifcant tiles in the dyadic parabolic decomposition. We focus on explicit constructions and corresponding numerical analysis that directly reflect the nature of FIOs. We express our separated representation 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]. An efficient algorithm to compute PSWFs was proposed by [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. We derive our discretization from the inverse transform based on discrete almost symmetric wave packets [21].

The connection of our algorithm to discrete almost symmetric wave packets is important in imaging and inverse scattering applications. In these applications, the FIO acts on the data (u in the above). The wave packets can aid in regularizing the process of partitioning the data in ξ-space starting from a finite set of samples through sparsity promoting optimization (instead of standard interpolation). Moreover, in the context of directional pointwise regularity analysis [2, 26, 29, 30, 31, 32, 39], the mentioned connection enables the numerical estimation and study of propagation of scaling exponents by the FIO extending the corresponding results for Calderón-Zygmund operators using wavelets [40].

Indeed, imaging and inverse scattering of seismic reflection data can be 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 scattering operators; see Stolk and De Hoop [52, 53, 54]. First-order evolution equations and associated propagators also play a role in implementations of wave-equation imaging and inverse scattering; we mention time and depth extrapolation (or downward continuation), and velocity continuation. Also, extended imaging can be described in terms of solving a Cauchy initial value problem for an evolution equation (Duchkov and De Hoop [23]). Parametrices of such evolution equations are natural examples of the FIOs considered here. In the present contribution, we consider situations without caustics. Extension to the caustic case will be the subject of a forthcoming paper.

Our analysis is based on the work of Seeger, Sogge and Stein [45]. The fast computation of FIOs (in dimension $d = 2$) was recently considered by Candès, Demanet and Ying [14]. In this 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 in (1.1) 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 [17] – and a polar coordinates Taylor (or Maclaurin) expansion of its phase 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 [15], a butterfly algorithm is 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 [10] for the computation of oscillatory integrals related to acoustic wave scattering. As far as applications in reflection seismology are concerned, we mention "beam" migration [1, 27, 41] and "beamlet" propagators [60].

The outline of this paper is as follows. Below we give a brief introduction to the dyadic parabolic decomposition of phase space, the co-partition of unity, and wave packets. In Section 2, we summarize the multi-scale operator expansion proposed in [18], construct the separated expansion of the complex exponential in (1.1) using PSWFs, and analyze the rank properties of this expansion. Details on PSWFs and on their numerical evaluation are collected in Appendices A and B. Section 3 proceeds with constructing the discretization of the operator expansion. We begin with a summary of discrete almost symmetric wave packets and establish the connection with the approximate operator action. Then we discuss the deformation of phase space discretization under the operator, leading to strategies for choosing the oversampling factors and for the discrete evaluation of the canonical transformation by USFFT, and we relate wave packet decay properties to reduction of the size of the calculation domain. Based on the theoretical developments described in Section 2, we propose a box (frequency tile) based algorithm, an individual packet based algorithm, and a hybrid packet-box algorithm, and investigate their computational complexity and properties. We analyze the algorithms in the example of solution operators of evolution equations represented as Trotter products and establish the relationship with a phase-space localized approximation. In Section 4 we detail, as an example within the class of FIOs considered in the present work, application to parametrices of evolution equations. We establish the explicit relationship with paraxial ray theory for this case, the expansion terms of the phase of the complex exponential being obtained as combinations of submatrices of the propagator matrix of the Hamilton-Jacobi system along paraxial rays. The resulting computational procedure is a "one step" algorithm for (potentially very) large time steps. Section 5 provides numerical illustrations for evolution-equation based propagation and imaging. We compare the proposed algorithms in examples of wave propagation in heterogeneous isotropic medium, and we illustrate imaging with an homogeneous anisotropic Hamiltonian. Finally, in Section 6, we conclude on the present work and discuss future perspectives.

Wave packets. We briefly discuss the (co)frame of wave packets [13, 21, 51]. Let $u \in L^2(\mathbb{R}^d)$ represent a (seismic) velocity field. We consider the Fourier transform, $\hat{u}(\xi) = \int u(x)\, \exp[-\mathrm{i}\langle x, \xi\rangle]\, \mathrm{d}x$. We begin with an overlapping covering of the positive ξ_1 axis ($\xi' = \xi_1$) by boxes of the form

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

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

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

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

$$B_{\nu,k} = \Theta_{\nu,k}^{-1} B_k. \tag{1.3}$$

The $B_{\nu,k}$ are illustrated in Fig. 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, we encounter 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 co-partition of unity

$$\hat{\chi}_0(\xi)\hat{\beta}_0(\xi) + \sum_{k\geq 1}\sum_{\nu} \hat{\chi}_{\nu,k}(\xi)\hat{\beta}_{\nu,k}(\xi) = 1, \tag{1.4}$$

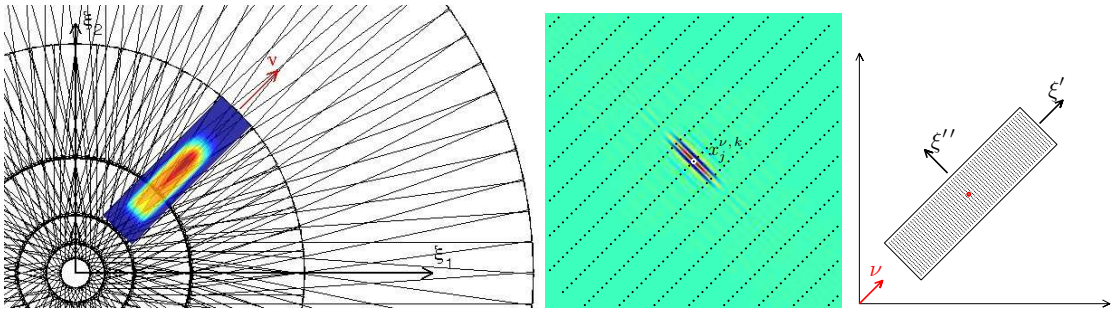

FIG. 1. *Geometry for 2d wave packets: Frequency domain boxes and window function $\hat{\chi}_{\nu,k}(\xi)$ for one particular box for scale $k = 3$ with orientation ν (left). One wave packet corresponding to the box highlighted in the subfigure on the left and central locations of wave packets in this box (center). Orientations of ξ' and ξ'' in the Taylor series expansion of $S(y,\xi)$ (right).*

and satisfy the estimates

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

A function $\hat{\chi}_{\nu,k}$ is plotted in color in Fig. 1 (left). We then form

$$\hat{\psi}_{\nu,k}(\xi) = \rho_k^{-1/2} \hat{\beta}_{\nu,k}(\xi), \quad \hat{\varphi}_{\nu,k}(\xi) = \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi), \tag{1.5}$$

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

$$\left. \begin{array}{l} |\varphi_{\nu,k}(x)| \\ |\psi_{\nu,k}(x)| \end{array} \right\} \leq C_N 2^{k(d+1)/4} \, (\, 2^k |\langle \nu, x \rangle| + 2^{k/2} \|x\| \,)^{-N} \tag{1.6}$$

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

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

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

$$\hat{\varphi}_\gamma(\xi) = \hat{\varphi}_{\nu,k}(\xi) \, \exp[-\mathrm{i}\langle x_j^{\nu,k}, \xi \rangle], \quad \gamma = (x_j^{\nu,k}, \nu, k), \tag{1.7}$$

and similarly for $\hat{\psi}_\gamma(\xi)$. A function $\varphi_{\nu,k}$, as well as the corresponding lattice with points $x_j^{\nu,k}$, are plotted in Fig. 1 (middle). We obtain the transform pair

$$u_\gamma = \int u(x) \overline{\psi_\gamma(x)} \, \mathrm{d}x, \qquad u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x) \tag{1.8}$$

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 Fourier integral operators.

2.1. Dyadic parabolic decomposition and separated representation. In this section, we summarize a result in [18] that will underly our discretization. Let $\varphi_\gamma(x)$, $\gamma = (j, \nu, k)$, denote a wave packet with central position $x_j^{\nu,k}$ and orientation ν at scale k.
The action of the operator F on a single wave packet is given by:

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

since $\hat{\varphi}_\gamma(\xi) = \rho_k^{-1/2}\hat{\chi}_{\nu,k}(\xi)\exp[-\mathrm{i}\langle\xi, x_j\rangle]$ is the Fourier transform of φ_γ. In [18], an approximation of $(F\varphi_\gamma)(y)$ to order $\mathcal{O}(2^{-k/2})$ is obtained via Taylor expansions of $S(y,\xi)$ and $a(y,\xi)$ 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 $\mathcal{O}(2^{-k/2})$ ([18], Lemma 3.1). By homogeneity in ξ of $S(y,\xi)$, the first order Taylor expansion 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 $S^0_{\frac{1}{2},\mathrm{rad}}$ on $\mathbf{1}_{\nu,k}(\xi)$ (cf. [18], (22)). We introduce the *"coordinate transform"*:

$$y \to T_{\nu,k}(y) = \frac{\partial S}{\partial \xi}(y,\nu), \tag{2.2}$$

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 $\left\langle \xi, T_{\nu,k}(y) - x_j^{\nu,k} \right\rangle$ in (2.1) results in the approximation:

$$(F\varphi_\gamma)(y) = a(y,\nu)\varphi_\gamma(T_{\nu,k}(y)) + \mathcal{O}(2^0), \tag{2.3}$$

with error term in the same smoothness class as the wave packet φ_γ itself. We will refer to this approximation as the *zero-order approximation*. Note that it differs from what is commonly referred to as rigid motion since the flow out is governed by all rays with wave vector $2^k\nu$ and origin on the support of φ_γ, not only by the ray connecting the central positions $x_j^{\nu,k}$ and $y_j^{\nu,k} = T_{\nu,k}^{-1}(x_j^{\nu,k})$.

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 $S(y,\xi)$ (cf. Fig. 1 (right) for an illustration of the expansion directions). Using again homogeneity in ξ, this yields 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 $S^{-\frac{1}{2}}_{\frac{1}{2},\mathrm{rad}}$ on $\mathbf{1}_{\nu,k}(\xi)$ (cf. [18], (22)). Approximation of the complex exponential $\exp\left[\mathrm{i}\frac{1}{2\xi'}\left\langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y,\nu)\ \xi'' \right\rangle\right]$ for arguments bounded by c by a polynomial function leads to a tensor-product representation separating the y and ξ variables. This yields the result [18, Theorem 4.1]:

THEOREM 2.1. *With functions $T_{\nu,k}(y)$ defined by (2.2), functions $\alpha_{\nu,k}^{(r)}(y)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ such that*

$$\exp\left[\mathrm{i}\frac{1}{2\xi'}\left\langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y,\nu)\ \xi'' \right\rangle\right]\mathbf{1}_{\nu,k}(\xi) \approx \sum_{r=1}^{R} \alpha_{\nu,k}^{(r)}(y)\hat{\vartheta}_{\nu,k}^{(r)}(\xi), \tag{2.4}$$

one may express

$$(F\varphi_\gamma)(y) = a(y,\nu)\sum_{r=1}^{R} \alpha_{\nu,k}^{(r)}(y)(\vartheta_{\nu,k}^{(r)} * \varphi_\gamma)(T_{\nu,k}(y)) + 2^{-k/2} f_\gamma, \tag{2.5}$$

with $R \sim k/\log(k)$, where f_γ is a "curvelet"-like function (cf. [18], (23)) centered at $\chi(\gamma)$.

Hence, an approximation of $(F\varphi_\gamma)(y)$ to order $\mathcal{O}(2^{-k/2})$ is obtained 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 $T_{\nu,k}(y)$. This expansion can be extended to any order.

Further approximations. 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 alternative result [18, Theorem 4.2]: With

$$\hat{\vartheta}_\gamma(\xi) = \exp\left[\mathrm{i}\frac{1}{2\xi'}\left\langle \xi'', \frac{\partial^2 S}{\partial \xi''^2}(y_j^{\nu,k},\nu)\,\xi''\right\rangle\right]\mathbf{1}_{\nu,k}(\xi), \tag{2.6}$$

one may express:

$$(F\varphi_\gamma)(y) = a(y_j^{\nu,k},\nu)\,(\vartheta_\gamma * \varphi_\gamma)\,(T_{\nu,k}(y)) + 2^{-k/2} f_\gamma, \tag{2.7}$$

where f_γ is a "curvelet"-like function centered at $\chi(\gamma)$ (cf. [18], (23)).

Furthermore, the change of coordinates $T_{\nu,k}$ can be approximated by expansion of $S(y,\nu)$ about $(y_j^{\nu,k},\nu)$, yielding the approximation [18, Theorem 4.3]: One may express

$$(F\varphi_\gamma)(y) = a(y_j^{\nu,k},\nu)\,(\vartheta_\gamma * \varphi_\gamma)\left(DT_\gamma(y-y_j^{\nu,k}) + M_\gamma\cdot(y-y_j^{\nu,k})^2\right) + 2^{-k/2} f_\gamma, \tag{2.8}$$

where f_γ is a "curvelet"-like function centered at $\chi(\gamma)$ (cf. [18], (23)) , $DT_\gamma = \frac{\partial T_{\nu,k}}{\partial y}(y_j^{\nu,k}) = \frac{\partial^2 S}{\partial \xi \partial y}(y_j^{\nu,k},\nu)$ and $M_\gamma = \frac{1}{2}\frac{\partial^2 S}{\partial y^2}(y_j^{\nu,k},\nu)\nu$. In this approximation of $T_{\nu,k}$, the quadratic term in y corresponds to the curvature of an infinitesimal plane wave attached to φ_γ under the the underlying canonical transformation, and the linear term is composed of a 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 2.5.

2.2. Prolate spheroidal wave functions (PSWFs) and tensor products. Here, we revisit (2.4). The phase of the exponential on the left-hand side consists of terms each of which reveals a separation of variables in phase space. We discuss how to construct a separated representation as expressed by the right-hand side of (2.4).

2.2.1. PSWFs. We make use of the eigenfunctions ψ of the integral operator F^c with kernel $\exp[\mathrm{i}cxz]$:

$$(F^c\psi)(x) = \int_{-1}^{1} \exp\left[\mathrm{i}cxz\right]\psi(z)dz, \quad c\in\mathbb{R}^+, \quad x\in[-1,1].$$

These eigenfunctions ψ turn out to be the prolate spheroidal wave functions, or Slepian functions. We refer to Appendix A and to the references in this paragraph for a detailed treatment on these functions. Here, we give a brief summary. Originally, PSWFs were introduced and studied in the late 19th century in a classical mathematical physics context as the continuous eigenfunctions of the differential operator:

$$D^c = (1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx} - c^2x^2, \quad c\in\mathbb{R}^+, \tag{2.9}$$

on the interval $[-1,1]$. In the early 60s, in a series of seminal papers [36, 37, 48, 49, 50], it was discovered that PSWFs are also the eigenfunctions ψ of the above integral operator. It is this later property that has drawn the interest of researchers from different fields, and will also be most useful to us here. An extension of PSWFs to general dimensions, termed generalized PSWFs, has been introduced in [48]: 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

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

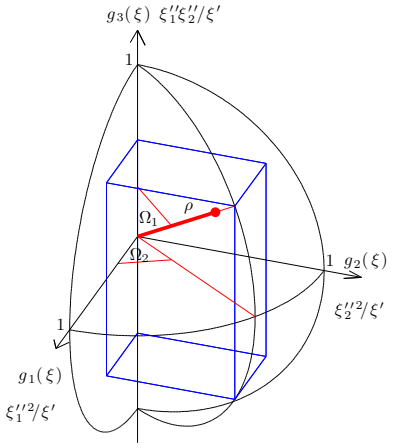

FIG. 2. *Illustration of PSWF coordinates for $g(\xi)$ and $\mathcal{D} = 3$ ($d = 3$). The Cartesian boxes $f(y)$ and $g(\xi)$ need to be included in the unit ball $\mathcal{R}$ on which $\psi_\kappa^c(\rho, \Omega)$ form an orthonormal basis.*

has a continuous solution on $\mathcal{R}$, where $\mathcal{R}$ is the unit (hyper-)ball in $\mathcal{D}$ dimensions, and κ a multi-index. These functions ψ_κ^c are the generalized PSWFs[1]. They are bounded, purely real, orthonormal and complete in $L^2(\mathcal{R})$. It can be shown that these functions are also the eigenfunctions of a differential operator similar to (2.9), which allows to construct efficient procedures for their numerical evaluation. Their eigenvalue spectrum consists of only very few eigenvalues with significant magnitude (the precise number depending on the bandwidth parameter c) and then decays exponentially fast to values close to zero [37, 35, 38]: For instance, for $\mathcal{D} = 1$, the spectrum contains roughly $2c/\pi$ eigenvalues with magnitude close to $\sqrt{2\pi/c}$, then has exponential decay.

PSWFs have lately received considerable attention in the computational community and have been used in an important number of applications in various contexts (e.g. [9, 57]), mostly due to the fact that a numerical tool for evaluating them for practically all values of c encountered in practice has become available [46, 62] (see also e.g. [42, 61] for asymptotic results and approximations). In dimension $\mathcal{D} = 1$, PSWFs can be numerically constructed by expansions in Legendre polynomials. For $\mathcal{D} \geq 2$, they are constructed in polar coordinates (ρ, Ω), in which their radial parts separate from their angular parts:

$$\psi_\kappa^c = \psi_\kappa^c(\rho, \Omega) = \Psi_\kappa^c(\rho) S_\kappa(\Omega).$$

The radial functions $\Psi_\kappa^c(\rho)$ are obtained as expansions in Jacobi polynomials, while the angular portions $S_\kappa(\Omega$ are given by complete sets of orthonormal surface harmonics (in the practically most important case $\mathcal{D} = 3$, these are the spherical harmonics). The corresponding eigenvalues can be obtained by numerical integration of (2.10) (or, for $\mathcal{D} \geq 2$, its counterpart for the radial functions $\Psi_\kappa^c(\rho)$). We refer to Appendix A for more details on the numerical construction of PSWFs.

2.2.2. Tensor product. The kernel of operator (2.10) admits the representation:

$$\exp[\mathrm{i}c\langle x, z\rangle] = \sum_\kappa \lambda_\kappa^c \psi_\kappa^c(x) \psi_\kappa^c(z), \quad ||x||, ||z|| \leq 1. \tag{2.11}$$

Motivated by the above mentioned compactness of the eigenvalue spectrum, we will now use (2.11) in the construction of the tensor-product (2.4). We start with defining appropriate functions of $\frac{\partial^2 S}{\partial \xi_j'' \partial \xi_l''}(y, \nu)$ and $\frac{\xi_j'' \xi_l''}{\xi'}$ mapping onto the unit ball $\mathcal{R}$. Then we establish the relation between the

[1]In what follows, we will use the term PSWF for both PSWFs in one dimension, and generalized PSWFs.

tensor product terms $\alpha^{(r)}_{\nu,k}(y)$ and $\hat{\vartheta}^{(r)}_{\nu,k}(\xi)$ and the PSWFs. Finally, we investigate the numerical rank R of the tensor product (2.4) when constructed from (2.11).

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 $\tilde{f}_\nu : \mathbb{R}^d \to \mathbb{R}^{\mathcal{D}(d)}$ $\tilde{g} : \mathbb{R}^d \to \mathbb{R}^{\mathcal{D}(d)}$:

$$\tilde{f}_\nu(y) = \left[(2-\delta_{jl})\frac{\partial^2 S}{\partial\xi''_j\partial\xi''_l}(y,\nu)\right], \qquad \tilde{g}(\xi',\xi'') = \left[(2-\delta_{jl})\frac{\xi''_j\xi''_l}{\xi'}\right],$$

where

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

due to symmetry in partial derivatives and in $\xi''_j\xi''_l$. Let:

$$c = c(\nu) = \frac{1}{2}\sup_{\mathbf{1}_{\nu,k}(\xi)}|\tilde{g}(\xi',\xi'')|\sup_y\left|\tilde{f}_\nu(y)\right|, \tag{2.12}$$

and define:

$$f_\nu(y) = \tilde{f}_\nu(y)/\sqrt{c}, \qquad g(\xi',\xi'') = \tilde{g}(\xi',\xi'')/\sqrt{c}. \tag{2.13}$$

By elementary manipulations of the left hand side of (2.4), we see that:

$$\exp\left[\mathrm{i}\frac{1}{2\xi'}\left\langle\xi'',\frac{\partial^2 S}{\partial\xi''^2}(y,\nu)\,\xi''\right\rangle\right]\mathbf{1}_{\nu,k}(\xi) = \exp\left[\mathrm{i}\frac{1}{2}\sum_{j,l=2}^{d}\frac{\xi''_j\xi''_l}{\xi'}\frac{\partial^2 S}{\partial\xi''_j\partial\xi''_l}(y,\nu)\right]\mathbf{1}_{\nu,k}(\xi) = \tag{2.14}$$

$$\exp\left[\mathrm{i}\frac{1}{2}\sum_{m=1}^{\mathcal{D}(d)}\tilde{g}_m(\xi',\xi'')\tilde{f}_m(y)\right]\mathbf{1}_{\nu,k}(\xi) = \exp\left[\mathrm{i}c\langle f(y),g(\xi',\xi'')\rangle\right]\mathbf{1}_{\nu,k}(\xi) =$$

$$= \sum_\kappa \lambda^c_\kappa\psi^c_\kappa(f(y))\psi^c_\kappa(g(\xi',\xi''))\,\mathbf{1}_{\nu,k}(\xi).$$

The (transformed) Cartesian boxes $f(y)$ and $g(\xi',\xi'')$ in (2.13) have been included in the unit ball $\mathcal{R}$ by normalization with the bandwidth parameter c since PSWFs are defined on (hyper-) balls (cf. illustration in Fig. 2). We note that in view of the parabolic scaling, c in (2.13) is (asymptotically) independent of scale since $\sup_{j,l,\mathbf{1}_{\nu,k}(\xi)}\xi''_j\xi''_l/\xi' = \sup_{j,l,\mathbf{1}_{\nu,1}(\xi)}(\xi''_j 2^{k/2})(2^{k/2}\xi''_l 2^{k/2})/\xi'/(\xi' 2^k) = \sup_{j,l,\mathbf{1}_{\nu,1}(\xi)}\xi''_j\xi''_l/\xi'$ and $\frac{\partial^2 S}{\partial\xi''^2}(y,\nu)$ are scale independent.
Now let the sequence of multi-indices $\kappa_1,\kappa_2,\cdots$ correspond to the sorted sequence of eigenvalues $|\lambda^c_{\kappa_1}| \geq |\lambda^c_{\kappa_2}| \geq \cdots$. We truncate the infinite sum over the multi-index κ at the Rth term, to within precision $\varepsilon(k)$:

$$\exp\left[\mathrm{i}\frac{1}{2\xi'}\left\langle\xi'',\frac{\partial^2 S}{\partial\xi''^2}(y,\nu)\,\xi''\right\rangle\right]\mathbf{1}_{\nu,k}(\xi) = \sum_{r=1}^{R_{\nu,k}}\lambda^c_{\kappa_r}\psi^c_{\kappa_r}(f(y))\psi^c_{\kappa_r}(g(\xi',\xi''))\,\mathbf{1}_{\nu,k}(\xi) + \varepsilon(k) \tag{2.15}$$

$$= \sum_{r=1}^{R_{\nu,k}}\alpha^{(r)}_{\nu,k}(y)\hat{\vartheta}^{(r)}_{\nu,k}(\xi) + \varepsilon(k),$$

where 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 complete our construction (2.15) of the tensor-product (2.4) by identifying:

$$\alpha^{(r)}_{\nu,k}(y) = \psi^c_{\kappa_r}(f(y)), \tag{2.16}$$

$$\hat{\vartheta}^{(r)}_{\nu,k}(\xi) = \lambda^c_{\kappa_r}\psi^c_{\kappa_r}(g(\xi',\xi'')). \tag{2.17}$$

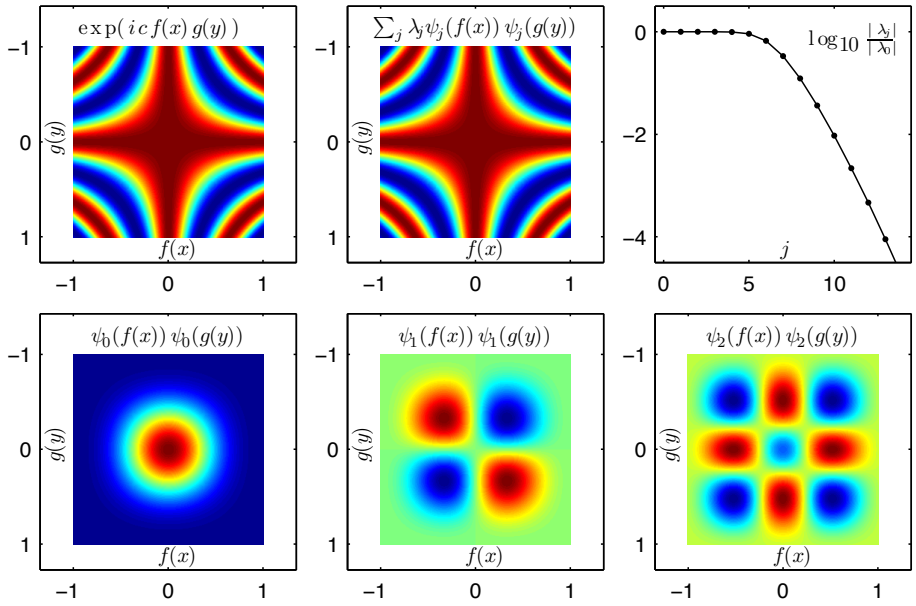

FIG. 3. *Finite rank tensor-product approximation of complex exponential for $\mathcal{D} = 1$ and bandwidth $c = 10$ (depicted are the real parts only): Exponential (top left), tensor-product approximation with $R = 15$ terms (top center), eigenvalue spectrum $|\lambda^c_{\kappa_j}|$ (top right). First three terms of tensor-product approximation (bottom).*

Here, the eigenvalues λ^c_κ could alternatively be absorbed in either of the functions[2] $\alpha^{(r)}_{\nu,k}(y)$ and $\hat{\vartheta}^{(r)}_{\nu,k}(\xi)$. Fig. 3 depicts an example finite rank tensor-product approximation of $\exp[\mathrm{i}cf(x)g(y)]$ for $\mathcal{D} = 1$ and bandwidth $c = 10$.

It is important to note that in our application of (2.15), both the expansion coefficients for constructing the functions $\psi^c_{\kappa_r}$ and the eigenvalues can be pre-computed and tabulated for given discrete sets of bandwidth parameters c, say, $c_1 < c_2 < \cdots$, for the highest desired accuracy ε. For actual bandwidth c given by (2.12), the bandwidth used in (2.15) can then be chosen as the smallest c_i such that $c_i \geq c$, which guarantees that the values of the functions (2.13) are confined to the unit ball. In order to keep the rank of (2.15) close to the smallest possible rank, the set c_i has to be chosen dense enough, for instance such that $R(c_{i+1}) = R(c_i) + 1$. In the following paragraph, we proceed with a more detailed analysis of the rank properties of tensor product (2.15).

Rank estimates. The rank R of approximation (2.15) is controlled by the desired precision ε and the bandwidth parameter c, which is in turn determined by the precise choice of the frequency tiling (1.3), and by the largest value of $|\frac{\partial^2 S}{\partial \xi'^2}(y,\nu)|$ attained on y. The exponentially fast decay of the eigenvalue spectrum, beyond a small number of eigenvalues $\lambda^c_{\kappa_1}, \cdots, \lambda^c_{\kappa_L}$ with significant magnitude [37, 35, 38, 47, 62], and the orthonormality of the functions ψ^c_j guarantee the fast convergence of (2.15) and finite rank R for arbitrarily small finite precision ε. We revisit here bounds on the precision ε for given rank R for $\mathcal{D} = 1$ ($d = 2$). We have:

$$|\lambda^c_r| = \frac{\sqrt{\pi}c^r(r!)^2}{(2r)!\Gamma(r+\frac{3}{2})} \exp\left[\int_0^c \left(\frac{2(\psi^b_r(1))^2 - 1)}{2b} - \frac{r}{b}\right) db\right] \leq \frac{\sqrt{\pi}c^r(r!)^2}{(2r)!\Gamma(r+\frac{3}{2})} \tag{2.18}$$

and $|\psi^c_r(1)| < \sqrt{r+1/2}$ [47], hence:

$$|\lambda^c_r| \leq \frac{\sqrt{\pi}c^r(r!)^2}{(2r)!\Gamma(r+\frac{3}{2})} \leq \frac{\sqrt{\pi}c^r r!}{(2r)!} \leq \sqrt{\pi}c^r 2^{-r\log_2(r)} = \sqrt{\pi}2^{-r[\log_2(r)-\log_2(c)]}. \tag{2.19}$$

[2]We omit explicit reference to the bandwidth parameter c here for convenience of notation.

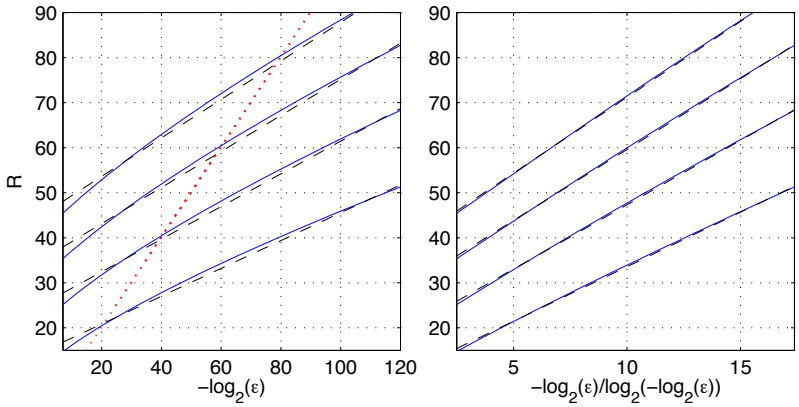

FIG. 4. *Plots of numerical evaluation of* (2.23) *for* $c = \{10, 20, 30, 50\}$ *(blue solid line) in* $-\log(\varepsilon)$ *(left) and in* $-\log(\varepsilon)/\log(-\log(\varepsilon))$ *(right) vs. R coordinates. Dashed solid lines correspond to linear fits in the respective coordinates. Plot of bound* (2.21) *(left, red dotted line).*

Furthermore, for $r \geq 2c$,

$$|\lambda_r^c| \leq 2^{-r+1}. \tag{2.20}$$

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 following L^∞ bound, valid for $R \geq 2c$ [47]:

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

We use the fact that the functions ψ_r form an orthonormal basis on the unit ball to obtain the L^2 bound:

$$\varepsilon(R) = \left\| \sum_{r=R+1}^{\infty} \lambda_r \psi_r^c(x)\psi_r^c(z) \right\|_{L^2(-1,1)} = \sqrt{\sum_{r=R+1}^{\infty} |\lambda_r|^2} \leq \frac{4}{\sqrt{3}} 2^{-(R+1)}, \tag{2.21}$$

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

$$R(\varepsilon) \geq -\log_2(\varepsilon) + \log_2(4/\sqrt{3}) - 1. \tag{2.22}$$

Note that these bounds are based on (2.20), which allows us to obtain closed form estimates, but is very conservative. We aim at obtaining a finer estimate on the order of $R(\varepsilon)$ and proceed with investigating the right most inequality in (2.19). We evaluate numerically:

$$\varepsilon(R) = \sqrt{\sum_{r=R+1}^{\infty} |\lambda_r|^2} \leq \sqrt{\pi} \sqrt{\sum_{r=R+1}^{\infty} 2^{-2r[\log_2(r)-\log_2(c)]}} \tag{2.23}$$

for different bandwidths c. Results are plotted in Fig. 4, together with (2.22). Clearly, the numerical results indicate that:

$$R(\varepsilon) = \mathcal{O}(-\log(\varepsilon)/\log(-\log(\varepsilon))). \tag{2.24}$$

For accuracy $\varepsilon(k) = \mathcal{O}(2^{-k/2})$ we therefore have:

$$R(k) = \mathcal{O}(k/\log(k)) \tag{2.25}$$

in agreement with Theorem 2.1.

3. Discretization. We develop an algorithm for the evaluation of the approximate action of F on a function u at discrete frequency and space points ξ_l and y_n, respectively, based on the approximation (2.5) for a single wave packet and tensor product (2.15). Our choice of discretization will closely match the structure of the discrete wave packet transform [21]. The motivation for this is that data can be efficiently and effectively stored, compressed, regularized, and pre-processed in the wave packet domain (i.e., in form of wave packet coefficients). Our overlay of discretizations will enable to switch from wave packet coefficients to data in the frequency domain – the input to (1.1) – efficiently through standard FFTs. We assume in this section that the generating function $S(y,\xi)$ – more specifically, its partial derivatives $\frac{\partial^2 S}{\partial \xi'^2}(y,\nu)$ and the functions $T_{\nu,k}(y)$ and $T^{-1}_{\nu,k}(x)$ – are known. The issue of how they can be computed for parametrices of evolution equations will be the subject of Section 4. As a special case, we will revisit discretization of evolution time into a sequence of short time steps ("thin-slab" propagation) in Section 3.5.

3.1. Discrete almost symmetric wave packets and approximate FIO action. We begin with writing the convolutions $(\vartheta^{(r)}_{\nu,k} * \varphi_\gamma)(T_{\nu,k}(y))$ in (2.5) in the Fourier domain:

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

In what follows, we will absorb the amplitudes $a(y,\nu)$ in the functions $\alpha^{(r)}_{\nu,k}(y)$. The structure of (3.1) is reminiscent of the (adjoint) discrete wave packet transform:

$$u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x) = \sum_\xi \sum_{\nu,k} e^{\mathrm{i}\langle x,\xi\rangle} \hat{u}(\xi)\hat{\beta}_{\nu,k}(\xi)\hat{\chi}_{\nu,k}(\xi), \tag{3.2}$$

since:

$$(Fu)(y) = \sum_\gamma u_\gamma \tilde{\phi}_\gamma(y) = \sum_{\nu,k}\sum_{r=1}^{R_{\nu,k}} \alpha^{(r)}_{\nu,k}(y) \sum_{\xi\in \mathbf{1}_{\nu,k}} e^{\mathrm{i}\langle T_{\nu,k}(y),\xi\rangle} \hat{u}(\xi)\hat{\beta}_{\nu,k}(\xi)\hat{\chi}_{\nu,k}\hat{\vartheta}^{(r)}_{\nu,k}(\xi), \tag{3.3}$$

and we will indeed use the same discretization, which we recall here briefly for convenience (see [21] for details). We assume that the data $u(x_i)$ are given in discrete form at sampling points $x_i = N^{-1}i$, $i \in \mathbb{R}^d$, $-\frac{N}{2} \le i_n < \frac{N}{2}$. Following the discretization of the "inner" forward transform:

$$\tilde{u}_{j,\nu,k} = \frac{1}{\rho_k^{1/2}} \frac{1}{\sigma'_k(\sigma''_k)^{n-1}} \sum_l \hat{u}(\xi^{\nu,k}_l)\hat{\beta}_{\nu,k}(\xi^{\nu,k}_l) \exp[2\pi\mathrm{i}\langle x^{\nu,k}_j, \xi^{\nu,k}_l\rangle] \approx u_\gamma, \tag{3.4}$$

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:

$$\hat{u}(\xi^{\nu,k}_l)\hat{\beta}_{\nu,k}(\xi^{\nu,k}_l)\hat{\chi}_{\nu,k}(\xi^{\nu,k}_l) = \rho_k^{-1/2} \left(\sum_j \tilde{u}_{j,\nu,k} \exp\left[-2\pi\mathrm{i}\langle x^{\nu,k}_j, \xi^{\nu,k}_l\rangle\right] \right) \hat{\chi}_{\nu,k}(\xi^{\nu,k}_l). \tag{3.5}$$

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

$$\Xi^k = \left\{ l \in \mathbb{Z}^d \;\middle|\; -\frac{N'_k}{2} \le l_1 < \frac{N'_k}{2}, \ldots, -\frac{N''_k}{2} \le l_d < \frac{N''_k}{2} \right\}. \tag{3.6}$$

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. We choose the $\xi_l^{\nu,k}$ (covering the box $B_{\nu,k}$) as

$$\xi_l^{\nu,k} = \Theta_{\nu,k}^{-1}\left(D_k S_k^{-1}\Xi_l^k + \xi_k' e_1\right), \tag{3.7}$$

where the matrix S_k is defined as

$$S_k = \begin{pmatrix} N_k' & 0_{1\times d-1} \\ 0_{d-1\times 1} & N_k'' I_{d-1} \end{pmatrix}.$$

The dot product in the phase of the exponential in (3.5) then becomes

$$\langle x_j^{\nu,k}, \xi_l^{\nu,k}\rangle = \left(D_k S_k^{-1}\Xi_l^k + \xi_k' e_1\right)^t D_k^{-1} X_j = \frac{j_1\xi_k'}{L_k'} + \left(\frac{j_1 l_1}{N_k'} + \frac{j_2 l_2 + \ldots + j_d l_d}{N_k''}\right). \tag{3.8}$$

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 $\dot{\tilde{u}}_{j,\nu,k}$ (cf. (3.5)) for $l \in \Xi^k$:

$$\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\exp(2\pi \mathrm{i} j_1\xi_k'/L_k') = \rho_k^{-1/2} N_k'(N_k'')^{d-1}\sum_j \tilde{u}_{j,\nu,k}\exp\left[-2\pi\mathrm{i}\langle x_j,\xi_l\rangle\right]. \tag{3.9}$$

where $\xi_l = l$ and $x_j = S_k^{-1} j$ with $j \in \Xi^k$, while $(N_k'(N_k'')^{d-1}) = \det S_k$. The number N_ξ^k of discrete frequency points $\xi_l^{\nu,k}$ in (3.9) is of order:

$$N_\xi^k \sim \sigma_k'(\sigma_k'')^{d-1} 2^k 2^{\frac{k(d-1)}{2}}, \qquad N_\xi^{k_{max}} \sim \sigma_k'(\sigma_k'')^{d-1}\left(N/2c_0\right)^{\frac{d+1}{2}}.$$

One can use an n-dimensional FFT for the fast 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 at $x_j^{\nu,k}$, i.e., from the forward wave packet transform of the data. The discrete outer transform:

$$u(x_i) \approx \sum_{\nu,k}\sum_{l\in\Xi^k} e^{2\pi\mathrm{i}\langle x_i,\xi_l^{\nu,k}\rangle}\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \tag{3.10}$$

completes the discretization of (3.2). It is evaluated by USFFT [4, 24, 25] from irregularly spaced points $\xi_l^{\nu,k}$ to regularly spaced x_i.

We proceed with the discretization of (3.3). First, note that the (arguments of) the generating function $S(y,\xi)$ bear physical units, and consequently the same is the case for the functions $\alpha_{\nu,k}^{(r)}(y)$, $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ and $x = T_{\nu,k}(y)$. If 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 $\xi_n^{phys} = \xi_{l_n}/(N\Delta_n^x)$. We absorb the normalization factors in the functions $\alpha_{\nu,k}^{(r)}(y)$, $\hat{\vartheta}_{\nu,k}^{(r)}(\xi)$ and $T_{\nu,k}$ and continue, with slight abuse of notation, to use the same symbols for them. Now let $y_i = T_{\nu,k}^{-1}(x_i)$. Then,

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

and we obtain the discretization of (3.3):

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

We note that in contrast to (3.10), the "outer" transform $\xi_l^{\nu,k} \to x_i$ (USFFT) has to be evaluated per box (ν,k), since the functions $T_{\nu,k}(y)$, $\alpha_{\nu,k}^{(r)}(y)$ are different for each box: (3.11) is organized as

$$(Fu)(y_i) \approx \sum_{\nu,k}(Fu_{\nu,k})(y_i), \tag{3.12}$$

where $u_{\nu,k}$ denotes the data component of the box (ν, k),

$$u_{\nu,k}(x_i) = \sum_{\gamma':k'=k,\nu'=\nu} u_{\gamma'}\varphi_{\gamma'}(x_i). \tag{3.13}$$

For use in estimates below, let us be more specific on the general procedure for tiling with boxes $B_{\nu,k}$ outlined in Section 1, yielding discrete almost symmetric wave packets (see [21] for more details): First, the radial direction is partitioned according to $|\xi'| = c_0 2^k,\ k = 1, \cdots, k_{max}$, where the parameter c_0 defines the radius of the coarsest scale $k = 0$ and $k_{max} = \left\lceil \log_2\left(\frac{N}{2c_0}\right)\right\rceil$. Then, for $d \geq 3$, the set of rotated boxes (1.3) is constructed for each $k > 0$ with the cubed sphere, the d dimensional cube consisting of $2d$ $(d-1)$-dimensional sides. On each side, a grid is constructed, with $N_{c,k} = \left[c_1 2^{(k-1)/2}\right]$ points in one direction. The number of orientations ν at scale k is then given by $\nu_{max,k} = 2d(N_{c,k})^{d-1} \approx \left\lceil (\sqrt{2}dc_1)^{d-1} 2^{\frac{k(d-1)}{2}}\right\rceil \sim 2^{\frac{k(d-1)}{2}}$, the parameter c_1 defining the number of orientations for $k = 1$, and the total number of boxes is:

$$\mathcal{N}_{(\nu,k)} \sim N^{\frac{d-1}{2}}. \tag{3.14}$$

3.2. Deformation, compression and oversampling. The action of F on (3.13) is twofold: Modification of its spatial support, resulting from application of the frequency windows $\hat{\vartheta}^{(r)}_{\nu,k}$, and deformation under the transformation $y \to T_{\nu,k}(y)$. Both need to be accounted for in the discretization by introduction of additional oversampling factors.

Oversampling and calculation domain. Let $\mathcal{E}(x) = T^{-1}_{\nu,k}(x)$ denote the co-moving frame of reference of $Fu_{\nu,k}$, and consider the data component (3.13) with support:

$$\mathcal{U}_{\nu,k} = \operatorname{supp} u_{\nu,k}(x) \subset \left(-\frac{1}{2}, \frac{1}{2}\right]^d.$$

Under the action of F:

$$\tilde{\mathcal{U}}_{\nu,k} = \operatorname{supp}(Fu_{\nu,k})(\mathcal{E}(x)) \subset \left(-\zeta\frac{1}{2}, \zeta\frac{1}{2}\right]^d, \qquad \zeta \geq 1, \qquad \mathcal{U}_{\nu,k} \subseteq \tilde{\mathcal{U}}_{\nu,k},$$

since the functions $\tilde{\phi}_\gamma(\mathcal{E}(x)) = (F\varphi_\gamma)(\mathcal{E}(x))$ constituting $(Fu_{\nu,k})(\mathcal{E}(x))$ have enlarged spatial support w.r.t. $\varphi_\gamma(x)$ in the ξ'' directions. Consequently, the sampling density in ξ has to be increased by a factor $\zeta \geq 1$ w.r.t. the original discretization $\xi_l^{\nu,k}$. To account for this, we construct the above discretization for zero-padded data, i.e. discretization is performed for data $u^{zp}(x_i)$ obtained from the original data $u(x_i)$ by augmenting in each direction with $\lceil(\zeta - 1)N\rceil$ zeros. We denote the resulting box data components (3.13) by $u^{zp}_{\nu,k}$. Note that the frequency support of $\tilde{\phi}_\gamma(\mathcal{E}(x))$ and hence of $(Fu_{\nu,k})(\mathcal{E}(x))$ remains in $\mathbf{1}_{\nu,k}(\xi)$, as is clear from (3.11) and the fact that $\hat{\chi}_{\nu,k}(\xi)$ and $\hat{\chi}_{\nu,k}(\xi)$ are supported on $\mathbf{1}_{\nu,k}(\xi)$.

The amount of spreading of $\tilde{\phi}_\gamma(\mathcal{E}(x))$ in the ξ'' directions can be related to the parameters c_0, c_1 of the wave packet transform and to $\frac{\partial^2 S}{\partial \xi''^2}$, by geometrically imposing connectivity of wave packets at neighboring orientations under the action of F. For instance, let l''_k and $\tilde{l}''_k$ be measures of the width of $\varphi_\gamma(x)$ and $\tilde{\phi}_\gamma(\mathcal{E}(x))$, respectively, for $d = 2$. Then, $\tilde{l}''_k \approx \max(l''_k, 2\frac{\partial^2 S}{\partial \xi''^2}\tan(\mathcal{C}\pi/\nu_{max,k}))$, where $\mathcal{C}$ is a constant depending on the overlap of two neighboring boxes.

In general, only a fraction of the wave-packets $\varphi_{\gamma'}$, $\gamma' : k' = k, \nu' = \nu$ will yield numerically significant contributions to $u_{\nu,k}$ and $(Fu_{\nu,k})$, resulting in effective compression in the wave packet domain [12, 51]. Together with the decay properties (1.6), we obtain an effective reduction of the calculation domain on which $(Fu_{\nu,k})$ actually needs to be evaluated. Indeed, a wave packet $\varphi_\gamma(x)$ has, to precision ε, support in a box $b_{j,\nu,k}$:

$$\operatorname{supp}^\varepsilon \varphi_\gamma = \inf_{x\in X}\{|\varphi_\gamma(x)| \geq \varepsilon\} \subseteq b_{j,\nu,k}, \qquad ||b_{j,\nu,k}|| = l'_k \times (l''_k)^{d-1}, \qquad l'_k \sim 2^{-k}, \quad l''_k \sim 2^{-k/2}, \tag{3.15}$$

and the support of effectively non-zero $u_{\nu,k}$ is therefore:

$$\mathcal{V}_{\nu,k} = \bigcup_j b_{j,\nu,k}, \qquad \mathcal{V}_{\nu,k} \subseteq \mathcal{U}_{\nu,k}. \tag{3.16}$$

Consequently, for $\tilde{\phi}_{j,\nu,k}(x)$ and $(Fu_{\nu,k})(x)$, we have:

$$\begin{aligned} \mathrm{supp}^{\varepsilon}\, \tilde{\phi}_{j,\nu,k} \subseteq \tilde{b}_{j,\nu,k}, \qquad ||\tilde{b}_{j,\nu,k}|| = \tilde{l}'_k \times (\tilde{l}''_k)^{d-1}, \qquad \tilde{l}'_k \sim 2^{-k}, \quad \tilde{l}''_k \sim 2^{-k/2}, \\ \tilde{\mathcal{V}}_{\nu,k} = \bigcup_j \tilde{b}_{j,\nu,k}, \qquad \tilde{\mathcal{V}}_{\nu,k} \subseteq \tilde{\mathcal{U}}_{\nu,k}, \quad \tilde{\mathcal{V}}_{\nu,k} \subseteq \mathcal{V}_{\nu,k}. \end{aligned}$$

Note that the volume of effective supports of $\varphi_\gamma(x)$ and $\tilde{\phi}_{j,\nu,k}(\mathcal{E}(x))$ decay with increasing frequency scale k as: $\mathcal{O}\left(2^{-k}2^{-k\frac{d-1}{2}}\right)$. Oversampling and compression in the evaluation of (3.11) are schematically depicted in Fig. 5.

Deformation and spatial grid resolution. Application of the coordinate transform $y \longrightarrow x = T_{\nu,k}(y)$, $\mathcal{V}_{\nu,k} \longrightarrow \mathcal{W}_{\nu,k} = T^{-1}_{\nu,k}(\mathcal{V}_{\nu,k})$ maps the co-moving reference frame $\mathcal{E}(x)$ onto y and results in a translation and deformation in space of $\tilde{\phi}_{j,\nu,k}(y)$ and $\tilde{\mathcal{V}}$. This yields irregularly spaced samples y_i in $\tilde{\mathcal{W}}_{\nu,k}$ from regularly spaced samples x_i in $\tilde{\mathcal{V}}_{\nu,k}$ (and vice versa, irregularly spaced samples x_n in $\tilde{\mathcal{V}}_{\nu,k}$ from regularly spaced samples y_n in $\tilde{\mathcal{W}}_{\nu,k}$) and may induce a local change in frequency in the output domain y, since the map $x \longrightarrow T^{-1}_{\nu,k}(x)$ can locally contract or expand. Indeed, for two points $\tilde{x}$ and $\tilde{y}$ connected by $\tilde{y} = T^{-1}_{\nu,k}(\tilde{x})$, it follows from (2.8) that:

$$\left.\frac{\partial x}{\partial y}\right|_{\tilde{x}} = \frac{\partial^2 S(\tilde{y},\nu)}{\partial \xi \partial y}. \tag{3.17}$$

The efficient evaluation of the sum over boxes $\sum_{\nu,k}$ requires $(Fu_{\nu,k})(y)$ to be evaluated on (for convenience of visualization, regularly spaced) points y_n on a grid that is common for each box. The simplest choice is given by definition of an (arbitrary) common reference point $y_{n,0}$ and global sampling density Δy, chosen to be at least as dense as the densest local sampling $T^{-1}_{\nu,k}(x_j)$: $\Delta y = \Delta_{min} y = \inf_{j,\nu,k} |T^{-1}_{\nu,k}(x_j + \Delta x) - T^{-1}_{\nu,k}(x_j)|$. Alternatively, we can locally adapt the grid resolution through a hierarchical set of resolution levels $\{\Delta_l y\}$, reflecting (3.17) and constructed, for instance, in a multiresolution manner as $\{\Delta_l y = 2^l \Delta_{min} y\}$, $l = 0, 1, \cdots$, in order to keep the visualization of $(Fu)(y_n)$ efficient.

Once the grid is fixed, the coordinate transform is evaluated by USFFT from discrete frequencies $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to irregularly spaced discrete samples $x_n = T_{\nu,k}(y_n)$ in $\tilde{\mathcal{V}}_{\nu,k}$.
We note that the rate of compression is data-dependent, and so is the resulting reduction in calculation domain $\tilde{\mathcal{V}}_{\nu,k}$, i.e. the effective number of points $x_n = T_{\nu,k}(y_n)$ that actually need to be evaluated per box (ν, k). Also, the output sampling density Δy is controlled by the derivatives of S and hence problem-dependent. In our analyses of computational complexity given below, the resulting alterations of number of operations will be absorbed in the oversampling factor ζ.

3.3. "Box" algorithm. Here we analyze the sequence of operations that need to be performed per box in the evaluation of (3.11). We start from the "inner adjoint" discrete transform (3.9), $\hat{u}^{zp}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$, obtained for zero-padded data $u^{zp}(x_i)$, and perform the following flow of operations:

1. evaluate tensor product functions $\alpha^{(r)}_{\nu,k}(y_n)$ and $\hat{\vartheta}^{(r)}_{\nu,k}(\xi_l^{\nu,k})$, $r = 1, \cdots, R_{\nu,k}$, $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}$
2. for each tensor product term:
   (a) window $\longrightarrow \hat{\vartheta}^{(r)}_{\nu,k}(\xi_l^{\nu,k})\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$
   (b) adjoint USFFT from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = T_{\nu,k}(y_n) \longrightarrow \sum_j \left(\vartheta^{(r)}_{\nu,k} * \varphi_{j,\nu,k}\right)(x_n)$
   (c) multiply with amplitudes $\alpha^{(r)}_{\nu,k}(y_n)$

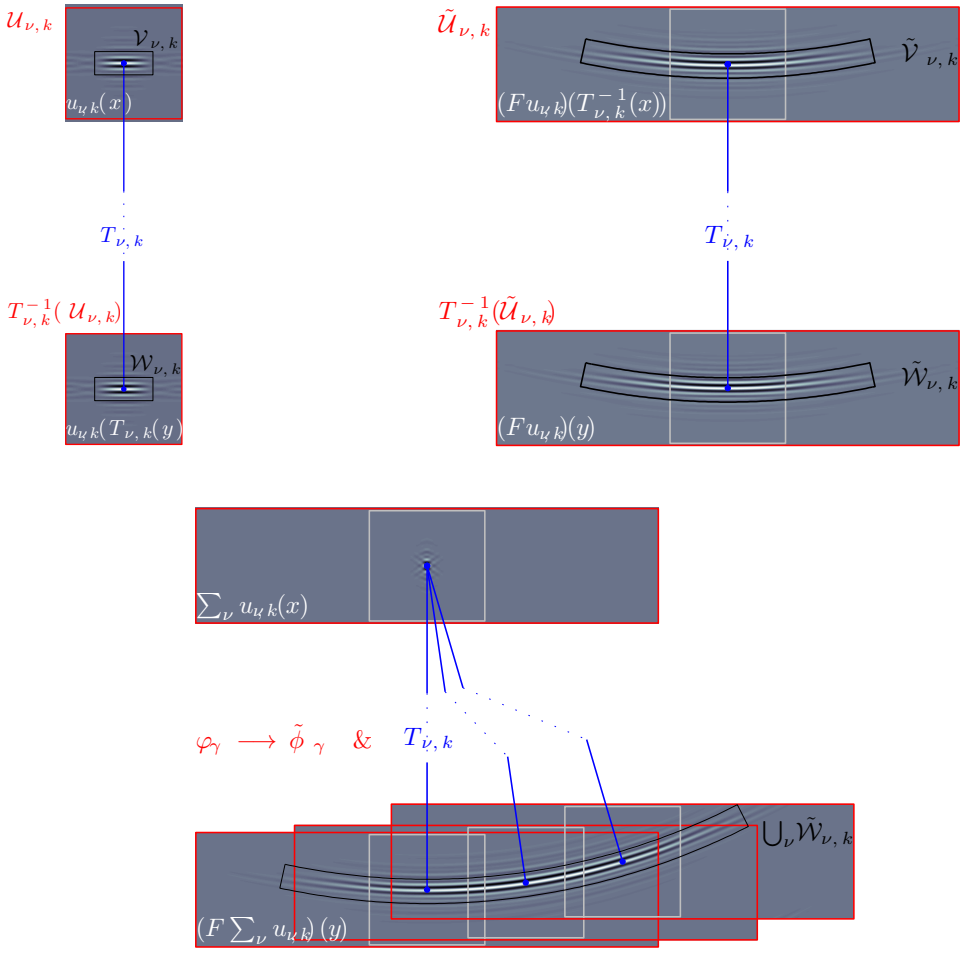

FIG. 5. *Illustration of oversampling and "compression" of calculation domain for one single wave packet (top; left: zero order approximation, right: approximations to $\mathcal{O}(2^{-k/2})$) and for three wave packets with common central position and frequency scale and different orientations (bottom; approximation to $\mathcal{O}(2^{-k/2})$.*

3. stack contributions of tensor-product terms
   $\longrightarrow (Fu_{\nu,k})(y_n) \approx \sum_j \sum_r \alpha^{(r)}_{\nu,k}(y_n) \left(\vartheta^{(r)}_{\nu,k} * \varphi_{j,\nu,k}\right)(x_n)$

The operations count per box (ν, k), including explicitly the constants involved, yields:

- evaluation of $R_{\nu,k}$ PSWFs[3] at $(\zeta N)^d$ spatial points: $\mathcal{O}(cR_{\nu,k}(\zeta N)^d)$
- evaluation of $R_{\nu,k}$ PSWFs at N_ξ frequency points: $\mathcal{O}\left(cR_{\nu,k}cN^{\frac{d+1}{2}}\right)$
- $4R_{\nu,k}$ N_ξ point multiplications: $\mathcal{O}\left(R_{\nu,k}(\zeta N)^{\frac{d+1}{2}}\right)$
- $R_{\nu,k}$ $(\zeta N)^d$ point addition: $\mathcal{O}\left(R_{\nu,k}(\zeta N)^d\right)$
- $R_{\nu,k}$ USFFTs from $\xi^{\nu,k}_l \in \mathbf{1}_{\nu,k}$ to $(\zeta N)^d$ spatial points: $\mathcal{O}(dR_{\nu,k}c_u(\sigma_u\zeta N)^d \log(N))$, where σ_u is the oversampling factor of the USFFT.

Overall, the number of operations to be performed per box is hence:

$$\sim \mathcal{O}\left(dN^d \log(N)\right). \tag{3.18}$$

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

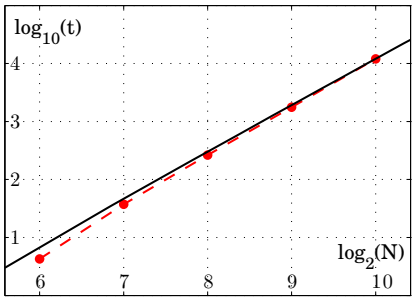

FIG. 6. *Computation time as a function of sample size N (red dots and broken line) and complexity estimate* (3.19) *(black solid) for parametrix of half-wave equation (cf. Section 5.1) in $d = 2$ dimensions in homogeneous medium ($v = 2km/s$). Evolution time is $\mathcal{T} = 5s$).*

We continue with describing a modification of the above algorithm in which we seek to reduce the number of USFFTs by substituting them with standard FFTs. Standard FFT has same computational complexity as USFFT but significantly smaller constants, hence requires less computation time. Here, we start from the "inner adjoint" discrete transform (3.9), $\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$, for *original* data $u(x_i)$. We first obtain the box contribution $u_{\nu,k}$ on the regularly spaced points x_i by USFFT and then zero-pad, use standard FFTs to obtain the tensor-product contributions $\alpha_{\nu,k}^{(r)}(y_i)\left(\vartheta_{\nu,k}^{(r)} * \varphi_{j,\nu,k}\right)(x_i)$, and evaluate the change of coordinates to $x_n = T_{\nu,k}(y_n)$ by one final USFFT:

1. compute 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 $x_i \longrightarrow u_{\nu,k}(x_i)$
2. zero-pad $\longrightarrow u_{\nu,k}^{zp}(x_i)$
3. compute FFT of $u_{\nu,k}^{zp}(x_i)$, inducing regularly spaced frequencies $\tilde{\xi} \longrightarrow \hat{u}(\tilde{\xi})\hat{\beta}_{\nu,k}(\tilde{\xi})\hat{\chi}_{\nu,k}(\tilde{\xi})$
4. compute tensor product functions $\alpha_{\nu,k}^{(r)}(y_i)$ and $\hat{\vartheta}_{\nu,k}^{(r)}(\tilde{\xi})$, $r = 1, \cdots, R_{\nu,k}$, $\tilde{\xi} \in \mathbf{1}_{\nu,k}(\xi)$
5. for each tensor product term:
   (a) window $\longrightarrow \hat{\vartheta}_{\nu,k}^{(r)}(\tilde{\xi})\hat{u}(\tilde{\xi})\hat{\beta}_{\nu,k}(\tilde{\xi})\hat{\chi}_{\nu,k}(\tilde{\xi})$
   (b) inverse FFT from $\tilde{\xi} \in \mathbf{1}_{\nu,k}$ to $x_j \longrightarrow \left(\vartheta_{\nu,k}^{(r)} * \varphi_{j,\nu,k}\right)(x_i)$
   (c) multiply with amplitudes $\alpha_{\nu,k}^{(r)}(y_i)$
6. stack contributions of tensor-product terms
   $\longrightarrow (Fu_{\nu,k})(y_i) \approx \sum_j \sum_r \alpha_{\nu,k}^{(r)}(y_i)\left(\vartheta_{\nu,k}^{(r)} * \varphi_{j,\nu,k}\right)(x_i)$
7. coordinate transform: FFT $x_i \to \tilde{\xi}$ and adjoint USFFT $\tilde{\xi} \in \mathbf{1}_{\nu,k}(\xi) \to x_n = T_{\nu,k}(y_n)$

This requires only one USFFT from irregularly spaced $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}$ to N^d regularly spaced points x_j, and one USFFT from regularly spaced $\tilde{\xi} \in \mathbf{1}_{\nu,k}$ to $(\zeta N)^d$ irregularly spaced points x_n. In addition, $R_{\nu,k} + 2$ FFTs between $\tilde{\xi} \in \mathbf{1}_{\nu,k}$ and $(\zeta N)^d$ spatial points x_i have to be evaluated. The last item in the operations count for the previous algorithm is now replaced by:

- one USFFT without and one with additional oversampling ζ: $\mathcal{O}(d(1+\zeta^d)c_u(\sigma_u N)^d \log(N))$
- $R_{\nu,k} + 2$ FFTs: $\mathcal{O}(d(R_{\nu,k}+2)c_f(\zeta N)^d \log(N))$

The computational complexity remains the same as for the previous algorithm and is given by (3.18).

Finally, the accumulative complexity for the evaluation of all $\mathcal{N}_{(\nu,k)}$ boxes (cf. (3.14)) is:

$$\sim \mathcal{O}\left(dN^{\frac{3d-1}{2}} \log(N)\right). \tag{3.19}$$

Actual computation time as a function of problem size N for $d = 2$ ($\mathcal{D} = 1$) is plotted in Fig. 6 and compared to the complexity estimate (3.19).

3.4. Further approximations: "Packet" algorithms. We proceed with analyzing the discretization of approximations (2.7) and (2.8). In approximation (2.7), the functions $\vartheta_\gamma(\xi)$ are in general different for each data wave packet $\varphi_\gamma(x)$. Therefore, the computation of $\tilde{\phi}_{j,\nu,k}(\mathcal{E}(x_i)) = (\vartheta_\gamma * \varphi_{j,\nu,k})(\mathcal{E}(x_i))$ must be performed for each packet individually. The change of coordinates $T_{\nu,k}$ can still be evaluated for all packets of a box (ν,k) at once, since $T_{\nu,k}$ is independent of index j. We start from the set of data wave packet coefficients (3.4) for zero-padded data $u^{zp}(x_i)$ and obtain the following "hybrid packet-box" algorithm:

for each box (ν,k):
- for each $\gamma' : k' = k, \nu' = \nu$:
  1. FFT from $u_{\gamma'}$ to $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi) \longrightarrow \hat{u}_{\gamma'}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$
  2. window $\longrightarrow \hat{\vartheta}_{\gamma'}(\xi_l^{\nu,k})\hat{u}_{\gamma'}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$
  3. multiply with amplitude $\longrightarrow \hat{\tilde{\phi}}_{\gamma'}(\xi_l^{\nu,k}) = \hat{\vartheta}_{\gamma'}(\xi_l^{\nu,k})\hat{u}_{\gamma'}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$
- sum $\sum_{\gamma'} \hat{\tilde{\phi}}_{\gamma'}(\xi_l^{\nu,k})$
- adjoint USFFT from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $x_n = T_{\nu,k}(y_n)$
  $\longrightarrow (Fu_{\nu,k})(y_n) \approx \sum_j a(y_j^{\nu,k},\nu)\,(\vartheta_\gamma * \varphi_\gamma)\,(x_n)$

stack contributions of boxes (ν,k), yielding the final approximation (2.7) for $(Fu)(y_n)$.

In contrast, for approximation (2.8), the approximate change of coordinates also has to be evaluated packet per packet. We obtain a pure "packet" algorithm, with operations per individual packet:

1. FFT from u_γ to $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi) \longrightarrow \hat{u}_\gamma(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$
2. window $\longrightarrow \hat{\vartheta}_\gamma(\xi_l^{\nu,k})\hat{u}_\gamma(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$
3. multiply with amplitude $\longrightarrow \hat{\tilde{\phi}}_\gamma(\xi_l^{\nu,k}) =)\hat{\vartheta}_\gamma(\xi_l^{\nu,k})\hat{u}_\gamma(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k})$
4. adjoint USFFT from $\xi_l^{\nu,k} \in \mathbf{1}_{\nu,k}(\xi)$ to $\left(DT_\gamma(y_n - y_j^{\nu,k}) + M_\gamma \cdot (y_n - y_j^{\nu,k})^2\right) \longrightarrow (F\varphi_\gamma)(y_n)$

Summing the contributions from all data wave packets yields the action (Fu) under approximation (2.8). The operations count for evaluating approximations (2.7) and (2.8) for one single wave packet is:

- one N_ξ point FFT: $\mathcal{O}\left(d(\zeta N)^{\frac{d+1}{2}}\log(N)\right)$
- four N_ξ point multiplications and one multiplication with a constant: $\mathcal{O}\left((\zeta N)^{\frac{d+1}{2}}\right)$
- one adjoint USFFT from N_ξ frequency points to $(\zeta N)^d$ irregularly spaced points in space: $\mathcal{O}\left(dc_u(\sigma_u\zeta N)^d\log(N)\right)$

yielding:

$$\sim \mathcal{O}\left(dN^d\log(N)\right). \tag{3.20}$$

The total number of wave packet coefficients is roughly $\mathcal{O}(N^d)$, yet the number of data wave packets with practically non-zero coefficients strongly depends on the data and typically amounts to only a small fraction of this number. Let us nonetheless write out the complexity under the assumption that F needs to be evaluated for all frame elements, bearing in mind that this situation corresponds to limit cases such as a Dirac impulse or the absence of coherent structures (viz., random noise). For approximation (2.8), this gives:

$$\sim \mathcal{O}\left(dN^{2d}\log(N)\right), \tag{3.21}$$

and for the hybrid packet-box approximation (2.7):

$$\sim \mathcal{O}\left(dN^{\frac{3d+1}{2}}\log(N)\right), \tag{3.22}$$

which is above approximation (2.5), but below approximation (2.7) since the coordinate transform USFFT can be performed per box (ν,k).

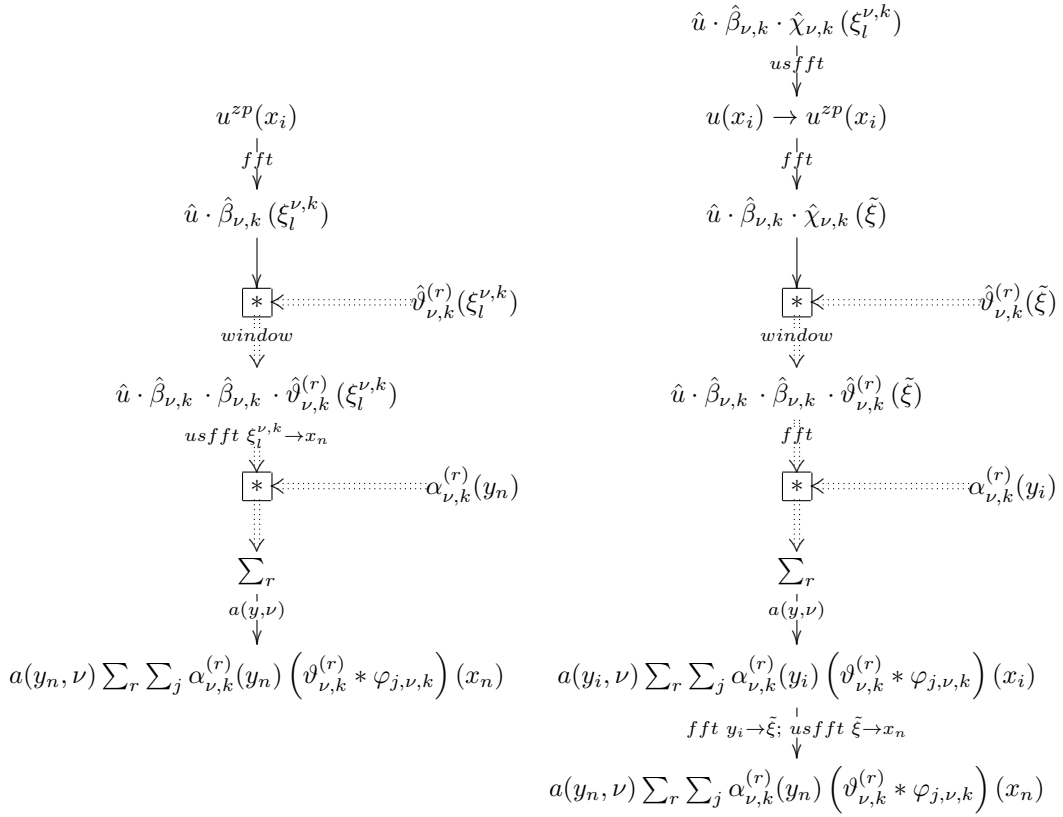

TABLE 3.1
"Box" algorithm (left), with FFTs replacing USFFTs (right), for one box (ν,k). Double arrows indicate operations performed for each individual tensor-product term, $r=1,\cdots,R_{\nu,k}$.

3.5. Example: Trotter product. Here, we analyze approximations (2.5), (2.7) and (2.8) for evolution equations for the specific case of discretization of evolution time into a sequence of small time steps. Consider the evolution equation

$$[\partial_t+\mathrm{i}P(t,x,D_x)]u=0,\quad u|_{t=t_0}=u_0, \tag{3.23}$$

on the interval $t\in[t_0,T]$, where P is a pseudodifferential operator with symbol in $S^1_{1,0}$, in $\mathbb{R}^d$ (in the case of the half wave equation, $P=P(x,\xi)=\sqrt{c(x)^2||\xi||^2}$). The solution operator, $F(t,t_0)$ say, can be written in the form of a Trotter product, resulting in a computational scheme driven by marching-on-in-t. If $t\geq t_N>t_{N-1}>\cdots>t_0$, 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 $N\to\infty$. 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

$$\bar{F}(t'+\Delta,t')u(t',.)(y)=(2\pi)^{-n}\int\exp[\mathrm{i}\,(P(t',y,\xi)\Delta-\langle\xi,y\rangle)]\,\hat{u}(t',\xi)\,\mathrm{d}\xi, \tag{3.24}$$

where $P(t',y,\xi)$ denotes the symbol of the operator P in (3.23). This is a Fourier integral operator of order 0 in the class considered in this paper, with the simple substitution

$$a(y,\xi)=1,\quad S(y,\xi)=P(t',y,\xi)\Delta-\langle\xi,y\rangle. \tag{3.25}$$

The associated canonical transformation is given by

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

$u_{\gamma'}$	$u_{\gamma'}$
$\vdots\ fft$	$\vdots\ fft$
$\hat{u}_{\gamma'} \cdot \hat{\beta}_{\nu,k}\left(\xi_l^{\nu,k}\right)$	$\hat{u}_{\gamma'} \cdot \hat{\beta}_{\nu,k}\left(\xi_l^{\nu,k}\right)$
$\vdots\ window, amp$	$\vdots\ window, amp$
$a(y_j^{\nu,k},\nu)\hat{u}_{\gamma'} \cdot \hat{\beta}_{\nu,k} \cdot \hat{\chi}_{\nu,k} \cdot \hat{\vartheta}_{\gamma'}\left(\xi_l^{\nu,k}\right)$	$a(y_j^{\nu,k},\nu)\hat{u}_{\gamma'} \cdot \hat{\beta}_{\nu,k} \cdot \hat{\chi}_{\nu,k} \cdot \hat{\vartheta}_{\gamma'}\left(\xi_l^{\nu,k}\right)$
$\vdots$	$\vdots$
$\sum_{\gamma':k'=k,\nu'=\nu}$	$usfft\ \xi_l^{\nu,k} \to \left(DT_\gamma \cdot (y_n - y_j^{\nu,k}) + M_\gamma \cdot (y_n - y_j^{\nu,k})^2\right)$
$\vdots\ usfft\ \xi_l^{\nu,k} \to x_n$	$\sum_{\gamma':k'=k,\nu'=\nu}$
$\sum_j a(y_j^{\nu,k},\nu)\left(\vartheta_\gamma * \varphi_\gamma\right)(x_n)$	$\sum_j a(y_j^{\nu,k},\nu)\left(\vartheta_\gamma * \varphi_\gamma\right)\left(DT_\gamma \cdot (y_n - y_j^{\nu,k}) + M_\gamma \cdot (y_n - y_j^{\nu,k})^2\right)$

TABLE 3.2

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

with the Hamilton system,

$$\frac{dx}{dt} = \frac{\partial P}{\partial \xi}(t,x,\xi)\,, \qquad \frac{d\xi}{dt} = -\frac{\partial P}{\partial x}(t,x,\xi)\,, \tag{3.26}$$

it follows that

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

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; this is discussed in more detail in Section 4.

In the case of depth extrapolation, 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

$$P(z,(x,t),(\xi,\tau)) = -\tau\sqrt{c(z,x)^{-2} - \tau^{-2}|\xi|^2}, \tag{3.27}$$

and

$$S((y,t),(\xi,\tau)) = P(z',(y,t),(\xi,\tau))\Delta - \langle \xi, y\rangle - \tau\, t. \tag{3.28}$$

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:

$$\frac{\partial P}{\partial \xi}(z',(y,t),\nu) = \frac{\tau_\nu^{-1}\xi_\nu}{\sqrt{c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2}}, \quad \frac{\partial P}{\partial \tau}(z',(y,t),\nu) = -\frac{c(z',y)^{-2}}{\sqrt{c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2}}, \tag{3.29}$$

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

$$\tau_\nu \frac{\partial^2 P}{\partial \xi^2}(z',(y,t),\nu) = \frac{[c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2]\, I - \tau_\nu^{-2}\xi_\nu \otimes \xi_\nu}{\left[c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}}, \tag{3.30}$$

$$\tau_\nu \frac{\partial^2 P}{\partial \tau^2}(z',(y,t),\nu) = -\frac{c(z',y)^{-2}\tau_\nu^{-2}|\xi_\nu|^2}{\left[c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}},$$

$$\tau_\nu \frac{\partial^2 P}{\partial \xi \partial \tau}(z',(y,t),\nu) = -\frac{c(z',y)^{-2}\tau_\nu^{-1}\xi_\nu}{\left[c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}}.$$

Hence, with $(\xi', \xi'') = R_\nu^{-1}(\xi,\tau)$ and $\xi'' = \tilde{R}_\nu^{-1}(\xi,\tau)$ [4], and with:

$$\partial_{(\xi',\xi'')} P(.,(.,.),R_\nu(\xi',\xi'')) = R_\nu^{-1}(\partial_{(\xi,\tau)}P)(.,(.,.),R_\nu(\xi',\xi'')),$$

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

the expression for the phase expansion of the operator is:

$$\left\langle \xi, \frac{\partial P}{\partial \xi}(z',(y,t),\nu)\right\rangle + \tau \frac{\partial P}{\partial \tau}(z',(y,t),\nu) + \frac{1}{2\xi'}\left\langle \xi'', \frac{\partial^2 P}{\partial \xi''^2}(z',(y,t),\nu)\, \xi''\right\rangle \tag{3.31}$$

$$= \frac{\langle \xi, \tau_\nu^{-1}\xi_\nu\rangle - \tau\, c(z',y)^{-2}}{\sqrt{c(z',y)^{-2} - \tau_\nu^{-2}|\xi_\nu|^2}} +$$

$$\frac{1}{2\xi'}\left\langle \xi'', \left(\tilde{R}_\nu^{-1}\left[\tau_\nu^{-1}\tilde{R}_\nu^{-1}\begin{pmatrix} \frac{[c(z',y)^{-2}-\tau_\nu^{-2}|\xi_\nu|^2]\, I - \tau_\nu^{-2}\xi_\nu\otimes\xi_\nu}{\left[c(z',y)^{-2}-\tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}} & -\frac{c(z',y)^{-2}\tau_\nu^{-1}\xi_\nu}{\left[c(z',y)^{-2}-\tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}} \\ -\frac{c(z',y)^{-2}\tau_\nu^{-1}\xi_\nu}{\left[c(z',y)^{-2}-\tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}} & -\frac{c(z',y)^{-2}\tau_\nu^{-2}|\xi_\nu|^2}{\left[c(z',y)^{-2}-\tau_\nu^{-2}|\xi_\nu|^2\right]^{3/2}} \end{pmatrix}^T\right]^T\right)\xi''\right\rangle.$$

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 a Fourier integral operator [5]; it converges in 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 $\frac{1}{2} \le \alpha$, $s = 1$, and balance of accuracies $\mathcal{O}(\Delta^{1/2})$ and $\mathcal{O}(2^{-k/2})$ requires $\Delta \sim 2^{-k}$ [16, 43]. Multi-composition of Fourier integral operators to approximate Cauchy IVPs were initially proposed by Kumano-go and Taniguchi; their approximation tends to the exact solution up to a regularizing operator. The underlying method is based on the computation and estimation of phase functions and amplitudes of the Fourier integral operators appearing in these multi-products, and is formalized in the Kumano-go-Taniguchi theorem.

We can now construct a process similar to 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 Fig. 7 (left), we show curvilinear coordinates particular to wave packets, which enable to define tubes to which the propagation is confined[6] (see e.g. [11], Fig. 1 and 2).

[4] That is, $\tilde{R}_\nu^{-1}$ is R_ν^{-1} without the first row.

[5] Geometrically. $\mathcal{W}_N(z,z_0)$ has some similarities with the wavefront construction method for computing the propagation of singularities.

[6] Here, 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.

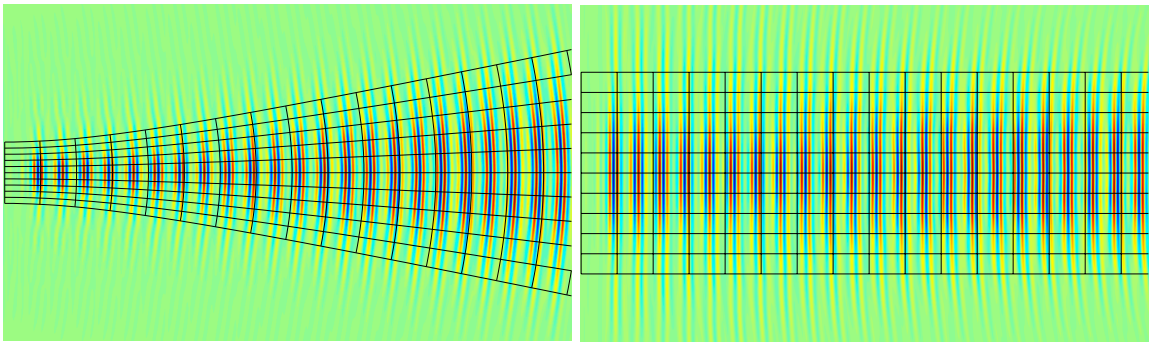

FIG. 7. *A "beam" of wave packets in homogeneous background under approximation* (2.5) *for the half-wave equation, in Cartesian coordinates* (x, z) *(left; z horizontal) and elliptic coordinates* $x = a\cosh(\mu)\cos(\varsigma)$, $z = a\sinh(\mu)\sin(\varsigma)$ *(right; μ horizontal); elliptic coordinate 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.*

4. Parametrix. In order to perform actual computations, the values of $\frac{\partial S}{\partial \xi}(y,\nu)$ and $\frac{\partial^2 S}{\partial \xi'^2}(y,\nu)$ need to be known. The generating function $S(y,\xi)$ is the solution to the Hamilton-Jacobi equation, which is in general not accessible in global closed form expression. Yet, the proposed algorithms do not require knowledge of $S(y,\xi)$ itself, but only of its first and second order derivatives at a finite number of discrete points (y,ν). Here, we detail how these derivatives can be obtained numerically for parametrices of evolution equations, which represent examples of the FIOs considered in the present work. Evolution equations play an important role in imaging and inverse scattering applications and generate extended imaging [22, 23]. As a special case, this includes "thin-slab" propagation as described in Section 3.5, in which straight rays and closed form expressions approximate the first and second order terms of the phase expansion, respectively, for small time steps. Here, we establish the connection with paraxial ray theory for (arbitrarily) large time steps and show that the derivatives of S can be obtained numerically from paraxial solutions to the Hamilton-Jacobi system in Fermi coordinates as specific combinations of blocks of propagator matrices. We note that effectively, this turns the numerical procedures described in the previous section into approximate one-step solvers for Cauchy initial value problems for evolution equations for (potentially very) large time steps.

4.1. First-order derivatives of generating function S. Coordinate transform. Let $\mathcal{H}(y,\eta)$ be the Hamiltonian governing the Hamiltonian flow associated with an evolution equation, and let evolution time be from $t = t_0$ to $t = \mathcal{T}$, where potentially $\mathcal{T} \gg t_0$. For convenience, we parametrize the initial conditions in this section by $(y_0, \eta_0) = (y(t = t_0), \eta(t = t_0))$; consistency with the notation in previous sections is obtained by setting $(x, \xi) = (y_0, \eta_0)$. Our choice of notation is summarized in the diagram Fig. 8. We will freely switch between these notations to link with expressions in previous sections. Let us denote the solution for fixed initial conditions by $\eta(t)) = (y(y_0, \eta_0, t), \eta(y_0, \eta_0, t))$. We have (cf. (1.2) and (2.2)):

$$\begin{aligned} x = \quad y_0 = \quad & \frac{\partial S}{\partial \xi}(y(\mathcal{T}),\nu) = T_{\nu,k}(y(\mathcal{T})) \\ \eta = \quad \eta(\mathcal{T}) = \quad & \frac{\partial S}{\partial y}(y(\mathcal{T}),\nu). \end{aligned} \tag{4.1}$$

The bi-characteristic $(y(t), \eta(t))$, commonly referred to as a ray, is the solution to the characteristic system:

$$\frac{d}{dt}\begin{pmatrix} y \\ \eta \end{pmatrix} = \begin{pmatrix} \frac{\partial \mathcal{H}(y,\eta)}{\partial \eta} \\ -\frac{\partial \mathcal{H}(y,\eta)}{\partial y} \end{pmatrix}. \tag{4.2}$$

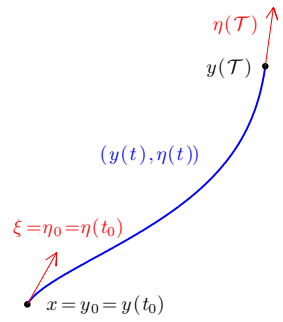

FIG. 8. *Illustration of evolution under the system* (4.2).

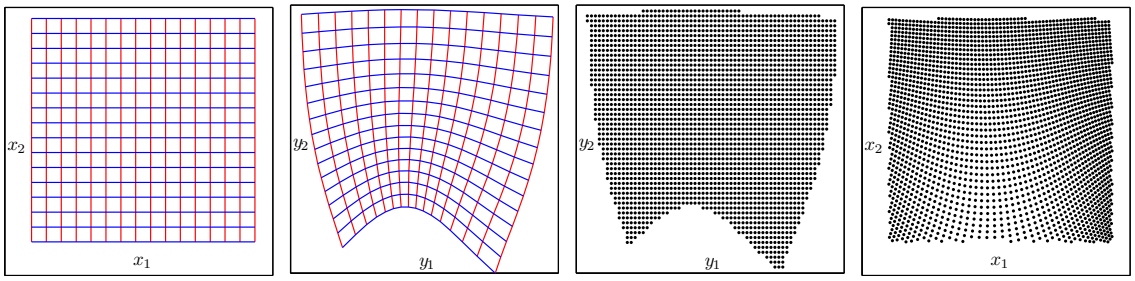

FIG. 9. *Schematic illustration of discrete evaluation of the coordinate transform* $T_{\nu,k}$*:* $y_m(\mathcal{T}) = T^{-1}_{\nu,k}(x_m)$ *from regularly spaced* x_m *(left); interpolation of* $x_m = T_{\nu,k}(y_m(\mathcal{T}))$ *at regularly spaced* y_n *gives* $x_n = T_{\nu,k}(y_n(\mathcal{T}))$ *(right).*

Integration of the system (4.2) yields the map:

$$y(\mathcal{T}) = T^{-1}_{\nu,k}(x). \tag{4.3}$$

and gives $y_j^{\nu,k} = y(\mathcal{T}) = T^{-1}_{\nu,k}(x_j^{\nu,k})$ for approximation (2.8). For numerical evaluation of approximations (2.5) and (2.7), the inverse relation $x_n = T_{\nu,k}(y_n(\mathcal{T}))$ needs to be evaluated for regularly spaced points y_n. It can be obtained by first solving (4.2) from t_0 to $\mathcal{T}$ with initial conditions (x_m, η_0), where x_m is a discrete set of points on $\tilde{\mathcal{V}}_{\nu,k}$, yielding the irregularly spaced set of points $y_m(\mathcal{T}) = T^{-1}_{\nu,k}(x_m)$ on the bi-characteristic. Then, $x_n = T_{\nu,k}(y_n(\mathcal{T}))$ is evaluated by back-tracking rays from (y_n, η_n) subject to $\xi_n = \xi$ and evolution from time t_0 to $\mathcal{T}$. Alternatively, x_n can be obtained by interpolation of the map $(x_m, \xi) \to (y_m, \eta_m)$ on (projections of) the Lagrangian of the operator (e.g. [34]). In the absence of caustics, this is equivalent to (standard) interpolation of $y_m \to x_m$ at regularly spaced points y_n (cf. Fig. 9).

4.2. Second-order derivatives of generating function S. The second-order derivatives of the generating function $S(y(\mathcal{T}), \nu)$ can be obtained from the derivatives:

$$\frac{\partial(y,\eta)}{\partial(y_0,\eta_0)}(y_0,\eta_0,\mathcal{T}) \tag{4.4}$$

evaluated in Fermi coordinates (cf. e.g. [33]) for standard plane wave and point source initial conditions. Fermi coordinates have similar properties as ray-centered coordinates, commonly employed in the geophysical literature (see e.g. in [56]), and are defined as follows. Let $f_2, \cdots, f_d$ be a set of orthonormal vectors in the plane tangent to the wave front at $t = t_0$, and let $f_1 = \frac{\partial y}{\partial t}(y_0, \eta_0, t_0)$. The subset f_I, $I = 2, \cdots, d$ can be chosen with arbitrary orientation in the tangent plane. Denote by $f_i(t)$ the coordinate system f_i transported parallel along the ray. The Fermi coordinates y_f are coordinates in this system, where $y_f^1 = t$ is time along ray, and $y_f^2, \cdots, y_f^d$ essentially describe the

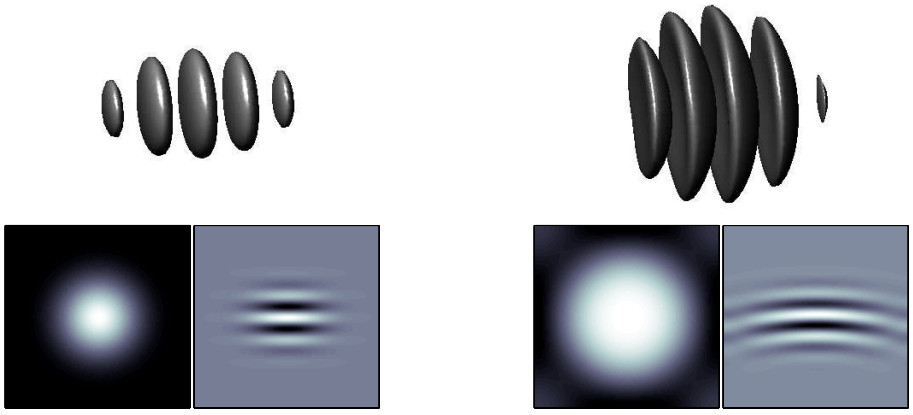

FIG. 10. *Parametrix of half-wave equation: A wave packet in $d = 3$ dimensions. Iso-amplitude plot of initial wave packet (top left) and resulting packet after propagation for ≈ 5 wave lengths in homogeneous background (top right); corresponding cross-sections (bottom line; normal to / in plane with axis of symmetry, respectively).*

distance from the ray $(y(t), \eta(t))$. The transformation matrices with global Cartesian coordinates read:

$$H_{ij}(t) = \frac{\partial y^i}{\partial y_f^j}(t,0) = f_j^i(t), \qquad \bar{H}_{jl}(t) = \frac{\partial y_f^j}{\partial y^l}(t,0), \qquad H\bar{H} = I_{d\times d}, \tag{4.5}$$

where f_j^i is the i-th component of f_j with corresponding cotangent vectors η_f^i. The components f_I in the tangent plane satisfy:

$$\frac{df_I(t)}{dt} = -1/\langle \eta, \eta\rangle \left\langle f_I, -\frac{\partial \mathcal{H}(x,\eta)}{\partial x}\right\rangle \eta, \tag{4.6}$$

and f_1 is known from the solution of (4.2). We fix f_I, $I = 2, \cdots, d$ to coincide with the unit vectors lying on the ξ'' axes (cf. Fig. 1, right).

In Cartesian coordinates, the derivatives (4.4) are the solution to the linear system:

$$\frac{d}{dt}\left(\frac{\partial(y,\eta)}{\partial((y_0,\eta_0)}(y_0,\eta_0,t)\right) = \begin{pmatrix} \frac{\partial^2\mathcal{H}(y,\eta)}{\partial\eta\partial y} & \frac{\partial^2\mathcal{H}(y,\eta)}{\partial\eta^2} \\ -\frac{\partial^2\mathcal{H}(y,\eta)}{\partial y^2} & -\frac{\partial^2\mathcal{H}(y,\eta)}{\partial y\partial\eta} \end{pmatrix} \cdot \frac{\partial(y,\eta)}{\partial((y_0,\eta_0)}(y_0,\eta_0,t). \tag{4.7}$$

Transformation to Fermi coordinates and reduction to the subsystem f_I, $I = 2, \cdots, d$, in the tangent plane yields (cf. e.g. [56]):

$$\frac{d}{dt}\left(\frac{\partial(y_{f_I},\eta_{f_I})}{\partial(y_{f_I,0},\eta_{f_I,0})}(0,0,t)\right) = \begin{pmatrix} A_{f_I} & B_{f_I} \\ C_{f_I} & D_{f_I} \end{pmatrix} \cdot \frac{\partial(y_{f_I},\eta_{f_I})}{\partial(y_{f_I,0},\eta_{f_I,0})}(0,0,t). \tag{4.8}$$

where

$$A_{f_I,MN} = \bar{H}_{Mn}H_{mN}\left(\frac{\partial^2\mathcal{H}(y,\eta)}{\partial\eta_n\partial y_m} - \frac{1}{\langle\eta,\eta\rangle}\eta_n\frac{\partial\mathcal{H}(y,\eta)}{\partial y_m}\right) \tag{4.9}$$

$$B_{f_I,MN} = \bar{H}_{Mn}\bar{H}_{Nm}\frac{\partial^2\mathcal{H}(y,\eta)}{\partial\eta_n\partial\eta_m} \tag{4.10}$$

$$C_{f_I,MN} = H_{nM}H_{mN}\frac{\partial^2\mathcal{H}(y,\eta)}{\partial y_n\partial y_m} \tag{4.11}$$

$$D_{f_I,MN} = H_{nM}\bar{H}_{Nm}\left(\frac{\partial^2\mathcal{H}(y,\eta)}{\partial y_n\partial\eta_m} - \frac{\partial\mathcal{H}(y,\eta)}{\partial y_n}\frac{\partial\mathcal{H}(y,\eta)}{\partial y_m}\right). \tag{4.12}$$

Let $\Pi_{\mathcal{T}}$ denote the fundamental matrix of (4.8):

$$\Pi_{\mathcal{T}} = \begin{pmatrix} W_1 & W_2 \\ W_3 & W_4 \end{pmatrix}, \tag{4.13}$$

$$\frac{\partial(y_{f_I}, \eta_{f_I})}{\partial((y_{f_I,0}, \eta_{f_I,0})}(\mathcal{T}) = \Pi_{\mathcal{T}} \frac{\partial(y_{f_I}, \eta_{f_I})}{\partial((y_{f_I,0}, \eta_{f_I,0})}(t_0),$$

that is, $\Pi_{\mathcal{T}}$ is the solution to (4.8) with standard (plane wave + point source) initial conditions:

$$\frac{\partial(y_{f_I}, \eta_{f_I})}{\partial((y_{f_I,0}, \eta_{f_I,0})}(t_0) = \begin{pmatrix} I_{d-1\times d-1} & 0_{d-1\times d-1} \\ 0_{d-1\times d-1} & I_{d-1\times d-1} \end{pmatrix}.$$

Then, we have:

$$\frac{\partial^2 S}{\partial \xi''^2}(y(\mathcal{T}), \nu) = -W_1^{-1} W_2. \tag{4.14}$$

Packet algorithm. We finally give the explicit expressions for the matrices DT_γ and M_γ in the approximate coordinate transform:

$$DT_\gamma (y_n - y_j^{\nu,k}) + M_\gamma \cdot (y_n - y_j^{\nu,k})^2$$

in approximation (2.8). For convenience of notation, denote:

$$H^t = H(t), \qquad \bar{H}^t = (H(t))^{-1},$$

and let G^t, $\bar{G}^t$ be the respective transformation matrices between global Cartesian coordinates and the local Cartesian coordinate system $g_i(t) = \{g_1(t), f_2(t), \ldots, f_d(t)\}$, where $g_1(t)$ is the unit vector normal to the wavefront at its intersection with ray $(y(t), \eta(t))$. The matrix DT_γ consists of dilation terms in directions $g_i(\mathcal{T})$, shear terms in directions $g_I(\mathcal{T})$, and rotation from $g_i(\mathcal{T})$ to $g_i(t_0)$, or equivalently, dilations in $f_i(\mathcal{T})$ and transformation from $f_i(\mathcal{T})$ to $f_i(t_0)$. Let be:

$$T_{\gamma,\mathcal{T}} = \begin{bmatrix} 1 & 0_{1\times d-1} \\ 0_{d-1\times 1} & W_1 \end{bmatrix}^{-1}. \tag{4.15}$$

Then, DT_γ is given by:

$$DT_\gamma = \bar{H}^0 (H^{\mathcal{T}} T_{\gamma,\mathcal{T}}). \tag{4.16}$$

The matrix M_γ consists of quadratic terms in the $f_i(\mathcal{T})$ directions, and rotation from $g_i(\mathcal{T})$ to $g_i(t_0)$. With:

$$P_{\gamma,\mathcal{T}} = \begin{bmatrix} 0 & 0_{1\times d-1} \\ 0_{d-1\times 1} & W_3 W_1^{-1} \end{bmatrix}, \tag{4.17}$$

we obtain:

$$M_\gamma \cdot (y_n - y_j^{\nu,k})^2 = \bar{G}^0 \left((y_n - y_j^{\nu,k})^T (G^{\mathcal{T}} P_{\gamma,\mathcal{T}}) (y_n - y_j^{\nu,k}) g_1(\mathcal{T}) \right). \tag{4.18}$$

5. Applications and numerical examples. In this section, we illustrate and compare the proposed approximations in numerical examples. The first example consists in wave evolution in isotropic homogeneous medium and in isotropic heterogeneous medium in $d = 2$ dimensions. It will serve us as a basis for analyzing and comparing the different approximations proposed in the present work. The second example demonstrates evolution equation based imaging and involves an anisotropic, homogeneous Hamiltonian. In general, the formulation of imaging operators in terms of solution operators of evolution equations is obtained through extension to an extended image domain [22, 23], with at least $d = 3$ dimensions. For the purpose of illustration, we restrict ourselves here to $d = 2$ dimensions.

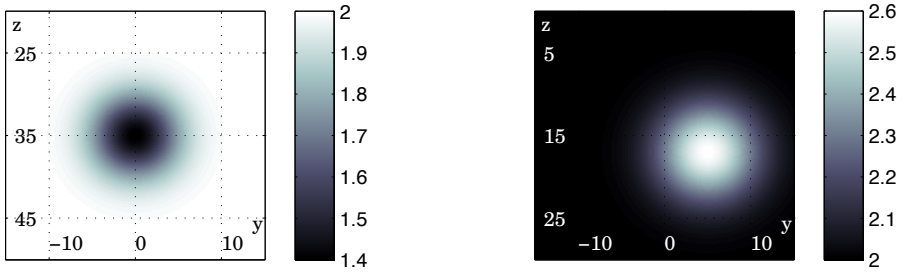

FIG. 11. *Lens* (5.4)*: low velocity lens (left) and high velocity lens (right) used in numerical examples in Section 5.1.*

5.1. Wave propagation – isotropic, heterogeneous case. Consider the initial value problem:

$$(\partial_t - iP)u = 0, \qquad u(x, t_0) = f(x), \tag{5.1}$$

for the half-wave equation, i.e.:

$$P(x, \xi) = \sqrt{c(x)^2 ||\xi||^2}, \tag{5.2}$$

where $c(x)$ is the medium velocity. We compare the accuracy of the "box" algorithm approximation (2.5), the "hybrid packet-box" algorithm approximation (2.7), and the "packet algorithm" approximation (2.8) to zero order approximation for (5.1) with band-limited Dirac initial conditions $u(x, t = 0)$ and large evolution time $\mathcal{T}$. We define the band-limited Dirac source in the ξ domain as:

$$\hat{u}(\xi, \omega) = \sum_{k'} \sum_{|\nu - \nu_c| \leq \Delta} \hat{\chi}_{\nu,k'}(\xi), \tag{5.3}$$

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

$$c(x) = c_0 + \mu \exp(-|x - x_0|^2/\sigma^2), \tag{5.4}$$

where $c_0 = 2km/s$, $\mu = -0.3km/s$, $\sigma = 5km$ and $x_0 = (0, 35)km$. The low velocity lens is depicted in Fig. 11 (left). The output spatial sampling density Δy is set equal to the initial sampling density Δx. We consider evolution time $\mathcal{T} = 30s$ for the homogeneous case, and $\mathcal{T} = 20s$ for the heterogeneous case. With this parameter setup, the boundaries of the calculation domains span roughly $N_1 \times N_2 = 1900 \times 300$ and 1100×300 samples for the homogeneous and for the heterogeneous case, respectively.

Fig. 12 (homogeneous case) and Fig. 13–14 (heterogeneous case) compare the different approximations of $(Fu)(y_n)$: zero-order approximation (top row); approximations (2.5), (2.7) and (2.8) (second row for homogeneous case, second to fourth row for heterogeneous case). The bottom row compares the amplitudes along the wavefront. The left columns correspond to initial (5.3) with $k' = 3$ only, the columns on the right with all frequency scales $k' = 1 - 4$. In Fig. 14 we include physical amplitudes $a(y, \nu)$, while in Fig. 13, we set $a(y, \nu) = 1$ for better visual comparison of the different approximations. Note that in the homogeneous case, approximations (2.5), (2.7) and (2.8) are equivalent since $\frac{\partial^2 S}{\partial \xi'^2}(y, \nu) = c_0 \mathcal{T}$ is independent of y, $T_{\nu,k}$ describes, for fixed (ν, k), parallel

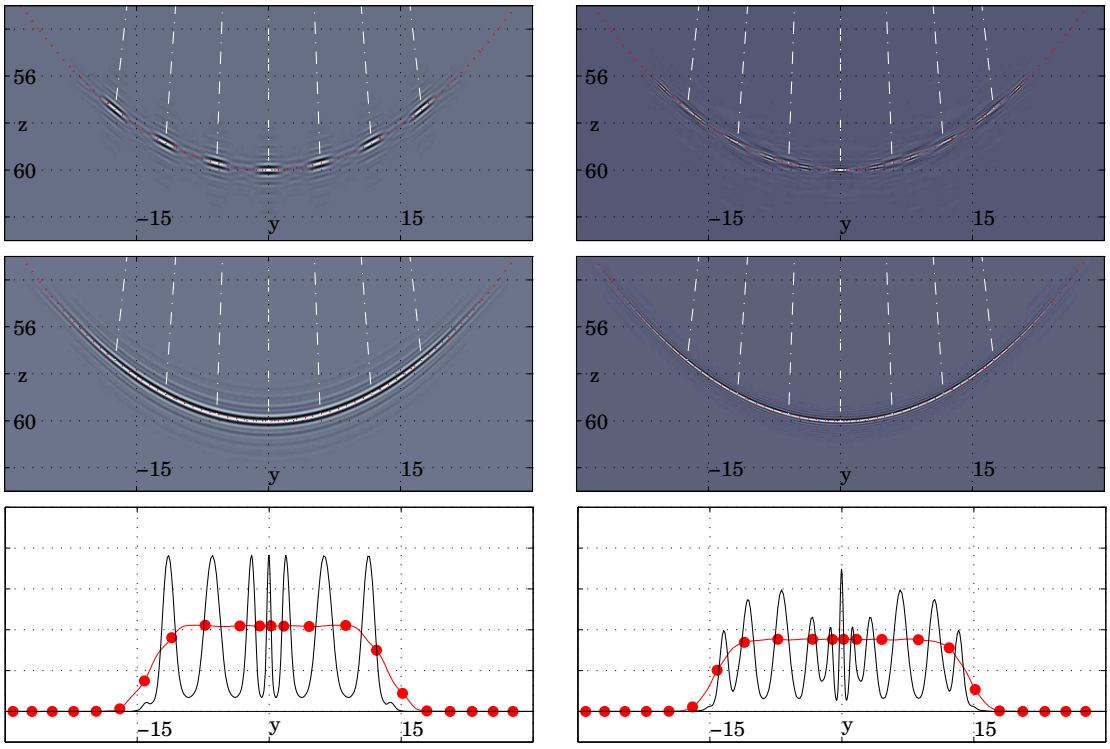

FIG. 12. *Wave propagation in isotropic, homogeneous medium for initial conditions* (5.3) *with* $k' = 2$ *(left column) and* $k' = 1 - 4$ *(right column): zero order approximation (top row), approximations* (2.5), (2.7) *and* (2.8) *(center row), and corresponding amplitudes along wave front (bottom 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 ration is not equal to one.*

straight rays of path length $c_0\mathcal{T}$, $DT_\gamma = I_{d\times d}$ and $M_\gamma = 0_{d\times d}$. Consequently, in this case, the zero order approximation is equal to rigid motion.

We start with investigating the homogeneous ("straight rays") case (cf. Fig. 12). As observed in [20], under the zero order approximation, the wave front breaks apart, with constituting wave packets at given scale ending up disconnected (top row): The wave packets do not receive any deformation and are merely de-placed data wave packets, resulting in large gaps in the wave front due to the geometry of propagation when $c_0\mathcal{T}$ 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. We note that 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): Whereas amplitudes under approximations (2.5), (2.7) and (2.8) are essentially constant, zero-order approximation results in strong fluctuations, regardless of scale k.

Let us now turn our attention to the heterogeneous case, cf. Fig. 13 and 14. As above, the wave front breaks apart under the zero order approximation (top rows), with error not decreasing with increasing scale k. Yet, unlike rigid motion, the coordinate transformation $T_{\nu,k}$ accounts for the full deformation resulting from the underlying canonical transformation. Still, as in the homogeneous case, only center points $y_l^{\nu,k}$ sit precisely on the singularity. In contrast, under approximation (2.5) (second rows), 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

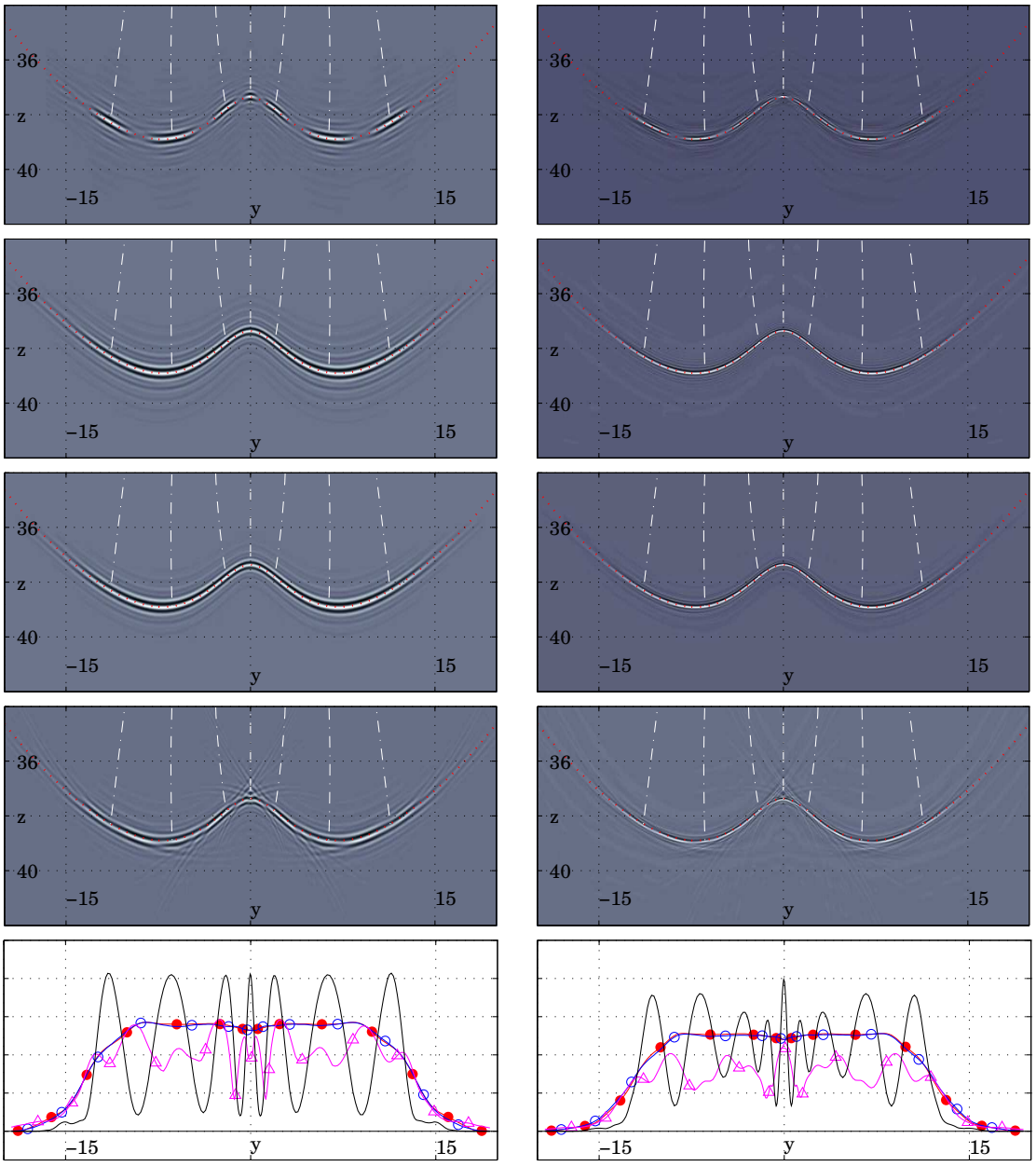

FIG. 13. *Wave propagation in isotropic, heterogeneous medium for initial conditions* (5.3) *with* $k' = 2$ *(left column) and* $k' = 1 - 4$ *(right column), and physical amplitudes set to* $a(y, \nu) = 1$*: zero order approximation (top row), approximation* (2.5) *(second row),* (2.7) *(third row) and* (2.8) *(fourth row); corresponding amplitudes along wave front (bottom row): zero order approximation (solid black), approximation* (2.5) *(red dot), approximation* (2.7) *(black circle), approximation* (2.8) *(triangle). The white dot-dashed lines indicate rays of seven wave-packets at scale* $k = 3$*. Note that the aspect ration is not equal to one.*

at $x = 0$ caused by the low velocity lens, resulting in packets being "squeezed" in their direction of propagation. Results obtained under approximation (2.7) (third rows) are very similar, despite the further approximation of $\frac{\partial^2 S}{\partial \xi'^2}(y, \nu)$ by the center location values $\frac{\partial^2 S}{\partial \xi'^2}(y_j^{\nu,k}, \nu)$. The reason for this lies in the fact that in this example, the dependence of $\frac{\partial^2 S}{\partial \xi'^2}(y, \nu)$ on y is very weak within

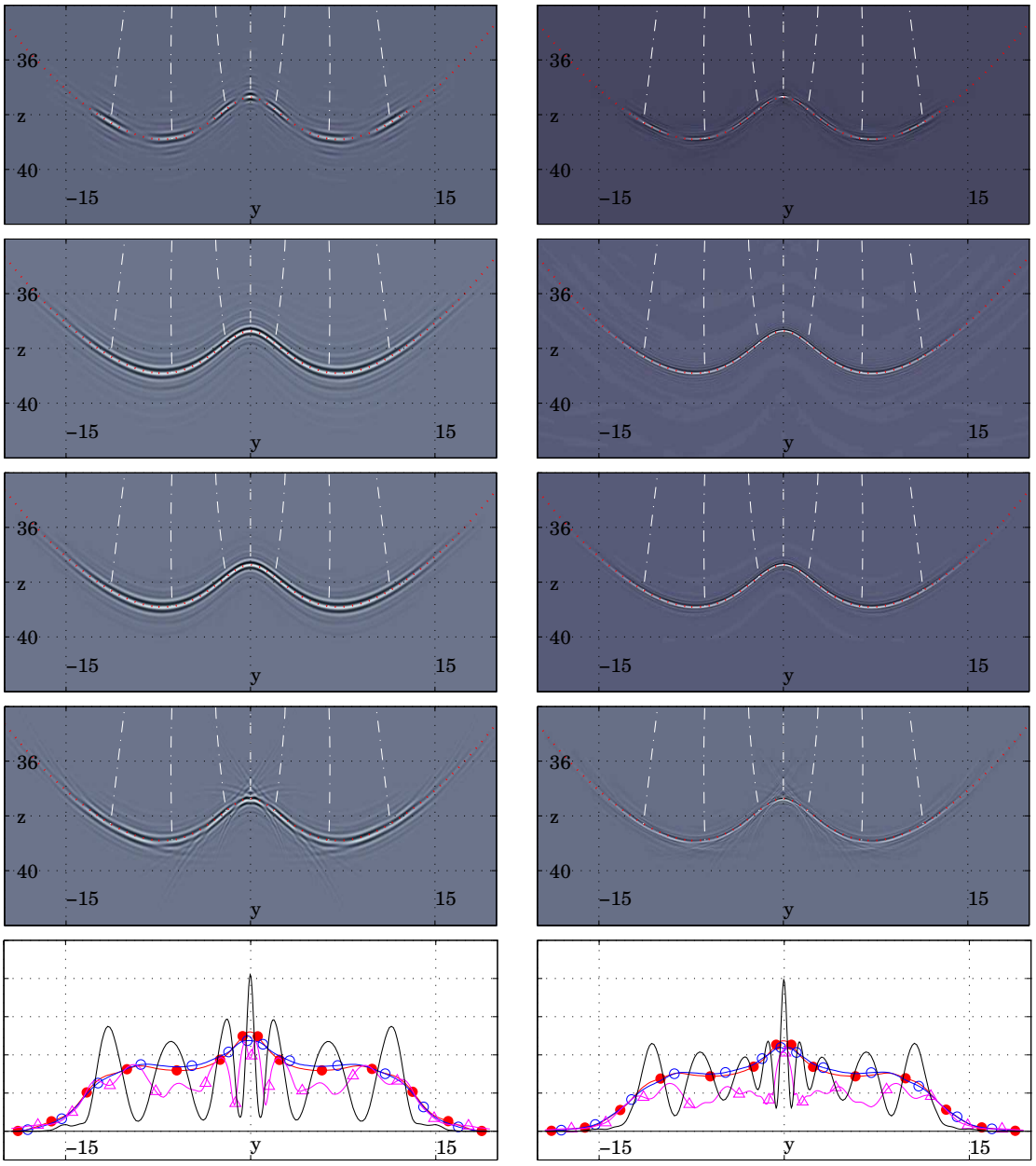

FIG. 14. *Wave propagation in isotropic, heterogeneous medium for initial conditions* (5.3) *with* $k' = 2$ *(left column) and* $k' = 1 - 4$ *(right column) including physical amplitudes* $a(y, \nu)$*: zero order approximation (top row), approximation* (2.5) *(second row),* (2.7) *(third row) and* (2.8) *(fourth row); corresponding amplitudes along wave front (bottom row): zero order approximation (solid black), approximation* (2.5) *(red dot), approximation* (2.7) *(black circle), approximation* (2.8) *(triangle). The white dot-dashed lines indicate rays of seven wave-packets at scale* $k = 3$*. Note that the aspect ration is not equal to one.*

$\tilde{b}_{j,\nu,k}$, the domain of support of the individual wave packets. In contrast, under approximation (2.8) (fourth rows), significant artifacts result from the additional approximate (second order) expansion of the coordinate transform. Since the spatial extent of the modified wave packets is large w.r.t the spatial extent of the perturbation of the background, the error of the approximate coordinate transform is small only close to the center locations $y_j^{\nu,k}$. In particular, we observe artifacts from

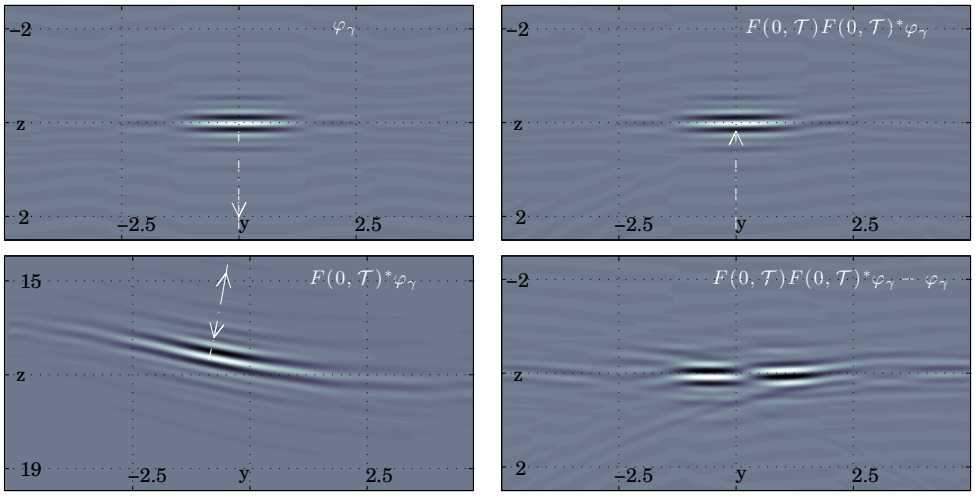

FIG. 15. *Retrofocus experiment. Top row: initial single wave packet $\varphi_\gamma(x)$ at scale $k = k_0 = 3$ (left), and retrofocussed wave packet $\tilde{\psi}_\gamma(x) = (FF^*\varphi_\gamma)(x)$ (right). Bottom row: downwards propagated wave packet $\tilde{\phi}_\gamma(y) = (F^*\varphi_\gamma)(y)$ (left) and difference $((FF^* - I)\varphi_\gamma)(x)$ (right, magnified by a factor 8).*

wave packets that "stick out" of the wave front into regions towards the vertical symmetry axis, close to which the coordinate transform gradually contracts more and more violently due to the low velocity lens at position $(0, 35)km$. Nevertheless, approximation (2.8) appears to produce a more accurate approximation of wave front than zero-order approximation. The above statements are further confirmed by investigation of the amplitudes along the wave fronts (bottom rows): zero order approximation produces large gaps, while amplitudes under approximation (2.5) and (2.7) are nearly fluctuation free. We note that, unlike zero-order approximation, amplitude fluctuation under approximation (2.8) decrease for finer scales.

In Fig. 14, we include the physical amplitudes $a(y, \nu)$. As expected, we observe an amplitude peak close to the focus point of the lens (contraction of flow), followed by a dip (expansion of flow), and again increase when reaching (close to) homogeneous medium far off the location of the low velocity lens, and amplitudes dying out eventually due to the directional band-limiting of the input $u(x)$.

Limited aperture array retrofocussing via phase space localization. We apply the proposed "box algorithm" approximation (2.5) in a retrofocus experiment for one single wave packet φ_γ:

$$F(0,\mathcal{T})\,(F(0,\mathcal{T})^*\varphi_\gamma)\,(x), \tag{5.5}$$

where F is the solution operator to (5.1)–(5.2). As the background c we use the high velocity lens depicted in Fig. 11 (right), given by (5.4) 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 vertical direction, and is depicted in Fig. 15 (top left). The initial data are discretized at $N \times N = 512 \times 512$ sample points, resulting in maximum frequency scale $k_{max} = 5$. Spatial sampling density Δy is set to equal the initial sampling density Δx, and the evolution time is $\mathcal{T} = 8s$.

We begin with evaluating $\tilde{\phi}_\gamma(y) = (F(0,\mathcal{T})^*\varphi_\gamma)(y)$, plotted in Fig. 15 (second row, left). Then, we compress $\tilde{\phi}_\gamma(y)$ by simple hard thresholding of wave packet coefficients below 10% of the magnitude of the largest coefficient. We note that this also effectively results in a band-limiting operation, significant boxes being concentrated in a narrow cone about the central wave vector of $\tilde{\phi}_\gamma(y)$. Finally, we evaluate $\tilde{\psi}_\gamma(x) = (F(0,\mathcal{T})\tilde{\phi}_\gamma)(x)$ on the limited aperture array detected by $\tilde{\phi}_\gamma)(x)$, and obtain

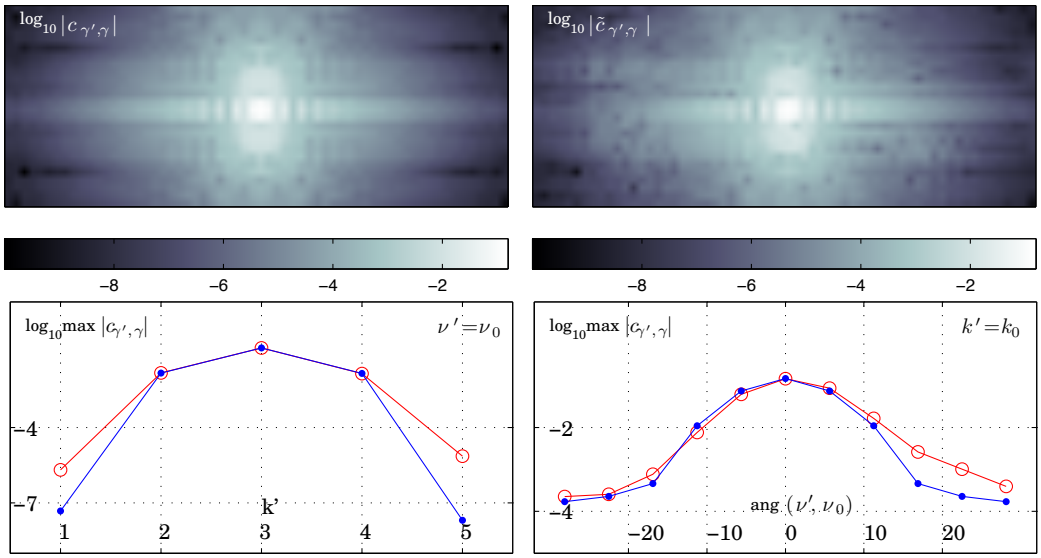

FIG. 16. *Decay properties of* $\tilde{c}_{\gamma',\gamma} = \langle \varphi_{\gamma'}, FF^*\varphi_\gamma \rangle$ *versus* $c_{\gamma',\gamma} = \langle \varphi_{\gamma'}, \varphi_\gamma \rangle$. *Top row: logarithmic magnitude of* $c_{\gamma',\gamma}$ *(left) and* $\tilde{c}_{\gamma',\gamma}$ *(right) for the box* $B_{\nu=\nu_0,k=k_0}$. *Bottom row: decay of coefficients* $\max_j |c_{\gamma',\gamma}|$ *and* $\max_j |\tilde{c}_{\gamma',\gamma}|$ *for* $\nu' = \nu_0$ *fixed as a function of* $k' = k_0 \pm [0,1,2]$ *(left), and for* $k' = k_0$ *fixed as a function of* $\mathrm{ang}(\nu', \nu)$ *(right); blue dots correspond to* $c_{\gamma',\gamma}$, *red circles to* $\tilde{c}_{\gamma',\gamma}$.

the retrofocussed wave packet (Fig. 15, top right). Fig. 15 (second row, right) depicts the difference $\tilde{\psi}_\gamma(x) - \varphi_\gamma(x)$ between retrofocused and the original wave packet, i.e., $(FF^* - I)\varphi_\gamma(x)$ (magnified by a factor 8). In Fig. 16, 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 $c_{\gamma',\gamma}$ (top left) and $\tilde{c}_{\gamma',\gamma}$ (top right) 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' ($\nu' = \nu_0$, bottom left) and of orientation ν' ($k' = k_0$, bottom right). Note that this corresponds to the analysis of the decay properties of the kernel of the pseudo-differential operator FF^*.

We observe that the propagated wave packet $\tilde{\phi}_\gamma(y) = (F^*\varphi_\gamma)(y)$ remains well-localized in space. The original and retrofocussed 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 [59], interferometry [44] and partial reconstruction [18].

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 [22], isochrons defined by imaging operators are 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 pre-stack 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

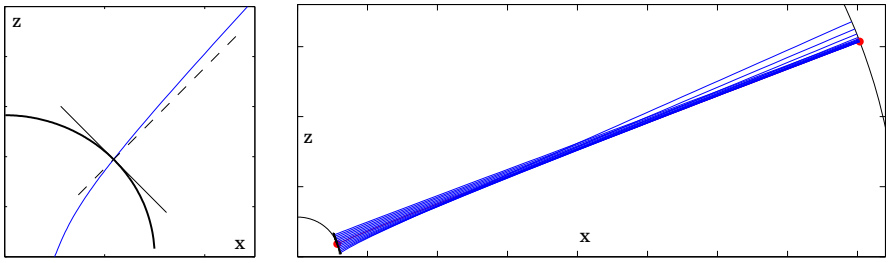

FIG. 17. *Illustration of the geometrical properties of Hamiltonian* (5.6). *Left: isochron front (black thick solid), isochron ray (blue solid), wave front tangent (black solid) and normal (black dashed), illustrating the anisotropy of* (5.6). *Right: two isochrons for small and large evolution times (curved black solid lines) and evolution under Hamiltonian* (5.6) *of a piece of plane wave tangent to the "early isochron" (thick black line segment and bundle of rays in blue solid): Hamiltonian* (5.6) *can create caustics for initial conditions differing from an isochron.*

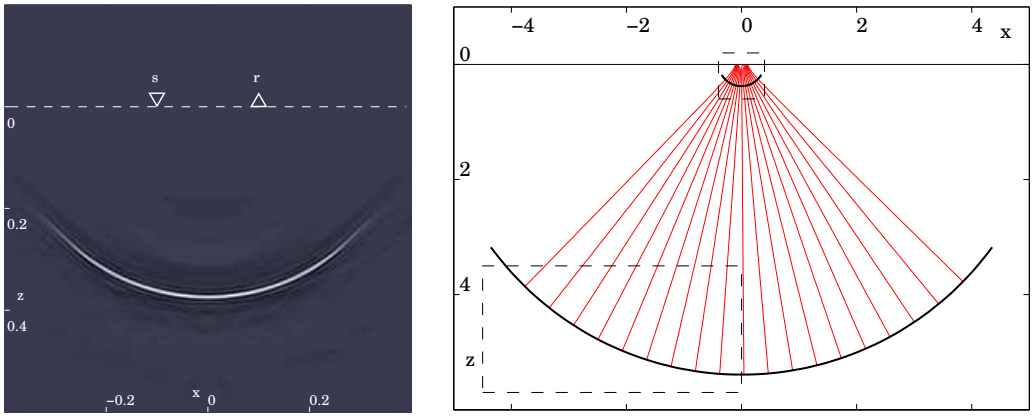

FIG. 18. *Source-receiver geometry and initial band-limited isochron front (left). Geometry of evolution of initial isochron front under the flow of Hamiltonian* (5.6) *(right): Initial isochron and front after evolution for* $\mathcal{T} = 5s$ *(small and large black solid curves, respectively) and 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 considered in our example.*

is in this case given by:

$$\mathcal{H}(y,z,\omega,k_y,k_z) = \omega - \frac{c}{k_y z\sqrt{2}}\left(\frac{\sqrt{Q_-Q_+}}{\sqrt{Q_-}+\sqrt{Q_+}}\right), \tag{5.6}$$

$$Q_\pm = z^2(k_y^2+k_z^2)^2 + (2hk_yk_z \pm z(k_y^2-k_z^2))q_\pm,$$

$$q_\pm = 2hk_yk_z \pm \sqrt{4h^2k_y^2k_z^2 + z^2(k_y^2+k_z^2)^2}.$$

Restriction from the pre-stack imaging volume formulation to $d = 2$ is chosen here for illustration purposes and implies that the Hamiltonian (5.6) has a singularity at $z = 0$. Hence, evolution must be initialized at $z > 0$, i.e. the initial conditions have to be isochrons at early two-way travel times $\mathcal{T}_0$ [28]. Hamiltonian (5.6) is anisotropic and can create caustics for initial conditions that differ from isochron fronts. This is illustrated in Fig. 17.

We use background $c = c_0 = 2km/s$ and half-offset $h = 100m$. As initial condition, we use an isochron front for $\mathcal{T}_0 = 0.39s$ (resulting in maximum initial depth $z < 400m$). The initial isochron is band-limited and compressed by hard-thresholding of its wave packet coefficients, and is plotted in Fig. 18 (left). We evaluate evolution of the isochron front under approximation (2.8) for $\mathcal{T} = 5s$ and compare it to evolution under the zero order approximation. The geometry of the problem is

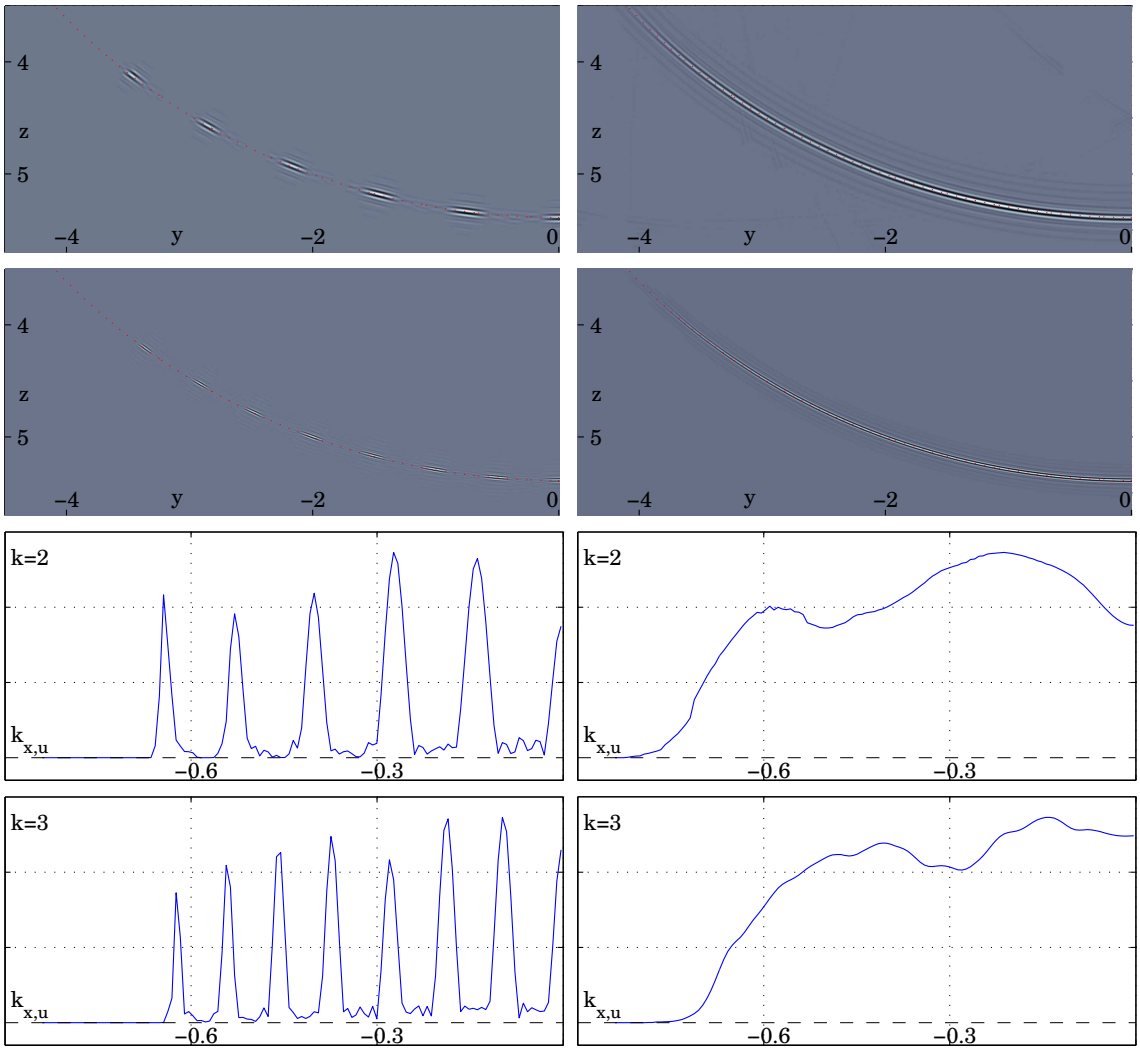

FIG. 19. *Evolution of initial isochron front Fig. 18 (left) under Hamiltonian* (5.6) *for zero order approximation (left column) and approximation* (2.8)*: isochron front band-limited to frequencies scales* $k = 2$ *(top row) and* $k = 3$ *(second row) and corresponding amplitudes along the fronts (third and bottom row, respectively). The red dotted line indicates the theoretical position of the isochron.*

depicted in Fig. 18 (right). Note that the area including the calculation domain is significantly larger than the area including the initial conditions. Results are plotted in Fig. 19 for frequency scales $k = 2$ and $k = 3$, and in Fig. 20 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 along the isochron. Resorting to finer scales or including all scales does not improve the image. In contrast, approximation (2.8) produces a very satisfactory image isochron that is sharply aligned along the theoretical position of the front, without gaps or major amplitude variations. We note that the 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 Fig. 19 (bottom left) that different box orientations do not contribute equal energy, and that amplitude fluctuations under approximation (2.8) depicted in Fig. 19 (bottom right) merely reflect these variations.

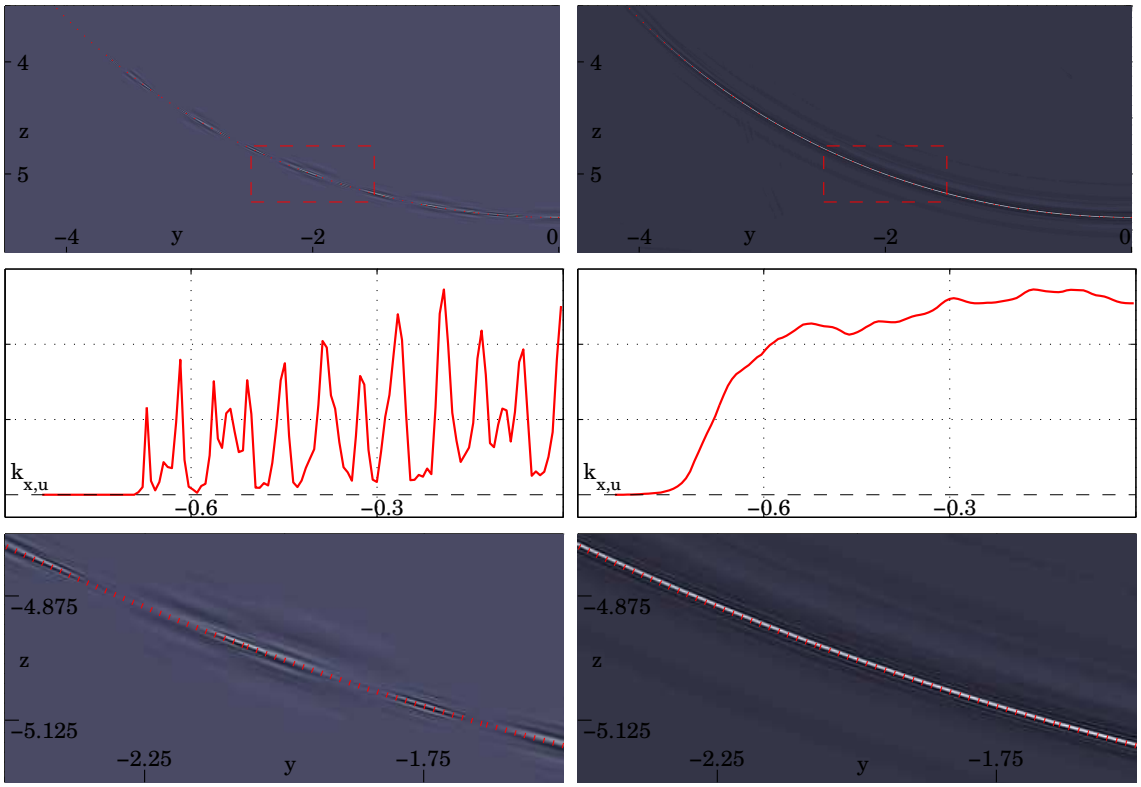

FIG. 20. *Evolution of initial isochron front Fig. 18 (left) under Hamiltonian* (5.6) *for zero order approximation (left column) and approximation* (2.8)*: isochron fronts obtained using all frequencies scales k (first row) and corresponding amplitudes along the fronts (second row). Third row: magnification of the image regions indicated in the first row by red dashed rectangles. The red dotted line indicates the theoretical position of the isochron.*

6. Discussion. In the present work, 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 finite dimension. Discretization being initiated from 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. This accounts for the inherent practical flexibility common to all of the three approximations (2.5), (2.7) and (2.8): Indeed, 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 and in illumination analysis, partial reconstruction or target-oriented imaging applications. Another consequence is that in the case of parametrices of evolution equations, which also generate extended imaging, we obtain effective one-step procedures for (arbitrarily) large time step that are insensitive to numerical dispersion and error accumulation. 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. Here, we promote the idea of geometry. We note that approximation (2.5) can in principle be used to find arbitrarily accurate solutions to evolution equations through iterations of Volterra equations.

Any of the described algorithms can be efficiently parallelized, since computations per individual

box, tensor product term or packet can be performed independently. Also, the procedures offer the possibility for incorporating "real-time" image visualization: Computations can be hierarchically organized such that useful intermediary results – for instance, successively finer and finer scales, and updated output points for successively refined grid resolution – can be visualized on the fly as they become available during computations.

When all boxes contribute to the output, the computational complexity of the "box" algorithm (2.5) is $\mathcal{O}(RN^{(d-1)/2}\log(N))$ above diagonal approximation. The additional factor R is fundamental in this approximation, 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 frequency boxes that have to be evaluated separately. This factor demands further comments: 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 very 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.

The "hybrid packet-box" and "packet" algorithms for approximations (2.7) and (2.8) appear very attractive at first, since no tensor product representation as in approximation (2.5) needs to be obtained. 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 complicated 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-per-packet, worsening overall complexity. In particular, while being the simplest of the proposed procedures when applied to individual packets, approximation (2.8) loses any organization of computations by boxes.

We note that these procedures have similarities with beam migration (e.g. [1, 27, 41, 58]), where coherent data components are extracted and migrated individually. Here, we obtain "beams" as data wave packets which are one-step propagated using a phase-space localized paraxial approximation. For approximations (2.5) and (2.7), we can also form "boxed beams" as data wave packets that share the same frequency scale and dip. We finally mention the effective reduction in calculation domain resulting from the spatial decay properties of individual wave packets.

Future work will include the 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 uni-directional separated representations has been put forward for computations in high dimensions by Beylkin et al. [5, 9]. Incorporation of such strategies promises computational advantages, in particular for dimensions $d > 3$, while typically resulting in purely numerical algorithms. We also mention the development of quadratures for special functions [8, 62]. Finally, we envision the adaptation of the present work to Gaussian wave packets, which offer significantly better decay properties than almost symmetric wave packets, and potentially very high compression rates through sparsity promoting optimization. Also, since the Fourier transform and many elementary manipulations of a Gaussians yield Gaussians, we expect to be able to speed-up the proposed approximations by replacing certain computations by analytic calculus.

REFERENCES

[1] T. Alkhalifah. Gaussian beam depth migration for anisotropic media. *Geophysics*, 60(5):1474–1484, 1995.

[2] 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(3):567–600, 2000.

[3] G. Bao and W.W. Symes. Computation of pseudo-dierential operators. *SIAM J. Sci. Comput.*, 17:416–429, 1996.

[4] G. Beylkin. On the fast Fourier transform of functions with singularities. *Appl. Comput. Harmon. Anal.*, 2:363–381, 1995.
[5] G. Beylkin, V. Cheruvu, and F. Pérez. Fast adaptive algorithms in the non-standard form for multidimensional problems. *Appl. Comput. Harmon. Anal.*, 24:354–377, 2008.
[6] G. Beylkin, R. Coifman, and V . Rokhlin. Fast wavelet transforms and numerical algorithms i. *Comm. Pure Appl. Math.*, 44(2):141–183, 1991.
[7] G. Beylkin and M.J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM J. Sci. Comput.*, 26(6):2133–2159, 2005.
[8] G. Beylkin and L. Monzon. On generalized Gaussian quadratures for exponentials and their applications. *Appl. Comput. Harmon. Anal.*, 12:332–373, 2002.
[9] G. Beylkin and K. Sandberg. Wave propagation using bases for bandlimited functions. *Wave Motion*, 41:263–291, 2005.
[10] B. Bradie, R. Coifman, and A.Grossmann. Fast numerical computations of oscillatory integrals related to accoustic scattering, i. *Appl. Comput. Harmon. Anal.*, 1(1):94–99, 1993.
[11] S. Brandsberg-Dahl and J. Etgen. Beam-wave imaging. In *Proceedings of SEG*, 2003.
[12] E. Candès and L. Demanet. Curvelets and Fourier integral operators. *C. R. Acad. Sci. Paris*, I(336):395–398, 2003.
[13] E. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Model. Simul.*, 5(3):861–899, 2006.
[14] E. Candès, L. Demanet, and L. Ying. Fast computation of Fourier integral operators. *SIAM J. Sci. Comput.*, 29(6):2464–2493, 2007.
[15] E. Candès, L. Demanet, and L. Ying. A fast buttery algorithm for the computation of Fourier integral operators. *SIAM Multiscale Model. Simul.*, 7:1727–1750, 2009.
[16] 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.*, 153:52–74, 2003.
[17] 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*, 31:43–70, 2000.
[18] 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(2):025005+, 2009.
[19] L. Demanet and L.Ying. Discrete symbol calculus. *preprint*, 2008.
[20] H. Douma and M.V. de Hoop. Leading-order seismic imaging using curvelets. *Geophysics*, 72(6):231–248, 2007.
[21] A. Duchkov, F. Andersson, and M.V. de Hoop. Discrete almost symmetric wave packets and multi-scale representation of (seismic) waves. *IEEE T. Geosci. Remote Sensing*, submitted, 2010.
[22] A.A. Duchkov and M.V. de Hoop. Extended isochron rays in shot-geophone (map) migration. *Geophysics*, 2009. Submitted.
[23] A.A. Duchkov, M.V. de Hoop, and A. Sá Baretto. Evolution-equation approach to seismic image, and data, continuation. *Wave Motion*, 45:952969, 2008.
[24] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.*, 14(6):1368–1393, 1993.
[25] A. Dutt and V. Rokhlin. Fast Fourier transforms for nonequispaced data II. *Appl. Comput. Harmon. Anal.*, 2:85–100, 1995.
[26] M. Fedi. Global and local multiscale analysis of magnetic susceptibility data. *Pure appl. geophys.*, 160:2399–2417, 2003.
[27] N.R. Hill. Prestack gaussian beam depth migration. *Geophysics*, 66:1240–1250, 2001.
[28] E. Iversen. The isochron ray in seismic modeling and imaging. *Geophysics*, 69:1053–1070, 2004.
[29] S. Jaffard. Exposants de Hölder en des points données et coéfficients dondelettes. *C. R. Acad. Sci. Paris*, 308:79–81, 1989.
[30] S. Jaffard. Multifractal formalism for functions part i: results valid for all functions. *SIAM J. Math. Anal.*, 28(4):944–970, 1997.
[31] S. Jaffard. Multifractal formalism for functions part ii: Self-similar functions. *SIAM J. Math. Anal.*, 28(4):971–998, 1997.
[32] S. Jaffard. Wavelet techniques in multifractal analysis. In *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus et M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics*, volume 72(2), pages 91–152. AMS, 2004.
[33] W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, 1995.
[34] D. Kraaijpoel. *Seismic ray fields and ray field maps: theory and algorithms*. PhD thesis, Universiteit Utrecht, 2003.
[35] H.J. Landau. On Szegö's eigenvalue distribution theorem and non-hermitean kernels. *J. Anal. Math.*, 28:335–357, 1975.
[36] H.J. Landau and H.O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty–II. *Bell Syst. Tech. J.*, January:65–94, 1961.
[37] 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 Syst. Tech. J.*, July:1295–1337, 1962.
[38] H.J. Landau and H. Widom. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.*, 77:469–481, 1980.

[39] C.F. Li and C. Liner. Wavelet-based detection of singularities in acoustic impedance from surface seismic reflection data. *Geophysics*, 73(1):V1–V9, 2008.
[40] Y. Meyer and R. Coifman. *Wavelets: Caldéron-Zygmund and multilonear operators*. Cambridge University Press, 1997.
[41] M.I. Protasov and V.A. Cheverda. True-amplitude seismic imaging. *Doklady Earth Sciences*, 407A:441–445, 2006.
[42] 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.*, 22:105–123, 2007.
[43] J.H. Le Rousseau. Fourier-integral-operator approximation of solutions to first-order hyperbolic pseudodierential equations i: convergence in Sobolev spaces. *Comm. PDE*, 31:867–906, 2006.
[44] G. Schuster. *Seismic Interferometry*. Cambridge University Press, 2009.
[45] A. Seeger, C.D. Sogge, and E.M. Stein. Regularity properties of Fourier integral operators. *Annals of Math.*, 134:231251, 1991.
[46] Y. Shkolnisky. Prolate spheroidal wave functions on a disc—integration and approximation of two-dimensional bandlimited functions. *Appl. Comput. Harmon. Anal.*, 22:235–256, 2007.
[47] Y. Shkolnisky, M. Tygert, and V. Rokhlin. Approximation of bandlimited functions. *Appl. Comput. Harmon. Anal.*, 21(3):413–420, 2006.
[48] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty–IV: extensions to many dimensions, generalized prolate spheroidal wave functions. *Bell Syst. Tech. J.*, November:3009–3057, 1964.
[49] D. Slepian. Prolate spheroidal wave functions, Fourier analysis and uncertainty–V: the discrete case. *Bell Syst. Tech. J.*, 57:1371–1430, 1978.
[50] D. Slepian and H.O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty–I. *Bell Syst. Tech. J.*, January:43–63, 1961.
[51] H. Smith. A parametrix construction for wave equations with $c^{1,1}$ coefficients. *Ann. Inst. Fourier, Grenoble*, 48:797–835, 1998.
[52] C.C. Stolk and M. V. de Hoop. Modeling of seismic data in the downward continuation approach. *SIAM J. Appl. Math.*, 65:1388–1406, 2005.
[53] C.C. Stolk and M. V. de Hoop. Seismic inverse scattering in the downward continuation approach. *Wave Motion*, 43:579–598, 2006.
[54] C.C. Stolk and M.V. de Hoop. Microlocal analysis of seismic inverse scattering in anisotropic, elastic media. *Comm. Pure Appl. Math.*, 55:261–301, 2002.
[55] M.E. Taylor. *Partial Differential Equations III: Nonlinear Equations*. Springer, 1996.
[56] V. Červený. *Seismic ray theory*. Cambridge University Press, Cambridge, UK, 2001.
[57] G.G. Walter and X. Shen. Wavelets based on prolate spheroidal wave functions. *J. Fourier Analysis and Applications*, 10(1):1–26, 2004.
[58] B. Wang, C. Mason, K. Yoon, J. Ji, J. Cai, and S. Suh. Complex-salt model building using combination of interactive beam migration and localized rtm. In *Proceedings of SEG*, 2009.
[59] S. Wang, M. de Hoop, and B. Ursin. Illumination analysis of wave-equation imaging with "curvelets". In *Proceedings of GMIG*, pages 71–90, 2009.
[60] R.S. Wu, Y. Wang, and M. Luo. Beamlet migration using local cosine basis. *Geophysics*, 73:207–221, 2008.
[61] H. Xiao and V. Rokhlin. High-frequency asymptotic expansions for certain prolate spheroidal wave functions. *J. Fourier Analysis and Applications*, 9(6):575–596, 2003.
[62] H. Xiao, V. Rokhlin, and N. Yarvin. Prolate spheroidal wave functions, quadrature and interpolation. *Inverse problems*, 17:805–838, 2001.

Appendix A. Prolate spheroidal wave functions.

For $c > 0$ real, define the operator $F^c : L^2(\mathcal{R}) \to L^2(\mathcal{R})$ on the unit (hyper)ball $\mathcal{R}$ in $\mathcal{D}$ dimensions:

$$(F^c\phi)(x) = \int_{\mathcal{R}} \exp\left[\mathrm{i}c\langle x, z\rangle\right] \phi(z) dz. \tag{A.1}$$

Denote by $\lambda_{\kappa_1}, \lambda_{\kappa_2}, \ldots$ its eigenvalues in decreasing order $|\lambda_{\kappa_{j-1}}| \geq |\lambda_{\kappa_j}|$, and by ψ_{κ_j} the corresponding eigenfunctions:

$$\lambda_{\kappa_j}\psi_{\kappa_j}(x) = \int_{\mathcal{R}} \exp\left[\mathrm{i}c\langle x, z\rangle\right] \psi_j(z) dz, \quad x \in \mathcal{R}. \tag{A.2}$$

We adhere here to the normalization convention $||\psi_j||_{L^2(\mathcal{R})} = 1$. The eigenvalues are non-zero and either real or imaginary. The eigenfunctions ψ_j are real, orthonormal and complete in $L^2(\mathcal{R})$, either even or odd. For $\mathcal{D} = 1$, they turn out to be the the PSWFs from classical mathematical physics [36, 50]; for $\mathcal{D} \geq 2$, they are referred to as generalized PSWFs [48]. In mathematics, PSWFs are known as a set of functions obtained by a time-limiting, low-passing, and second time-limiting operation.

A.1. PSWFs on an interval ($\mathcal{D} = 1$). The functions ψ_j are also the eigenfunctions of the self-adjoint operator [36, 50]:

$$(Q^c\phi)(x) = \frac{1}{\pi}\int_{[-1,1]} \frac{\sin(c(x-z))}{x-z}\phi(z)dz, \quad x \in [-1,1] \tag{A.3}$$

with eigenvalues $\mu_j = \frac{c}{2\pi}|\lambda_j|^2$, and of the differential operator:

$$D^c = (1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx} - c^2x^2,$$

i.e., they are the bounded and continuous solutions of:

$$(1-x^2)\frac{d^2\psi(x)}{dx^2} - 2x\frac{d\psi(x)}{dx} + (\chi_j - c^2x^2)\psi(x) = 0, \tag{A.4}$$

on $[-1,1]$, where χ_j is a sequence of strictly increasing positive real numbers [36, 50]. Each function $\psi_j(x)$ can be expanded in a Legendre series on the interval $[-1,1]$:

$$\psi_j(x) = \sum_{k=0}^{\infty} d_k^j \tilde{P}_k(x), \tag{A.5}$$

where $\tilde{P}_k(x)$ are normalized Legendre polynomials (cf. B.1). It can be shown that the coefficients d_k^j have superalgebraic decay (cf. e.g. [62], Lemma 3.3 and Theorem 3.4).

Construction. Substitution of (A.5) into (A.4) gives a three-term recursion for the coefficients d_k^j. This recursion can be restated in matrix form as [62]:

$$(B^N - \chi_j \cdot I)d^j = 0,$$

where $B^N_{k,k} = k(k+1) + \frac{2k(k+1)-1}{(2k+3)(2k-1)}c^2$ and $B^N_{k,k+2} = B^N_{k+2,k} = \frac{(k+2)(k+1)}{(2k+3)\sqrt{(2k+1)(2k+5)}}c^2$. The matrix B^N is infinite, but the eigenvectors d^j of interest (corresponding to the expansion coefficients of the first m PSWFs) lie almost entirely in the leading rows and columns of B^N ([62], Theorem 3.4). The eigenvalues λ_j can be obtained by numerical evaluation of, for instance, (A.2) or (A.3) for fixed x, or of the relation ([62], Corollary 3.2):

$$\frac{\lambda_m^2}{\lambda_n^2} = \frac{\int_{[-1,1]}\psi_n'(x)\psi_m(x)dx}{\int_{[-1,1]}\psi_m'(x)\psi_n(x)dx}.$$

A.2. Generalized PSWFs on a disc ($\mathcal{D} = 2$). In polar coordinates (ρ,θ), the eigenfunctions and eigenvalues of (A.1) are given by [48]:

$$\begin{aligned}\psi_{0,n}(\rho,\theta) &= \frac{1}{\sqrt{2\pi}}\Psi_{0,n}(\rho); \qquad \lambda_{0,n} = 2\pi\beta_{0,n}\\ \psi_{N,n,l}(\rho,\theta) &= \frac{1}{\sqrt{\pi}}\Psi_{N,n}(\rho)S_N^l(\theta); \quad \lambda_{N,n} = 2\pi \mathrm{i}^N\beta_{N,n}\end{aligned}$$

with:

$$\beta_{N,n}\Psi_{N,n}(\rho) = \int_0^1 J_N(c\rho\rho')\Psi_{N,n}(\rho')d\rho',$$

where $J_N(x)$ is the Bessel function of the first kind, and:

$$\begin{aligned}S_N^0(\theta) &= \cos(N\theta)\\ S_N^1(\theta) &= \sin(N\theta)\end{aligned}$$

Substituting $\alpha_{N,n} = \sqrt{c}\beta_{N,n}$ and $\varphi_{N,n}(\rho) = \sqrt{\rho}\Psi_{N,n}(\rho)$ gives the equivalent equation:

$$\alpha_{N,n}\varphi_{N,n}(\rho) = \int_0^1 J_N(c\rho\rho')\sqrt{c\rho\rho'}\varphi_{N,n}(\rho')d\rho'. \tag{A.6}$$

Similar to the case $\mathcal{D} = 1$, the solutions of (A.6) are solutions to a differential equation:

$$L^c\varphi(x) = -\kappa\varphi(x)$$

with self-adjoint differential operator L^c:

$$L^c\varphi(x) = (1-x^2)\frac{d^2\varphi(x)}{dx^2} - 2x\frac{d\varphi(x)}{dx} + \left(\frac{\frac{1}{4}-N^2}{x^2} - c^2x^2\right)\varphi(x). \tag{A.7}$$

Construction. The strategy consists in expanding $\varphi(x)$ in series [46, 48]:

$$\varphi_{N,n}(x) = \sum_{k=0}^{\infty} d_k^{N,n} V_{N,k}(x), \tag{A.8}$$

with polynomials $V_{N,k}(x)$ defined by:

$$V_{N,k}(r) = \sqrt{2(2k+N+1)}r^{N+1/2}P_k^{(N,0)}(1-2r^2); \quad r \in [0,1],$$

where $P_k^{\alpha,\beta}(x)$ are Jacobi polynomials (see B.3). The polynomials $V_{N,k}$ are orthonormal on $[0,1]$ w.r.t. k,

$$\int_0^1 V_{N,k}(\rho)V_{N,l}(\rho)d\rho = \delta_{kl}.$$

They are the eigenfunctions of the differential operator (A.7) with $c = 0$, i.e. L^0, with eigenvalues:

$$\omega_{N,k} = (N+2k+\frac{1}{2})(N+2k+\frac{3}{2}).$$

Expansion coefficients. To construct the matrix of L^c w.r.t. the basis $V_{N,k}$, we note that $L^cV_{N,k} = -\omega_{N,k}B_{N,k} - x^2c^2V_{N,k}$. To obtain $x^2V_{N,k}(x)$ as a linear combination of $V_{N,k}(x)$, we substitute $x \to 1-2x^2$ in the recursion for Jacobi polynomials. This gives a three term recursion for $x^2V_{N,k}(x)$ [46, 48]. The matrix B^N of L^c has coefficients [46, 62]:

$$B^N_{k,l} = \langle V_{N,k}, L^cV_{N,l}\rangle$$

which are zero except for the diagonal elements and the elements above and below the diagonal, given by:

$$\begin{aligned} B^N_{k,k} &= -\omega_{N,k} - c^2\frac{2k(k+1)+N(2k+N+1)}{(2k+N)(2k+N+2)} \\ B^N_{k+1,k} = B^N_{k,k+1} &= c^2\frac{(k+1)(k+N+1)}{(2k+N+1)(2k+N+2)}\sqrt{\frac{2k+N+1}{2k+N+3}}. \end{aligned}$$

Now

$$\langle L^c\varphi_{N,n}, V_{N,l}\rangle = -\chi_{N,n}d_l^{N,n}$$

and since L^c is self adjoint [46, 48]:

$$\langle L^c\varphi_{N,n}, V_{N,l}\rangle = \sum_{k=0}^{\infty}\langle V_{N,k}, L^cV_{N,l}\rangle.$$

Hence:

$$(B^N + \chi_{N,n} I) d^{N,n} = 0,$$

which is an eigenproblem similar to the one dimensional case. The elements of the eigenvectors $d^{N,n}$ are the expansion coefficients $d_k^{N,n}$ for $\varphi_{N,n}(x)$ in (A.8).
Eigenvalues. The eigenvalues $\lambda_{N,n}$ are given by:

$$\lambda_{N,n} = \mathrm{i}^N \frac{2\pi}{\sqrt{c}} \alpha_{N,n}$$

where $\alpha_{N,n}$ can be obtained by numerical evaluation of (A.6) for fixed ρ.

A.3. Generalized PSWFs on a (hyper-)ball ($\mathcal{D} \geq 3$). Write $x = \rho\Omega$ and $z = r\Phi$, where Ω and Φ are unit vectors in $\mathcal{D}$ dimensions, and let be:

$$\mathcal{D} = 2 + p, \qquad p = 1, 2, \cdots .$$

It can be shown [48] that the eigenfunctions and eigenvalues of (A.1) are given by:

$$\begin{aligned} \psi_{N,n,l}(\rho, \Omega) &= \Psi_{N,n}(\rho) S_N^l(\Omega) \\ \lambda_{N,n,l} &= i^N (2\pi)^{1+p/2} \beta_{N,n}, \end{aligned}$$

where

$$S_N^l(\Omega), \qquad l = 1, \ldots, (2N+p)\frac{(N+p-1)!}{p!N!}$$

are a complete set of orthonormal surface harmonics of degree N, and:

$$\begin{aligned} \beta_{N,n} &= \gamma_{N,n} c^{-\frac{p+1}{2}} \\ \Psi_{N,n}(\rho) &= \varphi_{N,n}(\rho) \rho^{-\frac{p+1}{2}}, \end{aligned}$$

and $\varphi_{N,n}(\rho)$, $\gamma_{N,n}$ are the solutions of:

$$\gamma_{N,n} \varphi_{N,n}(\rho) = \int_0^1 J_{N+\frac{p}{2}}(c\rho\rho') \sqrt{c\rho\rho'} \varphi_{N,n}(\rho') d\rho'. \tag{A.9}$$

This is, however, (A.6) for the two dimensional case with N replaced by $N + p/2$. The expressions in A.2 for the construction of $\varphi_{N,n}(\rho)$ can hence be taken over exactly with N replaced by $\tilde{N} = N + p/2$ [48].

In Fig. 21, we depict radial functions $\Psi_{N,n}(\rho)$ for several choices of c, N and n (left), and eigenvalue spectra for $c = 1$ and $c = 10$ in logarithmic scale (center and right; on the right we depict cross-section of the eigenvalue spectra).

A.4. SUMMARY: Numerical recipe for the construction of PSWFs.

Dimension $\mathcal{D} = 1$.

1. Construct the normalized Legendre polynomials $\tilde{P}_k(x)$.
2. Construct K rows and columns of the matrix B^N with non-zero entries $B^N_{k,k} = k(k+1) + \frac{2k(k+1)-1}{(2k+3)(2k-1)} \cdot c^2$ and $B^N_{k,k+2} = B^N_{k+2,k} = \frac{(k+2)(k+1)}{(2k+3)\sqrt{(2k+1)(2k+5)}} \cdot c^2$
3. Solve the eigenvalue problem $B^N d^j = \chi_j d^j$, yielding the expansion coefficients d_k^j.
4. Compute the PSWFs: $\psi_j(x) = \sum_k d_k^j \tilde{P}_k(x)$.
5. Compute the eigenvalues by evaluation of $\lambda_j \psi_j(x) = \int_{-1}^1 \exp(icxz)\, \psi_j(z) dz$ for fixed x.

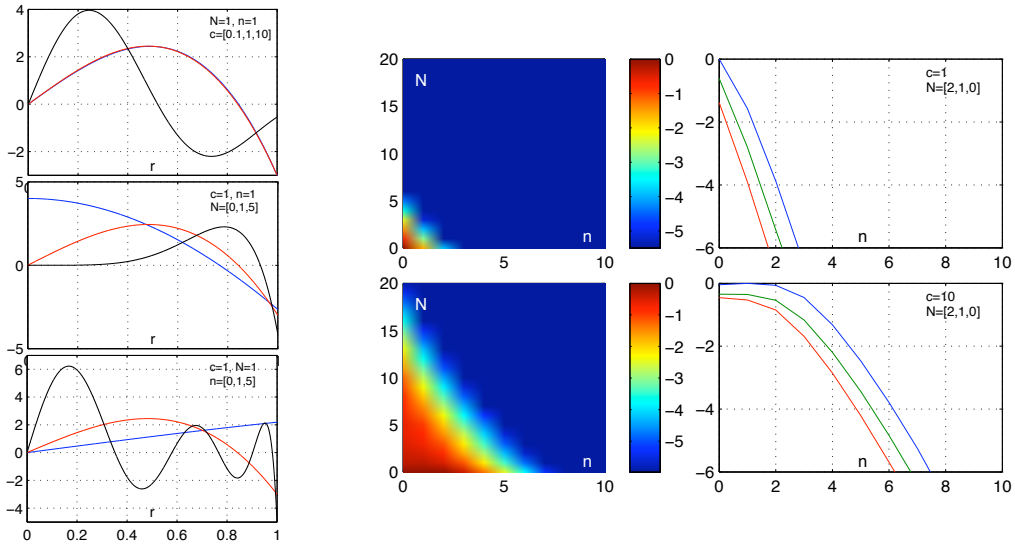

FIG. 21. *Radial functions $\Psi_{N,n}(\rho)$ (left), logarithmic plots of eigenvalue spectra (center), and cross-sections of the eigenvalue spectra (right) for different bandwidths and $\mathcal{D} = 3$.*

Dimension $\mathcal{D} \geq 2$. Let $p = \mathcal{D} - 2$ and $\tilde{N} = N + p/2$.

1. Construct the Jacobi polynomials $P_k^{(\tilde{N},0)}(x)$.
2. Construct the polynomials $V_{\tilde{N},k}(\rho) = \sqrt{2(2k+\tilde{N}+1)} r^{\tilde{N}+1/2} P_k^{(\tilde{N},0)}(1-2\rho^2)$; $\quad \rho \in [0,1]$,.
3. Construct K rows and columns of the matrix $B^{\tilde{N}}$ with non-zero entries $B^{\tilde{N}}_{k,k} = -(\tilde{N}+2k+\frac{1}{2})(\tilde{N}+2k+\frac{3}{2}) - c^2 \frac{2k(k+1)+\tilde{N}(2k+\tilde{N}+1)}{(2k+\tilde{N})(2k+\tilde{N}+2)}$ and $B^{\tilde{N}}_{k+1,k} = B^{\tilde{N}}_{k,k+1} = c^2 \frac{(k+1)(k+\tilde{N}+1)}{(2k+\tilde{N}+1)(2k+\tilde{N}+2)} \sqrt{\frac{2k+\tilde{N}+1}{2k+\tilde{N}+3}}$.
4. Solve the eigenproblem $B^{\tilde{N}} d^{\tilde{N},n} = \chi_j d^{\tilde{N},n}$, wielding the expansion coefficients $d_k^{\tilde{N},n}$.
5. Compute the functions: $\varphi_{\tilde{N},n}(\rho) = \sum_k d_k^{\tilde{N},n} T_{\tilde{N},k}(\rho)$.
6. Compute the set of orthonormal surface harmonics $S_N^l(\Omega)$.
7. Compute the PSWFs: $\psi_{N,n,l}(\rho,\Omega) = \rho^{-(p+1)/2} \varphi_{\tilde{N},n}(\rho) S_N^l(\Omega)$.
8. Compute the eigenvalues $\lambda_{N,n} = i^N (2\pi)^{1+p/2} c^{-(p+1)/2} \alpha_{\tilde{N},n}$ by evaluation of $\alpha_{\tilde{N},n} \varphi_{\tilde{N},n}(\rho) = \int_0^1 J_{\tilde{N}}(cr\rho) \sqrt{cr\rho} \varphi_{\tilde{N},n}(r) dr$ for fixed ρ.

For $\mathcal{D} = 2$, the hyperparameter is $\Omega = \theta$ (polar angle), and the functions $S_N^l(\Omega)$ are given by:

$$S_0^{\{0,1\}} = \frac{1}{2\sqrt{2\pi}}, \qquad S_N^l = \frac{1}{\sqrt{\pi}} \cos(N\theta - l\frac{\pi}{2}), \quad N = 1, 2, \ldots, \quad l = 0, 1.$$

For $\mathcal{D} = 3$, the hyperparameter is $\Omega = (\theta, \phi)$ (polar and azimuthal angle), and the functions $S_N^l(\Omega)$, $l = 1, \ldots, 2N+1$ are given by the spherical harmonics of degree N (cf. B.4).

Evaluation of PSWFs. The eigenvalues λ_{κ_j} and expansion coefficients d_k^j can be precomputed and stored for any given value of c. Subsequent evaluation of a PSWF then takes $\mathcal{O}(c)$ operations.

Appendix B. Special functions.

B.1. Legendre polynomials. The Legendre polynomials $P_k(x)$ are given by the recursion:

$$P_{k+1}(x) = \frac{2k+1}{k+1} x P_k(x) - \frac{k}{k+1} P_{k-1}(x),$$

with $P_0(x) = 1$ and $P_1(x) = x$. They are orthogonal over the interval $[-1, 1]$:

$$\int_{-1}^{1} P_k(x) P_l(x) dx = \delta_{kl} \frac{1}{2k+1}.$$

The normalized Legendre polynomials are given by:

$$\tilde{P}_k(x) = P_k(x)\sqrt{k+\frac{1}{2}}, \qquad ||\tilde{P}_k||_{L^2[-1,1]} = 1.$$

B.2. Associated Legendre polynomials. For positive m, the associated Legendre polynomials are defined in terms of Legendre polynomials as:

$$P_N^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\frac{d^m}{dx^m}P_N(x).$$

For negative m, they are defined as:

$$P_N^{-m}(x) = (-1)^m\frac{(N-m)!}{(N+m)!}P_N^m(x).$$

They also obey the recurrence relation:

$$P_N^m(x) = \frac{2N-1}{N-m}xP_{N-1}^m(x) - \frac{N+m-1}{N-m}P_{N-2}^m(x).$$

They are part of the spherical harmonics (cf. B.4) and orthogonal over $[-1,1]$ with the weighting function 1 w.r.t. N, $\int_{-1}^{1}P_N^m(x)P_{N'}^m(x)dx = 2(N+m)!/((2N+1)(N-m)!)\delta_{N,N'}$, and orthogonal over $[-1,1]$ w.r.t. m with the weighting function $(1-x^2)^{-1}$, $\int_{-1}^{1}P_N^m(x)P_N^{m'}(x)(1-x^2)^{-1}dx = (N+m)!/(m(N-m)!)\delta_{m,m'}$.

B.3. Jacobi polynomials. Jacobi polynomials $P_k^{(\alpha,\beta)}(x)$ are defined by the three term recursion:

$$P_{k+1}^{(\alpha,\beta)}(x) = \frac{a_{3k}x - a_{2k}}{a_{1k}}P_k^{(\alpha,\beta)}(x) - \frac{a_{4k}}{a_{1k}}P_{k-1}^{(\alpha,\beta)}(x)$$

with[7] $P_0^{(\alpha,\beta)}(x) = 1$ and $P_1^{(\alpha,\beta)}(x) = \frac{1}{2}(\alpha-\beta+(\alpha+\beta+2)x)$ $[P_0^{(\alpha,0)}(x) = 1$ and $P_1^{(\alpha,0)}(x) = \frac{1}{2}(\alpha+(\alpha+2)x)]$ and

$$\begin{aligned}
a_{1k} &= 2(k+1)(k+\alpha+\beta+1)(2k+\alpha+\beta) && [= 2(k+1)(k+\alpha+1)(2k+\alpha)]\\
a_{2k} &= -(2k+\alpha+\beta+1)(\alpha^2-\beta^2) && [= -(2k+\alpha+1)\alpha^2]\\
a_{3k} &= (2k+\alpha+\beta)(2k+\alpha+\beta+1)(2k+\alpha+\beta+2) && [= (2k+\alpha)(2k+\alpha+1)(2k+\alpha+2)]\\
a_{4k} &= 2(k+\alpha)(k+\beta)(2k+\alpha+\beta+2) && [= 2k(k+\alpha)(2k+\alpha+2)].
\end{aligned}$$

They are orthogonal on $[-1,1]$ w.r.t. the weight function $w(x) = (1-x)^\alpha(1+x)^\beta$:

$$\int_{-1}^{1}P_k^{(\alpha,\beta)}(x)P_l^{(\alpha,\beta)}(x)w(x)dx = \delta_{kl}\frac{2^{\alpha+\beta+1}}{2k+\alpha+\beta+1}\frac{\Gamma(k+\alpha+1)\Gamma(k+\beta+1)}{k!\Gamma(k+\alpha+\beta+1)} \quad \left[= \delta_{kl}\frac{2^{\alpha+1}}{2k+\alpha+1}\right]$$

and satisfy the differential equation:

$$(1-x^2)\frac{d^2P_k^{(\alpha,\beta)}(x)}{dx^2} + (\beta-\alpha-(\alpha+\beta+2)x)\frac{dP_k^{(\alpha,\beta)}(x)}{dx} + k(k+\alpha+\beta+1) = 0$$
$$\left[(1-x^2)\frac{d^2P_k^{(\alpha,0)}(x)}{dx^2} - (\alpha+(\alpha+\beta+2)x)\frac{dP_k^{(\alpha)}(x)}{dx} + k(k+\alpha+1) = 0\right].$$

Their derivatives are obtained as $\frac{dP_k^{(\alpha,\beta)}(x)}{dx} = \frac{1}{2}(k+\alpha+\beta)P_{k-1}^{(\alpha+1,\beta+1)}(x)$.

[7] We state in square brackets $[\cdot]$ the special case $\beta = 0$ used in A.2 and A.3.

B.4. Spherical harmonics. Spherical harmonics of order N are given by:

$$S_N^m(\theta,\phi) = \sqrt{\frac{2L+1}{4\pi}\frac{(N-m)!}{(N+m)!}} P_N^m(\cos(\theta)) e^{\mathrm{i}m\phi}, \quad m = -N,\ldots,N-1,N$$

where $\theta \in [0,\pi]$ is the polar (colatitudinal) $\phi \in [0,2\pi)$ is the azimuthal (longitudinal) coordinate, and $P_N^m(x)$ are the associated Legendre polynomials. They are the angular portion of the solution to Laplace's equation in spherical coordinates. The normalization is usually chosen such that:

$$\int_0^{2\pi}\int_0^{\pi} S_N^m(\theta,\phi)\bar{S}_{N'}^{m'}(\theta,\phi)\sin(\theta)d\theta d\phi = \delta_{NN'}\delta_{mm'}.$$

They are a L^2 basis on the sphere and obey:

$$S_N^{-m}(\theta,\phi) = (-1)^m \bar{S}_N^m(\theta,\phi).$$