Discrete almost symmetric wave packets and multi-scale geometrical representation of (seismic) waves.

Anton A. Duchkov, Fredrik Andersson, and Maarten V. de Hoop

Abstract—We discuss a multi-scale geometrical representation of (seismic) waves via a decomposition into wave packets. Wave packets can be thought of as certain localized “fat” plane waves. As a starting point we take the frame of continuous “curvelets” and associated transform. Different discretizations, and approximations, of this transform lead to different discrete wave packets. One such discretization leads to the introduction of digital curvelets. Here, we follow a different discretization while constructing discrete almost symmetric wave packets and making use of the USFFT. The resulting discrete transform is unitary. In our discretization, we directly approximate the underlying dyadic parabolic decomposition, its (rotational) symmetry, and approximate all the necessary decay estimates for continuous “curvelets” in phase space with arbitrary accuracy. Another relevant aspect of our discretization is the appearance of parameters that control the tiling of phase space corresponding with the dyadic parabolic decomposition, preserving the relative parabolic scaling, while adapting to the physical problem at hand. We consider applications in exploration seismology and global seismology. For these, we need transforms in higher dimensions, that is, in dimension $n = 3, 4, 5$.

Index Terms—wave packets, dyadic parabolic decomposition, curvelet transform, seismology

I. INTRODUCTION

The theory of second microlocalization, or dyadic parabolic decomposition, has been an important development in harmonic analysis [1]. Over the recent years, it has led to the development of new multi-scale (computational) techniques, not only in image processing [2], but also in partial differential equations [3] and inverse problems encompassing inverse scattering [4], [5] and wave-equation tomography [6]. However, this class of techniques also sheds light on various physical aspects of solutions to wave equations and inverse problems. For example, Andersson et al. [7] prove a concentration of wave packets result for hyperbolic evolution equations with limited smoothness, De Hoop et al. [5] develop the notion of illumination and partial reconstruction in inverse scattering, and Brytik et al. [6] reveal the imprint of geometrical optics and associated ray perturbation through wave packets in the sensitivity analysis of wave-equation or “finite-frequency travel-time” tomography [8], [9].

At the basis of all these developments is the frame of continuous “curvelets” introduced by Smith [3], and further developed by Candès and Donoho [2], and the associated class of transforms. The general concept has its roots in the theory of coherent wave packets and Fourier integral operators [10], and parabolic cutoffs dating back to Boutet de Monvel [11]. Furthermore, the “curvelet” transform can be related to the Fourier-Bros-Iagolnitzer (FBI) transform [12].

Different discretizations, and approximations, of the transforms in the above mentioned class lead to different discrete wave packets. One such discretization leads to the introduction of digital curvelets. Here, we follow a different discretization while constructing discrete almost symmetric wave packets and making use of the USFFT as developed by Dutt and Rokhlin [13] and Beylkin [14]. Numerically, our transform is unitary. Another characteristic of our discretization is the appearance of parameters that control the tiling of phase space corresponding with the dyadic parabolic decomposition, while adapting to the physical problem at hand. Moreover, the algorithm can readily adapt, locally, to other scalings, such as cubic scaling efficient in the vicinity of certain caustics (Smith, personal communication). The decomposition corresponding with the resulting discrete transform leads to a multi-scale geometrical representation of (seismic) waves.

We hasten to mention the algorithm – without wrapping – developed by Candès et al. [15] for the digital curvelet transform [16], [17]. In dimension 2, our algorithm has various elements in common. However, differences include the symmetries important in multi-scale analysis of inverse problems applications, see Figure 1, and the construction of an explicit inverse (USFFT based discretization and unitarity). In this paper, we have applications in exploration seismology (wide azimuths towed streamer surveys) and global seismology (US-Array) in mind. For these, we develop discrete transforms in higher dimensions, while the data sets are inherently very large, which necessitates the development of algorithms for massively parallel computing.

We illustrate the multi-scale geometrical representation in dimension 3 using synthetic data containing caustics. Future applications of the multi-scale, discrete transform developed here, include higher-dimensional data regularization ($l^1$ optimization [18]), denoising [19] and visualization, computational directional regularity analysis, and feature extraction.
II. DYADIC PARABOLIC DECOMPOSITION

Here, we discuss the (co)frame of continuous “curvelets”. Let \( u \in L^2(\mathbb{R}^n) \) represent a (seismic) wavefield. We consider the Fourier transform, \( \hat{u}(\xi) = \int u(x) \exp[-2\pi i (x, \xi)] \, dx \). We begin with an overlapping covering of the positive \( \xi_1 \) axis by rectangles of the form

\[
B_k = \left[ \xi_k' - \frac{L_k'}{2}, \xi_k' + \frac{L_k'}{2} \right] \times \left[ -\frac{L_k''}{2}, \frac{L_k''}{2} \right]^{n-1},
\]

where the centers \( \xi_k' \), as well as the side lengths \( L_k' \) and \( L_k'' \), satisfy the parabolic scaling condition

\[
\xi_k' \sim 2^k, \quad L_k' \sim 2^k, \quad L_k'' \sim 2^{k/2}, \quad \text{as} \ k \to \infty.
\]

For \( k = 0 \), \( B_0 \) is a cube centered at \( \xi_0 = 0 \), with \( L_k' = L_k'' \). Next, for each \( k \geq 1 \), let \( \nu \) vary over a set of approximately \( 2^{k(n-1)/2} \) vectors, close to uniformly distributed over the \((n-1)\)-sphere, \( S^{n-1} \). (We adhere to the convention that \( \epsilon_1 \) aligned with the \( \xi_1 \)-axis belongs to the set of vectors \( \nu \), for all \( k \)). Vectors \( \nu \) define the radial orientations of boxes (that is, their “axes”); to define the box orientation fully one needs an additional \((n-2)\)-dimensional parameter set defining rotations of the boxes around their axes.

Let \( \Theta_{\nu,k} \) denote the product of the above mentioned \( 1 + (n-2) \) rotations. We set

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

In particular, \( \nu = \Theta_{\nu,k}^{-1} \epsilon_1 \) for each \( k \). The parameters \( \xi_k', L_k', L_k'' \), and \( \nu \) are chosen so that the \( B_{\nu,k} \) amply cover \( \mathbb{R}^n \), in the sense that with \( L_k' \) and \( L_k'' \) multiplied by some fixed \( r < 1 \), the interiors would still cover \( \mathbb{R}^n \); see Figure 1 a). We consider \( n \geq 3 \).

A. Continuous “curvelets”

Following [3], [7], we introduce two sequences of smooth functions, \( \hat{\chi}_{\nu,k} \) and \( \hat{\beta}_{\nu,k} \), on \( \mathbb{R}^n \), each supported in \( B_{\nu,k} \), so that

\[
\hat{\chi}_0(\xi) \hat{\beta}_0(\xi) + \sum_{k \geq 1} \sum_{\nu} \hat{\chi}_{\nu,k}(\xi) \hat{\beta}_{\nu,k}(\xi) = 1,
\]

yielding a co-partition of unity. The window functions are designed to satisfy the estimates

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

(4)

One then forms

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

(5)

with \( \rho_k = |B_k| = L_k' (L_k'')^{n-1} \). These functions satisfy the estimates

\[
\left| \frac{\varphi_{\nu,k}(x)}{\psi_{\nu,k}(x)} \right| \leq C N 2^{k(n+1)/4} \left( 2^k |\langle \nu, x \rangle| + 2^{k/2} |x| \right)^{-N}
\]

(6)

for all \( N \). Let \( \{x_{j,k}^{\nu,k}\} \) denote a set of points in \( \mathbb{R}^n \), depending on \( (\nu, k) \), to be specified below. We obtain a frame/co-frame pair in \( L^2(\mathbb{R}^n) \) by subjecting \( \varphi_{\nu,k} \) and \( \psi_{\nu,k} \) to translations over \( x_{j,k}^{\nu,k} \), resulting in \( \varphi_{\nu,k}(x - x_{j,k}^{\nu,k}) \) and likewise for \( \psi_{\nu,k} \).

Introducing triplets \( \gamma = (j, \nu, k) \), we get \( \varphi_\gamma(x) = \varphi_{\nu,k}(x - x_{j,k}^{\nu,k}) \) or

\[
\varphi_\gamma(\xi) = \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi) \exp[-2\pi i (x_{j,k}^{\nu,k}, \xi)], \quad k \geq 1,
\]

(7)

and similarly for \( \psi_\gamma(\xi) \). The translation factor \( \exp[-2\pi i (x_{j,k}^{\nu,k}, \xi)] \) is representative of a Fourier basis (with frequencies \( x_{j,k}^{\nu,k} \)) for functions of \( \xi \). Indeed, the (compact) support of \( \hat{\chi}_{\nu,k} \) determines an orthonormal basis defining a Fourier series. We introduce the lattice

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

(8)

and capture the scaling of \( B_k \) in the dilation matrix

\[
D_k = \begin{pmatrix} L_k' & 0_{1 \times n-1} \\ 0_{n-1 \times 1} & L_k'' \end{pmatrix}.
\]

Indeed, choosing

\[
x_{j,k}^{\nu,k} = \Theta_{\nu,k}^{-1} D_k^{-1} X_j,
\]

(9)

yields an orthonormal basis, \( \exp[-2\pi i (x_{j,k}^{\nu,k}, \xi)] \), for functions supported in \( B_{\nu,k} \).
The weight function \( B. \) Transform pair

\[ u_\gamma = \langle \psi_\gamma, u \rangle = \int u(x) \overline{\psi_\gamma(x)} \, dx = \int \hat{u}(\xi) \overline{\psi_\gamma(\xi)} \, d\xi; \quad (10) \]

then

\[ u(x) = \sum_{\gamma} u_\gamma \varphi_\gamma(x). \quad (11) \]

The \( \psi_\gamma(x) \) are sometimes referred to as Riesz representers [15]. For each fixed wedge indexed by \( \nu, k \), it holds true that

\[ \sum_{\gamma; k'=k, \nu' = \nu} u_\gamma \varphi_{\gamma'}(x) = \hat{u}(\xi) \tilde{\beta}_{\nu,k}(\xi) \varphi_{\nu,k}(x). \quad (12) \]

Upon substituting (7) into (10), we get

\[ u_\gamma = \rho_k^{-1/2} \int \hat{u}(\xi) \tilde{\beta}_{\nu,k}(\xi) \exp[2\pi i (x^{\nu,k}_j, \xi)] \, d\xi \quad (13) \]

and recognize that the computation of \( u_\gamma \) can be interpreted as an inverse Fourier transform of \( \hat{u}(\xi) \tilde{\beta}_{\nu,k}(\xi) \) evaluated on the grid points \( x^{\nu,k}_j \). Upon substituting (7) into (12), we get

\[ \hat{u}(\xi) \tilde{\beta}_{\nu,k}(\xi) \hat{\varphi}_{\nu,k}(\xi) = \rho_k^{-1/2} \left( \sum_j u_{j,\nu,k} \exp[-2\pi i (x^{\nu,k}_j, \xi)] \right) \hat{\varphi}_{\nu,k}(\xi). \quad (14) \]

which represents the Fourier series expansion of \( \hat{u}(\xi) \tilde{\beta}_{\nu,k}(\xi) \) multiplied by \( \hat{\varphi}_{\nu,k}(\xi) \). We refer to (13)-(14) as the “inner” transform pair. The transform is unitary in the case of a tight frame, when \( \tilde{\beta}_{\nu,k} = \hat{\varphi}_{\nu,k} \). We refer to the Fourier transform mapping \( u \) to \( \hat{u} \) and its inverse as the “outer” transform pair.

We introduce the notation \( C \) for the “curvelet” transform (analysis): \( u_\gamma = (Cu)_\gamma \), and also define \( C^{-1}\{u_\gamma\} = \sum_{\gamma} u_\gamma \varphi_\gamma \) for the inverse transform (synthesis). It holds true that \( C^{-1}C = I \) on \( L^2(\mathbb{R}^n) \), and that \( C^{-1} \equiv \Pi \) is a projection operator of \( \ell^2_\gamma \) onto the range of the analysis operator \( C \); here,

\[ \Pi_{\gamma'} \gamma = \langle \psi_{\gamma'}, \varphi_\gamma \rangle. \quad (15) \]

Let \( d \) denote the pseudodistance on the cosphere bundle \( S^*(\mathbb{R}^n) \) introduced in [20, Definition 2.1]

\[ d(x, \nu; x', \nu') = \left| (\nu, x - x') \right| + \left| (\nu', x - x') \right| + \min\{|x-x'|, |x-x'|^2 + |\nu - \nu'|^2\}. \quad (16) \]

If \( \gamma = (x, \nu, k) \) (identifying \( j \) with \( x = x^{\nu,k}_j \) and \( \gamma' = (x', \nu', k') \), let [7]

\[ d(\gamma; \gamma') = 2^{-\min(k, k')} + d(x, \nu; x', \nu'). \quad (17) \]

The weight function \( \mu_\delta(\gamma, \gamma') \) introduced in [3] is given by

\[ \mu_\delta(\gamma, \gamma') = (1 + |k' - k|^2)^{-\delta} \times 2^{-\delta|n + \delta|k'-k}(2^{-\delta|n + \delta|\min(k', k)})d(\gamma, \gamma')^{-\delta(n + \delta)}. \quad (18) \]

Then \( \Pi \) satisfies the estimate,

\[ \|\Pi_{\gamma'} \gamma\| \leq C(\delta) \mu_\delta(\gamma', \gamma), \quad \text{for all} \quad \delta > 0. \quad (19) \]

III. DISCRETE TRANSFORM PAIR

Here, we discretize \( C \) defined by (10). We ensure that the discrete transform is unitary within a prescribed accuracy.

A. Forward Fourier transform, Fundamental cell

In the applications considered, the original function, \( u \), is always subjected to a cutoff yielding a compactly supported function. According to Shannon’s sampling theorem, we can represent such a function by means of sampling its Fourier transform on a properly chosen equally spaced grid. If \( \text{supp}(u) \subset D = \{ x \in \mathbb{R}^n \mid |x| < \frac{1}{2} = x_{\text{max}} \} \), it is sufficient to sample \( \hat{u} \) at integer points, \( \xi \text{, } k \in \mathbb{Z}^n \), to be able to reconstruct \( u(x) \).

In the discrete setting, we assume that \( u \) is sampled at equally spaced points, \( x_i = N^{-1}i \), with \( i \in X^N \) and

\[ X^N = \{ i \in \mathbb{Z}^n \mid -\frac{N}{2} \leq i_1 < \frac{N}{2}, \ldots, -\frac{N}{2} \leq i_n < \frac{N}{2} \}. \quad (20) \]

We encounter the Fourier series,

\[ \hat{u}(\xi) = \frac{1}{N^n} \sum_{i \in X^N} u(x_i) \exp[-2\pi i (\xi, x_i)]. \]

In the process of computing the “curvelet” coefficients, \( \hat{u} \) is multiplied by window functions \( \tilde{\beta}_{\nu,k} \) thus bandlimiting the original function according to scale \( k \). A maximum scale considered, \( k_{\text{max}} \), defines a maximum frequency \( \xi_{\text{max}} \). With \( c_0 \) the parameter defining the coarsest scale (that is, the units), we have the relationship,

\[ \frac{\xi_{\text{max}}}{c_0} = 2^{k_{\text{max}}}. \quad (21) \]

The maximum frequency and the sampling density in the space domain are related according to

\[ N = [2\xi_{\text{max}}]; \]

the fundamental cell is then given by \( \frac{\sqrt{2}}{2} \frac{1}{2^{k_{\text{max}}}} = \frac{N}{2} \).

We note that \( u_{j,\nu,k} = u_{j,\nu,k} = \rho_k^{-1/2} (u \ast \tilde{\beta}_{\nu,k}(x^{\nu,k}_j)) \), cf. (13). The function \( u \ast \tilde{\beta}_{\nu,k} \) no longer has compact support. However, the fast decay of \( \tilde{\beta}_{\nu,k} \) (cf. (6)) is inherited by \( u \ast \tilde{\beta}_{\nu,k} \). Hence, \( u \) will be very small for large \( x^{\nu,k}_j \)'s, and in practice it will be sufficient to keep and use only the lattice points located in a neighborhood of \( D \) for representation (11) to be accurate.

B. Discretization and unitarity

It appears to be advantageous to introduce normalized coefficients, that is, if \( \alpha_n \) is a judiciously chosen normalization constant (dependent on dimension \( n \)), the transform pair is developed for \( \hat{u}_{j,\nu,k} \) such that

\[ \alpha_n^{-1} \hat{u}_{j,\nu,k} \approx u_{j,\nu,k}, \quad (22) \]

following the discretization of the “inner” transform (13):

\[ \hat{u}_{j,\nu,k} = \frac{1}{\rho_k^{1/2} \sigma_{\nu}^{-1} \sigma_{\nu}^{-1}} \sum_l \hat{u}(\xi_l^{\nu,k}) \tilde{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[2\pi i (x^{\nu,k}_j, \xi_l^{\nu,k})], \quad (23) \]
The points $\xi_{\nu,k}$ are chosen on a (regular) rotated grid. Specifically, we let
\[
\Xi^k = \left\{ l \in \mathbb{Z}^n \mid -\frac{N'_k}{2} \leq l_1 < \frac{N'_k}{2}, \ldots, -\frac{N''_k}{2} \leq l_n < \frac{N''_k}{2} \right\}.
\]

(24)
The points in this set are denoted by $\Xi^k$. 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, denoting the accuracy of approximation (23) to the inverse Fourier transform. We choose the $\xi_{\nu,k}^r$ (covering the box $B_{\nu,k}$) as
\[
\xi_{\nu,k}^r = \Theta^{-1}_{\nu,k} (D_k S_k^{-1} \Xi^k + \xi_{\nu,k}^r e_1),
\]
where the matrix $S_k$ is defined as
\[
S_k = \left( \begin{array}{c} N'_k \\ 0_{n-1 \times 1} \end{array} \right).
\]

The dot product in the phase of the exponential in (23) then becomes
\[
\langle x_j, \xi_{\nu,k}^r \rangle = (D_k S_k^{-1} \Xi^k + \xi_{\nu,k}^r e_1)^t D_k^{-1} X_j = j_1 l_1 N'_k + j_2 l_2 + \ldots + j_n l_n.
\]

(26)
Thus, the specific choice of points $\xi_{\nu,k}^r$ allows for a fast evaluation of $\hat{u}_{j,\nu,k}$ (cf. (23)) for $j \in \Xi^k$,
\[
\hat{u}_{j,\nu,k} \exp(-2\pi i j l \xi_{\nu,k}^r / N'_k) = \frac{\alpha_n}{\sigma'_k (\sigma''_k)^{n-1} (N'_k)^n - 1} \sum_{l \in \Xi^k} \hat{u} (\xi_{\nu,k}^r) \hat{\beta}_{\nu,k} (\xi_{\nu,k}^r) \exp[2\pi i (j l \xi_{\nu,k}^r / N'_k)],
\]
where $\xi_l = l$ and $x_j = S_k^{-1} j$ with $j \in \Xi^k$, while $1/(N'_k (N''_k)^{n-1}) = \det S_k^{-1}$ is finite. (Compare this with equation (14) where we need infinitely many coefficients, cf. (8).) Indeed, we only encounter translations over
\[
\tilde{z}_{\nu,k}^r = \Theta^{-1}_{\nu,k} D_k^{-1} \Xi^k,
\]
(28)
cf. (26), defining a finite subset of the grid given in (9). We denote the discrete “inner” transforms (cf. (23)) by $G_{\nu,k}$, that is, $\hat{u}_{j,\nu,k} = (G_{\nu,k} \{ \hat{u} (\xi_{\nu,k}^r) \hat{\beta}_{\nu,k} (\xi_{\nu,k}^r) \}_{l \in \Xi^k}) (j)$. (We note that the normalization factor has been absorbed in $G_{\nu,k}$. One can use an n-dimensional FFT for the fast evaluation of $\hat{u}_{j,\nu,k}$ in (27) if the values for $\hat{u}$ and $\hat{\beta}_{\nu,k}$ are given at $\xi_{\nu,k}^r$. For the fast evaluation of these, we invoke the USFFT for the discrete “outer” transform, cf. (20), completing the discretization $\tilde{C}$ of C:
\[
\hat{u}_{j,\nu,k} = \frac{1}{\alpha_n} \left\{ \hat{\beta}_{\nu,k} (\xi_{\nu,k}^r) \right\} \sum_{l \in \Xi^k} u(x_l) \exp(-2\pi i (\xi_{\nu,k}^r / N'_k) x_l).
\]

(29)

We proceed with constructing the adjoint of the discretization of C, with the objective to construct a discretization of $C^{-1}$, while adapting the normalization factor $\alpha_n$ in order to obtain a discretely unitary transform pair.

We begin with revisiting (12) evaluated at points $\xi_{\nu,k}^r$, $m \in \Xi^k$, which concerns the “inner” transform. Given the (normalized) coefficients, we evaluate
\[
\frac{1}{\rho_k} \sum_{j \in \Xi^k} \hat{u}_{j,\nu,k} \exp(-2\pi i (\xi_{\nu,k}^r / N'_k) x_j) \hat{\chi}_{\nu,k}(\xi_{\nu,k}^r).
\]

which approximates the left-hand side of (12). Upon substituting (27) into this expression, we obtain
\[
\hat{\beta}_{\nu,k} (\xi_{\nu,k}^r) \frac{\alpha_n^2}{\sigma'_k (\sigma''_k)^{n-1}} \sum_{l \in \Xi^k} \hat{u} (\xi_{\nu,k}^r) \hat{\beta}_{\nu,k} (\xi_{\nu,k}^r) \exp(2\pi i (x_l \xi_{\nu,k}^r / N'_k)) \hat{\chi}_{\nu,k}(\xi_{\nu,k}^r) = \frac{1}{\rho_k} \sum_{j \in \Xi^k} \hat{u}_{j,\nu,k} \exp(-2\pi i (x_j \xi_{\nu,k}^r / N'_k)) \hat{\chi}_{\nu,k}(\xi_{\nu,k}^r),
\]
and recover the right-hand side of (12).

At this stage, the sample points in frequency change with each box, $B_{\nu,k}$, and hence the sum over $\nu,k$ is intimately connected to the inverse discrete “outer” Fourier transform to recover $u(x_i)$. Indeed, for the inverse Fourier transform of $\hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi)$, we apply the same discretization as in (23):
\[
\int \hat{u}(\xi) \hat{\beta}_{\nu,k}(\xi) \hat{\chi}_{\nu,k}(\xi) \exp[2\pi i (x_i, \xi)] d\xi \approx \frac{1}{\sigma'_k (\sigma''_k)^{n-1}} \sum_{l \in \Xi^k} \hat{u}(\xi_{\nu,k}^r) \hat{\beta}_{\nu,k}(\xi_{\nu,k}^r) \hat{\chi}_{\nu,k}(\xi_{\nu,k}^r) \exp[2\pi i (x_i, \xi_{\nu,k}^r)].
\]

We arrive at
\[
u, k \approx \frac{1}{\alpha_n^2} \sum_{\nu, k} \sum_{l \in \Xi^k} (G_{\nu,k} \{ \hat{u}_{j,\nu,k} \}_{j \in \Xi^k}) (l) \times \hat{\chi}_{\nu,k}(\xi_{\nu,k}^r) \exp[2\pi i (x_i, \xi_{\nu,k}^r)].
\]

(31)
\[ \tilde{C}^{-1} = \tilde{C}^*, \text{ whence the discrete transform pair within the discretization error becomes unitary. Moreover, (31) implies that } \]

\[ (\tilde{C}^{-1}\tilde{C}u)(x_i) \approx u(x_i). \]

The discrete transform and its adjoint (approximate inverse) can be summarized as

\[ u(x_i) \xrightarrow{\text{USFFT}} \tilde{u}(\xi_{l}^{\nu,k}) \xrightarrow{\text{windowing}} \tilde{u}(\xi_{l}^{\nu,k}) \tilde{\beta}_{\nu,k}(\xi_{l}^{\nu,k}) \xrightarrow{\text{IFFT}} \tilde{u}_\gamma, \]

and

\[ u(x_i) \xrightarrow{\text{adjoint USFFT}} \tilde{u}(\xi_{l}^{\nu,k}) \tilde{\beta}_{\nu,k}(\xi_{l}^{\nu,k}) \tilde{x}_\nu, \]

respectively. The multi-scale geometric nature of the discrete transform appears through identifying \( \gamma \) with points \((x_j, 2^bl)\) in phase space.

C. “Outer” transform revisited, USFFT

Because the points \( \xi_{l}^{\nu,k} \) are unequally spaced, fast evaluation of the sum, \( \sum_{x \in X^N} \), in (29) and the sum, \( \sum_{\nu,k} \sum_{l \in \Xi^k} \), in (31) pertaining to the “outer” Fourier transform, is achieved by introducing the USFFT. We summarize the key steps of this procedure here.

We choose Gaussian functions as in [13]. Given a prescribed accuracy, \( \epsilon_1 \), choosing \( a < -\pi / \log \epsilon_1 \), we let

\[ \lambda_a(x) = \exp(-ax^2). \]

We will omit the subscript \( a \) in our notation here. We introduce

\[ \Lambda^N(x) = \left( \sum_{x \in \Xi^k} \hat{\lambda}(N^{-1}x_1 + i_1)^2 \cdots \hat{\lambda}(N^{-1}x_n + i_n)^2 \right)^{-1/2}, \]

and observe that \( \Lambda^N > 0 \) and that \( \Lambda^N \) is periodic with period \( N \), that is, \( \Lambda^N(x_1 + N, ..., x_n + N) = \Lambda^N(x_1, ..., x_n) \). Because \( \hat{\lambda} \) is concentrated around zero, we have \( \Lambda^N(x) \approx \hat{\lambda}(N^{-1}x)^{-1} \cdots \hat{\lambda}(N^{-1}x)^{-1} \) as \( x \) is close to zero. We generate a cube of samples of size \( (2N)^n \) by padding the original data cube, \( u(x_i), i \in X^N \) (cf. (20)), with zeros; that is, \( u(x_i) = 0 \) if \( \frac{1}{2} < |x_i| \leq 1 \).

First, we multiply the regularly sampled function by \( \Lambda^{2N} \):

\[ U(x_i) = \Lambda^{2N}(N x_i) u(x_i). \]

This multiplication will compensate for the (Fourier) smearing driving the interpolation below. Secondly, we apply an FFT0 (the subscript \( 0 \) indicating the absence of normalization factors) to obtain \( \hat{U}(\xi_l) \) with \( \xi_l = l, l \in X^{2N} \) as in (20). Thirdly, we smear out the equally space transform onto the Fourier points \( \xi_{l}^{\nu,k} \) by computing the convolutions

\[ \hat{\nu}(\xi_l) = \sum_{\{l' : |2\xi_l - l'| < d_\lambda(c_\nu)\}} \hat{\lambda}(2\xi_l - l'_1) \times \cdots \times \hat{\lambda}(2\xi_n - l'_n) \hat{U}(\xi_l'), \]

where \( d_\lambda(c_\nu) \) is the smallest number such that \( \lambda(x) < \epsilon_1 \) if \( |x| > d_\lambda(\epsilon_1) \); \( \hat{\nu} \) is an approximation to the sum, \( \sum_{x \in X^N} \), in (29). We note that, upon considering the points \( \xi_{l}^{\nu,k}, l \in \Xi^k \), covering the box \( B_{\nu,k} \), we need to include a neighborhood, \( B_{\nu,k} \) say, of \( B_{\nu,k} \) determined by \( d_\lambda(c_\nu) \) to carry out the proper interpolation.

Conversely, to evaluate the sum, \( \sum_{\nu,k} \sum_{l \in \Xi^k} \), in (31) at equally spaced points, \( x_i \), from \( \hat{\nu}(\xi_{l}^{\nu,k}) = (G_{\nu,k}^{*} \{ \hat{u}_{j,x_{\nu,k}} \}_{j \in \Xi^k})(l) \hat{\chi}_{\nu,k}(\xi_{l}^{\nu,k}) \), given at unequally spaced points, \( \xi_{l}^{\nu,k} \), we carry out the following steps. First, we apply smearing to an equally spaced grid by computing the convolutions,

\[ \hat{U}(\xi_l) = \sum_{\{l' : |2\xi_l - l'| < d_\lambda(c_\nu)\}} \lambda(2\xi_l^{\nu,k} - l'_1) \times \cdots \times \lambda(2\xi_n^{\nu,k} - l'_n) \hat{\nu}(\xi_{l}^{\nu,k}), \]

\[ l' \in X^{2N}. \]

Secondly, we apply an inverse FFT0 to obtain \( U(x_i) \) with \( x_i = N^{-1}x, l \in X^{2N} \). Thirdly, we compensate for the (Fourier) smearing by multiplying by \( \Lambda^{2N} \):

\[ \alpha_n^2 u(x_i) \approx \Lambda^{2N}(N x_i) U(x_i), \]

IV. TILING AND PARTITION OF UNITY

Here we specify the general procedure of the tiling with boxes \( B_{\nu,k} \) outlined in Section II. In the radial direction, we have

\[ |\xi_{l}^{\nu,k}| = c_0 2^k, \quad k = 1, \ldots, k_{\text{max}}, \]

cf. (21). Here, \( k_{\text{max}} \) is either given or can be estimated from the size, \( N^n \), of the sampled data cube:

\[ k_{\text{max}} = \left\lfloor \log_2 \left( \frac{N}{c_0} \right) \right\rfloor; \]

\( c_0 \) becomes a parameter defining the radius of the coarsest scale (\( k = 0 \)). (One can use an alternative convention of defining \( \xi_{l}^{\nu,k} \), namely, \( \xi_{l}^{\nu,k'} = 2^k, k' = k_0, \ldots; \) then \( c_0 = 2^{k_0} \) and \( k = k' - k_0). \)

A. Cubed sphere

For each \( k > 0 \) we have to construct a set of rotated boxes (cf. (3)). We begin with introducing a set of points close to evenly distributed on the \((n-1)\)-dimensional unit sphere that will define the \( \nu \) for given \( k \); each \( \nu \) defines the radial orientation of a box, \( B_{\nu,k} \). We use the cubed sphere [21] to define these points. The \( n \)-dimensional cube consists of \( 2n \) \((n-1)\)-dimensional sides (see Figure 2 a)). On each side of the cube, we define a grid that is characterized by \( N_{c,k} \), the number of points in one direction. The grid points from each side will be mapped into \( 2n \) sectors on the \((n-1)\)-dimensional sphere, symmetrically distributed around the coordinate axes (see Figure 2 a, c)).

Let \( N_{c,k} \) denote the total number of box rotations, that is, the number of points on \( S^{n-1} \), for scale \( k \). Following the dyadic parabolic scaling in (2), we take

\[ N_{c,k} = [c_1 2^{(k-1)/2}]_{\text{odd}}, \]

\[ \nu_{\text{max},k} = 2n (N_{c,k})^{n-1} \sim 2^{k(n-1)/2}, \]
For all other sectors. We denote the maximum value of the symmetries, the same holds for the "vertical" direction, and orientations for cubed sphere grid; in this case, the relative difference between sector. This plot provides a measure of non-uniformity of the distance between neighboring points in the "horizontal" directions, see Figure 2 c). In Figure 2 d) we show the distribution of points on the sphere; see Figure 3 a) where Nc,k is rounded to the nearest odd number in order for all the basis vectors ei, 1 ≤ i ≤ n, to be contained in the set of orientations ν; c1 is a parameter defining the number of orientations for k = 1. Then the number of orientations for k > 1 is defined by (35).

Grid points on one side of the cube are shown in Figure 2 c) (for Nc,k = 9). Note that the grid is chosen to be symmetric but not equally spaced so as to achieve more uniformity in the distribution of points on the sphere; see Figure 3 a) where points from three sides of the cube are mapped back onto the unit sphere (in the center of the plot one can see a junction of three sectors). We note that the cubed sphere provides us with a natural notion of point neighbor in the "vertical" and "horizontal" directions, see Figure 2 c). In Figure 2 d) we show the distance between neighboring points in the "horizontal" direction after mapping them onto the unit sphere (for a given sector). This plot provides a measure of non-uniformity of the cubed sphere grid; in this case, the relative difference between maximum and minimum distance is 18%. (Due to the imposed symmetries, the same holds for the "vertical" direction, and for all other sectors.) We denote the maximum value of the neighboring distance function as d(k), and use it to define the half width of the boxes Bν,k.

1) Box orientations: Each box Bν,k is a rectangular region in n-space and its orientation is defined by n orthogonal unit vectors, ω1, . . . , ωn, which also form a (local) orthogonal basis within the box. The radial direction (orientation of the long side of the box) is obtained from the grid points following the cubed sphere construction; we set ω1 ≡ ν. The orientations of the short sides are found from solving a set of n linear equations: Let ωp for p = 1, . . . , q − 1 be given, then ωq is a solution of

\[
(ω_q, ω_p) = 0, \quad (ω_q, e_i) = 0, \quad p = 1, \ldots, q - 1, \quad i = 1, \ldots, n - q, \quad q > 1,
\]

subject to the normalization |ωq| = 1. For n = 3, writing ω1 = (ω11, ω12, ω13)T, our choice of directions follows to be

\[
ω_1 ≡ ν,
ω_2 = (0, ω_13, -ω_12)^T \frac{\text{sign}(ω_13)}{\sqrt{ω_1^2 + ω_1^3}},
ω_3 = ω_1 ∧ ω_2.
\]

Our choice of orientation for ω2 is shown schematically in Fig. 2 b) (indicated by arrows). The vectors ω2 and ω3 define the orientations of squares centered at the end points of ω1. One can amply cover the unit sphere with such squares, see Figure 3 b); these become the short sides of the boxes Bν,k.

In Figure 4 we show the covering by squares away and near junctions of the different sectors, at two different scales.

2) Box sizes: For boxes belonging to scale k, the “central” frequency is given in (33) as c02k. The outer radius of boxes belonging to scale k is thus c02k+1, while the inner radius is c02k−1. The half box width is obtained by multiplying d(k) with the outer radius; then

\[
L_k^F = c_02^{k+2}d(k),
\]

cf. (1). With the requirement that the box width is smaller then the inner radius, we arrive at the constraint, 2π < Nc,1, on the minimal number of orientations for the first scale, k = 1.

We reconsider the boxes Bν oriented along the e1-axis, as in (1). Asymptotically, that is, for large k, the inner and outer radii will define the left and right boundaries of the box (that is, in the radial direction). However, the left boundary of the box should be taken somewhat smaller than the earlier defined inner radius. This is illustrated in Figure 5 a), where arrows are indicating that the support function of the box – to be defined in polar coordinates – will be stretched in the transverse (polar) directions while filling out the entire box. Thus, due to the curvature of the inner sphere, the left boundary of the box should be chosen smaller, and can be estimated from the inner radius (~2k−1) and L′′k. We obtain the box length,

\[
L_k' = c_02^{k+1} - K(k), \quad K(k) = \sqrt{(c_02^k)^2 - (L_k'^2)/2}.
\]

In Figure 6 we illustrate how the dyadic parabolic decomposition is built up following the construction above based on the cubed sphere.

In summary, the parameters controlling the tiling are: the dimension n, the number of samples Nc,k defining the size of the cube associated with the coarsest scale (and thus the “units”), and e1 = Nc,1 defining the number of orientations for k = 1.

B. Construction of window functions, Partition of unity

We construct the functions ˆχν,k, and will choose, here, ˆβν,k = ˆχν,k (the frame is tight). We adopt local orthogonal coordinates on each box, using basis vectors ω1 =
Fig. 3. a) Points mapped from three cube sides (see Figure 2 c)) back onto the sphere defining orientations \( \nu \); b) covering the sphere with squares (based on the grid points shown in a).

\[
\omega_1(\nu,k), \ldots, \omega_n(\nu,k). \quad \text{The rotation matrix } \Theta_{\nu,k} \text{ in (3) maps } \omega_1(\nu,k), \ldots, \omega_n(\nu,k) \text{ to the global basis vectors } e_1, \ldots, e_n. \quad \text{In terms of matrices,}
\]

\[
\Theta_{\nu,k}^{-1} = (\omega_1(\nu,k), \ldots, \omega_n(\nu,k)). \quad (40)
\]

Let \( \xi_{\nu,k} \) denote the coordinates of covectors \( \xi \) with respect to the rotating basis \( \omega_1 = \omega_1(\nu,k), \ldots, \omega_n = \omega_n(\nu,k) \); we will also write \( \xi \) for the (co)vector of coordinates with respect to the global basis \( e_1, \ldots, e_n \). Then \( \xi_{\nu,k} = \Theta_{\nu,k} \xi \).

A transformation of coordinates \( \xi_{\nu,k} \) to coordinates \( \xi'_{\nu,k} \) is illustrated in Figure 5 b).

We then introduce “polar” coordinates,

\[
r = |\xi| = |\xi_{\nu,k}|, \quad \phi' = \arcsin \left[ \sqrt{(\xi_2^{\nu,k})^2 + \ldots + (\xi_n^{\nu,k})^2}/r \right]
\]

writing \( \xi_{\nu,k} = (\xi_1^{\nu,k}, \ldots, \xi_n^{\nu,k}) \), see Figure 7.

1) Auxiliary functions: Given the tiling presented above, we discuss how to compute the associated window functions numerically. We rely on the construction presented in [7].

We need auxiliary functions: A radial function, \( w(r) \), and its dilations \( w(2^{-k}r) \), windowing along the axes of the boxes, and transverse functions, \( \kappa_k(r, \phi) \), windowing in the transverse directions. For \( w(r) \), it is natural to follow the construction of Meyer wavelets, that is

\[
w(r) = \begin{cases} 
\sin \left( \frac{\pi}{2} a_m(2r - 1) \right), & \text{if } \frac{1}{2} \leq r < 1; \\
\cos \left( \frac{\pi}{2} a_m( r - 1) \right), & \text{if } 1 \leq r \leq 2; \\
0, & \text{otherwise .}
\end{cases} \quad (42)
\]

Here, \( a_m(r) \) is defined so that \( w \in C_0^{m+1} \) (for details, see [7, Section 6.2]). For the illustrations in this paper, we take
Given $w$ the appropriate decay properties within numerical precision. To accommodate this approach, we introduce polar coordinates $(r, \phi^\nu)$ in each $(\nu, \omega_{p+1})$-plane ($p = 1, \ldots, n - 1$) with
$$\phi_p^\nu = \arcsin(e^{\nu k}_{p+1}/r), \quad p = 1, \ldots, n - 1.$$ 

Then
$$\kappa_k(r, \phi^\nu_1, \ldots, \phi^\nu_{n-1}) = \prod_{p=1}^{n-1} \cos \left( \frac{\pi}{2} a_m \left( \frac{r}{2^{k+1}} \left| \phi_p^\nu \right| \right) \right).$$

2) Partition of unity: The radial window functions satisfy the admissibility condition,
$$|w_0(r)|^2 + \sum_{k\geq 0} |w(2^{-k}r)|^2 = 1.$$ 

We let
$$v_{\nu,k}(r, \phi^\nu) = \frac{\kappa_k(r, \phi^\nu)}{\sqrt{\sum_{p=1}^{n} \kappa_k(r, \phi^\nu - \phi^\nu_p)^2}},$$
then (cf. (41))
$$\hat{\chi}_{\nu,k}(\xi^\nu) = w(2^{-(k-1)}r(\xi^\nu)) v_{\nu,k}(r(\xi^\nu), \phi^\nu(\xi^\nu)), \quad k \geq 1.$$ 

Ideally, the construction would lead to window functions such that $\hat{\chi}_{\nu,k}$ can be obtained from $\hat{\chi}_{\nu,k}$ upon coordinate rotation. For $n \geq 3$ this cannot be achieved. Hence, the $\hat{\chi}_{\nu,k}$ should be precomputed and stored in memory, for each $B_{\nu,k}$, in applications.

We note the particular $r$ dependence in $\kappa_k(r, \phi^\nu)$. This dependence is designed to make the supports of $\hat{\chi}_{\nu,k}$ “fill out” the boxes $B_{\nu,k}$. As a consequence, the decay in estimate (6) is modified from $(2^k|\nu, x| + 2^{k/2}|x|)^{-N}$ to $(2^k|\nu, x| + 4 \cdot 2^{k/2}|x|)^{-N}$. This effect is, in practice, significant in as much as it decreases the required oversampling factors $\sigma_0^\nu$. However, the angular overlap between the windows (but not the total number of discretization points) increases.

We illustrate the functions $\beta_{\nu,k} = \hat{\chi}_{\nu,k}$ in Figure 8 for a $\nu$ away from a junction (top row), and for a $\nu$ near a junction (bottom row).

3) Finest scale: For the finest scale, we have $c_0 2^{k_{\text{max}}} < \frac{\pi}{2}$, that is, the largest annulus considered fits in the fundamental cell. The question remains how to fill in the residual region with window functions so as to satisfy the partition of unity over the entire fundamental cell. This will guarantee that property $C^{-1}C \approx I$ holds true for the discrete transform.

One option is to add one more scale ($k_{\text{max}} + 1$) that will reach Fourier space up to a distance $N$ from the origin (in the radial direction). This is computationally expensive, however. Another option is to use wavelets to cover the “boundary” region in Fourier space. However, this approach fails to meet the goal of performing directional analysis and decomposition. Here, we assume that there are essentially no Fourier data close to the boundaries of the fundamental cell. This is a reasonable assumption while dealing with band-limited wavefields. Then

We could also have taken tensor products of functions each defined on a $(\nu, \omega_{p+1})$ coordinate plane ($p = 1, \ldots, n - 1$). To accommodate this approach, we introduce polar coordinates $(r, \phi^\nu_p)$ in each $(\nu, \omega_{p+1})$-plane ($p = 1, \ldots, n - 1$) with
$$\phi_p^\nu = \arcsin(e^{\nu k}_{p+1}/r), \quad p = 1, \ldots, n - 1.$$
we modify the radial window function for scale $k_{\text{max}}$ as shown (dashed) in Figure 9. The residual is ignored, which results in a low-pass filtering of the wavefield before wave-packet decomposition; more precisely, the reconstruction yields a smoothed version of the original wavefield. While constructing the partition of unity for all scales, $0 \leq k < k_{\text{max}}$, we use the function $w(r)$ (cf. (42)) subjected to scaling (cf. (45)), which is also shown (solid line) in Figure 9.

4) Oversampling and local grids: Next, we evaluate $\hat{\beta}_{\nu,k} = \hat{\chi}_{\nu,k}$ on a regular grid, $\xi^{\nu,k}_l, l \in \Xi^k$, that is defined on the box $B_{\nu,k}$ as described by equations (25) and (24). Oversampling is related to the fact that the functions $\psi_\gamma$ are not compactly supported. To analyze accuracy, we assume a “test” oversampling described by factors $(\sigma'_{k}, \sigma''_{k})$ and corresponding parameters $(N'_k, N''_k)$. The coefficient evaluation can be described by a convolution with $\rho_{k}^{-1/2} \beta_{\nu,k} = \psi_{\nu,k}$, cf. (13). This convolution is discretized through equation (23), with $(N'_k, N''_k)$ now replaced by $(N'_k, N''_k)$. Using this discretization, via an inverse FFT, we obtain $\psi_{\nu,k}(x^{\nu,k}_j)$ with $x^{\nu,k}_j$ given in (28) and $j \in \tilde{\Xi}^k$ (cf. (24) with $(N'_k, N''_k)$ replaced by $(N'_k, N''_k)$ as before). A prescribed accuracy, $\epsilon_2 > 0$, then determines the choice of

$$|\psi_{\nu,k}(x^{\nu,k}_j)| < \epsilon_2 \quad \text{for all} \quad j \in \tilde{\Xi}^k.$$
The smallest $\tilde{N}_k^\ell$ and $\bar{N}_k^\ell$ satisfying this condition are $N_k^\ell = N_k^\ell(\epsilon_2)$ and $N_k^\ell = N_k^\ell(\epsilon_2)$, respectively. To obtain estimates for computational complexity one has to make the connection with estimate (6). In practice, we carry out the testing for $\psi_{\nu,k}$ with $\nu = \epsilon_1$ only.

The choice of locally orthogonal grids on each box results in orthogonal translation grids $x_j^{\nu,k}$. Translation grids are shown by white points in Figure 10 for the box $(\nu,k) = (7,3)$ (panel a)) and $(\nu,k) = (13,4)$ (panel b)). The proper rotation of translation grids is possible due to the use of the USFFT algorithm. Gray infilled points correspond to large coefficients $\hat{u}_\gamma$ for the synthetic wavefield. We observe that the large coefficients are concentrated along the wavefront, as they should.

V. DISCRETE ALMOST SYMMETRIC WAVE PACKETS

In Figure 11 we show isosurfaces for one discrete almost symmetric wave packet for $n = 3$. We illustrate the degree of rotational symmetry about $\nu$ of our discrete wave packets in panels a) - d) using values of 20, 10, 2.5, and 1% of their maximum, respectively. Symmetry persists up to panel c).

Decay estimates (cf. (6)) are numerically confirmed in Figure 12. Here, we show two-dimensional cross sections of a discrete almost symmetric wave packet $(n = 3)$ corresponding to scales $k = 1$ (panel a)) and $k = 2$ (panel b)). In panel c) we show the decay along the $\nu$ direction in logarithmic scale (solid line); the dashed lines represent exponential decays, $\exp[-2^k |(\nu, \nu)|]^\nu$. Indeed, within numerical precision, discrete almost symmetric wave packets decay superalgebraically and obey the dyadic scaling.

Due to differences in window functions (cf. Figure 8) wave packets corresponding to different orientations will have slightly different shapes. We illustrate the degree of uniformity in shape with orientation (for given scale) in Figures 13 and 14. In Figure 13 we show two-dimensional cross sections of discrete almost symmetric wave packets $(n = 3)$ for scale $k = 1$, corresponding with a window function from a “regular” box overlapping region (see Figure 4 top, panel a)) and a window function from the most “irregular” box overlapping region (see Figure 4 middle, panel b)), and their difference (panel c)). We have applied rotations to align them with the $x_1$ axis. In Figure 14 we illustrate this particular difference in three dimensions, for scales $k = 1$ (panel a)) and $k = 2$ (panel b)). The almost symmetric shapes of the differences reflect the fact that the main distortion induced by varying window functions arises from a change of wave-packet amplitude rather than shape.

VI. NUMERICAL APPLICATIONS

A. Clustering of coefficients and geometry

We consider a synthetic seismic wavefield data cube for $n = 3$ of size $N^3$ with $N = 128$. In Figure 15 a) we show a (subsurface) model consisting of a smooth low velocity lens centered at $(0,1)$ km and a discontinuity at 2 km depth. A finite-difference method was used to generate, in this model, the surface data cube, $d = d(s,r,t)$, where $s$ denotes the source location, $r$ denotes the receiver location, and $t$ the time: For each source, $s$, $d(s,r,t)$ is the solution to the acoustic wave equation recorded by an array of receivers located at the surface, $z = 0$. We show two cross sections of $d$ in Figure 15 b). For one source, $s$, we illustrate the propagation of singularities with rays for the incident field (Figure 15 c))
and rays for the reflected field (Figure 15 d)), demonstrating the formation of caustics. We apply our discrete transform to this data cube.

We illustrate the geometrical aspects of the decomposition, using the data cube shown in Figure 15, in Figures 16-17. The clustering of coefficients for one particular box ($\nu = 3, k = 2$) is shown in Figure 16: Isosurfaces delineate regions of large coefficients (top) while the absolute values of coefficients are plotted in color (bottom). The largest values (for each scale) occurring at indices $\gamma_0 = (j_0, \nu_0, k_0)$, say, provide a geometrical viewpoint via points $(\tilde{x}_{j_0,\nu_0,k_0}^{s_0}, \tilde{x}_{j_0,\nu_0,k_0}^{\nu_0})$ in phase space. In terms of the data cube, we identify $\tilde{x}_{j_0,\nu_0,k_0}^{s_0}$ with $(s_0, r_0, t_0)$ and $\tilde{x}_{j_0,\nu_0,k_0}^{\nu_0}$ with wavevectors and frequency, $(\xi_{s_0}, \xi_{r_0}, \tau_0)$. This information allows us to construct a pair of rays: An incident ray starting at source $s_0$ in the direction defined by $\xi_{s_0}/\tau_0$, and a reflected ray starting at receiver $r_0$ in the direction defined by $\xi_{r_0}/\tau_0$. Such rays are shown in Figure 17.

Upon mapping $j$ to $\tilde{x}_{j}^{\mu}\nu$ (cf. (28)) we can visualize the clustering of coefficients in $(s, r, t)$ (Figure 18): The isosurfaces correspond to those shown in Figure 16 top. It becomes clear that the clustering is aligned with the wave fronts, as it should.

Next, we test the compression capability of discrete almost symmetric wave packets. To this end, we apply the discrete transform to the data cube shown in Figure 15 and subject the resulting coefficients to thresholding; we use only the significant coefficients for the reconstruction. The quality of reconstruction is measured by

$$\text{SNR} = 20 \log_{10} \frac{\sqrt{N^{-3} \sum_{i \in X^N} |u(x_i)|^2}}{\sqrt{N^{-3} \sum_{i \in X^N} |u(x_i) - u_{CR}(x_i)|^2}}$$

(essentially, we have replaced $\max_{i \in X^N} |u(x_i)|$ in the definition of peak signal-to-noise ratio by $\sqrt{N^{-3} \sum_{i \in X^N} |u(x_i)|^2}$)

where $u$ represents the original data and $u_{CR}$ is its reconstruction corresponding to compression rate,

$$CR = \frac{N^3}{N_{WP}}$$

where $N_{WP}$ is the number of coefficients used for reconstruction, and $N^3$ is total number of samples in the original data cube.

In Figure 19, we show the dependence of reconstruction quality on compression rate. In Figure 20 we show two-dimensional cross sections of the original data as well as its reconstructions corresponding to compression rates $CR = 0.5, 7$ and $50$. We note that the geometry of wave fronts is preserved under significant compression rates. These reconstructions illustrate that wave packets are the natural building blocks of seismic data [4]; they are geometrically related to “contact elements”.

Fig. 12. Decay of discrete almost symmetric wave packets (real parts): a) $k = 1$; b) $k = 2$; c) decays in logarithmic scale in the $\nu$ direction (solid lines) and exponential decays (dashed lines).
Fig. 18. Clustering of (large) coefficients in a particular box ($\nu = 3$, $k = 2$ as in Figure 16) visualized in $(s, r, t)$ upon mapping $j$ to $\tilde{x}_{\nu,k}^j$ in Figure 16. The orientation of $\nu$ is indicated by an arrow. The clusters are aligned with wave fronts in the data.

B. Multi-scale analysis and pointwise regularity

First we consider the distribution, $\alpha = \delta(x_1)$, defined by

$$\int \alpha(x)f(x)\,dx = \int f(0, x_2, \ldots, x_n)\,dx_2 \ldots dx_n,$$

$f$ representing a real-valued test function, and verify the detection of its wavefront set, $\{(x,\xi) \mid x_1 = 0, |\xi| = e_1\}$.

Following supports, it is immediate that $\alpha_\gamma$ decays rapidly, that is, $O((2^k)^{-N})$ for all $N \in \mathbb{N}$ as $k$ becomes large, unless $\nu = e_1$. If $\nu = e_1$ the coordinates $\xi^{\nu,k}$ coincide with the usual...
Cartesian coordinates \((\Theta_{\nu, k} = I)\). Then
\[
\int \overline{\psi_\gamma}(\xi_1, 0, \ldots, 0) d\xi_1 = \\
\int \rho_k^{-1/2} \hat{\chi}_{\nu, k}(\xi_1, 0, \ldots, 0) \exp[2\pi i \xi_1 \langle x_{\nu, k}, e_1 \rangle] d\xi_1 = \\
\int \rho_k^{-1/2} w(2^{-k}(|\xi_1|) v_{\nu, k}(|\xi_1|, 0) \exp[2\pi i \xi_1 \langle x_{\nu, k}, e_1 \rangle] d\xi_1.
\]

Changing variable of integration, \(\xi_1' = 2^{-k} \xi_1\), yields
\[
\alpha_\gamma = \int \overline{\psi_\gamma}(\xi_1, 0, \ldots, 0) d\xi_1 = 2^k|B_k|^{-1/2} D(2^k(L'_k)^{-1}j_1),
\]
\[
D(2^k(L'_k)^{-1}j_1) = \int w(2^k|\xi_1'| v_{\nu, k}(2^k|\xi_1'|, 0) \exp[2\pi i \xi_1' 2^k(L'_k)^{-1}j_1] d\xi_1',
\]
where \(w(2^k|\xi_1'| v_{\nu, k}(2^k|\xi_1'|, 0)\) is independent of \(k\). If \(j_1 = 0\), \(\alpha_\gamma\) decays as \(2^k|B_k|^{-1/2} \sim 2^{(-\frac{n+1}{4})}\)
\(k\) becomes large.

First we test the decay numerically for \(\alpha = \delta(x_1)\) and \(n = 2\). Then, if \(\nu = e_1\) and \(j_1 = 0\), \(\alpha_\gamma\) decays as \(2^{1-\frac{n+1}{4}}\) as \(k\) becomes large. The original data containing the \(\delta\) are shown in Figure 21 a). In Figure 21 b) we show the dependence of the absolute values \(|\alpha_k|\) of coefficients \(\alpha_k\) (denoting \(\alpha_{j_1, e_1, k}\) by \(\alpha_k\)) with scale \(k\). The point \(j_1\) of a translation grid is chosen to be: right at the singularity (solid line), 2 samples away from the singularity (dashed line), and 6 samples away from the singularity (dotted line). Circles indicate the theoretical decay rate \(\sim 2^{k/4}\).

Second we test the decay numerically for the Heaviside function \(\alpha = H(x_1)\) and \(n = 2\) as shown in Figure 21 c)-d). Then, if \(\nu = e_1\) and \(j_1 = 0\), \(\alpha_\gamma\) decays as \(2^{1-\frac{n+1}{4}}\) as \(k\) becomes large. In Figure 21 d) we show the dependence of the absolute values \(|\alpha_k|\) with scale \(k\). The point \(j_1\) of a translation grid is chosen to be: right at the singularity (solid line), 2
samples away from the singularity (dashed line), and 6 samples away from the singularity (dotted line). Circles indicate the theoretical decay rate \( \sim 2^{-3k/4} \).

C. Phase-space criterion for wavefield matching

Let \( u \) denote a wavefield as before, and \( u_\lambda \) a perturbation of this wavefield, parameterized by \( \lambda \). We write \( x = (t, y) \) and introduce the cross correlation, \( R_\chi(u, u_\lambda)(t, y) = \int (\chi u)(t + t', y) (\chi u_\lambda)(t', y) \mathrm{d}t' \), where \( \chi \) is a smooth cutoff function extracting a constituent phase in the wavefield. We introduce

\[
E_\alpha(\lambda; k_0) = \sum_\gamma |2^{k-k_0}|^\alpha \left( 2^{\min(k, k_0)} |d(\gamma, \tau_0)| \right)^\alpha \times (R_\chi(u, u_\lambda))_\gamma^2,
\]

cf. (17) with \( \gamma_0 = ((0, y_0), \epsilon_\lambda, k_0) \), where \( \epsilon_\lambda \) corresponds with \( \epsilon_1 \). \( E_\alpha(\lambda; k_0) \) measures whether \( \chi u \) is close to \( \chi u_\lambda \). In Figure 22 we show an example of forming a cross correlation if \( u \) represents a reflection off a flat (horizontal) reflector and \( \lambda \) parameterizes a perturbation in medium velocity. In Figure 23 we plot \( E_\alpha(\lambda; k_0) \) as a function of \( \lambda \) for \( \alpha = 1 \); \( k_0 \) is chosen to be the scale carrying most of the energy. Indeed \( E_\alpha(\lambda; k_0) \) provides a quasi-geometrical criterion for wavefield matching.

![Fig. 22. Top: \( u \) superimposed on \( u_\lambda \), where \( u \) represents a reflection off a flat (horizontal) reflector and \( \lambda \) parameterizes a perturbation in medium velocity. Bottom: \( R_\chi(u, u_\lambda) \); the trace on the right shows the waveform at one particular location (value of \( y = h \)).](image)

![Fig. 23. The functional \( E_\alpha(\lambda; k_0) \) for \( u \) and \( u_\lambda \) as in Figure 22.](image)

**VII. CONCLUSIONS**

We have constructed a novel discretization of the continuous “curvelet” transform in dimension \( n \geq 3 \) leading to the introduction of discrete almost symmetric wave packets, and an associated discrete transform pair which is unitary. A further characteristic of our discretization is the appearance of parameters that control the tiling of phase space corresponding with the dyadic parabolic decomposition, while adapting to the (wave) physical problem at hand.

The computational complexity of our algorithm is determined by the computational complexity of the USFFT.

Discrete almost symmetric wave packets are well suited to develop a wave-ray duality in computational propagation and scattering problems. This duality is apparent in the clustering of wave-packet coefficients of wavefields and corresponding compression following wavefront sets.

**ACKNOWLEDGMENT**

This research was supported in part under NSF CMG grant DMS 0724644, and in part by the members, BP, ConocoPhillips, ExxonMobil, StatoilHydro, Total, of the Geophysical Imaging Group.

**REFERENCES**


Fig. 21. Numerical estimation of pointwise regularity using wave packets: a) $\delta(x_1)$; b) coefficient “decay” with scale $k$ for $\delta(x_1)$: at the singularity (solid line), 2 samples away from the singularity (dashed line), and 6 samples away (dotted line). The theoretical rate (at the singularity) is indicated by circles. c) $H(x_1)$; b) coefficient decay with scale $k$ for $H(x_1)$: at the singularity (solid line), 2 samples away from the singularity (dashed line), and 6 samples away (dotted line). The theoretical rate (at the singularity) is indicated by circles.


Anton A. Duchkov received his M.Sc.-Degree from Novosibirsk State University, Novosibirsk, Russia, in 1996. He received his Ph.D.-Degree in Geophysics from the Institute of Geophysics SB RAS, Novosibirsk, Russia, in 2005. Since 2006, he has been a Postdoctoral Research Fellow with the Center for Computational and Applied Mathematics (CCAM) at Purdue University, West Lafayette IN. His research interests are in seismic migration, imaging and migration velocity analysis.


Fredrik Andersson Biography text here.
Maarten V. de Hoop received his M.Sc.-Degree in theoretical physics (cum laude) at the University of Utrecht, the Netherlands in 1984. He received his Ph.D.-Degree in technical sciences (cum laude) at Delft University of Technology, the Netherlands, in 1992. De Hoop is a professor of Mathematics at, and the director of, the Center for Computational and Applied Mathematics (CCAM), and founder and director of the Geo-Mathematical Imaging Group (GMIG) at Purdue University, West Lafayette IN. He holds a (part-time) visiting professorship in the Department of Earth, Atmospheric and Planetary Sciences at MIT. He also holds a visiting professorship at the Graduate University of the Chinese Academy of Sciences (Beijing). He is a fellow of the Institute of Physics. He is the recipient of the 1996 J. Clarence Karcher Award of the Society of Exploration Geophysicists, and the award for young scientists of the International Society for Analysis, its Applications and Computation in 2001. He was a program leader and the chair of the Program on Random Media at SAMSI (the Statistical and Applied Mathematical Sciences Institute) in 2007-08. He is a member of the scientific committee of the Geomathematics Programme at PIMS (Pacific Insitute for the Mathematical Sciences).