Proceedings of the Project Review, Geo-Mathematical Imaging Group (Purdue University, West Lafayette IN), Vol. 1 (2008) pp. 1-22.

HIGHER-DIMENSIONAL DISCRETE SMITH CURVELET TRANSFORM: A PARALLEL ALGORITHM *

ANTON A. DUCHKOV†, STEVEN DONG‡, FREDRIK ANDERSSON§, AND MAARTEN V. DE HOOP¶

Abstract. We revisit Smith's frame of curvelets and discretize it making use of the USFFT as developed by Dutt and Roklin, and Beylkin. In the discretization, we directly approximate the underlying dyadic parabolic decomposition, its (rotational) symmetry, and approximate all the necessary decay estimates in phase space with arbitrary accuracy. Numerically, our transform is unitary. Moreover, if we apply the inverse transform after the forward transform, we approximate the identity matrix, as it should, and if we apply the forward transform after the inverse transform we recover the necessary decay estimates (of the matrix representing the identity operator). 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 $n = 3, 4, 5$, while the data sets and images are inherently very large. We propose a parallel algorithm suited for such large-scale problems in the realm of high performance computing.

Key words. curvelet transform, dyadic parabolic decomposition, computational harmonic analysis, compression

AMS subject classifications.

1. Introduction. The theory of second microlocalization, or dyadic parabolic decomposition, has been an important development in harmonic analysis [21]. Over the recent years, it has led to the development of new multi-scale computational techniques, not only in image processing [7], but also in partial differential equations [1] and inverse problems encompassing inverse scattering [15, 13] and wave-equation tomography [5]. However, this class of techniques also sheds light on various physical aspects of solutions to (hyperbolic) partial differential equations and inverse problems. For example, Andersson *et al.* [1] prove a concentration of wave packets result for hyperbolic evolution equations with limited smoothness, De Hoop *et al.* [13] develop the notion of illumination and partial reconstruction in inverse scattering, and Brytik *et al.* [5] 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 [11, 14].

At the basis of all these developments is the frame of curvelets introduced by Smith [20] and the associated class of transforms. The general concept has its roots in the theory of coherent wave packets and Fourier integral operators (Córdoba and Fefferman [10]), and parabolic cutoffs dating back to Boutet de Monvel [3]. Furthermore, the curvelet transform can be related to the Fourier-Bros-Iagolnitzer (FBI) transform (Bros and Iagolnitzer [4]).

Here, we revisit Smith's frame of curvelets and discretize it making use of the USFFT as developed by Dutt and Roklin [16] and Beylkin [2]. In the discretization, we directly approximate the underlying dyadic parabolic decomposition, its (rotational) symmetry, and approximate all the necessary decay estimates in phase space with arbitrary accuracy. Numerically, our transform is unitary. Moreover, if we apply the inverse transform after the forward transform, we approximate the identity matrix, as it should, and if we apply the forward transform after the inverse transform we recover the necessary decay estimates (of the matrix representing the identity operator [20]). Another relevant aspect of our discretization is the appearance of parameters that control the

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

†Center for Computational and Applied Mathemematics, and Geo-Mathematical Imaging Group, Purdue University, West Lafayette, IN 47907, USA (aduchkov@purdue.edu).

‡Center for Computational and Applied Mathemematics, and Geo-Mathematical Imaging Group, Purdue University, West Lafayette, IN 47907, USA (sdong@math.purdue.edu).

§Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden.

¶Center for Computational and Applied Mathemematics, and Geo-Mathematical Imaging Group, Purdue University, West Lafayette, IN 47907, USA (mdehoop@purdue.edu).

tiling of phase space corresponding with the dyadic parabolic decomposition, preserving the relative parabolic scaling, 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).

We hasten to mention the algorithm – without wrapping – developed by Ying *et al.* [6] for the digital curvelet transform [8, 9]. In dimension 2, our algorithm has certain elements in common. However, differences include the construction of an explicit inverse (USFFT based discretization and unitarity), the symmetries important in multi-scale analysis of inverse problems applications, and the parallel structure necessary for higher-dimensional and large data set applications. In this paper, we have applications in exploration seismology (wide azimuths towed streamer surveys) and global seismology USArray) in mind. For these, we indeed need transforms in higher dimensions, that is, in dimension $n = 3, 4, 5$, while the data sets and images are inherently very large, which necessitates the development of algorithms for massively parallel computing.

Future applications of the multi-scale, discrete transform developed here, include higher-dimensional data or image regularization (ℓ^1 optimization [12]), denoising [17] and visualization, computational directional regularity analysis and multi-scale characterization of boundary layers, and the further development of computational spectral and microlocal analysis.

2. Dyadic parabolic decomposition. We begin with an overlapping covering of the positive ξ_1 axis by rectangles of the form

$$B_k = \left[\xi_k' - \frac{L_k'}{2}, \xi_k' + \frac{L_k'}{2}\right] \times \left[-\frac{L_k''}{2}, \frac{L_k''}{2}\right]^{n-1}, \tag{2.1}$$

where the centers ξ_k', as well as the side lengths L_k' and L_k'', satisfy the parabolic scaling condition

$$\xi_k' \sim 2^k, \quad L_k' \sim 2^k, \quad L_k'' \sim 2^{k/2}, \quad \text{as } k \to \infty. \tag{2.2}$$

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 ν vary over a set of approximately $2^{k(n-1)/2}$ vectors, close to uniformly distributed over S^{n-1}. (We adhere to the convention that e_1 aligned with the ξ_1-axis belongs to the set of vectors ν, for all k.) Vectors ν 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

$$B_{\nu,k} = \Theta_{\nu,k}^{-1} B_k. \tag{2.3}$$

In particular, $\nu = \Theta_{\nu,k}^{-1} e_1$ for each k. The parameters ξ_k', L_k', L_k'', and ν are chosen so that the $B_{\nu,k}$ amply cover $\mathbb{R}^n$, in the sense that with L_k' and L_k'' multiplied by some fixed $r < 1$, the interiors would still cover $\mathbb{R}^n$.

2.1. Smith's curvelets. Following [20, 1], 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

$$|\langle \nu, \partial_\xi\rangle^j\, \partial_\xi^\alpha \hat{\chi}_{\nu,k}(\xi)| + |\langle \nu, \partial_\xi\rangle^j\, \partial_\xi^\alpha \hat{\beta}_{\nu,k}(\xi)| \leq C_{j,\alpha}\, 2^{-k(j+|\alpha|/2)}. \tag{2.4}$$

One then forms

$$\hat{\psi}_{\nu,k}(\xi) = \rho_k^{-1/2}\,\hat{\beta}_{\nu,k}(\xi), \qquad \hat{\varphi}_{\nu,k}(\xi) = \rho_k^{-1/2}\,\hat{\chi}_{\nu,k}(\xi), \tag{2.5}$$

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

$$\left.\begin{array}{l} |\varphi_{\nu,k}(x)| \\ |\psi_{\nu,k}(x)| \end{array}\right\} \le C_N 2^{k(n+1)/4}\,(\,2^k|\langle\nu,x\rangle| + 2^{k/2}|x|\,)^{-N} \tag{2.6}$$

for all N. Let $\{x_j^{\nu,k}\}$ denote a set of points in $\mathbb{R}^n$, depending on (ν,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 $x_j^{\nu,k}$, resulting in $\varphi_{\nu,k}(x - x_j^{\nu,k})$ and likewise for $\psi_{\nu,k}$. Introducing triplets $\gamma = (j,\nu,k)$, we get $\varphi_\gamma(x) = \varphi_{\nu,k}(x - x_j^{\nu,k})$, $\psi_\gamma(x) = \psi_{\nu,k}(x - x_j^{\nu,k})$ or [1]

$$\hat\varphi_\gamma(\xi) = \rho_k^{-1/2}\,\hat\chi_{\nu,k}(\xi)\,\exp[-2\pi\mathrm{i}\langle x_j^{\nu,k},\xi\rangle], \quad k \ge 1, \tag{2.7}$$

and similarly for $\hat\psi_\gamma(\xi)$. The translation factor $\exp[-2\pi\mathrm{i}\langle x_j^{\nu,k},\xi\rangle]$ is representative of a Fourier basis (with frequencies $x_j^{\nu,k}$) for functions of ξ. The (compact) support of $\hat\chi_{\nu,k}$ admits an orthonormal basis defining a Fourier series. We introduce the lattice

$$X_j := (j_1,\ldots,j_n) \in \mathbb{Z}^n, \tag{2.8}$$

and capture the scaling of B_k in the dilation matrix

$$D_k = \begin{pmatrix} L'_k & 0_{1\times n-1} \\ 0_{n-1\times 1} & L''_k I_{n-1} \end{pmatrix}.$$

Choosing

$$x_j^{\nu,k} = \Theta_{\nu,k}^{-1} D_k^{-1} X_j, \tag{2.9}$$

yields an orthogonal basis, $\exp[-2\pi\mathrm{i}\langle x_j^{\nu,k},\xi\rangle]$, for functions supported in $B_{\nu,k}$.

2.2. Transform pair. The normalized curvelet transform is given by

$$u_\gamma = \int u(x)\overline{\psi_\gamma(x)}\,\mathrm{d}x = \int \hat u(\xi)\overline{\hat\psi_\gamma(\xi)}\,\mathrm{d}\xi; \tag{2.10}$$

then

$$u(x) = \sum_\gamma u_\gamma \varphi_\gamma(x). \tag{2.11}$$

The $\psi_\gamma(x)$ are sometimes referred to as Riesz representers. In the discretization to be introduced below, it appears to be advantageous to introduce normalized curvelet coefficients, that is, if $\mathcal{N}$ is a judiciously chosen normalization constant, the transform pair is developed for $u'_\gamma := \mathcal{N}^{-n/2} u_\gamma$. For each fixed wedge indexed by ν, k, it holds true that

$$\sum_{\gamma':\ k'=k,\ \nu'=\nu} u_{\gamma'}\hat\varphi_{\gamma'}(\xi) = \hat u(\xi)\,\hat\beta_{\nu,k}(\xi)\,\hat\chi_{\nu,k}(\xi). \tag{2.12}$$

Upon substituting (2.7) into (2.10), we get

$$u_\gamma = \rho_k^{-1/2} \int \hat u(\xi)\hat\beta_{\nu,k}(\xi)\exp[2\pi\mathrm{i}\langle x_j^{\nu,k},\xi\rangle]\,\mathrm{d}\xi \tag{2.13}$$

[1] Our Fourier transform convention is: $\hat u(\xi) = \int u(x)\,\exp[-2\pi\mathrm{i}\langle x,\xi\rangle]\,\mathrm{d}x$; the inverse transform is $u(x) = \int \hat u(\xi)\,\exp[2\pi\mathrm{i}\langle x,\xi\rangle]\,\mathrm{d}\xi$.

and recognize that computation of u_γ can be interpreted as an inverse Fourier transform of $\hat{u}(\xi)\hat{\beta}_{\nu,k}(\xi)$ evaluated on the grid points $x_j^{\nu,k}$. Upon substituting (2.7) into (2.12), we get

$$\hat{u}(\xi)\hat{\beta}_{\nu,k}(\xi)\ \hat{\chi}_{\nu,k}(\xi) = \rho_k^{-1/2}\ \Big(\ \sum_j u_{j,\nu,k}\exp[-2\pi\mathrm{i}\langle x_j^{\nu,k},\xi\rangle]\ \Big)\ \hat{\chi}_{\nu,k}(\xi), \tag{2.14}$$

which represents the Fourier series expansion of $\hat{u}(\xi)\hat{\beta}_{\nu,k}(\xi)$. We refer to (2.13)-(2.14) as the "inner" transform pair. The transform is unitary in the case of a tight frame, when $\hat{\beta}_{\nu,k} = \hat{\chi}_{\nu,k}$. 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): $u_\gamma = (\mathrm{C}u)_\gamma$, and also define $\mathrm{C}^{-1}\{u_\gamma\} = \sum_\gamma u_\gamma\varphi_\gamma$ for the inverse transform (synthesis). It holds true that $\mathrm{C}^{-1}\mathrm{C} = I$ on $L^2(\mathbb{R}^n)$, and that $\mathrm{C}\mathrm{C}^{-1} \equiv \Pi$ is a projection operator of ℓ^2_γ onto the range of the analysis operator C; here,

$$\Pi_{\gamma'\gamma} = \langle\psi_{\gamma'},\varphi_\gamma\rangle. \tag{2.15}$$

Let d denote the pseudodistance on $S^*(X)$ introduced in [19, Definition 2.1]

$$d(x,\nu;x',\nu') \ = \ |\langle\nu,x-x'\rangle| + |\langle\nu',x-x'\rangle| + \min\{|x-x'|,|x-x'|^2\} + |\nu-\nu'|^2. \tag{2.16}$$

If $\gamma = (x,\nu,k)$ and $\gamma' = (x',\nu',k')$, let

$$\overline{d}(\gamma;\gamma') = 2^{-\min(k,k')} + d(x,\nu;x',\nu'). \tag{2.17}$$

The weight function $\mu_\delta(\gamma,\gamma')$ introduced in [20] is given by

$$\mu_\delta(\gamma,\gamma') = (1+|k'-k|^2)^{-1}2^{-(\frac{1}{2}n+\delta)|k'-k|}2^{-(n+\delta)\min(k',k)}\overline{d}(\gamma,\gamma')^{-(n+\delta)}.$$

Then Π satisfies the estimate,

$$|\Pi_{\gamma'\gamma}| \le C(\delta)\,\mu_\delta(\gamma',\gamma), \quad \text{for all } \delta>0. \tag{2.18}$$

3. Discrete transform pair. We begin with discretizing C defined by (2.10). We ensure that the discrete transform is unitary within a prescribed accuracy.

3.1. 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 $\mathrm{supp}(u) \subset \mathcal{D} = \{x\in\mathbb{R}^n \mid |x| < \frac{1}{2} = x_{\max}\}$, it is sufficient to sample $\hat{u}$ at integer points, $\xi_l = l \in \mathbb{Z}^n$, 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

$$X^N = \left\{ i\in\mathbb{Z}^n \ \middle|\ -\frac{N}{2} \le i_1 < \frac{N}{2},\dots,-\frac{N}{2}\le i_n<\frac{N}{2}\right\}.$$

The Fourier transform of u then reads

$$\hat{u}(\xi) = \frac{1}{N^n}\sum_{i\in X^N} u(x_i)\exp[-2\pi\mathrm{i}\langle\xi,x_i\rangle]. \tag{3.1}$$

By virtue of the unequally spaced FFT (USFFT), we can evaluate sums of this form rapidly, cf. Appendix A.

In the process of computing the curvelet coefficients, $\hat{u}$ is multiplied by windows $\hat{\beta}_{\nu,k}$ thus bandlimiting the original function according to scale k. A maximum scale considered, $k_{\max}$, defines

a maximum frequency $\xi_{\max}$. With c_0 the parameter defining the coarsest scale (that is, the units), we have the relationship,

$$\frac{\xi_{\max}}{c_0} = 2^{k_{\max}}. \tag{3.2}$$

Let

$$N = \lceil 2\xi_{\max} \rceil ;$$

the fundamental cell is then given by $\frac{N}{2}\frac{1}{2x_{\max}} = \frac{N}{2}$. N defines the sampling density in the space domain.

We now note that $u_\gamma = \rho_k^{-1/2}(u * \beta_{\nu,k})(x_j^{\nu,k})$. The function $u * \beta_{\nu,k}$ no longer has compact support. However, the fast decay of $\beta_{\nu,k}$ (cf. (2.6)) is inherited by $u * \beta_{\nu,k}$. Hence, u_γ will be very small for large $x_j^{\nu,k}$'s, and in practice it will be sufficient to keep and use only the lattice points located in a neighborhood of $\mathcal{D}$ for representation (2.11) to be accurate.

3.2. Inner products. A discretization of (2.13) leads to the approximation, $\tilde{u}_\gamma$, of u_γ,

$$\tilde{u}_\gamma = \frac{1}{\rho_k^{1/2}} \frac{\mathcal{N}^{n/2}}{\sigma_k'(\sigma_k'')^{n-1}} \sum_l \hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[2\pi \mathrm{i}\langle x_j^{\nu,k}, \xi_l^{\nu,k}\rangle], \tag{3.3}$$

applying the normalization of coefficients introduced below (2.11). We introduce the map, $\mathcal{G}_{\nu,k}$, according to $\mathcal{G}_{\nu,k}\{\hat{u}(\xi_l^{\nu,k})\} = \tilde{u}_\gamma$, $\gamma = (j,\nu,k)$, using (3.3).

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

$$\Xi^k = \left\{ l \in \mathbb{Z}^n \;\middle|\; -\frac{N_k'}{2} \le l_1 < \frac{N_k'}{2}, -\frac{N_k''}{2} \le l_2 < \frac{N_k''}{2}, \ldots, -\frac{N_k''}{2} \le l_n < \frac{N_k''}{2} \right\}. \tag{3.4}$$

The points in this set are denoted by Ξ_l^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, determining the accuracy of approximation (3.3) to the inverse Fourier transform [2]. We choose the $\xi_l^{\nu,k}$ (covering the box $B_{\nu,k}$) as

$$\xi_l^{\nu,k} = \Theta_{\nu,k}^{-1}\left(D_k S_k^{-1}\Xi_l^k + \xi_k' e_1\right). \tag{3.5}$$

The matrix S_k is defined as

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

The dot product in the phase of the exponential in (3.3) then becomes

$$\langle x_j^{\nu,k}, \xi_l^{\nu,k}\rangle = \left(D_k S_k^{-1}\Xi_l^k + \xi_k' e_1\right)^t D_k^{-1} X_j = \frac{j_1\xi_k'}{N_k'} + \left(\frac{j_1 l_1}{N_k'} + \frac{j_2 l_2 + \ldots + j_n l_n}{N_k''}\right). \tag{3.6}$$

Thus, the specific choice of points $\xi_l^{\nu,k}$ allows for a fast evaluation of $\tilde{u}_\gamma$ (cf. (3.3)) for $j \in \Xi^k$:

$$\tilde{u}_\gamma \exp(-2\pi \mathrm{i} j_1 \xi_k'/N_k') = \frac{\mathcal{N}^{n/2}}{[\sigma_k'(\sigma_k'')^{n-1}N_k'(N_k'')^{n-1}]^{1/2}} \sum_{l\in\Xi^k} \hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[2\pi \mathrm{i}\langle x_j, \xi_l\rangle], \tag{3.7}$$

[2] The error in the approximation $u_\gamma \approx \tilde{u}_\gamma$ can be estimated by the numerical integration error associated with the trapezoidal rule. The approximation error decays fast subject to the condition that $N_k' > L_k'$ and $N_k'' > L_k''$. The oversampling factors will need to be comparatively large for small k but decay as k grows due to the parabolic scaling property of the support of $\hat{\beta}_{\nu,k}$. Note that the oversampling factors are related to the number of points $(x_j^{\nu,k})$ where u_γ is evaluated, but not to their density, i.e. all extra points $(x_j^{\nu,k})$ appearing due to oversampling will be outside $\mathcal{D}$.

where $\xi_l = l$ and $x_j = S_k^{-1} j$ with $j \in \Xi^k$, while $1/(N_k'(N_k'')^{n-1}) = \det S_k^{-1}$. We note that the number of coefficients needed to reconstruct $\hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})$, $l \in \Xi^k$, is finite. (Compare this with equation (2.14) where we need infinitely many coefficients, cf. (2.8).) Indeed, we only encounter translations over

$$\tilde{x}_j^{\nu,k} = \Theta_{\nu,k}^{-1} D_k^{-1} \Xi_j^k, \tag{3.8}$$

defining a finite subset of the grid given in (2.9). One can use an n-dimensional FFT for the fast evaluation of $\tilde{u}_\gamma = \tilde{u}_{j,\nu,k}$ if the values for $\hat{u}$ and $\hat{\beta}_{\nu,k}$ are given at $\xi_l^{\nu,k}$.

The adjoint transform is built using the same sampling. A Fourier transformation of (2.11) gives

$$\hat{u}(\xi) = \sum_{\nu,k} \frac{1}{\rho_k^{1/2}} \Big(\sum_j \frac{u_{j,\nu,k}}{\mathcal{N}^{n/2}} \exp[-2\pi \mathrm{i} \langle x_j^{\nu,k}, \xi \rangle] \Big) \ \hat{\chi}_{\nu,k}(\xi).$$

The rapid decay of u_γ allows for a truncation of the above sums to sums over (3.8), that is,

$$\hat{u}(\xi) \approx \sum_{\nu,k} \frac{1}{\rho_k^{1/2}} \Big(\sum_{j \in \Xi^k} \frac{u_{j,\nu,k}}{\mathcal{N}^{n/2}} \exp[-2\pi \mathrm{i} \langle \tilde{x}_j^{\nu,k}, \xi \rangle] \Big) \ \hat{\chi}_{\nu,k}(\xi). \tag{3.9}$$

Replacing u_γ with the discrete approximation (3.3) and using (3.7) with (3.8) gives

$$\begin{aligned} \hat{u}(\xi) \approx \sum_{\nu,k} \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi) \sum_{j \in \Xi^k} \exp\Big[-2\pi \mathrm{i} (\langle \tilde{x}_j^{\nu,k}, \xi \rangle + j_1 \xi_k'/N_k'\Big] \\ (\sigma_k'(\sigma_k'')^{n-1} N_k'(N_k'')^{n-1})^{-1/2} \sum_{l \in \Xi^k} \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp\left[2\pi \mathrm{i} (l_1 j_1/N_k' + (l_2 j_2 + \ldots l_n j_n)/N_k'')\right] \end{aligned} \tag{3.10}$$

Now, if we evaluate the right hand side of (3.10) for a fixed pair (ν, k) at the points $\xi_m^{n,k}$, and using (3.6) we get

$$\begin{aligned} \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi_m^{n,k}) \sum_{j \in \Xi^k} \exp\left[-2\pi \mathrm{i} (m_1 j_1/N_k' + (m_2 j_2 + \ldots m_n j_n)/N_k''\right] \\ (\sigma_k'(\sigma_k'')^{n-1} N_k'(N_k'')^{n-1})^{-1/2} \sum_{l \in \Xi^k} \hat{u}(\xi_l^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[2\pi \mathrm{i} (l_1 j_1/N_k' + (l_2 j_2 + \ldots l_n j_n)/N_k'')] \\ = \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi_m^{n,k}) (\sigma_k'(\sigma_k'')^{n-1})^{-1/2} (N_k'(N_k'')^{n-1})^{1/2} \hat{u}(\xi_m^{\nu,k}) \hat{\beta}_{\nu,k}(\xi_m^{\nu,k}) \\ = \hat{u}(\xi_m^{\nu,k}) \hat{\chi}_{\nu,k}(\xi_m^{n,k}) \hat{\beta}_{\nu,k}(\xi_m^{\nu,k}), \end{aligned} \tag{3.11}$$

by virtue of the discrete Fourier transform.

Using (3.9) we may approximate $u(x)$ as

$$\begin{aligned} u(x) \approx \sum_{\nu,k} \int_{\mathbb{R}^n} \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi) \sum_{j \in \Xi^k} \frac{1}{\mathcal{N}^{n/2}} u_\gamma \exp(-2\pi \mathrm{i} \langle \tilde{x}_j^{\nu,k}, \xi \rangle) \exp(2\pi \mathrm{i} \langle x, \xi \rangle) \, d\xi \\ \approx \sum_{\nu,k} \sum_l \exp(2\pi \mathrm{i} \langle x, \xi_l^{\nu,k} \rangle) \frac{1}{\sigma_k'(\sigma_k'')^{n-1}} \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \sum_{j \in \Xi^k} \frac{1}{\mathcal{N}^{n/2}} u_\gamma \exp(-2\pi \mathrm{i} \langle \tilde{x}_j^{\nu,k}, \xi_l^{\nu,k} \rangle), \end{aligned} \tag{3.12}$$

where we have used the fact that $\hat{\chi}_{\nu,k}$ and $\hat{\beta}_{\nu,k}$ are smooth functions with compact support in the last approximation, in the same manner as in (3.3).

Using (3.11), we see that this, as expected, can be reformulated as

$$u(x) \approx \sum_{\nu,k} \frac{1}{\sigma_k'(\sigma_k'')^{n-1}} \sum_l \hat{u}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{n,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp(2\pi i\langle \xi_l^{\nu,k}, x\rangle),$$

Moreover, in the case where $\hat{\chi}_{\nu,k} = \hat{\beta}_{\nu,k}$, the right most part of (3.12), can be written

$$\frac{1}{\sigma_k'(\sigma_k'')^{n-1}} \rho_k^{-1/2} \hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \sum_{j\in\Xi^k} \frac{1}{\mathcal{N}^{n/2}} u_\gamma \exp(-2\pi \mathrm{i}\langle \tilde{x}_j^{\nu,k}, \xi_l^{\nu,k}\rangle) = \frac{1}{\mathcal{N}^n} \mathcal{G}_{\nu,k}^* u_\gamma,$$

yielding

$$u(x) \approx \sum_{\nu,k} \sum_l \exp(2\pi \mathrm{i}\langle x, \xi_l^{\nu,k}\rangle) \frac{N^{n/2}}{\mathcal{N}^n} \mathcal{G}_{\nu,k}^* u_\gamma. \tag{3.13}$$

In addition, note that

$$\mathcal{G}_{\nu,k}^* \mathcal{G}_{\nu,k} \hat{u}(\xi_l^{\nu,k}) = \mathcal{N}^n \hat{u}(\xi_l^{\nu,k}) \hat{\chi}_{\nu,k}^2(\xi_l^{\nu,k}). \tag{3.14}$$

Let us recapitulate the discrete transform pair. The discrete transform can be written as

$$\begin{aligned}(\mathrm{C}u)_\gamma = \mathcal{N}^{n/2} \frac{1}{\rho_k^{1/2}} \frac{1}{\sigma_k'(\sigma_k'')^{n-1}} \sum_l \\ \frac{1}{N^n} \sum_m u(x_m) \exp(-2\pi \mathrm{i}\langle x_m, \xi_l^{\nu,k}\rangle) \hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \exp[2\pi \mathrm{i}\langle x_j^{\nu,k}, \xi_l^{\nu,k}\rangle] \\ = \frac{\mathcal{N}^{n/2}}{N^n} \mathcal{G}_{\nu,k} \sum_m u_(x_m) \exp(-2\pi \mathrm{i}\langle x_m, \xi_l^{\nu,k}\rangle)\end{aligned} \tag{3.15}$$

On the other hand, using the reconstruction formula (3.13), we see that if we define

$$(\tilde{\mathrm{C}} u_\gamma)(x_m) = \sum_{\nu,k} \sum_l \exp(2\pi \mathrm{i}\langle x_m, \xi_l^{\nu,k}\rangle) \frac{N^{n/2}}{\mathcal{N}^n} \mathcal{G}_{\nu,k}^* u_\gamma, \tag{3.16}$$

then it is readily verified through (3.12) that

$$(\tilde{\mathrm{C}}\mathrm{C}u)(x_m) \approx u(x_m),$$

and it is also easy to see that $\tilde{\mathrm{C}} = \mathrm{C}^*$ with the choice $\mathcal{N} = N$. Hence, it follows that C is approximately unitary.

Smith's discrete curvelet transform and its adjoint (approximate inverse) can be summarized as

$$u(x_i) \xrightarrow{\text{USFFT}} \hat{u}(\xi_l^{\nu,k}) \xrightarrow{\text{Windowing}} \hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \xrightarrow{\text{IFFT}} \tilde{u}_\gamma,$$

and

$$u(x_i) \xleftarrow{\text{AdjointUSFFT}} \hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k})\hat{\chi}_{\nu,k}(\xi_l^{\nu,k}) \xleftarrow{\text{Windowing}} \hat{u}(\xi_l^{\nu,k})\hat{\beta}_{\nu,k}(\xi_l^{\nu,k}) \xleftarrow{\text{FFT}} \tilde{u}_\gamma,$$

respectively.

3.3. "Outer" transform, USFFT. In order to have a fast transform pair we will need to evaluate the sums in (3.15) and (3.16), pertaining to the "outer" transform pair, rapidly. Since the points $\xi_l^{\nu,k}$ are unequally spaced, we resort to the USFFT (see Appendix A). We summarize the key steps of the procedure here.

For each direction, let γ be a bump function, cf. Definition A.1. We will choose Gaussian functions as in [16]. Given a prescribed accuracy, ϵ, let $\lambda < -\pi/\log\epsilon$; we set

$$\gamma_\lambda(x) = \exp(-\lambda x^2).$$

We introduce

$$\Gamma^N(x) = \left[\sum_{i\in\mathbb{Z}^n} \hat\gamma(N^{-1}x_1 + i_1)^2 \cdots \hat\gamma(N^{-1}x_n + i_n)^2 \right]^{-1/2}; \tag{3.17}$$

we observe that $\Gamma^N > 0$ and that Γ^N is periodic with period, that is, $\Gamma^N(x_1 + N, \ldots, x_n + N) = \Gamma^N(x_1, \ldots, x_n)$. If $\hat\gamma$ is concentrated around zero, then $\Gamma^1(x) \approx \hat\gamma(x_1)^{-1} \cdots \hat\gamma(x_n)^{-1}$ as x is close to zero. We generate a cube of samples of size $(2N)^n$ by padding $u(x_i)$, $i \in X^N$ (cf. (3.1)) with zeros.

First, we multiply the regularly sampled function by Γ^{2N}:

$$U(x_i) = \Gamma^{2N}(x_i)\, u(x_i).$$

This multiplication will compensate for the (Fourier) smearing driving the interpolation below. Secondly, we apply an FFT to obtain $\hat U(\xi_l)$ with $\xi_l = l$, $l \in X^{2N}$ as in (3.1). Thirdly, we smear out the equally space transform onto the Fourier points $\xi_l^{\nu,k}$ by computing the convolutions

$$\hat U(\xi) = \sum_{\{l' :\ |2N\xi - l'| < d_\gamma(\epsilon)\}} \gamma(2N\xi_1 - l'_1) \cdots \gamma(2N\xi_n - l'_n)\, \hat U(\xi_{l'}), \quad \xi = \xi_l^{\nu,k},$$

where $d_\gamma(\epsilon)$ is defined in (A.10). 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, $\bar B_{\nu,k}$ say, of $B_{\nu,k}$ determined by $d_\gamma(\epsilon)$ to carry out the proper interpolation.

Conversely, to evaluate u on equally spaced points, x_j, from $\hat U$ 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,

$$\hat U(\xi_{l'}) = \sum_{\{l :\ |2\xi_l^{\nu,k} - l'| < d_\gamma(\epsilon)\}} \gamma(2\,(\xi_l^{\nu,k})_1 - l'_1) \cdots \gamma(2\,(\xi_l^{\nu,k})_n - l'_n)\, \hat U(\xi_l^{\nu,k}).$$

Secondly, we apply an inverse FFT to obtain $\hat U(x_j)$ with $x_i = (2N)^{-1} i$, $i \in X^{2N}$ as in (3.1). Thirdly, we compensate for the (Fourier) smearing by multiplying by Γ^{2N}:

$$u(x_i) = \Gamma^{2N}(x_i)\, U(x_i).$$

4. Tiling and partition of unity. Here we specify the general procedure of the tiling with boxes $B_{\nu,k}$ outlined in Section 2. In the radial direction, we have

$$|\xi'_k| = c_0 2^k, \quad k = 1, \ldots, k_{\max}, \tag{4.1}$$

cf. (3.2). Here, $k_{\max}$ is either given or can be estimated from the size, N^3, of the sampled data cube:

$$k_{\max} = \left\lceil \log_2 \left(\frac{N}{2c_0} \right) \right\rceil; \tag{4.2}$$

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

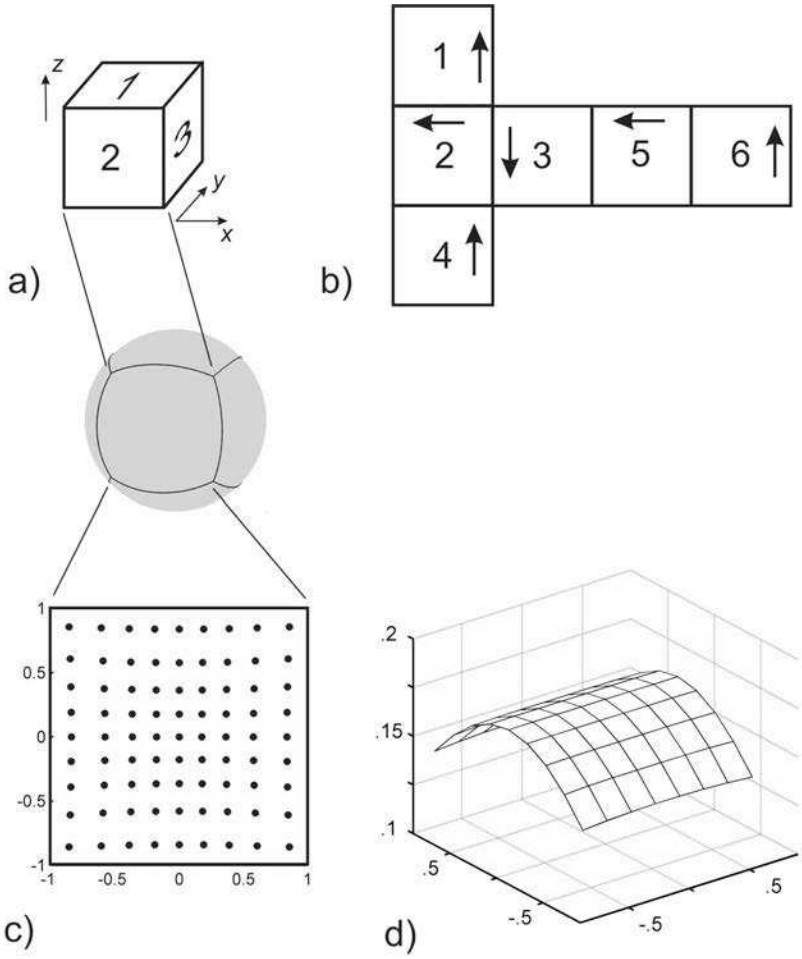

FIG. 1. *Defining grid on the unit sphere; cubed sphere construction for $n = 3$. (a) Cube in 3D; (b) unfolded cube, numbering of the sides, and local orientations (indicated by arrows) of the ω_2 vector in each sector. (c) Grid points on one side of the cube, for $N_{c,k} = 9$ (the grid is not equally spaced but is symmetric); (d) distance between neighboring points (in the "horizontal" direction) after mapping the grid points back onto the unit sphere (measure of non-uniformity of the cubed sphere grid).*

4.1. Cubed sphere. For each $k > 0$ we have to construct a set of rotated boxes (cf. (2.3)). We begin with introducing a set of points close to evenly distributed on the $(n-1)$-dimensional unit sphere that will define the ν for given k; each ν defines the radial orientation of a box, $B_{\nu,k}$. We use the cubed sphere [18] to define these points. The n-dimensional cube consists of $2n$ $(n-1)$-dimensional sides (see Figure 1 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 1 a), c)).

Let $\nu_{\max,k}$ 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 (2.2), we take

$$N_{c,k} = [c_1 2^{(k-1)/2}]_{odd}, \quad \nu_{\max,k} = 2n\,(N_{c,k})^{n-1} \sim 2^{k(n-1)/2}, \tag{4.3}$$

where $N_{c,k}$ is rounded to the nearest odd number in order for the basis vectors $e_1, \ldots, e_n$ to be contained in the set of orientations ν; c_1 is a parameter defining the number of orientations for $k = 1$. Then the number of orientations for $k > 1$ is defined by (4.3).

Grid points on one side of the cube are shown in Figure 1 c) (for $N_{c,k} = 9$). 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 2 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 1 c). In Figure 1 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 neigboring distance function as $d(k)$, and use it to define the half width of the boxes $B_{\nu,k}$.

4.1.1. Box orientations. Each box $B_{\nu,k}$ is a rectangular region in n-space and its orientation is defined by n orthogonal unit vectors, $\omega_1, \ldots, \omega_n$, 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 $\omega_1 \equiv \nu$. The orientations of the short sides are found from solving a set of n linear equations: Let ω_p for $p = 1, \ldots, q-1$ be given, then ω_q is a solution of

$$(\omega_q, \omega_p) = 0, \;\; (\omega_q, e_i) = 0, \quad p = 1, \ldots, q-1, \;\; i = 1, \ldots, n-q, \;\; q > 1, \tag{4.4}$$

subject to the normalization $|\omega_q| = 1$. For $n = 3$, writing $\omega_1 = (\omega_{11}, \omega_{12}, \omega_{13})^t$, our choice of directions follows to be

$$\begin{aligned} &\omega_1 \equiv \nu, \\ &\omega_2 = (0, \omega_{13}, -\omega_{12})^t \, \mathrm{sign}(\omega_{13}) / \sqrt{\omega_{12}^2 + \omega_{13}^2}, \\ &\omega_3 = \omega_1 \wedge \omega_2. \end{aligned} \tag{4.5}$$

Our choice of orientation for ω_2 is shown schematically in Fig. 1 b) (indicated by arrows). The vectors ω_2 and ω_3 define the orientations of squares centered at the end points of ω_1. One can amply cover the unit sphere with such squares, see Figure 2 b); these become the short sides of the boxes $B_{\nu,k}$. In Figure 3 we show the covering by squares away and near junctions of the different sectors, at two different scales.

4.1.2. Box sizes. For boxes belonging to scale k, the "central" frequency is given in (4.1) as $c_0 2^k$. The outer radius of boxes belonging to scale k is thus $c_0 2^{k+1}$, while the inner radius is $c_0 2^{k-1}$. The half box width is obtained by multiplying $d(k)$ with the outer radius; then

$$L_k'' = c_0 2^{k+2} d(k), \tag{4.6}$$

cf. (2.1). With the requirement that the box width is smaller then the inner radius, we arrive at the constraint, $2\pi < N_{c,1}$, on the minimal number of orientations for the first scale, $k = 1$.

We reconsider the boxes B_k ariented along the e_1-axis, as in (2.1). 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 4 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 ($\sim 2^{k-1}$) and L_k''. We obtain the box length,

$$L_k' = c_0 2^{k+1} - K(k), \quad K(k) = \sqrt{(c_0 2^k)^2 - (L_k''/2)^2}. \tag{4.7}$$

In Figure 5 we illustrate how the dyadic parabolic decomposition is built up following the construction above based on the cubed sphere.

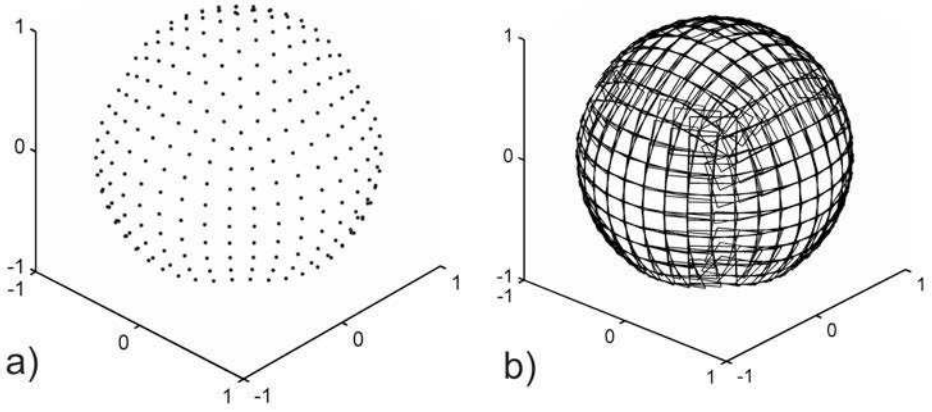

FIG. 2. *(a) Grid on the unit sphere. The points were mapped from three cube sides (see Figure 1 c)) back onto the sphere; (b) covering the sphere with squares (based on the grid points shown in a)).*

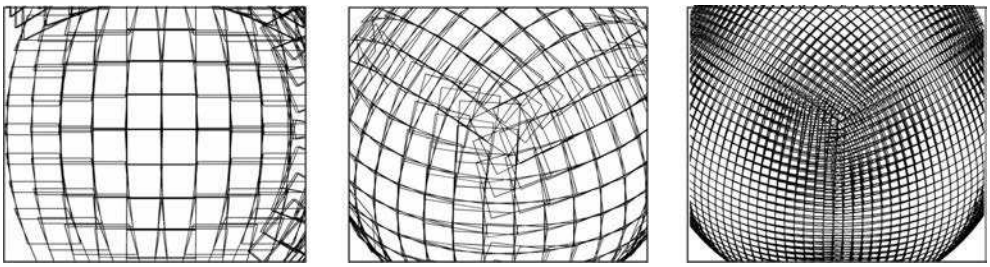

FIG. 3. *Covering of the unit sphere by squares, away from (left) and near (middle) junctions of the different sectors; right: same as the middle figure, but at a finer scale.*

In summary, the parameters controlling the tiling are: the dimension n, the number of samples N, c_0 defining the size of the cube associated with the coarsest scale (and thus the "units"), and $c_1 = N_{c,1}$ defining the number of orientations for $k = 1$.

4.2. Construction of window functions, Partition of unity. We construct the functions $\hat{\chi}_{\nu,k}$, and will choose, here, $\hat{\beta}_{\nu,k} = \hat{\chi}_{\nu,k}$ (the frame is tight). We adopt local orthogonal coordinates on each box, using basis vectors $\omega_1 = \omega_1(\nu, k), \ldots, \omega_n = \omega_n(\nu, k)$. The rotation matrix $\Theta_{\nu,k}$ in (2.3) maps $\omega_1(\nu, k), \ldots, \omega_n(\nu, k)$ to the global basis vectors $e_1, \ldots, e_n$. In terms of matrices,

$$\Theta_{\nu,k}^{-1} = (\omega_1(\nu, k), \cdots, \omega_n(\nu, k)) \,. \tag{4.8}$$

Let $\xi^{\nu,k}$ denote the (co)vector of coordinates of ξ with respect to the rotating basis $\omega_1 = \omega_1(\nu, k), \ldots, \omega_n = \omega_n(\nu, k)$; we will also write ξ 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 4 b).

We then introduce "polar" coordinates,

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

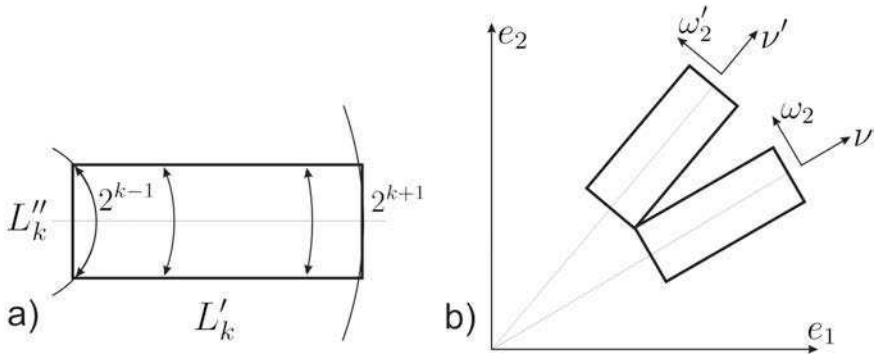

FIG. 4. *Box sizes and rotations. (a) Illustration of the calculation of $K(k)$; arrows indicating how the associated support function is stretched in polar coordinates to "fill out" the box; (b) local coordinate systems on the rotated boxes.*

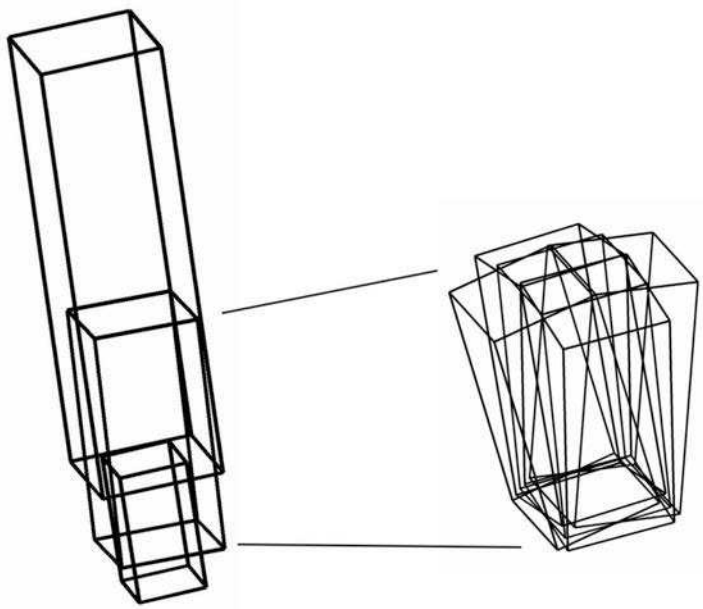

FIG. 5. *Building the dyadic parabolic decomposition for $n = 3$ using the cubed sphere; illustration for three successive scales.*

writing $\xi^{\nu,k} = (\xi_1^{\nu,k}, \ldots, \xi_n^{\nu,k})$.

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

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

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

Here, $a_m(r)$ is defined so that $w \in C_0^{m+1}$ (for details, see [1, Section 6.2]). For the illustrations in this paper, we take $m = 10$. This degree appears to be sufficient to display the appropriate decay properties within computational precision. Accompanying w is the coarsest scale ($k = 0$) function w_0, given by

$$w_0(r) = \cos\Big(\ \frac{\pi}{2} a_m(2r-1)\ \Big)\,,$$

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

For the transverse functions, we take

$$\kappa_k(r, \phi^\nu) = \cos\left(\frac{\pi}{2} a_m\left(\frac{r}{2^{k+1}} \frac{|\phi^\nu|}{\Delta\phi_k}\right)\right), \tag{4.9}$$

where

$$\Delta\phi_k = |\arctan(L_k''/2^{k+2})\,|.$$

We could also have taken tensor products of functions each defined on a (ν, ω_{p+1}) coordinate plane $(p = 1, \ldots, n-1)$. To accommodate this approach, we introduce polar coordinates (r, ϕ_p^ν) in each (ν, ω_{p+1})-plane $(p = 1, \ldots, n-1)$ with

$$\phi_p^\nu = \arcsin(\xi_{p+1}^{\nu,k}/r), \quad p = 1, \ldots, n-1.$$

Then

$$\kappa_k(r, \phi_1^\nu, \ldots, \phi_{n-1}^\nu) = \prod_{p=1}^{n-1} \cos\left(\frac{\pi}{2} a_m\left(\frac{r}{2^{k+1}} \frac{|\phi_p^\nu|}{\Delta\phi_k}\right)\right). \tag{4.10}$$

4.2.2. Partition of unity. The radial window functions satisfy the admissibility condition,

$$|w_0(r)|^2 + \sum_{k \ge 0} |w(2^{-k} r)|^2 = 1.$$

We let

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

then

$$\hat{\chi}_{\nu,k} = w(2^{-(k-1)} r)\, v_{\nu,k}(r, \phi^\nu), \quad k \ge 1. \tag{4.11}$$

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 $n \ge 3$ 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 (2.6) is modified from $(2^k|\langle \nu, x\rangle| + 2^{k/2}|x|)^{-N}$ to $(2^k|\langle \nu, x\rangle| + 4 \cdot 2^{k/2}|x|)^{-N}$. This effect, in practice, is significant in as much as it decreases the required oversampling factors σ_k''. However, the angular overlap between the windows (but not the total number of discretization points) increases. It is readily verified that $v_{\nu,k} \in C_0^{m+1}$, and that the resulting $\hat{\chi}_{\nu,k}$ satisfy (2.4) for $j, |\alpha| \le m+1$.

We illustrate the functions $\hat{\beta}_{\nu,k} = \hat{\chi}_{\nu,k}$ in Figure 6 for a ν away from a junction, and in Figure 7 for a ν near a junction.

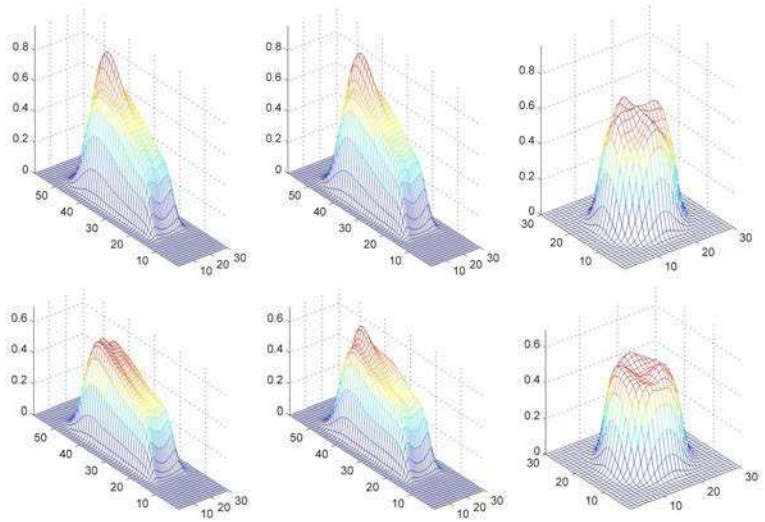

FIG. 6. *Illustration of $\hat{\chi}_{\nu,k}$, $k = 1$. We show three cross sections: the left two depict $\hat{\chi}_{\nu,k}$ keeping a coordinate with axis different from ν constant, while the rightmost depicts $\hat{\chi}_{\nu,k}$ keeping the coordinate in the ν-direction constant. The top row illustrates a window function in the most uniform region of the sphere, while the bottom row illustrates a window function where three sectors meet.*

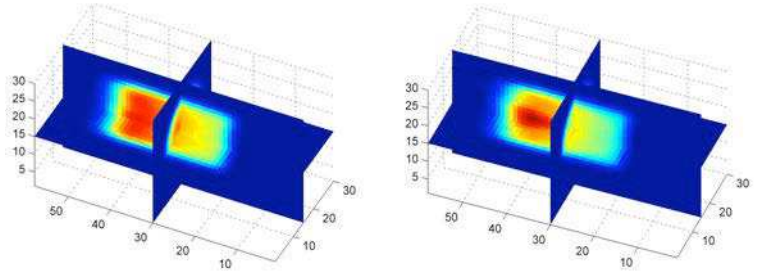

FIG. 7. *Illustration of $\hat{\chi}_{\nu,k}$, $k = 1$. We show three-dimensional views; the left figure corresponds with the bottom row of Figure 6 while the right figure corresponds with the top row of Figure 6.*

4.2.3. Finest scale. For the finest scale, we have $c_0 2^{k_{\max}} < \frac{N}{2}$, 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. One option is to decompose the residual function, that is, the inverse Fourier transform of the difference between the original, Fourier transformed function defined on the fundamental cell and the Fourier transformed function multiplied by the partition up to scale $k_{\max}$, further into wavelets. In our approach, we proceed with adding window functions $\beta_{\nu,k_{\max}+1} = \chi_{\nu,k_{\max}+1}$ that are collectively subjected to an anti-alias filter, here a (s-pole) Butterworth filter that optimizes "band flatness". Naturally, these finest scale curvelets will change shape significantly with ν.

4.2.4. Oversampling and local grids. Next, we evaluate $\hat{\beta}_{\nu,k} = \hat{\chi}_{\nu,k}$ on a regular grid, $\xi_l^{\nu,k}$, $l \in \Xi^k$, that is defined on the box $B_{\nu,k}$ as described by equations (3.5) and (3.4). Oversampling is related to the fact that the functions ψ_γ are not compactly supported. To analyze accuracy, let us assume a "test" oversampling described by factors $(\bar{\sigma}'_k, \bar{\sigma}''_k)$ and corresponding parameters $(\bar{N}'_k, \bar{N}''_k)$. The curvelet coefficient evaluation can be described by a convolution with $\rho_k^{-1/2}\beta_{\nu,k} = \psi_{\nu,k}$, cf. (2.13). This convolution is discretized through equation (3.3), with (N'_k, N''_k) now replaced by $(\bar{N}'_k, \bar{N}''_k)$. Using this discretization, via an inverse FFT, we obtain $\psi_{\nu,k}(\tilde{x}_j^{\nu,k})$ with $\tilde{x}_j^{\nu,k}$ given in

(3.8) and $j \in \bar{\Xi}^k$ (cf. (3.4) with (N'_k, N''_k) replaced by $(\bar{N}'_k, \bar{N}''_k)$ as before). A prescribed accuracy, $\epsilon > 0$, then determines the choice of $(\bar{N}'_k, \bar{N}''_k)$ through the condition

$$|\psi_{\nu,k}(\tilde{x}_j^{\nu,k})| < \epsilon \quad \text{for all } j \in \bar{\Xi}^k.$$

The smallest $\bar{N}'_k$ and $\bar{N}''_k$ satisfying this condition are $N'_k = N'_k(\epsilon)$ and $N''_k = N''_k(\epsilon)$, respectively. To obtain estimates for computational complexity one has to make the connection with estimate (2.6). In practice, we carry out the testing for $\psi_{\nu,k}$ with $\nu = e_1$ only.

In Figures 8-9 we show some examples of curvelets as they appear in our discrete transform pair.

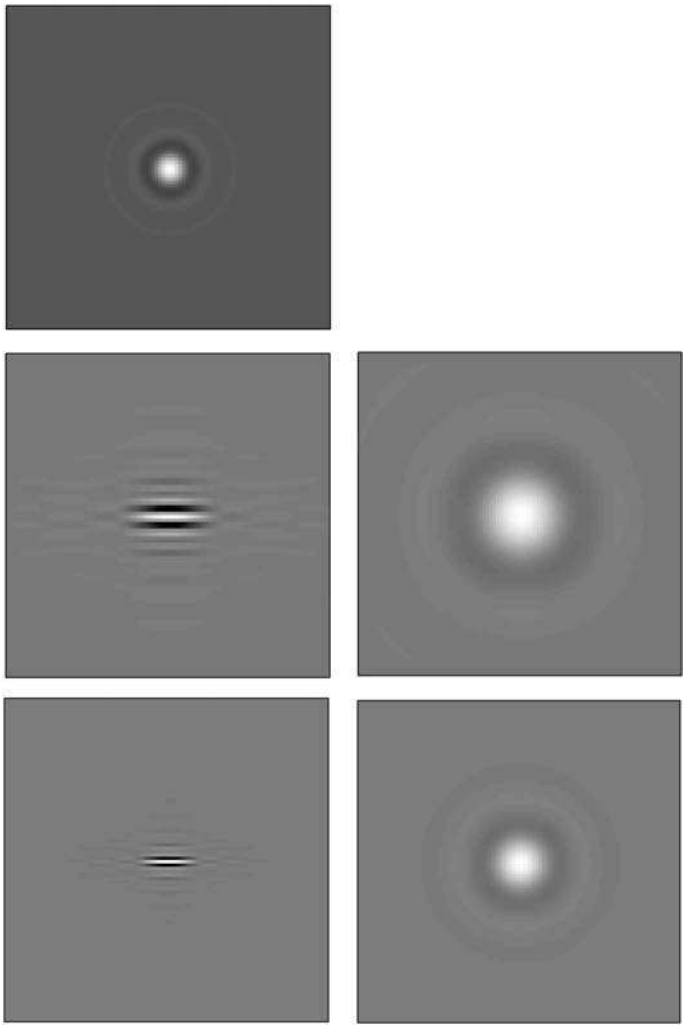

FIG. 8. *Smith curvelets: $k = 0$ (top), $k = 1$ (middle) and $k = 2$ (bottom); sections perpendicular to ν (right) and sections in a plane containing ν (left).*

5. Parallel algorithm. The typical initial data or image is large, and, hence, for the "outer" USFFT $u(x_i) \rightleftarrows \hat{u}(\xi_l^{\nu,k})$ we need a parallel algorithm. Further, we assume that all $B_{\nu,k}$ are small enough to fit a memory of a single CPU. Thus the "inner" FFT $\hat{u}(\xi_l^{\nu,k}) \rightleftarrows \tilde{u}_\gamma$ can be implemented locally and thus is straightforwardly parallelizable.

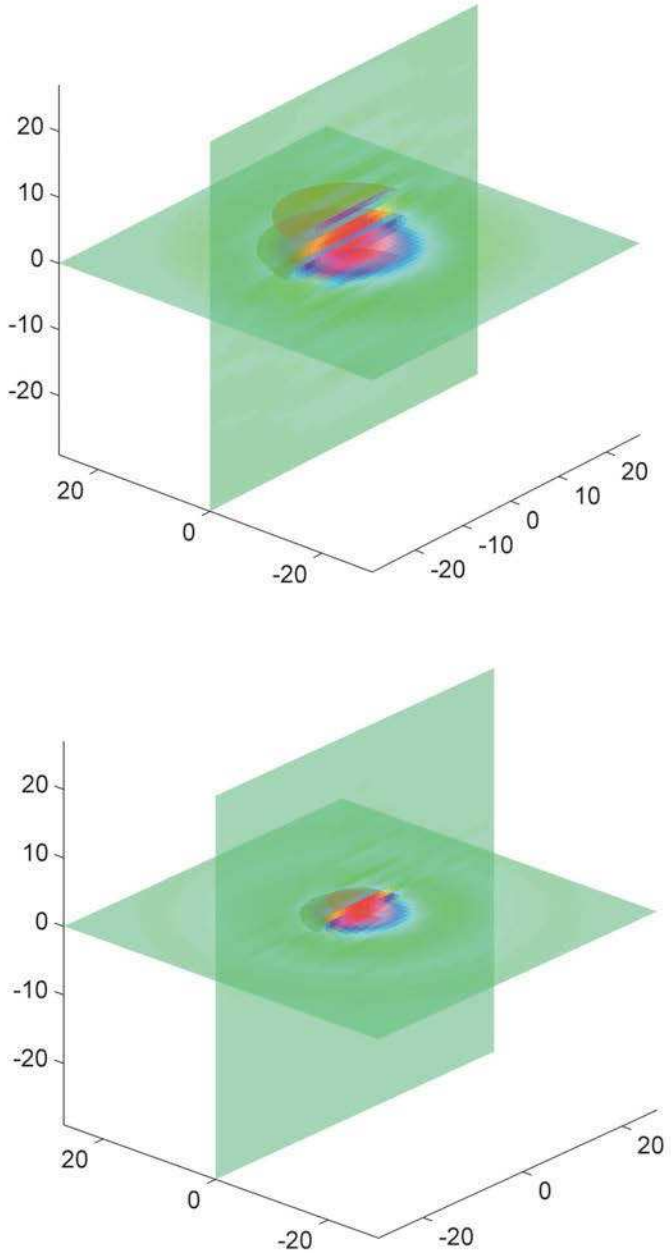

FIG. 9. *Smith curvelets (real parts): $k = 1$ (top) and $k = 2$ (bottom).*

The USFFT algorithm involves two main substeps: an FFT of the whole data cube and smearing out data from a global regular grid onto local rotated grids (or in the opposite direction). For the first substep we use a standard distributed-memory FFT implementation (the FFTW library) that involves communication (sending and receiving data) between CPUs while the output, $\hat{u}(\xi_l)$, will be distributed over different CPUs. The following procedure of smearing out data onto a local rotated grid $\xi_l^{\nu,k}$ will also need communication between CPUs, because points from the global grid ξ_l contained in the box $B_{\nu,k}$ will generically be stored on different CPUs (that is, a box $B_{\nu,k}$ is generically crossing the layout of $\hat{u}(\xi_l)$).

The smearing out procedure has no standard implementation and, in fact, can be organized

in different ways. In our implementation it will be carried out in a sequence of steps equal to the number of CPUs used. Consider a CPU that is working with box $B_{\nu,k}$. This CPU is storing a portion of the $\hat{u}(\xi_l)$ cube. We find points in this portion that are contained in the box $B_{\nu,k}$ and perform smearing out for those points storing the partial result. Then all $\hat{u}(\xi_l)$ coefficients are shifted to a neighboring CPU and the partial USFFT computation is repeated on a new portion of $\hat{u}(\xi_l)$. We proceed until each CPU had the entire $\hat{u}(\xi_l)$ array available for computations. Indeed, the smearing out is the most critical component from a performance viewpoint as it is the most expensive in terms of communication.

Appendix A. The USFFT.

We summarize the USFFT and its adaption to our discrete transform pair. For simplicity of notation we discuss the one-dimensional case.

A.1. Estimates and notation. DEFINITION A.1. *A non-negative function γ is said to be a bump function if it satisfies the following properties:*

$$\gamma(x) \quad \textit{is monotonically decreasing for } x > 0, \tag{A.1}$$

$$\frac{\hat{\gamma}(\xi+k)^2 + \hat{\gamma}(\xi-k)^2}{\hat{\gamma}(\xi+k)^2} \quad \textit{is monotonically increasing for } \xi > 0 \textit{ and } k \in \mathbb{Z}, \tag{A.2}$$

$$\hat{\gamma}(\xi) < \alpha^k, \quad 0 < \alpha < 1, \quad k \in \mathbb{Z}. \tag{A.3}$$

In the above, the constant α will typically be very small. If ϵ denotes the arithmetic precision, $\alpha \approx \epsilon^2$.

Let

$$\Gamma^M(\xi) = \Big(\sum_{j=-\infty}^{\infty} \hat{\gamma}^2\Big(\frac{\xi}{M} + j\Big)\Big)^{-\frac{1}{2}}, \tag{A.4}$$

and define the function φ through its Fourier transform by

$$\hat{\varphi}(\xi) = \Gamma^1(\xi)\hat{\gamma}(\xi). \tag{A.5}$$

LEMMA A.2. *The function φ defined by (A.5) is real-valued, even, and $\hat{\varphi}(\xi)$ is monotonically decreasing for $\xi > 0$.*

Proof. That φ is real and even follows from the fact that γ is real and even. To prove the statement of monotonicity, we note that

$$\hat{\varphi}(\xi) = \frac{\hat{\gamma}(\xi)}{\big(\sum_{k\in\mathbb{Z}} \hat{\gamma}(\xi+k)^2\big)^{\frac{1}{2}}} = \left(1 + \sum_{k\geq 1} \frac{\hat{\gamma}(\xi+k)^2 + \hat{\gamma}(\xi-k)^2}{\hat{\gamma}(\xi)^2}\right)^{-\frac{1}{2}}, \tag{A.6}$$

and that the denominator of (A.6) is monotonically increasing, since γ satisfies (A.2). ∎

LEMMA A.3. *Let $1 \leq p \leq \infty$, $f \in L^p(\mathbb{R})$ and $g \in \mathcal{S}$, where $\mathcal{S}$ denotes the Schwartz space. Then*

$$\sum_{k\in\mathbb{Z}} \hat{g}(\xi+k)\hat{f}(\xi+k) = \sum_{k\in\mathbb{Z}} (g * f)(k) e^{-2\pi i k\xi}, \tag{A.7}$$

Proof. From Poisson's summation formula,

$$\sum_{k\in\mathbb{Z}} f(\xi+k) = \sum_{k\in\mathbb{Z}} \hat{f}(k) e^{2\pi i k\xi}, \tag{A.8}$$

it follows that

$$\sum_{k\in\mathbb{Z}} \hat{g}(\xi+k)\hat{f}(\xi+k) = \sum_{k\in\mathbb{Z}} \widehat{\hat{g}h}(k)e^{2\pi ik\xi}.$$

Using that $\widehat{gf} = \hat{g} * \hat{f}$, and that $\hat{\hat{g}}(\xi) = g(-\xi)$, then yields the assertion of the lemma. □

We note that if f is bandlimited, $\mathrm{supp}(\hat{f}) \subset (-\frac{1}{2}, \frac{1}{2})$ say, then

$$\hat{f}(\xi) = \sum_{k\in\mathbb{Z}} \hat{f}(\xi+k)\chi_1(\xi+k). \tag{A.9}$$

By applying Lemma A.3, the Whittaker-Shannon sampling theorem follows. The next theorem deals with the case where the sinc function, corresponding with χ_1 in (A.9), is replaced by our constructed interpolation function, ϕ.

Let γ be a bump function satisfying the conditions of Definition A.1. We think of γ as a Gaussian (ignoring the fact that it does not have compact support). We denote by $d_\gamma(\epsilon)$ the smallest number such that $\gamma(x) < \epsilon$ if $|x| > d_\gamma(\epsilon)$, i.e.,

$$d_\gamma(\epsilon) = \inf_{|x|}\{x : \gamma(x) < \epsilon\}. \tag{A.10}$$

If the Gaussian is wide enough (which is quite narrow in practice), then the approximation

$$\int (\hat{f} * \gamma)(\xi)e^{2\pi i\xi x}\, d\xi \approx \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} (\hat{f} * \gamma)\left(\frac{k}{N}\right) e^{\frac{2\pi i\xi k}{N}} \tag{A.11}$$

will be sufficiently accurate. We emphasize that the required width of γ, in the above, depends on N. To compensate for the smearing effect caused by convolving with γ, we simply divide by $\hat{\gamma}$ in the Fourier domain.

Now, dividing by $\hat{\gamma}$ may seem strange and contradict intuition for what is allowed, since $\hat{\gamma}(x)$ typically becomes very small as $|x|$ increases. The key to avoiding that problem is to use oversampling. By using twice as many points in the trapezoidal rule as needed for the discrete (DFT) case, we can prevent $\hat{\gamma}(x)$ from becoming to small. We introduce the notation $\gamma_N(\xi) = \gamma(N\xi)$, and set

$$\Gamma^N(x) = \left(\sum_{j=-\infty}^{\infty} \hat{\gamma}^2(\frac{x}{N} + j)\right)^{-\frac{1}{2}}. \tag{A.12}$$

Note that $\Gamma > 0$ everywhere, and that if $\hat{\gamma}$ is concentrated around zero, then $\Gamma^1(x) \approx \hat{\gamma}(x)^{-1}$ when x is close to zero. It will prove more practical to work with Γ instead of $\hat{\gamma}^{-1}$ in obtaining error estimates. However, the difference between the two is small, and (depending on the sought for accuracy) Γ could in some cases be replaced by $\hat{\gamma}^{-1}$.

We collect the results of the discussion above in

THEOREM A.4. *For a given arithmetic precision ϵ, choose the bump function γ such that*

$$\frac{\hat{\gamma}(\frac{3}{4})}{\hat{\gamma}(\frac{1}{4})} < \epsilon \tag{A.13}$$

is satisfied. Suppose that $\hat{f}$ is either periodic or has support in a slightly narrower interval than $(-\frac{1}{2}, \frac{1}{2})$ (specifically such that $|(\gamma_{2N} * \hat{f})|(\xi)$ is below ϵ for $|\xi| > \frac{1}{2}$). Then we can accurately approximate f on the interval $\left[-\frac{N}{2}, \frac{N}{2}\right]$ by

$$\sup_{|x|\le\frac{N}{2}} \left| f(x) - \Gamma^{2N}(x) \sum_{l=-N}^{N-1} (\gamma_{2N} * \hat{f})\left(\frac{l}{2N}\right) e^{\frac{2\pi ilx}{2N}} \right| \le C\epsilon\|f\|_\infty, \tag{A.14}$$

where the constant C is independent of f.

REMARK A.5. *In the case where γ is a Gaussian, (A.13) explicitly becomes*

$$\lambda < \frac{\pi^2}{-2\log\epsilon}.$$

Proof. Let $g(x) = f\left(\frac{x}{N}\right)$. Instead of showing (A.14), it is equivalent to show that

$$\|\chi_{\frac{1}{2}}(\xi)(\hat{g}(\xi) - \Gamma^1(\xi)\sum_{k\in\mathbb{Z}}(\gamma * g)(k)e^{-2\pi ik\xi})\|_p \leq C\epsilon\|\hat{g}\|_p. \tag{A.15}$$

From the definition (A.5) of ϕ, and the fact that Γ^1 of (A.4) is 1-periodic, it follows that

$$\sum_{k\in\mathbb{Z}}\hat{\varphi}(\xi+k)\hat{g}(\xi+k) = \Gamma^1(\xi)\sum_{k\in\mathbb{Z}}\hat{\gamma}(\xi+k)g(\xi+k).$$

Using Lemma A.3, we note that

$$\sum_{k\in\mathbb{Z}}\hat{\varphi}(\xi+k)\hat{g}(\xi+k) = \Gamma^1(\xi)\sum_{k\in\mathbb{Z}}(\gamma * g)(k)e^{-2\pi ik\xi}. \tag{A.16}$$

Hence, by combining (A.16) with Minkowski's inequality, it holds true for the left hand side of (A.15) that

$$\begin{aligned}
&\|\chi_{\frac{1}{2}}(\xi)(\hat{g}(\xi) - \Gamma(\xi)\sum_{k\in\mathbb{Z}}(\gamma * g)(k)e^{-2\pi ik\xi})\|_p \\
&\leq \left(\int_{-\frac{1}{4}}^{\frac{1}{4}} |(1-\hat{\varphi}(\xi))\hat{g}(\xi)|^p\, d\xi\right)^{1/p} \\
&\quad + \sum_{k\in\mathbb{Z}\setminus\{0\}}\left(\int_{-\frac{1}{4}}^{\frac{1}{4}} |\hat{\varphi}(\xi+k)\hat{g}(\xi+k)|^p\, d\xi\right)^{1/p} \\
&\leq \left(\sup_{|\xi|\leq\frac{1}{4}}(1-\hat{\varphi}(\xi)) + 2\sum_{k=1}^{\infty}(\sup_{|\xi|\geq k-\frac{1}{4}}\hat{\varphi}(\xi))\right)\|\hat{g}\|_p \\
&\leq \left((1-\hat{\varphi}(\frac{1}{4})) + 2\sum_{k=0}^{\infty}\hat{\varphi}(\frac{3}{4}+k)\right)\|\hat{g}\|_p, \qquad \text{(A.17)}
\end{aligned}$$

where the monotonicity property of Lemma A.2 has been used in the last step.

Due to the 1-periodicity of Γ^1 the sum on the rightmost side of (A.17) can be rewritten as

$$\sum_{k=0}^{\infty}\hat{\varphi}(\frac{3}{4}+k) = \Gamma^1(\frac{3}{4})\sum_{k=0}^{\infty}\hat{\gamma}(\frac{3}{4}+k) \leq \frac{\sum_{k=0}^{\infty}\hat{\gamma}(\frac{3}{4}+k)}{\left(\sum_{k\in\mathbb{Z}}\hat{\gamma}(\frac{3}{4}+k)^2\right)^{\frac{1}{2}}} \tag{A.18}$$

Using the property (A.3) we can estimate (A.18) by

$$\sum_{k=0}^{\infty}\hat{\varphi}(\frac{3}{4}+k) \leq \frac{\hat{\gamma}(\frac{3}{4})\sum_{k=0}^{\infty}\alpha^k}{\left(\hat{\gamma}(\frac{1}{4})^2 + \sum_{k\in\mathbb{Z}\setminus\{-1\}}\hat{\gamma}(\frac{3}{4}+k)^2\right)^{\frac{1}{2}}} \leq \frac{\hat{\gamma}(\frac{3}{4})}{\hat{\gamma}(\frac{1}{4})}\sum_{k=0}^{\infty}\alpha^k \leq \frac{\epsilon}{1-\alpha}, \tag{A.19}$$

where the fact (A.13) is used in the last step. With regard to the term $(1-\hat{\varphi}(\xi))$ on the rightmost side of (A.17), we note that

$$1-\hat{\varphi}(\frac{1}{4}) = 1 - \frac{\hat{\gamma}(\frac{1}{4})}{\left(\sum_{k\in\mathbb{Z}}\hat{\gamma}(\frac{1}{4}+k)^2\right)^{\frac{1}{2}}} \leq \frac{\left(\sum_{k\in\mathbb{Z}}\hat{\gamma}(\frac{1}{4}+k)^2\right)^{\frac{1}{2}} - \hat{\gamma}(\frac{1}{4})}{\hat{\gamma}(\frac{1}{4})}.$$

Using the conjugate rule we obtain

$$1-\hat{\varphi}(\frac{1}{4}) \le \frac{\sum_{k\in\mathbb{Z}} \hat{\gamma}(\frac{1}{4}+k)^2 - \hat{\gamma}(\frac{1}{4})^2}{2\hat{\gamma}(\frac{1}{4})^2} = \frac{\sum_{k\ge 1} \hat{\gamma}(\frac{1}{4}+k)^2 + \hat{\gamma}(k-\frac{1}{4})^2}{2\hat{\gamma}(\frac{1}{4})^2}$$
$$\le \frac{\sum_{k\ge 0} \hat{\gamma}(\frac{3}{4}+k)^2}{\hat{\gamma}(\frac{1}{4})^2} = \frac{\hat{\gamma}(\frac{3}{4})^2}{\hat{\gamma}(\frac{1}{4})^2} \sum_{k\ge 0} \frac{\hat{\gamma}(\frac{3}{4}+k)^2}{\hat{\gamma}(\frac{3}{4})^2}.$$

By making use of (A.3) and (A.13) again, we obtain

$$1-\hat{\varphi}(\frac{1}{4}) \le \frac{\epsilon^2}{1-\alpha}. \tag{A.20}$$

Combining (A.19) and (A.20) finally yields the assertion of the theorem. □

A.2. Algorithms. USFFT: Evaluation of Fourier series at unequally spaced points

Data: Sample values f_j at equally spaced points $x_j = j$, $j = -\frac{N}{2}, \dots \frac{N}{2}$.

Output: An approximation of $\hat{\mathsf{f}}_\mathsf{k} = \sum_{\mathsf{j}=-\mathsf{N}/2}^{\mathsf{N}/2-1} \mathsf{f}_\mathsf{j} \mathsf{e}^{2\pi \mathsf{i}\mathsf{j}\xi_\mathsf{k}/\mathsf{N}}$ at unequally spaced Fourier points ξ_k, $k = 1, \dots, M$.

Step 1: Multiplying by Γ^{2N} in space:

$$\mathsf{h}_j = \begin{cases} \Gamma^{2N}(x_j)\mathsf{f}_j, & |j| < \frac{N}{2}; \\ 0, & \frac{N}{2} \le |j| \le N. \end{cases}$$

Step 2: Apply FFT to the equally spaced data $\hat{\mathsf{h}}$:

$$\hat{\mathsf{h}} = \mathrm{FFT}(\mathsf{h})$$

Step 3: Smear out the equally space data onto the Fourier points ξ_k by computing the convolutions

$$\hat{\mathsf{f}}_k = \sum_{\{l:|2N\xi_k - l| < d_\gamma(\epsilon)\}} \gamma(2N\xi_k - l)\hat{\mathsf{h}}_l$$

Adjoint USFFT: Evaluation at equally spaced points in space from unequally spaced Fourier data

Data: Sample values $\hat{\mathsf{f}}_k$ at unequally spaced points ξ_k, $k = 1, \dots, M$.

Output: An approximation of $\mathsf{f}_\mathsf{j} = \sum_{\mathsf{k}=-\mathsf{N}/2}^{\mathsf{N}/2-1} \mathsf{f}_\mathsf{k} \mathsf{e}^{2\pi \mathsf{i}\mathsf{j}\xi_\mathsf{k}}$.

Step 1: Smear out the data onto an equally spaced grid by computing the convolutions

$$\hat{\mathsf{h}}_l = \sum_{\{k:|2\xi_k - l| < d_\gamma(\epsilon)\}} \gamma(2\xi_k - l)\hat{\mathsf{f}}_\mathsf{k}, \quad -N \le l \le N-1.$$

Step 2: Apply (properly scaled) inverse FFT to the equally spaced data $\hat{\mathsf{h}}$:

$$\mathsf{h} = \mathrm{IFFT}(\hat{\mathsf{h}})$$

Step 3: Compensate for the (Fourier) smearing by multiplying by Γ^{2N} in space:

$$\mathsf{f}_j = \Gamma^{2N}(x_j)\mathsf{h}_j, \quad j = -\tfrac{N}{2}, \dots \tfrac{N}{2}.$$

A.3. Fast evaluation of Gaussians. We briefly discuss a splitting trick that can be employed to avoid intrinsic function calls in the implementation. It works similarly in one and higher

dimensions. See also [Greengard, accelerating the USFFT].

$$\begin{aligned}
\hat{\mathsf{f}}_k &= \sum_{\{l:|2\xi_k - l|<d_\gamma(\epsilon)\}} \gamma(2\xi_k - l)\hat{\mathsf{h}}_l \\
&= \sum_{\{l:|2\xi_k - l|<d_\gamma(\epsilon)\}} e^{-\lambda(2\xi_k - l)^2}\hat{\mathsf{h}}_l \\
&\approx \sum_{m=-\lfloor d_\gamma(\epsilon)\rfloor}^{\lfloor d_\gamma(\epsilon)\rfloor} e^{-\lambda(2\xi_k - [2\xi_k] - m)^2}\hat{\mathsf{h}}_l \\
&= \sum_{m=-\lfloor d_\gamma(\epsilon)\rfloor}^{\lfloor d_\gamma(\epsilon)\rfloor} e^{-\lambda((2\xi_k - [2\xi_k])^2 - 2m(2\xi_k - [2\xi_k]) + m^2)}\hat{\mathsf{h}}_l \\
&= a_k \sum_{m=-\lfloor d_\gamma(\epsilon)\rfloor}^{\lfloor d_\gamma(\epsilon)\rfloor} b_k^m c_m \hat{\mathsf{h}}_l,
\end{aligned}$$

where $a_k = e^{-\lambda(2\xi_k - [2\xi_k])^2}$, $b_k = e^{2\lambda(2\xi_k - [2\xi_k])}$, and $c_m = e^{-\lambda m^2}$.

REFERENCES

[1] F. Andersson, M.V. De Hoop, H.F. Smith, and G. Uhlmann, *A multi-scale approach to hyperbolic evolution equations with limited smoothness*, Comm. Partial Differential Equations, in print (2008).

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

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

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

[5] V. Brytik, M.V. De Hoop, and M. Salo, *Sensitivity analysis of wave-equation tomography: A multi-scale approach*, SIAM Review, (2008), p. in preparation.

[6] E.J. Candès, L. Demanet, D. Donoho, and L. Ying, *Fast discrete curvelet transforms*, SIAM Multiscale Model. Simul., 5-3 (2006), pp. 861–899.

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

[8] ———, *Continuous curvelet transform: I. Resolution of the wavefront set*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 162–197.

[9] ———, *Continuous curvelet transform: II. Discretization and frames*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 198–222.

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

[11] F.A. Dahlen, S.-H. Hung, and G. Nolet, *Fréchet kernels for finite-frequency traveltimes–I. Theory*, Geophys. J. Int., 141 (2000), pp. 157–174.

[12] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., LVII (2004), pp. 1413–1457.

[13] 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, (2008), p. in preparation.

[14] M.V. De Hoop and R.D. Van der Hilst, *On sensitivity kernels for 'wave-equation' transmission tomography'*, Geoph. J. Int., 160 (2005), pp. 621–633.

[15] H. Douma and M.V. De Hoop, *Leading-order seismic imaging using curvelets*, Geophysics, 72 (2007), pp. S231–S248.

[16] A. Dutt and V. Rokhlin, *Fast Fourier transforms for nonequispaced data*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1368–1393.

[17] F.J. Herrmann, P.P. Moghaddam, and C.C. Stolk, *Sparsity- and continuity-promoting seismic image recovery with curvelet frames*, Applied and Computational Harmonic Analysis, in print (2007).

[18] C. Ronchi, R. Ianoco, and P.S. Paolucci, *The "Cubed Sphere": a new method for the solution of partial differential equarions in spherical geometry*, J. Comput. Phys., 124 (1996), pp. 93–114.

[19] H.F. Smith, *A Hardy space for Fourier integral operators*, Jour. Geom. Anal., 8 (1998), pp. 629–654.

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

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