MULTISCALE MODEL. SIMUL.
Vol. 13, No. 1, pp. 1–42

FFRT: A FAST FINITE RIDGELET TRANSFORM FOR RADIATIVE TRANSPORT*

S. ETTER†, P. GROHS†, AND A. OBERMEIER†

Abstract. This paper introduces an FFT-based implementation of a fast finite ridgelet transform which we call FFRT. Inspired by recent work where it was shown that ridgelet discretizations of linear transport equations can be easily preconditioned by diagonal preconditioning, we use the FFRT for the numerical solution of such equations. Combining this FFRT-based method with a sparse collocation scheme, we construct a novel solver for the radiative transport equation which results in uniformly well-conditioned linear systems.

Key words. adaptive frame methods, radiative transport, ridgelets

AMS subject classifications. 35Q79, 42C15, 42C40, 65N35, 65T99

DOI. 10.1137/140977722

1. Introduction.

In the past two decades, applied harmonic analysis has had a big impact on image processing, computer science, and applied mathematics, primarily through the introduction of a wide range of multiscale systems, breaking ground with wavelets [10, 7] and continuing with ridgelets [2, 14], curvelets [6, 5, 4], shearlets [22, 21, 23, 20], contourlets [12], etc—the latter three of which fall into the framework of parabolic molecules [18], while all of the mentioned systems are encompassed by the even broader framework of α-molecules [17].

These systems share the property that they are very well adapted to representing certain classes of functions optimally (in the sense of the decay rate of the best N-term approximation)—functions with point singularities for wavelets, line singularities for ridgelets, and curved singularities for parabolic molecules. Since these classes make up the fundamental phenomenological features of most images in an extremely diverse set of applications, it is perhaps not surprising that many of the above-mentioned systems were originally investigated in view of their properties regarding image processing.

With a certain time lag, it is becoming apparent that these systems are also very suitable for solving partial differential equations (PDEs); again, wavelets were the first in this regard, for example, leading to provably optimal solvers for elliptic equations [8]. For differential equations with strong directional features—such as transport equations—it is intuitively clear that optimal solvers will need to take these features into account; however, the development of solvers based on directional systems is still in its infancy.

Following recent results [16] that ridgelets permit the construction of simple diagonal preconditioners for linear transport equations which arise in collocation-type discretization methods for kinetic transport equations (such as radiative transport), we intend this paper to be a further step toward establishing directional representation systems as a useful tool for solving PDEs.

Previous work in this direction includes [3], where curvelets are used to sparsify

*Received by the editors July 17, 2014; accepted for publication (in revised form) October 14, 2014; published electronically January 6, 2015.
http://www.siam.org/journals/mms/13-1/97772.html
†Seminar for Applied Mathematics, ETH Zürich, Zürich, Switzerland (ettersi@student.ethz.ch, philipp.grohs@sam.math.ethz.ch, axel.obermeier@sam.math.ethz.ch). The third author was supported by SNF project 146356.

the propagator matrix for linear wave equations with smooth coefficients, which is, however, quite a different problem from the stationary equation we consider. Furthermore, the paper [9] introduces an anisotropic bisection method and applies it to linear transport equations with possibly nonconstant convection fields. The reported numerical experiments exhibit an optimal convergence rate of N^{-1} in the L^2-norm for piecewise smooth solutions with curved singularities. A rigorous theoretical analysis of the method in [9] is, however, not yet available.

1.1. Toward establishing directional systems as a tool in numerical analysis. Perhaps the main reason for the success of wavelets in PDE solvers (which, as a long term goal, we would like to emulate) is that not only do they represent typical solutions efficiently, but—crucially—that they also simultaneously sparsify (in a suitable sense) the resulting system matrices corresponding to the differential operator and achieve uniformly well-conditioned matrices with simple preconditioning. The same is true for ridgelets with regard to transport equations; a detailed and rigorous treatment of these results is forthcoming in a separate paper [19].

The present paper serves two purposes: First we introduce an implementation of a fast FFT-based finite ridgelet transform FFRT (fast finite ridgelet transform). While previous public available implementations (cf. [11]) of the ridgelet transform have been based on the Radon transform, our implementation can exploit existing, fast FFT-algorithms and is conceptually simpler.

The second purpose of this paper is to use the FFRT for the numerical solution of kinetic transport equations arising in radiative transport. Using the preconditioner from [16] for linear transport equations together with a sparse discrete ordinates method similar to [15], we construct a solver which mitigates the curse of dimensionality and which results in uniformly well-conditioned linear systems which can be solved efficiently with conjugate gradiants (CGs).

1.2. Radiative transport equation. This paper is concerned with the numerical solution of the following model, described by the radiative transport equation (RTE),

$$\mathcal{A}u := \vec{s} \cdot \nabla u + \kappa\, u = f + \int_{\mathbb{S}^{d-1}} \sigma u \, \mathrm{d}\vec{s}'. \tag{1.1}$$

It is a steady state continuity equation describing the conservation of radiative intensity in an absorbing, emitting, and scattering medium; see, e.g., [25]. Let us assume that the following quantities are known at all locations $\vec{x} \in \Omega \subset \mathbb{R}^d$ and for all directions $\vec{s} \in \mathbb{S}^{d-1} := \left\{\vec{s} \in \mathbb{R}^d \,\middle|\, \|\vec{s}\|_2 = 1\right\}$:

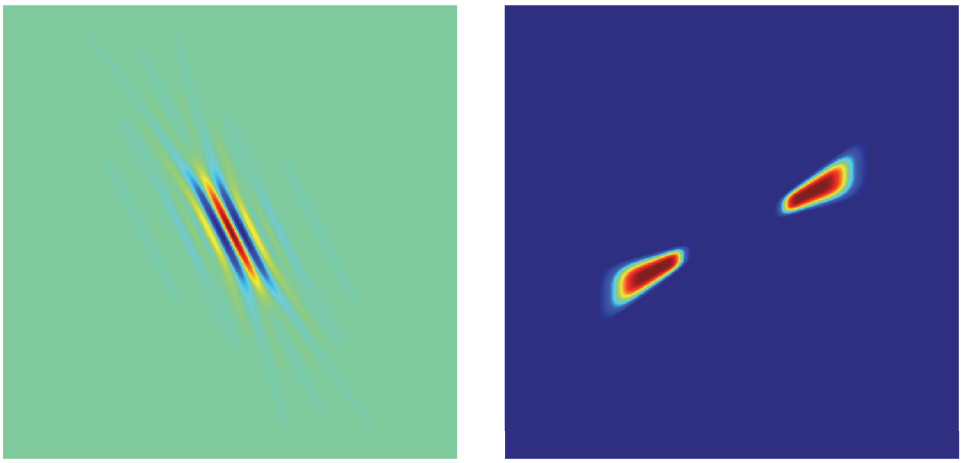
- absorption coefficient $\kappa(\vec{x}, \vec{s}) \geq \kappa_0 > 0$;
- source term $f(\vec{x}, \vec{s}) \in \mathbb{R}$;
- scattering kernel $\sigma\big(\vec{x}, \vec{s}, \vec{s}'\big) \in \mathbb{R}$.

Then the above equation allows us to find the unknown radiative intensity u as a function $\Omega \times \mathbb{S}^{d-1} \to \mathbb{R}$, which is the problem at the heart of this paper.

Although the RTE looks simple, standard numerical techniques for solving it do not perform well for two reasons:

- The transport term $s \cdot \nabla u$ leads to ill-conditioned systems of equations.
- With the dimension of the domain of u being 3 in 2-dimensional physical space and 5 in 3-dimensional space, the problem is fairly high-dimensional.

Both of these points make the accurate numerical solution of the RTE very costly or even impossible due to the memory and computational power limitations of today's hardware.



(a) Physical space (green denotes 0). (b) Fourier space (blue denotes 0).

FIG. 1. *An illustration of a ridgelet in the two relevant spaces (color available online).*

1.3. Ridgelets. Our proposed approach to solving (1.1), while addressing the above-mentioned problems, is to discretize the equation in physical space using ridgelets. At a glance, a ridgelet is a function which is located along a line, orthogonal to which it oscillates heavily and along which it varies only a little (see Figure 1(a) for an example). The idea is to build a basis (or rather, a frame) out of such ridgelets with varying locations, directions, and widths, with which we can represent a function whose features are located along curves by a linear combination of relatively few of them. Solutions of the RTE typically fall into this category of functions that can be efficiently represented by such a system, as the variations along the transport direction are smoothed out while the ones orthogonal to it are not—in particular, singularities in the input data may remain.

The above heuristic in physical space becomes less ambiguous by looking at it in Fourier space. In fact, the ridgelets are constructed from a particular partition of unity in Fourier space (in the same way that wavelets can be constructed from another partition of unity in Fourier space). There, a ridgelet is located at some line along which it has a length of $\mathcal{O}(2^j)$, while orthogonal to it its width is bounded by some constant independent of j (see Figure 1(b)). By discretizing in the right way, one can even make use of the FFT for the correspondence between physical and Fourier space, which will greatly accelerate the procedure.

1.4. Outline. We start with a brief overview of the model problem and ridgelets in the following subsections, before introducing several conventions related to notation and the definition of the finite Fourier transform in section 2. Our construction of the FFRT is described in detail in section 3, including proofs of complexity.

The subsequent section 4 discusses the applicability of ridgelets for the discretization of radiative transport equations. In particular we study a discrete ordinates method together with a source iteration scheme to include scattering. In addition, we cover convergence results and treatment of boundary conditions. The final section, section 5, reports numerical experiments.

At http://www.math.ethz.ch/∼pgrohs/FFRT/, the MATLAB codes of our im-

plementation can be downloaded freely.

2. Preliminaries.

2.1. Discrete intervals. Throughout this document, we will be working with functions sampled on equispaced grids over an interval. Therefore, it is useful to have some notation at hand to easily specify such a grid, which we call a discrete interval.

DEFINITION 2.1 (discrete intervals). *Let $a, b, c \in \mathbb{R}$ and $a < c$, $\frac{c-a}{b} \in \mathbb{N}_0$. Then*

$$[a : b : c] := \{a, a + b, \dots, c - b, c\}.$$

If b is omitted, then $b = 1$ is meant. Both the opening and the closing square brackets can be replaced by parentheses, and if done so, the last point at the respective end is excluded from the set.

Example 1. $[0{:}0.5{:}2] = \{0, 0.5, 1, 1.5, 2\}$, whereas $[0{:}0.5{:}2) = \{0, 0.5, 1, 1.5\}$.

2.2. Sobolev spaces. Sobolev spaces are an important tool for measuring the smoothness of a function. For our purposes, we will need a definition which is slightly different from the one usually given.

DEFINITION 2.2 ((periodic) Sobolev spaces). *Let $k \in \mathbb{N}$, $L_x, L_y \in \mathbb{R}^+$, and $\Omega = [0, L_x) \times [0, L_y)$ be some rectangular domain. Then the* Sobolev space H^k *is defined as*

$$H^k(\Omega) := \left\{ f \in V \;\middle|\; \int_\Omega \left| \frac{\partial^{k_x+k_y} f}{\partial^{k_x} x \partial^{k_y} y} \right|^2 \mathrm{d}x\,\mathrm{d}y < \infty \quad \forall k_x, k_y \in \mathbb{N},\, k_x + k_y \le k \right\},$$

where V is the subspace of $L^2(\Omega)$ of periodic functions (L_x-periodic in x and L_y-periodic in y).

For convenience we include the condition that our functions be periodic directly into the definition of the Sobolev spaces.

For characterizing Sobolev spaces (and their duals), as well as for other purposes, we introduce the *regularized absolute value*

$$\langle x \rangle := \sqrt{1 + |x|^2},$$

with which we can write

$$H^k(\widehat{\Omega}) := \left\{ \hat{f} \in L^2(\widehat{\Omega}) \;\middle|\; \left\langle \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} \right\rangle \hat{f}(\hat{x}, \hat{y}) \in L^2(\widehat{\Omega}) \right\}.$$

2.3. Fourier transform. Due to the many different versions of the Fourier transform, we introduce our notation and list the most important properties for easy reference.

DEFINITION 2.3 (Fourier transform). *Let $L_x, L_y \in \mathbb{R}^+$, $\Omega = [0, L_x) \times [0, L_y)$ be some rectangular domain and $f \in L^1(\Omega)$ a function. Then the function $\mathcal{F}[f] : \mathbb{Z}^2 \to \mathbb{C}$ given by*

$$\mathcal{F}[f](\hat{x}, \hat{y}) = \frac{1}{L_x L_y} \int_\Omega f(x, y)\, \mathrm{e}^{-2\pi \mathrm{i}\,(\hat{x}\frac{x}{L_x} + \hat{y}\frac{y}{L_y})}\, \mathrm{d}x\,\mathrm{d}y$$

is called the Fourier transform *of f.*

DEFINITION 2.4 (inverse Fourier transform). *Let L_x, L_y, and Ω be as above and let $\hat{f} \in \ell^1(\mathbb{Z}^2)$ (interpreted as a function). Then the function $\mathcal{F}^{-1}[\hat{f}] : \Omega \to \mathbb{C}$ given by*

$$\mathcal{F}^{-1}[\hat{f}](x,y) = \sum_{(\hat{x},\hat{y})\in\mathbb{Z}^2} \hat{f}(\hat{x},\hat{y})\, \mathrm{e}^{2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x}+\hat{y}\frac{y}{L_y})}$$

is called the inverse Fourier transform *of $\hat{f}$. To interpret this function for values outside of Ω, it is extended periodically (which is consistent with the definition). This is implicitly assumed in the rest of this article.*

FACT 2.5.

- *The inverse Fourier transform as defined lives up to its name:*

$$\begin{aligned} &\mathcal{F}^{-1}\mathcal{F}f = f \text{ for } f \in L^1(\Omega) \cap \mathcal{F}^{-1}(\ell^1(\mathbb{Z}^2)),\\ &\mathcal{F}\mathcal{F}^{-1}\hat{f} = \hat{f} \text{ for } \hat{f} \in \ell^1(\mathbb{Z}^2) \cap \mathcal{F}(L^1(\Omega)). \end{aligned} \tag{2.1}$$

- *The correspondence between translation and modulation behaves as usual:*

$$\mathcal{F}[f(\cdot - t_x, \cdot - t_y)](\hat{x},\hat{y}) = \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}\, \mathcal{F}[f](\hat{x},\hat{y}). \tag{2.2}$$

- *The Plancherel theorem holds:*

$$\int_\Omega f(x,y)\,\overline{g(x,y)}\,\mathrm{d}x\,\mathrm{d}y = L_x L_y \sum_{(\hat{x},\hat{y})\in\mathbb{Z}^2} \mathcal{F}[f](\hat{x},\hat{y})\,\overline{\mathcal{F}[g](\hat{x},\hat{y})}. \tag{2.3}$$

2.4. Finite Fourier transform. While the above definition of the Fourier transform provides a solid basis for theoretical work, it is useless if one wants to use it on a finite computing machine because of the integrals and the series. Therefore, what is needed is a finitely computable approximation to the ideal Fourier transform.

DEFINITION 2.6 (finite Fourier transform). *Let $L_x, L_y \in \mathbb{R}^+$, $N_x, N_y \in \mathbb{N}$,*

$$\Omega_{\mathrm{fin}} = \left[0 : \frac{L_x}{N_x} : L_x\right) \times \left[0 : \frac{L_y}{N_y} : L_y\right)$$

be an equispaced rectangular grid, let

$$\widehat{\Omega}_{\mathrm{fin}} = \left[-\left\lceil\frac{N_x-1}{2}\right\rceil : \left\lfloor\frac{N_x-1}{2}\right\rfloor\right] \times \left[-\left\lceil\frac{N_y-1}{2}\right\rceil : \left\lfloor\frac{N_y-1}{2}\right\rfloor\right]$$

be a part of the Fourier space, and let $f : \Omega_{\mathrm{fin}} \to \mathbb{C}$ be a function. Then the function $\mathtt{ft}[f] : \widehat{\Omega}_{\mathrm{fin}} \to \mathbb{C}$ given by

$$\mathtt{ft}[f](\hat{x},\hat{y}) = \frac{1}{N_x N_y} \sum_{(x,y)\in\Omega_{\mathrm{fin}}} f(x,y)\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x}+\hat{y}\frac{y}{L_y})}$$

is called the finite Fourier transform *of f.*

It is worth noting that `ft` corresponds to a trapezoidal rule approximation of the integral in the definition of $\mathcal{F}$.

DEFINITION 2.7 (inverse finite Fourier transform). *Let L_x, L_y, N_x, N_y, Ω_{fin}, and $\widehat{\Omega}_{\mathrm{fin}}$ be as above and let $\hat{f} : \widehat{\Omega}_{\mathrm{fin}} \to \mathbb{C}$ be a function. Then the function $\mathtt{ift}[\hat{f}] : \Omega_{\mathrm{fin}} \to \mathbb{C}$ given by*

$$\mathtt{ift}[\hat{f}](x,y) = \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\mathrm{fin}}} \hat{f}(\hat{x},\hat{y})\, \mathrm{e}^{2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x}+\hat{y}\frac{y}{L_y})}$$

is called the inverse finite Fourier transform *of $\hat{f}$.*

Instead of restricting $\texttt{ift}[\hat{f}]$ to Ω_{fin}, the above formula would also allow us to define $\texttt{ift}[\hat{f}]$ as a function $\Omega = [0, L_x) \times [0, L_y) \to \mathbb{C}$. In that case, we would have

$$\texttt{ift}[\hat{f}](x, y) = \mathcal{F}^{-1}\big[\mathcal{Z}[\hat{f}]\big] \quad \forall (x, y) \in \Omega, \tag{2.4}$$

where $\mathcal{Z}$ is defined as follows.

DEFINITION 2.8 (zero padding operator). *Let $\widehat{\Omega}_{\text{fin}}$, $\hat{f}$ be as in Definition 2.7. Then $\mathcal{Z}[\hat{f}] : \mathbb{Z}^2 \to \mathbb{C}$ is a function defined by*

$$\mathcal{Z}[\hat{f}](\hat{x}, \hat{y}) = \begin{cases} \hat{f}(\hat{x}, \hat{y}), & (\hat{x}, \hat{y}) \in \widehat{\Omega}_{\text{fin}}, \\ 0 & \textit{otherwise}, \end{cases}$$

and the symbol $\mathcal{Z}$ is called the zero padding operator*.*

Later on, we will also need an operator undoing the effect of $\mathcal{Z}$, which for symmetry we introduce here.

DEFINITION 2.9 (truncation operator). *Let $\widehat{\Omega}_{\text{fin}}$ be as in Definition 2.6, and let $\hat{f} : \mathbb{Z}^2 \to \mathbb{C}$. Then $\mathcal{T}[\hat{f}] : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$ is a function defined by*

$$\mathcal{T}[\hat{f}](\hat{x}, \hat{y}) = \hat{f}(\hat{x}, \hat{y}),$$

and the symbol $\mathcal{T}$ is called the truncation operator*.*

The advantage of defining `ift` the way we did is that only in this case is the transform truly finitely computable and all of the later statements about properties of the finite Fourier transforms correct. On the other hand, for the purpose of error estimation it will be useful to take the continuous viewpoint (2.4).

FACT 2.10.

- *The defined transformations are inverse to each other,*

  $$\texttt{ft} \circ \texttt{ift} = \texttt{ift} \circ \texttt{ft} = \mathbb{I},$$

  which is an immediate consequence of the summation property of roots of unity,

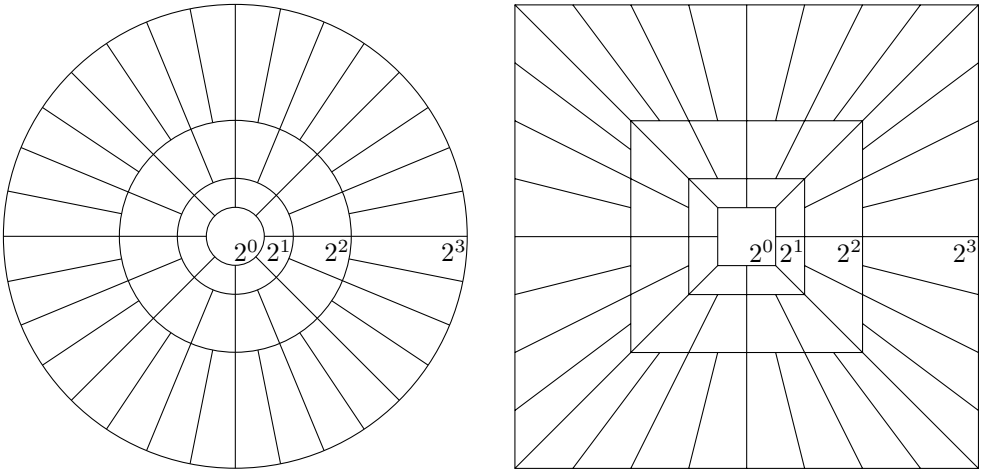
  $$\sum_{k=0}^{N-1} \mathrm{e}^{2\pi \mathrm{i}\, j \frac{k}{N}} = N\, \delta(j \bmod N) \quad \forall j \in \mathbb{Z}. \tag{2.5}$$

- `ft` *and* `ift` *can be computed using the* discrete Fourier transform*, more precisely,*

  $$\texttt{ft}[f](\hat{x}, \hat{y}) = \frac{1}{N_x N_y} \texttt{DFT}[F], \quad \textit{where} \quad F_{k_x, k_y} := f\left(L_x \frac{k_x}{N_x}, L_y \frac{k_y}{N_y} \right)_{\substack{k_x = 0, \dots, N_x - 1 \\ k_y = 0, \dots, N_y - 1}}.$$

- *As an immediate consequence of being able to calculate* `ft` *and* `ift` *using* `DFT`*, the complexity of evaluating them is $\mathcal{O}(N_x N_y \log(N_x N_y))$.*

3. Ridgelets. As briefly sketched in the introduction, the construction of the ridgelets is done via an appropriately chosen partition of unity in Fourier space (inspired by Littlewood–Paley-type dyadic partitions). In [16], this was based on a spherically symmetric partition (see Figure 2(a)), with scaling and rotation as the transformations relating the different elements to each other. Numerically, shearing



(a) Rotational partition from [16].

(b) Sheared partition used for implementation.

FIG. 2. *Two different partitions for ridgelets in Fourier space.*

is a much more convenient operation than rotation, and therefore we use a square partition in the following; see Figure 2(b).

To make smooth transitions possible—smoothness plays an important role in the properties of the Galerkin matrix—one ridgelet has to be supported on neighboring shears and scales. This is illustrated below in Figure 3.

In this section, we will show how these ridgelets can be constructed explicitly for a given transition function. Furthermore, we will show that the ridgelets form a frame in physical space. Finally, efficient methods will be presented which allow us to switch from an explicit representation of a function to its representation as a linear combination of ridgelets, and vice versa. Many ideas in this section are taken from [20] and [16].

We note that in the context of the frameworks that were mentioned in the introduction, ridgelets are α-molecules (cf. [17]) with $\alpha = 0$, but do not belong to "parabolic molecules" (which include shearlets and curvelets; cf. [18]), corresponding to $\alpha = \frac{1}{2}$.

3.1. Construction. The basis for the transition between neighboring ridgelets is a shape function $t : \mathbb{R} \to \mathbb{R}$ satisfying

$$t(0) = 0, \quad t(1) = 1, \quad t \in \mathcal{C}^m([0,1]) \text{ for an } m \geq 0.$$

Using this function, we construct a radial and a spherical window function:

$$w_r(x) := \begin{cases} \sin\big(\frac{\pi}{2}\, t(|x|-1)\big), & 1 \leq |x| \leq 2, \\ \cos\big(\frac{\pi}{2}\, t(\frac{1}{2}|x|-1)\big), & 2 < |x| < 4, \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

and

$$w_s(x) := \begin{cases} \sqrt{t(1+x)}, & x \leq 0, \\ \sqrt{t(1-x)}, & x > 0. \end{cases} \tag{3.2}$$

With these helper functions at hand, we can start defining our ridgelets.

DEFINITION 3.1 (basic ridgelet). *Let $\rho_x, \rho_y \in \mathbb{N}$ and $\Omega = [0, L_x) \times [0, L_y)$ be some rectangular domain. Then the* basic ridgelet $\psi_{(1,\mathtt{x},0)} : \Omega \to \mathbb{C}$ *is a function defined via its Fourier transform,*

$$\psi_{(1,\mathtt{x},0)} := \mathcal{F}^{-1}\big[\hat{\psi}_{(1,\mathtt{x},0)}\big],$$

where $\hat{\psi}_{(1,\mathtt{x},0)} : \mathbb{Z}^2 \to [0,1]$ is defined as

$$\hat{\psi}_{(1,\mathtt{x},0)}(\hat{x}, \hat{y}) := w_r\left(\frac{\hat{x}}{\rho_x}\right) w_s\left(\frac{\hat{y}/\rho_y}{\hat{x}/\rho_x}\right).$$

DEFINITION 3.2 (x-cone ridgelets). *Let Ω be as above. Then the* x*-cone ridgelets are a family of functions $\hat{\psi}_{(j,\mathtt{x},k)} : \mathbb{Z}^2 \to [0,1]$ parametrized by $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ defined in Fourier space by*

$$\hat{\psi}_{(j,\mathtt{x},k)}(\hat{x}, \hat{y}) := \hat{\psi}_{(1,\mathtt{x},0)}\left(\frac{\hat{x}}{2^{j-1}},\, \hat{y} - \frac{k}{2^{j-1}}\hat{x}\right),$$

and with corresponding physical space functions $\psi_{(j,\mathtt{x},k)} : \Omega \to \mathbb{C}$ given by

$$\psi_{(j,\mathtt{x},k)} := \mathcal{F}^{-1}\big[\hat{\psi}_{(j,\mathtt{x},k)}\big].$$

Note that the x-cone ridgelets correspond to a scaling (in the x-direction) with subsequent shearing (in the y-direction) of the basic ridgelet. Since we can only cover the y-axis in the limit $|x| \to \infty$ when shearing from $\hat{\psi}_{(1,\mathtt{x},0)}$, we only shear to the diagonal and cover the vertical cone by shearing from a vertical ridgelet (compare Figure 2(b)), or rather, by transposing the horizontal ridgelets.

DEFINITION 3.3 (y-cone ridgelets). *The* y*-cone ridgelets are a family of functions $\hat{\psi}_{(j,\mathtt{y},k)} : \mathbb{Z}^2 \to [0,1]$ parametrized by $j \in \mathbb{N}$ and $k \in \mathbb{Z}$ defined in Fourier space by*

$$\hat{\psi}_{(j,\mathtt{y},k)}(\hat{x}, \hat{y}) := \hat{\psi}_{(j,\mathtt{x},k)}(\hat{y}, \hat{x}),$$

and with corresponding physical space functions $\psi_{(j,\mathtt{y},k)} : \Omega \to \mathbb{C}$ given by

$$\psi_{(j,\mathtt{y},k)} := \mathcal{F}^{-1}\big[\hat{\psi}_{(j,\mathtt{y},k)}\big].$$

Although both variants are well-defined for arbitrary $k \in \mathbb{Z}$, having both variants means that we can restrict ourselves to $k \in [-2^{j-1}+1 : 2^{j-1}-1]$. The only missing part is the "diagonal" $|k| = 2^{j-1}$, which we need to define differently to achieve a continuous (but not necessarily differentiable) transition between the x- and y-cone ridgelets.

DEFINITION 3.4 (diagonal ridgelets). *Let ρ_x, ρ_y, and Ω be as above. Then the diagonal ridgelets are a family of functions $\hat{\psi}_{(j,\mathtt{d},k)} : \mathbb{Z}^2 \to [0,1]$ parametrized by $j \in \mathbb{N}$ and $k = \pm 2^{j-1}$ defined in Fourier space by*

$$\hat{\psi}_{(j,\mathtt{d},k)}(\hat{x}, \hat{y}) := \begin{cases} \hat{\psi}_{(j,\mathtt{x},k)}(\hat{x}, \hat{y}), & \frac{|\hat{x}|}{\rho_x} \geq \frac{|\hat{y}|}{\rho_y}, \\ \hat{\psi}_{(j,\mathtt{y},k)}(\hat{x}, \hat{y}), & \frac{|\hat{x}|}{\rho_x} < \frac{|\hat{y}|}{\rho_y}, \end{cases}$$

and with corresponding physical space functions $\psi_{(j,\mathtt{d},k)} : \Omega \to \mathbb{C}$ given by

$$\psi_{(j,\mathtt{d},k)} := \mathcal{F}^{-1}\big[\hat{\psi}_{(j,\mathtt{d},k)}\big].$$

With the ridgelets defined so far, we cover all directions. However, we still need some function that covers the low-frequency part (which corresponds to the mass of the function in physical space).

DEFINITION 3.5 (scaling function). *Given ρ_x, ρ_y, and Ω as above, the* scaling function $\hat{\psi}_{(0,\mathtt{s},0)} : \mathbb{Z}^2 \to [0,1]$ *is a function defined in Fourier space by*

$$\hat{\psi}_{(0,\mathtt{s},0)}(\hat{x},\hat{y}) := \begin{cases} 1, & \hat{z} < 1, \\ \cos\big(\frac{\pi}{2} t(\hat{z}-1)\big), & 1 \le \hat{z} < 2, \\ 0, & 2 \le \hat{z}, \end{cases} \quad \text{where } \hat{z} := \max\left\{ \frac{|\hat{x}|}{\rho_x}, \frac{|\hat{y}|}{\rho_y} \right\},$$

with corresponding physical space function $\psi_{(0,\mathtt{s},0)} : \Omega \to \mathbb{C}$ given by

$$\psi_{(0,\mathtt{s},0)} := \mathcal{F}^{-1}\big[\hat{\psi}_{(0,\mathtt{s},0)}\big].$$

With the scaling function, the set of ridgelets is now complete. In the remainder of this chapter, it will become clear that the most important property of a ridgelet is its well-defined support in the Fourier space; the supports for a few selected ridgelets are illustrated in Figure 3.

DEFINITION 3.6 (index set). *The* index set Λ *is defined as the set of all triples*

$$\Lambda = \left\{ \lambda = (\kappa, j, k) \;\middle|\; \begin{array}{ll} \kappa = \mathtt{s} & \wedge\, j = 0 \,\wedge\, k = 0 \\ \kappa = \mathtt{x,y} & \wedge\, j \in \mathbb{N} \wedge k \in [-2^{j-1}+1 : 2^{j-1}-1] \\ \kappa = \mathtt{d} & \wedge\, j \in \mathbb{N} \wedge k = \pm 2^{j-1} \end{array} \right\},$$

where j is called the scale parameter*, κ the* direction parameter*, and k the* shear parameter*.*

LEMMA 3.7 (partition of unity). *The $\hat{\psi}_\lambda$ with $\lambda \in \Lambda$ constitute a partition of unity, i.e.,*

$$\sum_{\lambda \in \Lambda} \hat{\psi}_\lambda^2(\hat{x},\hat{y}) = 1 \quad \forall (\hat{x},\hat{y}) \in \mathbb{Z}^2. \tag{3.3}$$

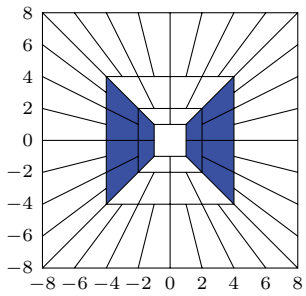
Proof. The full proof can be found in [16]; we mention the main points.

First, note that $w_s^2(x) + w_s^2(x-1) = 1$ for all $x \in [0,1]$. Therefore, we have that, e.g.,

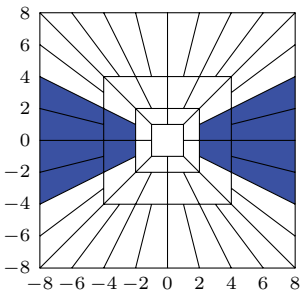
$$\hat{\psi}^2_{(j,\mathtt{x},k)}(\hat{x},\hat{y}) + \hat{\psi}^2_{(j,\mathtt{x},k+1)}(\hat{x},\hat{y}) = w_r^2\left(\frac{|\hat{x}|}{\rho_x 2^{j-1}} \right) \quad \forall\, (\hat{x},\hat{y}) \in \operatorname{supp} \hat{\psi}_{(j,\mathtt{x},k)} \cap \operatorname{supp} \hat{\psi}_{(j,\mathtt{x},k+1)}.$$

If we extend this argument to all ridgelets on the same scale j, we see that the dependence on the spherical part w_s completely drops out and that we are left with only the radial part w_r. But since we have that $w_r^2(x) + w_r^2(\frac{x}{2}) = 1$ for all $x \in [-4,-2] \cup [2,4]$, this will become one as well once we square and sum over all scales $j \in \mathbb{N}_0$. ∎

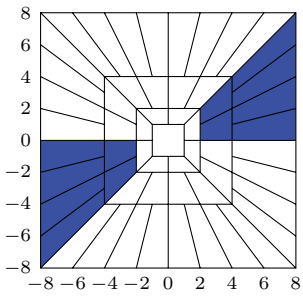
As in [20] and [16], this partition of unity property of the ridgelets will be the key point in proving the invertibility of the ridgelet transform in Theorem 3.18.



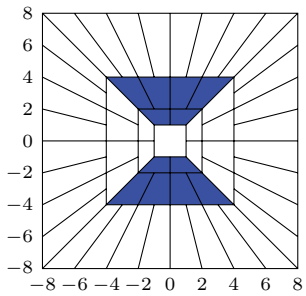
(a) Basic ridgelet $\hat{\psi}_{(1,\mathtt{x},0)}$.



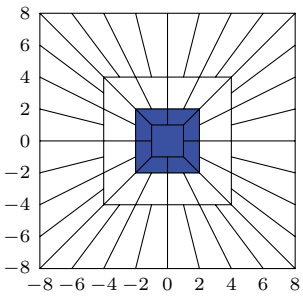
(b) $\mathtt{x}$-cone ridgelet $\hat{\psi}_{(2,\mathtt{x},0)}$ scaled from (a).



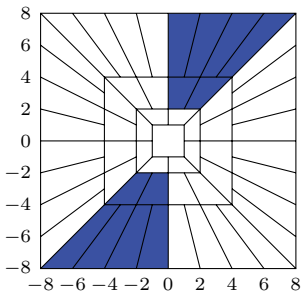
(c) $\mathtt{x}$-cone ridgelet $\hat{\psi}_{(2,\mathtt{x},1)}$ sheared from (b).



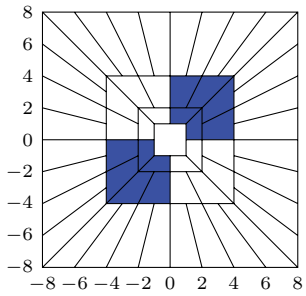
(d) $\mathtt{y}$-cone ridgelet $\hat{\psi}_{(1,\mathtt{y},0)}$ transposed from (a).



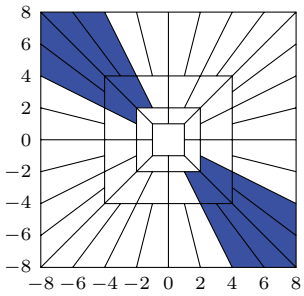
(e) Central ridgelet $\hat{\psi}_{(0,\mathtt{s},0)}$.



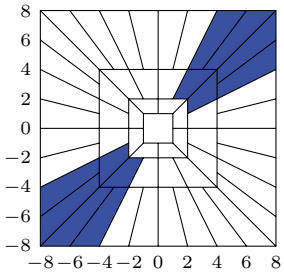
(f) $\mathtt{y}$-cone ridgelet $\hat{\psi}_{(2,\mathtt{y},1)}$ transposed from (c).



(g) Diagonal ridgelet $\hat{\psi}_{(1,\mathtt{d},1)}$ transition between (a) and (d).



(h) Diagonal ridgelet $\hat{\psi}_{(2,\mathtt{d},-2)}$.



(i) Diagonal ridgelet $\hat{\psi}_{(2,\mathtt{d},2)}$ transition between (c) and (f).

FIG. 3. *Fourier space supports of several ridgelets. Units are ρ_x for the x-axes and ρ_y for the y-axes.*

In the introduction, we specified the ridgelets as a set of functions with varying locations, directions, and width. So far, we have only covered direction and width. Next, we will thus have a look at the translations of the ridgelets.

For the ridgelets to form a frame, the different modulations of one ridgelet must form a basis on the support of the ridgelet. This immediately gives the necessary

resolution for each ridgelet.

DEFINITION 3.8 (translation sets). *Let ρ_x, ρ_y, and Λ be as above, and define the x- and y-resolutions as*

$$T_x^\lambda := \begin{cases} 4\rho_x & \text{if } \kappa = \mathtt{s}, \\ 2^{j+2}\rho_x & \text{if } \kappa = \mathtt{x}, \\ 8\rho_x & \text{if } \kappa = \mathtt{y}, \\ 2^{j+2}\rho_x & \text{if } \kappa = \mathtt{d} \end{cases} \quad \text{and} \quad T_y^\lambda := \begin{cases} 4\rho_y & \text{if } \kappa = \mathtt{s}, \\ 8\rho_y & \text{if } \kappa = \mathtt{x}, \\ 2^{j+2}\rho_y & \text{if } \kappa = \mathtt{y}, \\ 8\rho_y & \text{if } \kappa = \mathtt{d}. \end{cases}$$

Then the translation set T^λ *to a given parameter choice $\lambda \in \Lambda$ is defined as*

$$T^\lambda := \left[0 : \frac{L_x}{T_x^\lambda} : L_x\right) \times \left[0 : \frac{L_y}{T_y^\lambda} : L_y\right).$$

Here one of the advantages of using shearing instead of rotations comes to light: normally, the translations would have to be rotated/sheared differently for each parameter combination, but since shearing leaves our translation grid T^λ invariant, we can omit this.

We conclude this section by collecting all the defined functions in a set.

LEMMA 3.9 (ridgelet frame [16]). *The set of functions*

$$\{\psi_\lambda\}_{\Lambda,T^\lambda} := \left\{\psi_\lambda(\cdot - t_x, \cdot - t_y) \,\middle|\, \lambda \in \Lambda,\, t \in T^\lambda\right\}$$

is a frame. This implies that, suitably scaled, $\{\psi_\lambda\}_{\Lambda,T^\lambda}$ satisfy

$$\|f\|_2^2 \sim \sum_{\lambda\in\Lambda} \sum_{t\in T^\lambda} \left\langle f, \psi_\lambda(\cdot - t_x, \cdot - t_y)\right\rangle_2,$$

where $f : \Omega \to \mathbb{C}$ is some arbitrary function and the $\sim$ symbol means that the left side is bounded by a constant times the right side, and vice versa.

Remark 3.10. Compared to [16], the translations for ψ_λ do not have to be transformed, since the grid T^λ already incorporates the scaling and *is invariant to shearing.*

3.2. Ridgelet transform. In analogy to the Fourier transforms, the process of expressing a given function as a linear combination of ridgelets is called a ridgelet transform, and the opposite operation an inverse ridgelet transform. These transforms introduce a new space of coefficients which is called the ridgelet coefficient space.

DEFINITION 3.11 (ridgelet coefficient space). *The set of pairs*

$$\widetilde{\Omega} := \left\{(\lambda, t) \,\middle|\, \lambda \in \Lambda,\, t \in T^\lambda\right\}$$

is called the ridgelet coefficient space. *The notation is chosen in analogy to Ω and $\widehat{\Omega}$.*

DEFINITION 3.12 (ridgelet transform). *Let $\Omega = [0, L_x) \times [0, L_y)$ be some rectangular domain and $f : \Omega \to \mathbb{C}$ a function. Then the function $\mathcal{R}[f] : \widetilde{\Omega} \to \mathbb{C}$ given by*

$$\mathcal{R}[f](\lambda, t) := \frac{1}{L_x L_y} \frac{1}{T_x^\lambda T_y^\lambda} \int_\Omega f(x, y)\, \overline{\psi_\lambda(x - t_x, y - t_y)}\, \mathrm{d}x\, \mathrm{d}y$$

is called the ridgelet transform *of* f.

DEFINITION 3.13 (inverse ridgelet transform). *Let* $\Omega = [0, L_x) \times [0, L_y)$ *be some rectangular domain and* $\tilde{f} : \widetilde{\Omega} \to \mathbb{C}$ *a function. Then the function* $\mathcal{R}^{-1}[\tilde{f}] : \Omega \to \mathbb{C}$ *given by*

$$\mathcal{R}^{-1}[\tilde{f}](x, y) := \sum_{\lambda \in \Lambda} \sum_{t \in T^\lambda} \tilde{f}(\lambda, t)\, \psi_\lambda(x - t_x, y - t_y)$$

is called the inverse ridgelet transform *of* f.

As implied in the beginning, for both theoretical work as well as implementation, it is more useful to work with their Fourier transforms, however, since we can then discretize in a way that allows us to use FFT.

DEFINITION 3.14 (Fourier ridgelet transform). *Let* $\hat{f} : \mathbb{Z}^2 \to \mathbb{C}$ *be a function. Then the function* $\hat{\mathcal{R}}[\hat{f}] : \widetilde{\Omega} \to \mathbb{C}$ *given by*

$$\hat{\mathcal{R}}[\hat{f}](\lambda, t) := \mathcal{R}\big[\mathcal{F}^{-1}[\hat{f}]\big](\lambda, t)$$

is called the Fourier ridgelet transform *of* $\hat{f}$.

THEOREM 3.15. *The Fourier ridgelet transform is given by*

$$\hat{\mathcal{R}}[\hat{f}](\lambda, t) = \frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{x}, \hat{y}) \in \mathbb{Z}^2} \hat{f}(\hat{x}, \hat{y}) \hat{\psi}_\lambda(\hat{x}, \hat{y}) \mathrm{e}^{2\pi \mathrm{i}\, (\hat{x} \frac{t_x}{L_x} + \hat{y} \frac{t_y}{L_y})}.$$

Proof. The proof is a direct consequence of the Plancherel formula (2.3), the translation property of the Fourier transform (2.2), and the fact that the ridgelets $\hat{\psi}_\lambda$ are real. ∎

DEFINITION 3.16 (inverse Fourier ridgelet transform). *Let* $\tilde{f} : \widetilde{\Omega} \to \mathbb{C}$ *be a function. Then the function* $\hat{\mathcal{R}}^{-1}[\tilde{f}] : \mathbb{Z}^2 \to \mathbb{C}$ *given by*

$$\hat{\mathcal{R}}^{-1}[\tilde{f}](\hat{x}, \hat{y}) := \mathcal{F}\big[\mathcal{R}^{-1}[\tilde{f}]\big](\hat{x}, \hat{y})$$

is called the inverse Fourier ridgelet transform *of* $\tilde{f}$.

THEOREM 3.17. *The inverse Fourier ridgelet transform is given by*

$$\hat{\mathcal{R}}^{-1}[\tilde{f}](\hat{x}, \hat{y}) = \sum_{\lambda \in \Lambda} \sum_{t \in T^\lambda} \tilde{f}(\lambda, t) \hat{\psi}_\lambda(\hat{x}, \hat{y}) \mathrm{e}^{-2\pi \mathrm{i}\, (\hat{x} \frac{t_x}{L_x} + \hat{y} \frac{t_y}{L_y})}.$$

Proof. The proof is a direct consequence of the linearity of the Fourier transform and the translation property (2.2). ∎

Thanks to these Fourier ridgelet transforms, we are now able to prove the first main result of this paper.

THEOREM 3.18 (inverse property of the Fourier ridgelet transforms). *The inverse Fourier ridgelet transform is the left inverse of the Fourier ridgelet transform, i.e.,* $\hat{\mathcal{R}}^{-1} \circ \hat{\mathcal{R}} = \mathbb{I}$.

Proof. Let $\hat{f} : \mathbb{Z}^2 \to \mathbb{C}$. Then

$$\begin{aligned}
&\hat{\mathcal{R}}^{-1}\big[\hat{\mathcal{R}}[\hat{f}]\big](\hat{x}, \hat{y}) = \cdots \\
&= \sum_{\lambda \in \Lambda} \sum_{t \in T^\lambda} \frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{v}, \hat{w}) \in \mathbb{Z}^2} \hat{f}(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{x}, \hat{y}) \mathrm{e}^{2\pi \mathrm{i} \left((\hat{v} - \hat{x}) \frac{t_x}{L_x} + (\hat{w} - \hat{y}) \frac{t_y}{L_y}\right)} \\
&= \sum_{\lambda \in \Lambda} \sum_{(\hat{v}, \hat{w}) \in \mathbb{Z}^2} \hat{f}(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{x}, \hat{y}) \frac{1}{T_x^\lambda T_y^\lambda} \sum_{t \in T^\lambda} \mathrm{e}^{2\pi \mathrm{i} \left((\hat{v} - \hat{x}) \frac{t_x}{L_x} + (\hat{w} - \hat{y}) \frac{t_y}{L_y}\right)} \\
&= \sum_{\lambda \in \Lambda} \sum_{(\hat{v}, \hat{w}) \in \mathbb{Z}^2} \hat{f}(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{v}, \hat{w})\, \hat{\psi}_\lambda(\hat{x}, \hat{y})\, \delta\big((\hat{v} - \hat{x}) \bmod T_x^\lambda\big)\, \delta\big((\hat{w} - \hat{y}) \bmod T_y^\lambda\big).
\end{aligned}$$

The fact that $\frac{1}{T_x^\lambda T_y^\lambda} \sum_{t \in T^\lambda} \mathrm{e}^{2\pi \mathrm{i} \, ((\hat{v}-\hat{x})\frac{t_x}{L_x} + (\hat{w}-\hat{y})\frac{t_y}{L_y})}$ collapses to delta functions is due to the summation property of the root of unity; see (2.5).

Here, the choice of T_x^λ and T_y^λ in Definition 3.8 comes into play again, namely, that if $(\hat{x}, \hat{y}) \in \mathbb{Z}^2$ lies in the support of $\hat{\psi}_\lambda$, then all $(\hat{x} - n_x T_x^\lambda, \hat{y} - n_y T_y^\lambda), (n_x, n_y) \in \mathbb{Z}^2 \setminus \{(0,0)\}$, do not. Therefore, we do not have to bother with the moduli in the delta functions and can instead just write

$$\hat{\mathcal{R}}^{-1}\big[\hat{\mathcal{R}}[\hat{f}]\big](\hat{x}, \hat{y}) = \sum_{\lambda \in \Lambda} \hat{f}(\hat{x}, \hat{y}) \, \hat{\psi}_\lambda^2(\hat{x}, \hat{y}) = \hat{f}(\hat{x}, \hat{y}),$$

due to the partition of unity property (3.3) of the ridgelets. ∎

The above theorem implies that the Fourier ridgelet transform is injective and that the inverse Fourier ridgelet transform is surjective. However, it is important to stress that neither of them is bijective! An easy way to see this is to think about what happens if a ridgelet $\hat{\psi}_\lambda$ is transformed by $\hat{\mathcal{R}}$: Obviously, we have $\hat{\mathcal{R}}^{-1}[\tilde{f}] = \hat{\psi}_\lambda$ if we let $\tilde{f}(\mu, t) := \delta(\mu - \lambda)\, \delta(t)$ (here, $\mu - \lambda$ for $\mu, \lambda \in \Lambda$ is defined as zero iff $\mu = \lambda$, and anything different from zero otherwise). But $\tilde{f}$ is not the function produced by $\hat{\mathcal{R}}[\hat{\psi}_\lambda]$! Rather, since $\hat{\psi}_\lambda$ overlaps with some other ridgelets $\hat{\psi}_\mu$ in Fourier space, these $\hat{\psi}_\mu$ will have nonzero coefficients as well. In conclusion, we thus observe the following:

Fact 3.19. *The Fourier ridgelet transform is not surjective. The inverse Fourier ridgelet transform is not injective.*

Of course, both results hold equally for the original ridgelet transforms in physical space.

Corollary 3.20 (inverse property of the ridgelet transforms). *The inverse ridgelet transform is the left inverse of the ridgelet transform, i.e., $\mathcal{R}^{-1} \circ \mathcal{R} = \mathbb{I}$.*

Proof. By the definition of the (inverse) Fourier ridgelet transform (Definitions 3.14 and 3.16) and the inverse property of the Fourier transforms (2.1), we can write $\mathcal{R} = \hat{\mathcal{R}} \circ \mathcal{F}$ and $\mathcal{R}^{-1} = \mathcal{F}^{-1} \circ \hat{\mathcal{R}}^{-1}$. Then the claim follows by the inverse property of the Fourier transforms (2.1) and the above theorem. ∎

Fact 3.21. *The ridgelet transform is not surjective. The inverse ridgelet transform is not injective.*

3.3. Finite ridgelet transform. In the same way that we introduced the finite Fourier transform as finitely computable approximations to the ideal Fourier transforms, we will introduce here the finite ridgelet transforms as approximations to the ideal ridgelets transforms defined in the previous section. Since the formulae for the Fourier ridgelet transform already contain only series and sums, we only have to specify how we truncate the series.

Definition 3.22 (finite index set). *Let $J \in \mathbb{N}$. Then the* finite index set Λ_{fin} *is given by*

$$\Lambda_{\mathrm{fin}} = \big\{ \lambda \in \Lambda \,\big|\, j \leq J \big\}.$$

Definition 3.23 (finite ridgelet coefficient space). *The set of pairs*

$$\widetilde{\Omega}_{\mathrm{fin}} := \big\{ (\lambda, t) \,\big|\, \lambda \in \Lambda_{\mathrm{fin}},\, t \in T^\lambda \big\}$$

is called the finite ridgelet coefficient space.

Definition 3.24 (finite ridgelet transform). *Let*

$$\widehat{\Omega}_{\mathrm{fin}} = \big[-\rho_x \, 2^{J+1} : \rho_x \, 2^{J+1} - 1\big] \times \big[-\rho_y \, 2^{J+1} : \rho_y \, 2^{J+1} - 1\big]$$

be a subset of the Fourier space and $\hat{f} : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$ *a function. Then the function* $\texttt{rt}[\hat{f}] : \widetilde{\Omega}_{\text{fin}} \to \mathbb{C}$ *given by*

$$\texttt{rt}[\hat{f}](\lambda, t) = \frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\text{fin}}} \hat{f}(\hat{x},\hat{y})\, \hat{\psi}_\lambda(\hat{x},\hat{y})\, \mathrm{e}^{2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}$$

is called the finite (Fourier) ridgelet transform *of* $\hat{f}$*.*

Remark 3.25. The choice of $\widehat{\Omega}_{\text{fin}}$ corresponds to the finest possible grid in each direction, i.e., to

$$\Omega_{\text{fin}} = \bigcup_{\lambda\in\Lambda_{\text{fin}}} T^\lambda.$$

Thus the choice of J and ρ_x, ρ_y determines the N_x, N_y of Ω_{fin}.

DEFINITION 3.26 (finite inverse ridgelet transform). *Let* $\widehat{\Omega}_{\text{fin}}$ *be as above and* $\tilde{f} : \widetilde{\Omega}_{\text{fin}} \to \mathbb{C}$ *a function. Then the function* $\texttt{irt}[\tilde{f}] : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$ *given by*

$$\texttt{irt}[\tilde{f}](\hat{x},\hat{y}) = \sum_{\lambda\in\Lambda_{\text{fin}}} \sum_{t\in T^\lambda} \tilde{f}(\lambda, t)\, \hat{\psi}_\lambda(\hat{x},\hat{y})\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}$$

is called the finite inverse (Fourier) ridgelet transform *of* $\hat{f}$*.*

To achieve a well-localized discretization in space, the $\hat{\psi}_\lambda$ have to be smooth, which necessitates that the different scales mix in the partition of unity $\{\hat{\psi}_\lambda\}_\lambda$. Due to this fact, some unwelcome side effects at the highest scale J of the discretization are unavoidable, but depend on the specific treatment of the highest scale.

As we saw in the proof of Theorem 3.18 (and will see below), the invertibility of the transform is closely related to the partition of unity property. To have invertibility on the full range $\hat{\Omega}$, it would therefore be necessary to include scale $J+1$ as well, but, crucially, the functions on this scale would be cut off by $\hat{\Omega}$. Through the implicit periodicity of the Fourier transform, this cut-off would lead to substantial artifacts in the reconstruction after transforming back to physical space (as soon as one of corresponding coefficients is nonnegligible).

To prevent this possibility, we have decided to discard the contributions of the $(J+1)$st scale entirely, which, however, also means that we do not have full invertibility of the discrete transform.

LEMMA 3.27 (finite partition of unity). *The* $\hat{\psi}_\lambda$ *with* $\lambda \in \Lambda_{\text{fin}}$ *constitute a partition of unity on a part of* $\mathbb{Z}^2$*, namely,*

$$\sum_{\lambda\in\Lambda_{\text{fin}}} \hat{\psi}_\lambda^2(\hat{x},\hat{y}) = 1 \quad \forall(\hat{x},\hat{y}) \in \left[-\rho_x 2^J : \rho_x 2^J\right] \times \left[-\rho_y 2^J : \rho_y 2^J\right] =: \widehat{\Omega}_{\text{uni}}.$$

In particular, $\texttt{irt}[\texttt{rt}[\hat{f}]] = \hat{f}$ *holds only if* $\hat{f}$ *is zero outside of this region. For general* $\hat{f}$*,* $\texttt{irt}\circ\texttt{rt}$ *can be interpreted as a low-pass filter with a very high cut-off frequency.*

Proof. Proceeding exactly as in the proof of Theorem 3.18 for the full frame, we arrive at

$$\begin{aligned}\texttt{irt}[\texttt{rt}[\hat{f}]](\hat{x},\hat{y}) &= \sum_{\lambda\in\Lambda_{\text{fin}}} \hat{f}(\hat{x},\hat{y})\, \hat{\psi}_\lambda^2(\hat{x},\hat{y}) \\ &= \begin{cases} \hat{f}(\hat{x},\hat{y}), & (\hat{x},\hat{y}) \in \widehat{\Omega}_{\text{uni}}, \\ \hat{f}(\hat{x},\hat{y})\, w_r^2\Big(\max\Big\{\frac{|\hat{x}|}{\rho_x}, \frac{|\hat{y}|}{\rho_y}\Big\}\Big) & \text{otherwise.} \end{cases}\end{aligned}$$

∎

Remark 3.28. As is obvious from the overlap of the different ridgelets, there is some degree of redundancy in the ridgelet frame. By summing all contributions,

$$\sum_{\lambda_\in \Lambda_{\mathrm{fin}}} T_y^\lambda T_y^\lambda = \rho_x \rho_y \Big(16 + \frac{256}{3}(4^J - 1)\Big),$$

we see that the redundancy $\frac{|\widetilde{\Omega}_{\mathrm{fin}}|}{|\Omega_{\mathrm{uni}}|}$ is bounded by $\frac{64}{3}$ and $\frac{|\widetilde{\Omega}_{\mathrm{fin}}|}{|\Omega_{\mathrm{fin}}|}$ by $\frac{16}{3}$.

3.4. Implementation overview. The rest of this section will be dedicated to showing how the finite ridgelet transforms can be evaluated efficiently. The bottom line will be that both of them can be performed in $\mathcal{O}\big(\big|\widetilde{\Omega}_{\mathrm{fin}}\big| \log\big(\big|\widetilde{\Omega}_{\mathrm{fin}}\big|\big)\big)$, i.e., that up to a logarithmic factor, optimal computational complexity is achievable.

When trying to evaluate

$$\tilde{f}_\lambda(t) := \frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\mathrm{fin}}} \hat{f}(\hat{x},\hat{y})\, \hat{\psi}_\lambda(\hat{x},\hat{y})\, \mathrm{e}^{2\pi \mathrm{i}\,(\hat{x}\frac{t_x}{L_x} + \hat{y}\frac{t_y}{L_y})},$$

we can observe that the support of $\hat{\psi}_\lambda$ restricts the area where we need to evaluate from $\widehat{\Omega}_{\mathrm{fin}}$ to the following (actually the support is much more restricted—compare Figure 3—but this is the smallest quadrilateral which contains it). For example, for $\kappa = \mathtt{x}$, we have (since T_x^λ and T_y^λ are by definition always divisible by two)

$$\widehat{\Omega}_> := \Big[-\frac{T_x^\lambda}{2} : \frac{T_x^\lambda}{2} - 1\Big] \times \Big[-\frac{(|k|+1)\,T_y^\lambda}{2} : \frac{(|k|+1)\,T_y^\lambda}{2} - 1\Big].$$

This corresponds to having the following domain for $\mathtt{ift}[\hat{f}\hat{\psi}_\lambda]$;

$$\Omega_> := \Big[0 : \frac{L_x}{T_x^\lambda} : L_x\Big) \times \Big[0 : \frac{L_y}{(|k|+1)\,T_y^\lambda} : L_y\Big);$$

however, we only need to evaluate $\tilde{f}_\lambda$ on

$$\Omega_< := T^\lambda = \Big[0 : \frac{L_x}{T_x^\lambda} : L_x\Big) \times \Big[0 : \frac{L_y}{T_y^\lambda} : L_y\Big).$$

This motivates the following discussion of the interplay between such subgrids and $\mathtt{ft}$. The definitions and statements do not depend materially on the fact that we restrict ourselves to even N_x, N_y (which is always the case for us), but rather saves notational effort by eliminating the floor- and ceiling-operations.

Definition 3.29 (folding operations). *Let $n \in \mathbb{N}$, N_x, $N_y \in 2\mathbb{N}$,*

$$\widehat{\Omega}_> = \Big[-\frac{nN_x}{2} : \frac{nN_x}{2} - 1\Big] \times \Big[-\frac{N_y}{2} : \frac{N_y}{2} - 1\Big],$$
$$\widehat{\Omega}_< = \Big[-\frac{N_x}{2} : \frac{N_x}{2} - 1\Big] \times \Big[-\frac{N_y}{2} : \frac{N_y}{2} - 1\Big],$$

and $\hat{f}_> : \widehat{\Omega}_> \to \mathbb{C}$. Then $\mathtt{foldx}[\hat{f}_>, n, N_x] : \widehat{\Omega}_< \to \mathbb{C}$ is defined as

$$\mathtt{foldx}[\hat{f}_>, n, N_x](\hat{x},\hat{y}) := \begin{cases} \displaystyle\sum_{\hat{v}\in\left[-\lceil\frac{n-1}{2}\rceil : \lfloor\frac{n-1}{2}\rfloor\right]} \hat{f}_>(\hat{x} + N_x\hat{v}, \hat{y}), & \hat{x} \geq 0, \\ \displaystyle\sum_{\hat{v}\in\left[-\lfloor\frac{n-1}{2}\rfloor : \lceil\frac{n-1}{2}\rceil\right]} \hat{f}_>(\hat{x} + N_x\hat{v}, \hat{y}), & \hat{x} < 0 \end{cases}$$

(the only difference between the cases is in the rounding operations), and the symbol `foldx` *is called the* `x`*-folding operator.*

The `y`*-folding operator* `foldy` *is defined likewise.*

LEMMA 3.30 (folding lemma). *Let $n \in \mathbb{N}$, N_x, $N_y \in 2\mathbb{N}$, as well as L_x, $L_y \in \mathbb{R}^{>0}$ and*

$$\Omega_> = \left[0 : \frac{L_x}{nN_x} : L_x\right) \times \left[0 : \frac{L_y}{N_y} : L_y\right), \quad \widehat{\Omega}_> = \left[-\frac{nN_x}{2} : \frac{nN_x}{2} - 1\right] \times \left[-\frac{N_y}{2} : \frac{N_y}{2} - 1\right],$$
$$\Omega_< = \left[0 : \frac{L_x}{N_x} : L_x\right) \times \left[0 : \frac{L_y}{N_y} : L_y\right), \quad \widehat{\Omega}_< = \left[-\frac{N_x}{2} : \frac{N_x}{2} - 1\right] \times \left[-\frac{N_y}{2} : \frac{N_y}{2} - 1\right].$$

Furthermore, let $\hat{f}_> : \widehat{\Omega}_> \to \mathbb{C}$ and $f_< := \mathtt{ift}[\hat{f}_>]\big|_{\Omega_<}$. Then we have

$$\mathtt{ft}[f_<] = \mathtt{foldx}[\hat{f}_>, n, N_x].$$

Proof. First, assume $\hat{x} \geq 0$. Then

$$\begin{aligned}
&\mathtt{ft}[f_<](\hat{x}, \hat{y}) \\
&\quad = \frac{1}{N_x N_y} \sum_{(x,y)\in\Omega_<} \mathtt{ift}[\hat{f}_>](x,y)\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x} + \hat{y}\frac{y}{L_y})} \\
&\quad = \frac{1}{N_x N_y} \sum_{(x,y)\in\Omega_<} \underbrace{\alpha(n,x)}_{=1}\, \mathtt{ift}[\hat{f}_>](x,y)\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x} + \hat{y}\frac{y}{L_y})} + \cdots \\
&\qquad + \frac{1}{N_x N_y} \sum_{(x,y)\in\Omega_<} \sum_{v\in\Delta\Omega_x} \underbrace{\alpha(n,x+v)}_{=0}\, \mathtt{ift}[\hat{f}_>](x+v,y)\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x+v}{L_x} + \hat{y}\frac{y}{L_y})},
\end{aligned}$$

where $\Delta\Omega_x := (0 : \frac{L_x}{nN_x} : \frac{L_x}{N_x})$ and $\alpha(n,z) = \frac{1}{n}\sum_{\hat{v}\in\left[-\left\lceil\frac{n-1}{2}\right\rceil : \left\lfloor\frac{n-1}{2}\right\rfloor\right]} \mathrm{e}^{-2\pi\mathrm{i}\,N_x\hat{v}\frac{z}{L_x}}$. The first underbrace equals one because $N_x\frac{x}{L_x}$ is always an integer for $x \in \Omega_<$. On the other hand, $N_x\frac{x+v}{L_x}$ for $x \in \Omega_<$ and $v \in \Delta\Omega_x$ is always a proper fraction, and therefore the second underbrace collapses to zero due to the summation property of the roots of unity.

Note that

$$\sum_{(x,y)\in\Omega_<} g(x,y) + \sum_{(x,y)\in\Omega_<} \sum_{v\in\Delta\Omega_x} g(x+v,y) = \sum_{(x,y)\in\Omega_>} g(x,y).$$

Therefore, we can continue the above equations with

$$\begin{aligned}
\mathtt{ft}[f_<](\hat{x},\hat{y}) &= \frac{1}{n\,N_x N_y} \sum_{(x,y)\in\Omega_>} \sum_{\hat{v}\in\left[-\left\lceil\frac{n-1}{2}\right\rceil : \left\lfloor\frac{n-1}{2}\right\rfloor\right]} \mathtt{ift}[\hat{f}_>](x,y)\, \mathrm{e}^{-2\pi\mathrm{i}\left((\hat{x}+N_x\hat{v})\frac{x}{L_x} + \hat{y}\frac{y}{L_y}\right)} \\
&= \sum_{\hat{v}\in\left[-\left\lceil\frac{n-1}{2}\right\rceil : \left\lfloor\frac{n-1}{2}\right\rfloor\right]} \frac{1}{n\,N_x N_y} \sum_{(x,y)\in\Omega_>} \mathtt{ift}[\hat{f}_>](x,y)\, \mathrm{e}^{-2\pi\mathrm{i}\left((\hat{x}+N_x\hat{v})\frac{x}{L_x} + \hat{y}\frac{y}{L_y}\right)} \\
&= \sum_{\hat{v}\in\left[-\left\lceil\frac{n-1}{2}\right\rceil : \left\lfloor\frac{n-1}{2}\right\rfloor\right]} \mathtt{ft}[\mathtt{ift}[\hat{f}_>]](\hat{x}+N_x\hat{v},\hat{y}) \\
&= \sum_{\hat{v}\in\left[-\left\lceil\frac{n-1}{2}\right\rceil : \left\lfloor\frac{n-1}{2}\right\rfloor\right]} \hat{f}_>(\hat{x}+N_x\hat{v},\hat{y}) = \mathtt{foldx}[\hat{f}_>, n, N_x](\hat{x},\hat{y}),
\end{aligned}$$

recalling that we assumed $\hat{x} \geq 0$.

The proof for $\hat{x} < 0$ is the same, except that the range of $\hat{v}$ is shifted by $+1$ if n is even. Since $\mathrm{e}^{-2\pi\mathrm{i}\,(-\frac{n}{2})\frac{k}{n}} = \mathrm{e}^{-2\pi\mathrm{i}\,\frac{n}{2}\frac{k}{n}}$, the term which is dropped on the negative side in the sums over $\hat{v}$ in that case is equal to the new term on the positive side; therefore the above arguments showing that the first underbrace is one whereas the other is zero work out exactly the same. ∎

This result can now be used for evaluating the ridgelet transform for some fixed $\lambda \in \Lambda$.

LEMMA 3.31. *Let $\widehat{\Omega}_{\mathrm{fin}}$ be as in Definition 3.24 and let $\hat{f} : \widehat{\Omega}_{\mathrm{fin}} \to \mathbb{C}$. The complexity of evaluating*

$$\tilde{f}_\lambda(t) := \frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\mathrm{fin}}} \hat{f}(\hat{x},\hat{y})\,\hat{\psi}_\lambda(\hat{x},\hat{y})\,\mathrm{e}^{2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}$$

for some fixed $\lambda \in \Lambda$ and all $t \in T^\lambda$ is $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$.

Proof. We will show the lemma only for the case $\lambda = (j, \kappa = \mathtt{x}, k)$. For $\kappa \in \{\mathtt{y}, \mathtt{d}\}$, the proof is analogous, and for $\kappa = \mathtt{s}$ it is trivial.

Note that the above expression corresponds to $\tilde{f}_\lambda = \frac{1}{T_x^\lambda T_y^\lambda}\mathtt{ift}[\hat{f}\hat{\psi}_\lambda]$, except that the sum contains many more points than necessary. Because the support of $\hat{\psi}_\lambda$ is contained within

$$\widehat{\Omega}_> := \left[-\frac{T_x^\lambda}{2} : \frac{T_x^\lambda}{2} - 1\right] \times \left[-\frac{(|k|+1)\,T_y^\lambda}{2} : \frac{(|k|+1)\,T_y^\lambda}{2} - 1\right],$$

we can restrict the domain of $\hat{f}\hat{\psi}_\lambda$ from $\widehat{\Omega}_{\mathrm{fin}}$ to $\widehat{\Omega}_>$ such that the domain of $\mathtt{ift}[\hat{f}\hat{\psi}_\lambda]$ will be

$$\Omega_> := \left[0 : \frac{L_x}{T_x^\lambda} : L_x\right) \times \left[0 : \frac{L_y}{(|k|+1)\,T_y^\lambda} : L_y\right).$$

As mentioned in the beginning of this subsection, we only need $\tilde{f}_\lambda$ on

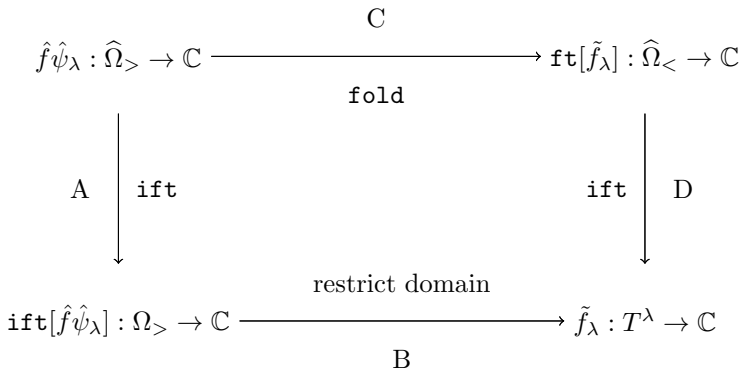
$$\Omega_< := T^\lambda = \left[0 : \frac{L_x}{T_x^\lambda} : L_x\right) \times \left[0 : \frac{L_y}{T_y^\lambda} : L_y\right);$$

thus we apply the folding lemma (Lemma 3.30) and compute $\tilde{f}_\lambda$ as

$$\tilde{f}_\lambda(t) = \frac{1}{T_x^\lambda T_y^\lambda}\,\mathtt{ift}\Big[\mathtt{foldy}\big[\hat{f}\hat{\psi}_\lambda,\, |k|+1,\, T_y^\lambda\big]\Big].$$

Since the number of points in the support of $\hat{\psi}_\lambda$ is bounded by $T_x^\lambda T_y^\lambda$, we can evaluate the sums in $\mathtt{foldy}$ for all $(\hat{x},\hat{y}) \in \big[-\lceil\frac{T_x^\lambda-1}{2}\rceil : \lfloor\frac{T_x^\lambda-1}{2}\rfloor\big] \times \big[-\lceil\frac{T_y^\lambda-1}{2}\rceil : \lfloor\frac{T_y^\lambda-1}{2}\rfloor\big]$ in only $\mathcal{O}(T_x^\lambda T_y^\lambda)$. The dominating computational effort is thus the $\mathtt{ift}$ whose complexity is known to be $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$; see Fact 2.10. ∎

The following diagram can be helpful for understanding the above arguments:



First, the folding lemma (Lemma 3.30) establishes arrow C by going along A, B, and the inverse of D. Then Lemma 3.31 proves that the overall complexity of going along arrows C and D is $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$. Note that this is less than going along A and B since in step A we would destroy the sparsity structure of $\operatorname{supp} \hat{f}\hat{\psi}_\lambda$.

With the work done so far, proving the overall complexity of the `rt` becomes easy.

THEOREM 3.32 (complexity of `rt`). *The complexity of evaluating* `rt` *for all* $\lambda \in \Lambda_{\text{fin}}$ *and* $t \in T^\lambda$ *is*

$$\mathcal{O}\big(\log(\rho_x\rho_y)\,\rho_x\rho_y(J4^J)\big).$$

Proof. By Lemma 3.31 and Definition 3.8, we know that the complexity of evaluating $\mathtt{rt}[\hat{f}](\lambda, t)$ for some fixed $\lambda \in \Lambda$ and all $t \in T^\lambda$ is $\mathcal{O}\big(\log(\rho_x\rho_y)\,\rho_x\rho_y\, j\, 2^j\big)$. On a fixed scale j we have $\mathcal{O}(2^j)$ ridgelets, and thus the overall complexity is

$$\sum_{j=0}^{J} \mathcal{O}\big(2^j\big)\,\mathcal{O}\big(\log(\rho_x\rho_y)\,\rho_x\rho_y\, j\, 2^j\big) = \mathcal{O}\big(\log(\rho_x\rho_y)\,\rho_x\rho_y(J4^J)\big). \qquad \square$$

The algorithm for evaluating the `irt` is derived in exactly the same way.

DEFINITION 3.33 (unfolding operations). *Let* $n \in \mathbb{N}$, N_x, $N_y \in 2\mathbb{N}$,

$$\widehat{\Omega}_> = \left[-\frac{nN_x}{2} : \frac{nN_x}{2} - 1\right] \times \left[-\frac{N_y 1}{2} : \frac{N_y}{2} - 1\right],$$
$$\widehat{\Omega}_< = \left[-\frac{N_x}{2} : \frac{N_x}{2} - 1\right] \times \left[-\frac{N_y 1}{2} : \frac{N_y}{2} - 1\right],$$

and $\hat{f}_< : \widehat{\Omega}_< \to \mathbb{C}$. *Then* $\mathtt{unfoldx}[\hat{f}_<, n, N_x] : \widehat{\Omega}_> \to \mathbb{C}$ *is defined as*

$$\mathtt{unfoldx}[\hat{f}_<, n, N_x](\hat{x}, \hat{y}) := \frac{1}{n} f_<\left(\hat{x} - N_x\left[\frac{\hat{x}}{N_x}\right], \hat{y}\right)$$

($[\cdot]$ *denotes rounding to the nearest integer, with tie-breaking in favor of the* next larger *number, i.e.,* $\left[-\frac{3}{2}\right] = -1$*) and the symbol* `unfoldx` *is called the* `x`-unfolding operator. *Note that, strictly speaking, the first argument is not the same as* $\hat{x} \bmod N_x$*, since the latter is in* $[0 : N_x)$ *instead of* $\widehat{\Omega}_<$.

The y-unfolding operator `unfoldy` *is defined likewise.*

LEMMA 3.34 (unfolding lemma). *Let* $n \in \mathbb{N}$, N_x, $N_y \in 2\mathbb{N}$, L_x, $L_y \in \mathbb{R}^{>0}$, $\Omega_>$, $\widehat{\Omega}_>$, $\Omega_<$, $\widehat{\Omega}_<$ *as in Lemma* 3.30*, and let* $\hat{f}_< : \widehat{\Omega}_< \to \mathbb{C}$*. Furthermore, let* $f_> : \Omega_> \to \mathbb{C}$ *be given by*

$$f_>(x,y) := \begin{cases} \mathtt{ift}[\hat{f}_<](x,y) & \text{if } (x,y) \in \Omega_<, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\mathtt{ft}[f_>] = \mathtt{unfoldx}[\hat{f}_<, n, N_x].$$

Proof. As before, we expand

$$\begin{aligned} \mathtt{ft}[f_>](\hat{x},\hat{y}) &= \frac{1}{n\,N_xN_y} \sum_{(x,y)\in\Omega_>} f_>(x,y)\,\mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x}+\hat{y}\frac{y}{L_y})} \\ &= \frac{1}{n\,N_xN_y} \sum_{(x,y)\in\Omega_<} \mathtt{ift}[\hat{f}_<](x,y)\,\mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{x}{L_x}+\hat{y}\frac{y}{L_y})} \\ &= \frac{1}{n}\frac{1}{N_xN_y} \sum_{(x,y)\in\Omega_<} \mathtt{ift}[\hat{f}_<](x,y)\,\mathrm{e}^{-2\pi\mathrm{i}\,\left((\hat{x}-N_x[\frac{\hat{x}}{N_x}])\frac{x}{L_x}+\hat{y}\frac{y}{L_y}\right)}. \end{aligned}$$

Again, the additional term in the exponential can be added because $N_x\frac{x}{L_x}$ is always an integer. Continuing from above,

$$\begin{aligned} \mathtt{ft}[f_>](\hat{x},\hat{y}) &= \frac{1}{n}\,\mathtt{ft}[\mathtt{ift}[\hat{f}_<]]\left(\hat{x} - N_x\Big[\frac{\hat{x}}{N_x}\Big],\,\hat{y}\right) \\ &= \frac{1}{n}\,\hat{f}_<\left(\hat{x} - N_x\Big[\frac{\hat{x}}{N_x}\Big],\,\hat{y}\right) \\ &= \mathtt{unfoldx}[\hat{f}_<, n, N_x](\hat{x},\,\hat{y}). \qquad \square \end{aligned}$$

LEMMA 3.35. *The complexity of evaluating*

$$\hat{f}_\lambda(\hat{x},\hat{y}) := \sum_{t\in T^\lambda} \tilde{f}(\lambda,t)\,\hat{\psi}_\lambda(\hat{x},\hat{y})\,\mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}$$

for some fixed $\lambda \in \Lambda$ *and all* $(\hat{x},\hat{y}) \in \operatorname{supp}\hat{\psi}_\lambda$ *is* $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$.

Proof. We will show the lemma only for the case $\lambda = (j, \kappa = \mathtt{x}, k)$. For $\kappa \in \{\mathtt{y},\mathtt{d}\}$, the proof is analogous, and for $\kappa = \mathtt{s}$ it is trivial.

Note that the above expression corresponds to $\hat{f}_\lambda = T_x^\lambda T_y^\lambda\,\mathtt{ft}[\tilde{f}(\lambda,\cdot)]$, except that we want the domain of $\mathtt{ft}[\tilde{f}(\lambda,\cdot)]$ to be not

$$\widehat{\Omega}_< := \left[-\frac{T_x^\lambda}{2} : \frac{T_x^\lambda}{2}\right] \times \left[-\frac{T_y^\lambda}{2} : \frac{T_y^\lambda}{2}\right]$$

but rather a subset of

$$\widehat{\Omega}_> := \left[-\frac{T_x^\lambda}{2} : \frac{T_x^\lambda}{2}\right] \times \left[-\frac{(|k|+1)\,T_y^\lambda}{2} : \frac{(|k|+1)\,T_y^\lambda}{2}\right].$$

We can achieve this by defining

$$T_>^\lambda := \left[0 : \frac{L_x}{T_x^\lambda} : L_x\right) \times \left[0 : \frac{L_y}{(|k|+1)\,T_y^\lambda} : L_y\right)$$

and

$$\tilde{f}_>(\lambda,\cdot) : T_>^\lambda \to \mathbb{C},\ t \mapsto \begin{cases} \tilde{f}(\lambda,t) & \text{if } t \in T^\lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Because all new terms in the `ft` sum are zero, writing $\hat{f}_\lambda = T_x^\lambda\,(|k|+1)\,T_y^\lambda\,\texttt{ft}[\tilde{f}_>(\lambda,\cdot)]$ is now correct both in terms of equal values as well as equal domains. Running the `ft` on a function which is mostly zero seems to be a waste of effort, however, and indeed the unfolding lemma (Lemma 3.34) shows that the above expression is equivalent to

$$\hat{f}_\lambda(\hat{x},\hat{y}) = T_x^\lambda\,(|k|+1)\,T_y^\lambda\,\texttt{unfoldy}\Big[\texttt{ft}[\tilde{f}(\lambda,\cdot)],\ |k|+1,\ T_y^\lambda\Big](\hat{x},\hat{y}).$$

We evaluate the `unfoldy` only on $\mathcal{O}(T_x^\lambda T_y^\lambda)$ points, and each point evaluation can be done in $\mathcal{O}(1)$ once $\texttt{ft}[\tilde{f}(\lambda,\cdot)]$ is available. The dominating computational effort is thus computing the `ft`, which can be done in $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$; see Fact 2.10. ∎

Again, it is helpful to visualize what is happening using a diagram:

$$\begin{array}{ccc} \tilde{f}(\lambda,\cdot) : T^\lambda \to \mathbb{C} & \xrightarrow[\text{extend domain}]{\text{C}} & \tilde{f}_>(\lambda,\cdot) : T_>^\lambda \to \mathbb{C} \\ \Big\downarrow \text{A}\ \texttt{ft} & & \texttt{ft}\ \text{D} \Big\downarrow \\ \texttt{ft}[\tilde{f}(\lambda,\cdot)] : \widehat{\Omega}_< \to \mathbb{C} & \xrightarrow[\text{B}]{\texttt{unfold}} & \hat{f}_\lambda : \operatorname{supp}\hat{\psi}_\lambda \subset \widehat{\Omega}_> \to \mathbb{C} \end{array}$$

The unfolding lemma (Lemma 3.34) establishes arrow B by going along the inverse of A and C, D. Then Lemma 3.35 proves that the overall complexity of going along arrows A and B is $\mathcal{O}\big(T_x^\lambda T_y^\lambda \log(T_x^\lambda T_y^\lambda)\big)$. Note that this is less than going along C and D since in step D we would have to compute the `ft` of the extended $\tilde{f}_>(\lambda,\cdot)$.

As before, the overall complexity of the `irt` simply follows from the above lemma.

THEOREM 3.36 (complexity of `irt`). *The complexity of evaluating* `irt` *for all* $(\hat{x},\hat{y})$ *in its domain is*

$$\mathcal{O}\big(\log(\rho_x\rho_y)\,\rho_x\rho_y(J4^J)\big).$$

Proof. The only thing different from the proof of the complexity of `rt` in Theorem 3.32 is that in the end we have to sum up all the $\hat{f}_\lambda$. Since every point in the Fourier space lies in the support of at most four ridgelets and we evaluate the `irt` on $\mathcal{O}\big(\rho_x\rho_y(4^J)\big)$ points, that sum does not dominate the overall complexity. ∎

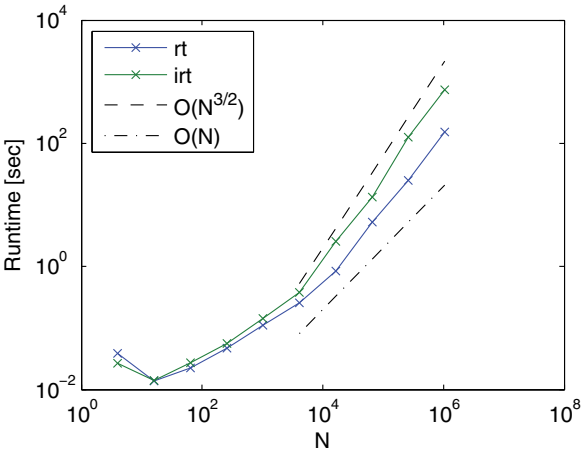


FIG. 4. *Scaling plot of* `rt` *and* `irt` *with* $\rho_x = \rho_y = 1$ *and* $J \in [1:10]$ *compared to* $N = 4^J$.

We have seen that the finite ridgelet transforms can in theory achieve an almost optimal computational complexity of $\mathcal{O}(N \log(N))$, where N is (proportional to) the number of input and output parameters. Achieving the same complexity in a practical implementation requires a lot of work, however, since the `fold` and `unfold` methods need to be able to exploit the special sparsity structures exhibited by the supports of the ridgelets [13]. The MATLAB implementation for this paper achieves only a theoretical scaling of $\mathcal{O}(N^{\frac{3}{2}})$, since the implementation was done using built-in MATLAB algorithms for folding and unfolding (see Figure 4); however, work on an implementation achieving $\mathcal{O}(N \log(N))$ is in progress.

3.5. Scalar products. To differentiate the different scalar products in the spaces we introduced, we mark them by the respective space they're operating in. The following collects their precise definitions and correspondences.

DEFINITION 3.37. *Let* N_x, N_y, *as well as* Ω_{fin}, $\widehat{\Omega}_{\text{fin}}$, *be as in Definition* 3.24 *of the finite Fourier transform, respectively, as in Remark* 3.25. *For* $f, g : \Omega_{\text{fin}} \to \mathbb{C}$, *the finite physical space scalar product of* f *and* g *is given by*

$$\langle f, g \rangle_{\Omega_{\text{fin}}} := \frac{1}{N_x N_y} \sum_{(x,y) \in \Omega_{\text{fin}}} f(x,y)\, \overline{g(x,y)}.$$

Furthermore, for $\hat{f}, \hat{g} : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$, *the finite Fourier space scalar product of* $\hat{f}$ *and* $\hat{g}$ *is given by*

$$\langle \hat{f}, \hat{g} \rangle_{\widehat{\Omega}_{\text{fin}}} := \sum_{(\hat{x},\hat{y}) \in \widehat{\Omega}_{\text{fin}}} \hat{f}(\hat{x},\hat{y})\, \overline{\hat{g}(\hat{x},\hat{y})}.$$

Let $\tilde{f}, \tilde{g} : \widetilde{\Omega}_{\text{fin}} \to \mathbb{C}$. *Then the finite ridgelet coefficient space scalar product of* $\tilde{f}$ *and* $\tilde{g}$ *is given by*

$$\langle \tilde{f}, \tilde{g} \rangle_{\widetilde{\Omega}_{\text{fin}}} := \sum_{\lambda \in \Lambda_{\text{fin}}} T_x^\lambda T_y^\lambda \sum_{t \in T^\lambda} \tilde{f}(\lambda, t)\, \overline{\tilde{g}(\lambda, t)}.$$

The above scalar products correspond to the standard ℓ^2 scalar product with some additional prefactors. The purpose of these prefactors is to assert that the scalar products are preserved under the Fourier and ridgelet transforms.

In (3.4), we list the correspondences between these different products, respectively, where they fail (in general):

$$\text{(3.4a)}\qquad \langle f, g\rangle_{\Omega_{\text{fin}}} = \langle \mathtt{ft}[f], \mathtt{ft}[g]\rangle_{\widehat{\Omega}_{\text{fin}}}, \qquad \langle \hat{f}, \hat{g}\rangle_{\widehat{\Omega}_{\text{fin}}} \neq \langle \mathtt{rt}[\hat{f}], \mathtt{rt}[\hat{g}]\rangle_{\widetilde{\Omega}_{\text{fin}}};$$

$$\text{(3.4b)}\qquad \langle f, \mathtt{ift}[\hat{g}]\rangle_{\Omega_{\text{fin}}} = \langle \mathtt{ft}[f], \hat{g}\rangle_{\widehat{\Omega}_{\text{fin}}}, \qquad \langle \hat{f}, \mathtt{irt}[\tilde{g}]\rangle_{\widehat{\Omega}_{\text{fin}}} = \langle \mathtt{rt}[\hat{f}], \tilde{g}\rangle_{\widetilde{\Omega}_{\text{fin}}};$$

$$\text{(3.4c)}\qquad \langle \mathtt{ift}[\hat{f}], \mathtt{ift}[\hat{g}]\rangle_{\Omega_{\text{fin}}} = \langle \hat{f}, \hat{g}\rangle_{\widehat{\Omega}_{\text{fin}}}, \qquad \langle \mathtt{irt}[\tilde{f}], \mathtt{irt}[\tilde{g}]\rangle_{\widehat{\Omega}_{\text{fin}}} \neq \langle \tilde{f}, \tilde{g}\rangle_{\widetilde{\Omega}_{\text{fin}}}.$$

By way of explanation, the first half of (3.4a) is just the finite analogue of the Plancherel formula (2.3) and is proved in exactly the same way with the corresponding finite counterparts. Since the finite Fourier transforms are mutually inverse, we can substitute f with $\mathtt{ift}[\hat{f}]$ or g with $\mathtt{ift}[\hat{g}]$, and the other equalities on the left-hand side follow immediately.

For the second equality of (3.4b), the proof is

$$\begin{aligned}
&\langle \hat{f},\, \mathtt{irt}[\tilde{g}]\rangle_{\widehat{\Omega}_{\text{fin}}} \\
&\quad= \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\text{fin}}} \hat{f}(\hat{x},\hat{y}) \overline{\sum_{\lambda\in\Lambda_{\text{fin}}} \sum_{t\in T^\lambda} \tilde{g}(\lambda,t)\, \hat{\psi}_\lambda(\hat{x},\hat{y})\, \mathrm{e}^{-2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})}} \\
&\quad= \sum_{\lambda\in\Lambda_{\text{fin}}} T_x^\lambda T_y^\lambda \sum_{t\in T^\lambda} \left(\frac{1}{T_x^\lambda T_y^\lambda} \sum_{(\hat{x},\hat{y})\in\widehat{\Omega}_{\text{fin}}} \hat{f}(\hat{x},\hat{y}) \hat{\psi}_\lambda(\hat{x},\hat{y})\, \mathrm{e}^{2\pi\mathrm{i}\,(\hat{x}\frac{t_x}{L_x}+\hat{y}\frac{t_y}{L_y})} \right) \overline{\tilde{g}(\lambda,t)} \\
&\quad= \langle \mathtt{rt}[\hat{f}],\, \tilde{g}\rangle_{\widetilde{\Omega}_{\text{fin}}}.
\end{aligned}$$

The right-hand equality of (3.4a) fails because while $\hat{\mathcal{R}}^{-1} \circ \hat{\mathcal{R}} = \mathbb{I}$, $\mathtt{irt} \circ \mathtt{rt}$ holds only on the region where the frame constitutes a partition of unity (compare Lemma 3.27), which in this case is not the entire domain of $\hat{g}$. For example, if we have a frame with highest scale J, $\rho_x = \rho_y = \rho$, and choose

$$\hat{g}(\hat{x},\hat{y}) = \delta\big(\hat{x} + \rho 2^{J+1}\big) + \delta\big(\hat{y} + \rho 2^{J+1}\big),$$

we have that $\mathtt{irt}[\mathtt{rt}[\hat{g}]](\hat{x},\hat{y}) = 0$ for all $(\hat{x},\hat{y})$ because no ridgelet covers that highest frequency part. With such a $\hat{g}$, $\langle \mathtt{rt}[\hat{f}], \mathtt{rt}[\hat{g}]\rangle_{\widetilde{\Omega}_{\text{fin}}}$ is zero for all $\hat{f}$, while $\langle \hat{f}, \hat{g}\rangle_{\widehat{\Omega}_{\text{fin}}}$ isn't necessarily.

Finally, the right-hand equality of (3.4c) would not even become an equality for the infinite ridgelet transforms, because due to the redundancy in the ridgelet coefficient space (compare Remark 3.28), the inverse ridgelet transform and its finite analogue must have a kernel, say K (for the latter). If we thus choose, e.g., $\tilde{f}, \tilde{g} \in K$, then $\langle \mathtt{irt}[\tilde{f}], \mathtt{irt}[\tilde{g}]\rangle_{\widehat{\Omega}_{\text{fin}}}$ is zero, while $\langle \tilde{f}, \tilde{g}\rangle_{\widetilde{\Omega}_{\text{fin}}}$ doesn't have to be.

4. Radiative transport equation.

4.1. Basic RTE solver. In order to get started, let us consider the following simplified version of the RTE

$$\text{(4.1)}\qquad \mathcal{A}u = \vec{s}\cdot\nabla u + \kappa\, u = f,$$

where $u, \kappa, f : \Omega \to \mathbb{R}$, i.e., we neglect scattering and consider only a fixed direction $\vec{s} \in \mathbb{S}^1$ (we set again $\Omega := [0, L_x) \times [0, L_y)$ and $\mathbb{S}^1 := \big\{\vec{s} \in \mathbb{R}^2 \,\big|\, \|\vec{s}\|_2 = 1\big\}$). In the following, we will write the solution u as the inverse Fourier transform $\mathcal{F}^{-1}[\hat{u}]$ of some

Fourier space function $\hat{u}$. A necessary consequence of this approach is that we must consider periodic boundary conditions

$$u(0,y) = u(L_x,y), \quad u(x,0) = u(x,L_y) \quad \forall x \in [0,L_x) \text{ and } y \in [0,L_y)$$

(recall the discussion on the inherent periodicity of the Fourier transforms after Definition 2.4). Although such boundary conditions are rarely physically justified, they do not prevent us from solving real-world problems either. For example, if we want to allow actual outflow, we can enlarge the domain slightly and artificially increase κ in this area such that all the radiation is absorbed before reentering Ω. Similarly, we can impose inflow boundary conditions by using appropriate forcing terms (combined with the absorption trick above), again in an enlarged domain; see subsection 5.3.

Due to the simple representation (and computability) of the operator $\mathcal{A}$ in Fourier space, we want to apply the operator in this form during the iteration.

DEFINITION 4.1 (Fourier space RTE operator). *Let $L_x, L_y \in \mathbb{N}$, $\vec{s} \in \mathbb{S}^1$, $\kappa : [0,L_x) \times [0,L_y) \to \mathbb{R}$, and $\hat{f} : \mathbb{Z}^2 \to \mathbb{C}$. Then $\widehat{\mathcal{A}}[\hat{f}] : \mathbb{Z}^2 \to \mathbb{C}$ is a function given by*

$$\widehat{\mathcal{A}}[\hat{f}](\hat{x},\hat{y}) := 2\pi \mathrm{i}\, \vec{s} \cdot \vec{\xi} \hat{f}(\hat{x},\hat{y}) + \mathcal{F}\big[\kappa\, \mathcal{F}^{-1}[\hat{f}]\big](\hat{x},\hat{y}),$$

where $\vec{\xi} = \left(\frac{\hat{x}}{L_x}, \frac{\hat{y}}{L_y}\right)^\top$ and the symbol $\widehat{\mathcal{A}}$ is called the Fourier space RTE operator.

To still be able to exploit the advantageous properties of the ridgelets, this makes it necessary to first transform from ridgelet coefficients to Fourier space, then apply $\widehat{\mathcal{A}}$ and then transform back to ridgelet coefficients.

DEFINITION 4.2 (ridgelet coefficient space RTE operator). *Let $\tilde{f} : \widetilde{\Omega} \to \mathbb{C}$. Then $\widetilde{\mathcal{A}}[\tilde{f}](\lambda,t) : \widetilde{\Omega} \to \mathbb{C}$ is a function given by*

$$\widetilde{\mathcal{A}}[\tilde{f}](\lambda,t) := \widehat{\mathcal{R}}\Big[\widehat{\mathcal{A}}\left[\widehat{\mathcal{R}}^{-1}[\tilde{f}]\right]\Big](\lambda,t)$$

and the symbol $\widetilde{\mathcal{A}}$ is called the ridgelet coefficient space RTE operator.

It can easily be verified that

$$s \cdot \nabla u + \kappa\, u = f \iff \widehat{\mathcal{A}}[\mathcal{F}[u]] = \mathcal{F}[f] \iff \widetilde{\mathcal{A}}\big[\widehat{\mathcal{R}}[\mathcal{F}[u]]\big] = \widehat{\mathcal{R}}[\mathcal{F}[f]]. \tag{4.2}$$

Since both $\mathcal{F}$ and $\widehat{\mathcal{R}}$ have left inverses, we can thus solve (4.1) by solving

$$\widetilde{\mathcal{A}}[\tilde{u}] = \widehat{\mathcal{R}}[\mathcal{F}[f]], \quad u = \mathcal{F}^{-1}\big[\widehat{\mathcal{R}}^{-1}[\tilde{u}]\big]. \tag{4.3}$$

In this abstract formulation, problem (4.2) reads

$$\mathcal{B}[u] = g, \tag{4.4}$$

where u and g are taken from some vector space V, and $\mathcal{B} : V \to V$ is a linear operator. For many iterative methods to work (including the CG method we want to use), $\mathcal{B}$ has to be self-adjoint and positive (semi)definite in the chosen norm, i.e., $\langle f, \mathcal{B}[g]\rangle = \langle \mathcal{B}[f], g\rangle$ and $\langle f, \mathcal{B}[f]\rangle \geq 0$ have to hold. If $\mathcal{B}$ does not yet satisfy these conditions, the standard approach is to solve the normal equations

$$\mathcal{B}^*\mathcal{B}[u] = \mathcal{B}^*[g]$$

instead of (4.4), where $\mathcal{B}^*$ denotes the adjoint of $\mathcal{B}$. Because $\mathcal{A}$ and $\widehat{\mathcal{A}}$ don't satisfy the "self-adjoint and positive semidefinite" criterion, we will have to use the normal equations as well.

Of course, also this problem cannot yet be solved numerically, as it involves infinitely many equations in infinitely many unknowns. But with the finite transforms developed in sections 2 and 3, the following discretization comes naturally.

DEFINITION 4.3 (finite Fourier space RTE operator). *Let* $L_x, L_y \in \mathbb{N}$, $s \in \mathbb{S}^1$, $N_x, N_y \in 2\mathbb{N}$,

$$\Omega_{\text{fin}} = \left[0 : \frac{L_x}{N_x} : L_x\right) \times \left[0 : \frac{L_y}{N_y} : L_y\right),$$
$$\widehat{\Omega}_{\text{fin}} = \left[-\frac{N_x}{2} : \frac{N_x}{2} - 1\right] \times \left[-\frac{N_y}{2} : \frac{N_y}{2} - 1\right],$$

$\kappa : \Omega_{\text{fin}} \to \mathbb{R}$, *and* $\hat{f} : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$. *Then* $\widehat{\mathcal{A}}_{\text{fin}}[\hat{f}] : \widehat{\Omega}_{\text{fin}} \to \mathbb{C}$ *is a function given by*

$$\widehat{\mathcal{A}}_{\text{fin}}[\hat{f}](\hat{x}, \hat{y}) := 2\pi \mathrm{i}\, \vec{s} \cdot \vec{\xi} \hat{f}(\hat{x}, \hat{y}) + \mathtt{ft}\big[\kappa\, \mathtt{ift}[\hat{f}]\big](\hat{x}, \hat{y})$$

and the symbol $\widehat{\mathcal{A}}_{\text{fin}}$ *is called the* finite Fourier space RTE operator.

To formulate the normal equations, we determine the adjoint, which is easily calculated,

$$\widehat{\mathcal{A}}^*_{\text{fin}}[\hat{f}](\hat{x}, \hat{y}) = -2\pi \mathrm{i}\, \vec{s} \cdot \vec{\xi} \hat{f}(\hat{x}, \hat{y}) + \mathtt{ft}\big[\bar{\kappa}\, \mathtt{ift}[\hat{f}]\big](\hat{x}, \hat{y}).$$

With these new operators, the normal equation to (4.3) becomes

$$\widetilde{\mathcal{B}}_{\text{fin}} := \mathtt{rt}\Big[\widehat{\mathcal{A}}^*_{\text{fin}} \widehat{\mathcal{A}}_{\text{fin}}\big[\mathtt{irt}[\tilde{u}]\big]\Big] = \mathtt{rt}\Big[\widehat{\mathcal{A}}^*_{\text{fin}}\big[\mathtt{ft}[f]\big]\Big], \qquad u = \mathtt{ift}[\mathtt{irt}[\tilde{u}]], \tag{4.5}$$

where here u and f mean the u and f from (4.1) sampled on the grid defined by the (i)ft. While it is clear that this equation corresponds to a linear system of equations, formulating this system explicitly can nevertheless be tedious. Luckily, though, doing so is neither necessary nor advisable. Rather, we can use iterative methods like the *conjugate gradient* (CG) iterations, which have the benefit that they are able to deal with vectors in an abstract sense, i.e., vectors which only satisfy the vector axioms but are not necessarily taken from $\mathbb{R}^n$, and abstract linear operators acting on such vectors.

It is easy to see, however, that the directional derivative term in $\widehat{\mathcal{A}}_{\text{fin}}$ leads to very ill conditioned operators. Therefore one needs to apply a preconditioner—which was constructed in [16]—such that the final equation reads

$$\widetilde{\mathcal{D}}_{\text{fin}} \widetilde{\mathcal{B}}_{\text{fin}} \widetilde{\mathcal{D}}_{\text{fin}} \tilde{u}_p = \widetilde{\mathcal{D}}_{\text{fin}}\Big[\mathtt{rt}\Big[\widehat{\mathcal{A}}^*_{\text{fin}}\big[\mathtt{ft}[f]\big]\Big]\Big], \qquad u = \mathtt{ift}\Big[\mathtt{irt}\Big[\widetilde{\mathcal{D}}_{\text{fin}} \tilde{u}_p\Big]\Big], \tag{4.6}$$

where the subscript p is intended to distinguish the solutions to the preconditioned problem from the solutions to (4.5). The preconditioner is defined as

$$\widetilde{\mathcal{D}}_{\text{fin}}[\tilde{f}](\lambda, t) := \frac{\tilde{f}(\lambda, t)}{1 + 2^j |\vec{s}_\lambda \cdot \vec{s}|},$$

where $\vec{s}$ is the transport direction in the RTE and

$$\vec{s}_\lambda = \begin{cases} (0,0)^\top & \text{if } \kappa = \mathtt{s}, \\ (1, \frac{k}{2^{j-1}})^\top & \text{if } \kappa = \mathtt{x}, \mathtt{d}, \\ (\frac{k}{2^{j-1}}, 1)^\top & \text{if } \kappa = \mathtt{y} \end{cases}$$

is the direction of the ridgelet $\hat{\psi}_\lambda$ in Fourier space. It was shown in [16] that this definition of the preconditioner $\widetilde{\mathcal{D}}_{\text{fin}}$ leads to a uniform bound on the condition number of the Galerkin matrix.

4.2. Convergence of basic RTE solver. The following discussion of the convergence only deals with the case of constant $\kappa > 0$, mainly in the interest of saving space, since the added technical difficulties of nonconstant κ are not very illuminating. Furthermore, a much more general convergence theory for ridgelet solvers is forthcoming in [19].

We begin with some anisotropic Sobolev spaces and a lemma about estimating a quantity that will occur later in these norms.

DEFINITION 4.4. *Let $\vec{s} \in \mathbb{S}^2$; then we define the* anisotropic Sobolev space

$$H^{k+\vec{s}}(\Omega) := \big\{ f \in L^2(\Omega) \,\big|\, (\vec{s}\cdot\nabla) f \in H^k(\Omega) \big\}.$$

It is equipped with the norm

$$\|f\|^2_{H^{k+\vec{s}}(\Omega)} := \|f\|^2_{H^k(\Omega)} + \|(\vec{s}\cdot\nabla) f\|^2_{H^k(\Omega)}.$$

We set $H^{\vec{s}} := H^{0+\vec{s}}$. These spaces are more easily characterized on the Fourier side,

$$H^{k+\vec{s}}(\widehat{\Omega}) := \Big\{ \hat{f} \in L^2(\widehat{\Omega}) \,\Big|\, \big\langle \vec{s}\cdot\vec{\xi} \big\rangle \big\langle \vec{\xi} \big\rangle^k \hat{f}(\hat{x},\hat{y}) \in L^2(\widehat{\Omega}) \Big\},$$

with norm

$$\big\|\hat{f}\big\|_{H^{k+\vec{s}}(\widehat{\Omega})} := \big\| \big\langle \vec{s}\cdot\vec{\xi} \big\rangle \big\langle \vec{\xi} \big\rangle^k \hat{f} \big\|_{L^2(\widehat{\Omega})}.$$

The finite-dimensional spaces $H^{k+\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})$ and their norm are defined accordingly.

LEMMA 4.5. *For $f \in H^k$, the solution u of $\mathcal{A}^*\mathcal{A}u = f$ is in $H^{k+\vec{s}}$, and we have that*

$$\big\| \hat{u}|_{\widehat{\Omega}\setminus[-\frac{N}{2}:\frac{N}{2})^2} \big\|_{H^{\vec{s}}(\widehat{\Omega})} \le \mathcal{O}\big(N^{-k}\big).$$

Proof. The added regularity can be shown by a variation of constants and a bootstrapping argument; however, for the case of constant κ it is trivial, as we can explicitly calculate the solution

$$\hat{u} = \frac{\hat{f}}{2\pi \mathrm{i} \vec{s}\cdot\vec{\xi} + \kappa}$$

in Fourier space, and the weight corresponding to the differentiation $\vec{s}\cdot\nabla$ is balanced by the denominator,

$$\frac{(1 + (\vec{s}\cdot\vec{\xi})^2)^{\frac{1}{2}}}{\kappa + 2\pi\mathrm{i}\vec{s}\cdot\vec{\xi}} \lesssim 1,$$

since $0 < \kappa < \infty$. Therefore, we can estimate

$$\big\| \hat{u}|_{\widehat{\Omega}\setminus[-\frac{N}{2}:\frac{N}{2})^2} \big\|_{H^{\vec{s}}(\widehat{\Omega})} \lesssim \big\| \underbrace{\big\langle \vec{\xi} \big\rangle^{-k}}_{=\mathcal{O}(N^{-k}) \text{ on } \widehat{\Omega}\setminus[-\frac{N}{2}:\frac{N}{2})^2} \overbrace{\big\langle \vec{s}\cdot\vec{\xi} \big\rangle \big\langle \vec{\xi} \big\rangle^k \hat{u}|_{\widehat{\Omega}\setminus[-\frac{N}{2}:\frac{N}{2})^2}}^{\in L^2(\widehat{\Omega})} \big\|_{L^2(\widehat{\Omega})} = \mathcal{O}(N^{-k}). \qquad \square$$

THEOREM 4.6. *Let $f \in H^k$ and let u be the solution to*

$$\mathcal{A}^*\mathcal{A}u = \mathcal{A}^* f. \tag{4.7}$$

Furthermore, let $N = 2^J$, where J is the highest scale in the selected subframe for the ridgelet solver. Then the output $\hat{u}_j$ of running the solver for the subframe up to scale $j \leq J$, satisfies the error

$$\left\|u - \mathcal{F}^{-1}\mathcal{Z}_N \hat{u}_j\right\|_{H^{\vec{s}}(\Omega)} \leq \mathrm{err}_N + \mathrm{err}_{\mathcal{F}} + \mathrm{err}_j,$$

where $\mathrm{err}_N = \mathcal{O}(N^{-k})$ *and* $\mathrm{err}_j = \mathcal{O}(2^{-jk})$.

The error $\mathrm{err}_{\mathcal{F}} =: \left\|\mathtt{ft}[f] - \mathcal{F}(f)\big|_{\widehat{\Omega}_{\mathrm{fin}}}\right\|_{L^2(\widehat{\Omega}_{\mathrm{fin}})}$ *stems from the discretization of the continuous Fourier transform, respectively, the FFT-algorithm. Numerical evidence suggests that* $\mathrm{err}_{\mathcal{F}} = \mathcal{O}(N^{-k})$ *as well (compare Table 2). However, we have no proof for this and didn't find a suitable reference which deals with Sobolev spaces (as opposed to classical derivatives or simply L^2-functions).*

Proof. We define the sesquilinear form

$$a(u,v) := \left\langle \mathcal{A}u, \mathcal{A}v\right\rangle_{L^2(\Omega)} = \left\langle \widehat{\mathcal{A}}\hat{u}, \widehat{\mathcal{A}}\hat{v}\right\rangle_{L^2(\widehat{\Omega})} =: \hat{a}(\hat{u},\hat{v})$$

and the linear functional

$$\ell(v) := \left\langle \mathcal{A}^* f, v\right\rangle_{L^2(\Omega)} = \left\langle \widehat{\mathcal{A}}^* \hat{f}, \hat{v}\right\rangle_{L^2(\widehat{\Omega})} =: \hat{\ell}(\hat{v}).$$

Both are continuous, and also a is coercive due to $\kappa > 0$. Then u from (4.7) is also the unique solution of

$$u \in H^{\vec{s}}(\Omega)\colon a(u,v) = \ell(v) \quad \forall v \in H^{\vec{s}}(\Omega).$$

By introducing $\widehat{V}_N := \left\{\hat{f}\colon [-\frac{N}{2} : \frac{N}{2})^2 \to \mathbb{C}\right\}$ and $V_N = \mathcal{F}^{-1}\mathcal{Z}_N\widehat{V}_N$, we are abusing notation somewhat (since the spaces don't actually correspond to each other via the Fourier transform), but this allows the following to be presented more concisely. Note that $V_N \subset H^{\vec{s}}(\Omega)$ and that $\widehat{V}_N$ is built over a finite domain. Let

$$u_N \in V_N\colon a(u_N, v_N) = \ell(v_N) \quad \forall v_N \in V_N. \tag{4.8}$$

The norm $\|\cdot\|_a$ induced by the inner product $a(\cdot,\cdot)$, is equivalent to $\|\cdot\|_{H^{\vec{s}}(\Omega)}$, and Céa's lemma yields

$$\|u - u_N\|_{H^{\vec{s}}(\Omega)} \lesssim \|u - u_N\|_a \leq \inf_{v_N \in V_N} \|u - v_N\|_a \lesssim \inf_{v_N \in V_N} \|u - v_N\|_{H^{\vec{s}}(\Omega)}.$$

Considering this difference on the Fourier side, we can choose $\hat{v}_N = \mathcal{F}(u)\big|_{[-N:N]^2}$, and thus the error can be estimated by Lemma 4.5,

$$\begin{aligned}\mathrm{err}_N &:= \|u - u_N\|_{H^{\vec{s}}(\Omega)} \\ &\leq \|u - v_N\|_{H^{\vec{s}}(\Omega)} = \|\hat{u} - \hat{v}_N\|_{H^{\vec{s}}(\widehat{\Omega})} = \left\|\hat{u}\big|_{\widehat{\Omega}\setminus[-\frac{N}{2}:\frac{N}{2})^2}\right\|_{H^{\vec{s}}(\widehat{\Omega})} \leq \mathcal{O}\big(N^{-k}\big).\end{aligned}$$

Next, consider the operator $\widehat{\mathcal{A}}_{\mathrm{fin}}$, which is equal to the restriction of $\widehat{\mathcal{A}}$ to $[-\frac{N}{2} : \frac{N}{2})^2$, since κ is constant. The solution of (4.8), u_N, is thus also a solution to

$$\hat{u}_N\colon \widehat{\mathcal{A}}_N^*\widehat{\mathcal{A}}_N\hat{u}_N = (-2\pi\mathrm{i}\vec{s}\cdot\vec{\xi} + \kappa)(2\pi\mathrm{i}\vec{s}\cdot\vec{\xi} + \kappa)\hat{u}_N = \widehat{\mathcal{A}}_N^*\hat{f}_N,$$

where $\hat{f}_N = \mathcal{F}(f)\big|_{[-\frac{N}{2}:\frac{N}{2})^2}$. We want to compare this to the solution u_{fin} of

$$\hat{u}_{\mathrm{fin}}\colon \widehat{\mathcal{A}}_N^*\widehat{\mathcal{A}}_N\hat{u}_{\mathrm{fin}} = \widehat{\mathcal{A}}_N^*\hat{f}_{\mathrm{fin}} = (-2\pi\mathrm{i}\vec{s}\cdot\vec{\xi} + \kappa)\mathtt{ft}(f).$$

Due to κ being constant, we can just calculate the solution by dividing by $2\pi\mathrm{i}s\cdot\vec{\xi}+\kappa$, thus

$$\begin{aligned}\big\|\hat{u}_N-\hat{u}_{\mathrm{fin}}\big\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})} &= \Big\|\underbrace{\frac{(1+(\vec{s}\cdot\vec{\xi})^2)^{\frac{1}{2}}}{\kappa+2\pi\mathrm{i}\vec{s}\cdot\vec{\xi}}}_{\lesssim 1}\big(\hat{f}_N(\vec{\xi})-\hat{f}_{\mathrm{fin}}(\vec{\xi})\big)\Big\|_{L^2(\widehat{\Omega}_{\mathrm{fin}})}\\ &\lesssim \big\|\hat{f}_N(\vec{\xi})-\hat{f}_{\mathrm{fin}}(\vec{\xi})\big\|_{L^2(\widehat{\Omega}_{\mathrm{fin}})} =: \mathrm{err}_{\mathcal{F}},\end{aligned}$$

which is the $L^2(\widehat{\Omega}_{\mathrm{fin}})$-error of the discretization of the continuous Fourier transform (up to a constant).

As the final step, we consider the error made by discretizing with ridgelets up to a certain scale. Let $N=2^J$, $j\le J$,

$$\hat{V}_j := \mathrm{span}\big\{\hat{\psi}_\lambda \,\big|\, \lambda\in\Lambda_j\big\}\subset\hat{V}_N,$$

as well as

$$\hat{a}_N(\hat{u}_N,\hat{v}_N)=\big\langle\widehat{\mathcal{A}}_N\hat{u}_N,\widehat{\mathcal{A}}_N\hat{v}_N\big\rangle \quad\text{and}\quad \hat{\ell}_{\mathrm{fin}}=\big\langle\widehat{\mathcal{A}}_N\hat{f}_{\mathrm{fin}},\hat{v}_N\big\rangle.$$

The sesquilinear form $\hat{a}_N$ is again continuous and coercive, and we see that $\hat{u}_{\mathrm{fin}}$ is the solution to

$$\hat{u}_{\mathrm{fin}}\colon \hat{a}_N(\hat{u}_{\mathrm{fin}},\hat{v}_N)=\hat{\ell}_{\mathrm{fin}}(\hat{v}_N)\quad\forall v_N\in\hat{V}_N.$$

Now we restrict once more to a subspace in the search for a solution, namely,

$$\hat{u}_j\colon a_N(\hat{u}_j,\hat{v}_j)=\hat{\ell}_{\mathrm{fin}}(\hat{v}_j)\quad\forall v_j\in\hat{V}_j.$$

Then Céa's lemma yields once more that (bearing in mind that the induced norm $\|\cdot\|_{\hat{a}_N}$ is again equivalent to the discrete analogue of the Sobolev norm $\|\cdot\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}$)

$$\|\hat{u}_{\mathrm{fin}}-\hat{u}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}\lesssim\|\hat{u}_{\mathrm{fin}}-\hat{u}_j\|_{\hat{a}_N}\le\inf_{\hat{v}_j\in\hat{V}_j}\|\hat{u}_{\mathrm{fin}}-\hat{v}_j\|_{\hat{a}_N}\lesssim\inf_{\hat{v}_j\in\hat{V}_j}\|\hat{u}_{\mathrm{fin}}-\hat{v}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}.$$

Since we have no precise control of the behavior of $\hat{u}_{\mathrm{fin}}$, we insert another $\hat{u}_N$ and, by the triangle inequality, this results in

$$\|\hat{u}_{\mathrm{fin}}-\hat{u}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}\le\underbrace{\big\|\hat{u}_{\mathrm{fin}}-\hat{u}_N\big\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}}_{\lesssim\mathrm{err}_{\mathcal{F}}}+\inf_{\hat{v}_j\in\hat{V}_j}\|\hat{u}_N-\hat{u}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}.$$

Since we can choose $\hat{v}_j=\mathtt{irt}[(\mathtt{rt}[\hat{u}_N])\big|_{\Lambda_j}]$ in the infimum, Lemmas 3.27 and 4.5 (up to scale j) imply

$$\inf_{\hat{v}_j\in\hat{V}_j}\|\hat{u}_N-\hat{u}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}\le\underbrace{\big\|\hat{u}_N\big|_{\widehat{\Omega}_{\mathrm{fin}}\setminus\widehat{\Omega}_{\mathrm{uni}(j)}}\big\|}_{=:\mathrm{err}_j}=\mathcal{O}\big(2^{-jk}\big).$$

Putting everything together, we see that

$$\begin{aligned}&\big\|u-\mathcal{F}^{-1}\mathcal{Z}_N\hat{u}_j\big\|_{H^{\vec{s}}(\Omega)}\\ &\quad\le\|u-u_N\|_{H^{\vec{s}}(\Omega)}+\big\|u_N-\mathcal{F}^{-1}\mathcal{Z}_N\hat{u}_{\mathrm{fin}}\big\|_{H^{\vec{s}}(\Omega)}+\big\|\mathcal{F}^{-1}\mathcal{Z}_N\hat{u}_{\mathrm{fin}}-\mathcal{F}^{-1}\mathcal{Z}_N\hat{u}_j\big\|_{H^{\vec{s}}(\Omega)}\\ &\quad=\|u-u_N\|_{H^{\vec{s}}(\Omega)}+\|\hat{u}_N-\hat{u}_{\mathrm{fin}}\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}+\|\hat{u}_{\mathrm{fin}}-\hat{u}_j\|_{H^{\vec{s}}(\widehat{\Omega}_{\mathrm{fin}})}\\ &\quad\lesssim\mathrm{err}_N+\mathrm{err}_{\mathcal{F}}+\mathrm{err}_j,\end{aligned}$$

with error behavior (for err_N and err_j) as claimed. ∎

4.3. Discrete ordinates method. As a next step, we consider the same equation

$$\vec{s} \cdot \nabla u + \kappa\, u = f, \tag{4.9}$$

but this time we let $\vec{s} \in \mathbb{S}^1$ also be an independent variable such that $u, \kappa, f : \Omega \times \mathbb{S}^1 \to \mathbb{R}$. The *discrete ordinates method* (DOM) as outlined in [15, section 2] solves this problem in the following way:

- Choose some directions $\{\vec{s}_i\}_{i=1}^{N_s} \subset \mathbb{S}^1, N_s \in \mathbb{N}$.
- Solve (4.9) for these fixed directions, yielding the one-directional solutions $u_i'(x, y)$.
- Interpolate the $(\vec{s}_i, u_i')$ to get a solution for the full domain $\Omega \times \mathbb{S}^1$.

In this report, we will use equispaced directions $s_i := 2\pi \frac{i-1}{N}$ and linear interpolation for simplicity. Step two is done with the ridgelet-based solver developed in the previous section.

Introducing the equispaced periodic linear interpolation operator

$$\overset{N}{\underset{i=1}{\boldsymbol{\mathcal{I}}}}_S[f_i](x, y, \vec{s}) := \begin{cases} f_{\ell+1}(x, y) & \text{if } \ell := \frac{N\varphi(\vec{s})}{2\pi} \in \mathbb{Z}, \\ (\lceil \ell \rceil - \ell) f_{\lfloor \ell \rfloor + 1}(x, y) + \cdots & \\ \quad + (\ell - \lfloor \ell \rfloor) f_{(\lceil \ell \rceil \bmod N)+1}(x, y) & \text{otherwise} \end{cases}$$

($\varphi(\vec{s})$ denotes the angle between $\vec{s} \in \mathbb{S}^1$ and the positive x-axis in the usual mathematical convention), we can write the solution u' produced by the DOM as

$$u'(x, y, \vec{s}) = \overset{N_s}{\underset{i=1}{\boldsymbol{\mathcal{I}}}}_S [u_i'](x, y, \vec{s}).$$

If we further let $u : \Omega \times \mathbb{S}^1$ be the exact solution and $\delta u_i(x, y) := u(x, y, \vec{s}_i) - u_i'(x, y)$ the error in the approximate solutions u_i', we get for the total error

$$\begin{aligned} \|u - u'\|_{L^2(\Omega \times \mathbb{S}^1)} &= \left\| u - \overset{N_s}{\underset{i=1}{\boldsymbol{\mathcal{I}}}}_S [u_i + \delta u_i] \right\|_{L^2(\Omega \times \mathbb{S}^1)} \\ &\le \left\| u - \overset{N_s}{\underset{i=1}{\boldsymbol{\mathcal{I}}}}_S [u_i] \right\|_{L^2(\Omega \times \mathbb{S}^1)} + \left\| \overset{N_s}{\underset{i=1}{\boldsymbol{\mathcal{I}}}}_S [\delta u_i] \right\|_{L^2(\Omega \times \mathbb{S}^1)}. \end{aligned}$$

The first term describes a pure interpolation error, which for linear interpolation is known to be $\mathcal{O}\big(N_s^{-2}\big)$ if $u(x, y, \cdot) \in \mathcal{C}^2$. The second term is the error due to the basic RTE solver, which in the previous section was shown to be $\mathcal{O}(2^{-Jk})$. In conclusion, we thus have

$$\|u - u'\|_{L^2(\Omega \times \mathbb{S}^1)} = \mathcal{O}\big(N_s^{-2}\big) + \mathcal{O}\Big(N_s 2^{-J(k-1)}\Big).$$

As we can see, we have to choose $N_s \sim b^J$ with $b = 2^{\frac{k}{3}}$ in order for the angular and spatial errors to be balanced. Assuming a uniform bound on the number of CG iterations, the outlined DOM thus scales as $\mathcal{O}\big(N_s\, 4^J\big) = \mathcal{O}\big((4b)^J\big)$, which can quickly become prohibitively expensive.

4.4. Sparse discrete ordinates method. In order to mitigate the scaling problem of the (full) discrete ordinates method, the *sparse discrete ordinates method* (SDOM) was developed in [15, section 4]. Adapted to the situation here, the SDOM reads as follows: Given a finite ridgelet frame with J and a constant $2 \ge b \in \mathbb{N}$, let $j = 1, \ldots, J$ and $i = 1, \ldots, b^{J-j+1}$. Then, for each pair (j, i), we solve the RTE in

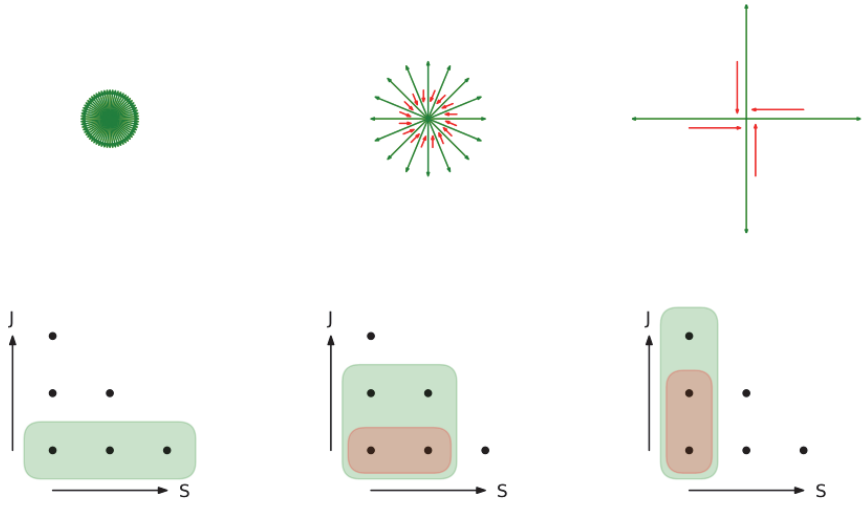


FIG. 5. *Illustration of the SDOM for $J = 3$ and $b = 4$. In the upper row, the lengths of the arrows represent the number of scales that were used, whereas their number and directions indicate how many and which directions are used for angular interpolation. The bottom row shows in which detail spaces the functions obtained in this way live (the J-arrow denotes increasing frame size, the S-arrow denotes increasing number of angular interpolation points).*

direction $\vec{s}_{j,i} := \left(\cos\left(2\pi\frac{i-1}{b^{J-j+1}}\right), \sin\left(2\pi\frac{i-1}{b^{J-j+1}}\right)\right)^\top$ using the ridgelet RTE solver developed in subsection 4.1 with a subframe of the original frame where $J = j$. These partial solutions are stored in $u'_{j,i}(x, y)$, and eventually the full solution $u'(x, y, s)$ is computed as

$$u'(x,y,\vec{s}) = \bigboxplus_{i=1}^{b^J} \boldsymbol{\mathcal{I}}_S\left[u'_{1,i}\right](x,y,\vec{s}) + \sum_{j=2}^{J}\left(\bigboxplus_{i=1}^{b^{J-j+1}} \boldsymbol{\mathcal{I}}_S\left[u'_{j,i}\right](x,y,\vec{s}) - \bigboxplus_{i=1}^{b^{J-j+1}} \boldsymbol{\mathcal{I}}_S\left[u'_{j-1,\,b\,(i-1)+1}\right](x,y,\vec{s})\right).$$

A graphical representation of the SDOM is given in Figure 5.

In [15, Lemma 4.3], Grella and Schwab show that in their setting the convergence of the SDOM deteriorates only by a logarithmic factor compared to the full DOM with physical and angular resolutions equal to the *highest* resolutions used in the SDOM. The same proof works in our case as well, but since the basic RTE solver is not a mere projection of the true solution to some subspace, our notation will be somewhat different.

Let $u : \Omega \times \mathbb{S}^1$ be the exact solution and $\delta u_{j,i}(x, y) := u(x, y, \vec{s}_i) - u'_{j,i}(x, y)$. Furthermore, let $\delta u_j(x, y, \vec{s}) := u(x, y, s) - u'_j(x, y, \vec{s})$, where u'_j is the solution obtained with the basic RTE solver for an arbitrary direction $\vec{s} \in \mathbb{S}^1$. We can then write

$$\begin{aligned}\left\|u - u'\right\|_{L^2(\Omega\times\mathbb{S}^1)} = \Bigg\| u - \bigboxplus_{i=1}^{b^J} \boldsymbol{\mathcal{I}}_S[u(\vec{s}_{1,i}) + \delta u_{1,i}] - \sum_{j=2}^{J}\Bigg(\bigboxplus_{i=1}^{b^{J-j+1}} \boldsymbol{\mathcal{I}}_S\left[u(\vec{s}_{j,i}) + \delta u_{j,i}\right] - \cdots \\ - \bigboxplus_{i=1}^{b^{J-j+1}} \boldsymbol{\mathcal{I}}_S\left[u(\vec{s}_{j-1,b\,(i-1)+1}) + \delta u_{j-1,\,b\,(i-1)+1}\right]\Bigg)\Bigg\|_{L^2(\Omega\times\mathbb{S}^1)},\end{aligned}$$

where $u(\vec{s})$ is shorthand notation for $u(\cdot, \cdot, \vec{s})$. Since $\vec{s}_{j,i} = \vec{s}_{j,b\,(i-1)+1}$, the u's in the

sum cancel, and the remaining terms can be rearranged to (all norms are $\|\cdot\|_{L^2(\Omega\times\mathbb{S}^1)}$)

$$\begin{aligned}\|u-u'\| &\le \left\|u-\overset{b^J}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} u(\vec{s}_{1,i})\right\| + \sum_{j=1}^{J-1}\left\|\overset{b^{J-j+1}}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{j,i} - \overset{b^{J-j}}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{j,b\,(i-1)+1}\right\| + \left\|\overset{b}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{J,i}\right\| \\ &\le \left\|u-\overset{b^J}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} u(\vec{s}_{1,i})\right\| + \sum_{j=1}^{J-1}\left\|\delta u_j - \overset{b^{J-j+1}}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{j,i}\right\| + \cdots \\ &\quad + \sum_{j=1}^{J-1}\left\|\delta u_j - \overset{b^{J-j}}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{j,b\,(i-1)+1}\right\| + \left\|\overset{b}{\underset{i=1}{\boldsymbol{\mathcal{I}}_S}} \delta u_{J,i}\right\|.\end{aligned}$$

Inserting the known convergence rates for linear interpolation and the basic RTE solver, and estimating the sums by their largest term, we get

$$\begin{aligned}\|u-u'\|_{L^2(\Omega\times\mathbb{S}^1)} &= \mathcal{O}\big(b^{-2J}\big) + \sum_{j=1}^{J-1} b^{-2(J-j+1)}\mathcal{O}\big(2^{-jk}\big) + \cdots \\ &\quad + \sum_{j=1}^{J-1} b^{-2(J-j)}\mathcal{O}\big(2^{-jk}\big) + \mathcal{O}\big(2^{-Jk}\big) \\ &= \mathcal{O}\big(J\big(b^{-2J}+2^{-Jk}\big)\big).\end{aligned}$$

Note that these estimates assume that the solution u *and* the error functions δu_j are at least $\mathcal{C}^2$ in $\vec{s}$ (compared to the convergence estimate for the DOM, which only required $u \in \mathcal{C}^2(\mathbb{S}^1)$).

The following complexity estimate is nothing but a reformulation of [15, Lemma 3.1].

THEOREM 4.7 (complexity of SDOM). *The SDOM has* $\mathcal{O}\big(J^{\delta(b-4)}\max\{b,4\}^J\big)$ *degrees of freedom.*

Proof. Simply sum the numbers of degrees of freedom of the $u_{j,i}$,

$$\sum_{j=1}^{J} b^{J-j+1}\,\mathcal{O}(4^j) = \mathcal{O}\bigg(b^{J+1}\sum_{j=1}^{J}\Big(\frac{4}{b}\Big)^j\bigg).$$

If we assume $b \neq 4$, we can continue with

$$\mathcal{O}\bigg(b^{J+1}\sum_{j=1}^{J}\Big(\frac{4}{b}\Big)^j\bigg) = \mathcal{O}\bigg(\frac{4^{J+1}-b^{J+1}}{4-b}\bigg) = \mathcal{O}\big(\max\{b,4\}^J\big).$$

Otherwise, we get

$$\mathcal{O}\bigg(b^{J+1}\sum_{j=1}^{J}\Big(\frac{4}{b}\Big)^j\bigg) = \mathcal{O}\big(J\,4^J\big). \qquad \square$$

Thus, if we assume the above convergence estimate to be correct, we see that the SDOM achieves a speedup of

$$\frac{\mathcal{O}\big(J^{\delta(b-4)}\max\{b,4\}^J\big)}{\mathcal{O}(J(b^{-2J}+2^{-Jk}))} \Big/ \frac{\mathcal{O}((4b)^J)}{\mathcal{O}(b^{-2J}+2^{-Jk})} = \mathcal{O}\bigg(\frac{\min\{b,4\}^J}{J^{1+\delta(b-4)}}\bigg)$$

compared to the DOM.

4.5. Source iterations. Finally, we are able to tackle the complete RTE, including the scattering term:

$$\vec{s} \cdot \nabla u + \kappa\, u = f + \int_{\mathbb{S}^1} \sigma u \,\mathrm{d}\vec{s}'.$$

This problem can be solved using the *source iteration* method, which is as follows:

- Set $u'^{(0)}(x, y, \vec{s}) = 0$.
- For $t = 1, \ldots, T$, solve

$$\vec{s} \cdot \nabla u'^{(t)} + \kappa\, u'^{(t)} = f + \int_{\mathbb{S}^1} \sigma u'^{(t-1)} \,\mathrm{d}\vec{s}'$$

  using, e.g., the DOM or SDOM based on the basic ridgelet RTE solver.

The idea of the source iterations is that the $u^{(t)}$ will converge to the true solution u for large enough t, which is what we observe numerically; see subsection 5.5.

5. Numerical experiments. Some parameters of the previously developed theory remain constant throughout this chapter. In order to avoid repeating them over and over again, we introduce them here:

- The physical space domain is the unit square $\Omega = [0, 1]^2$ (i.e., $L_x = L_y = 1$).
- Frames with highest scale J, as well as $\rho_x = \rho_y = 1$ and square finite Fourier spaces with $N_x = N_y = N$, are used.
- The transition function for the radial and spherical window functions (see (3.1), (3.2)), $t : [0, 1] \to [0, 1]$, is given by

$$t(x) := 35x^4 - 84x^5 + 70x^6 - 20x^7.$$

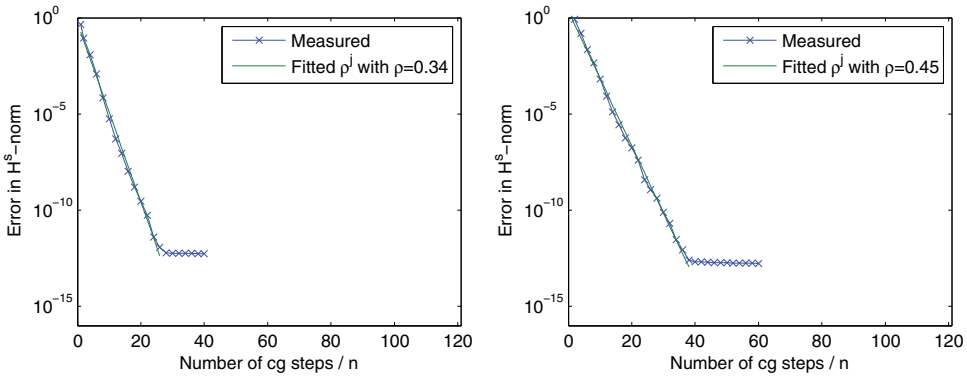
  This choice is the same as the one made in [20], which in turn took the polynomial v from [24].
- Since 3-dimensional functions are difficult to visualize, we will only look at the *incident radiation*

$$G[u](x, y) := \int_{\mathbb{S}^1} u(x, y, s) \,\mathrm{d}s$$

  when solving the multidirectional RTE. Note that both the DOM and SDOM produce solutions which are piecewise linear in s, and therefore we can compute the above integral exactly for these functions.

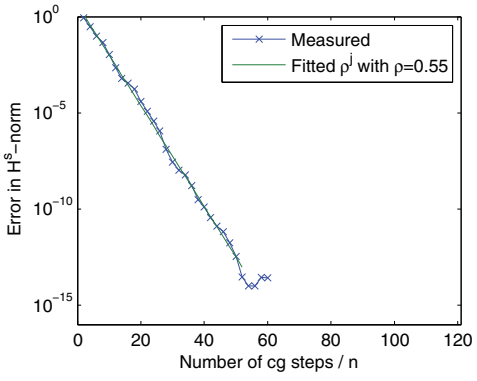
5.1. Convergence of CG. We know from [16] that the operator can be preconditioned such that for the full continuous frame, the condition remains bounded. We verify this numerically in Figure 6 and observe that the direction $\vec{s}$ affects the speed of convergence substantially.

A possible explanation for this behavior could be the clustering of eigenvalues: CG is known to converge much faster if the eigenvalues of the coefficient matrix are clustered around some few points (cf. [1]). For the Fourier case, this can be calculated explicitly, and it becomes apparent that if, e.g., $\vec{s} = (1, 0)^\top$, this clustering occurs, because then all Fourier space points with same $\hat{x}$ lead to the same eigenvalue. On the other hand, if the least common multiple of the denominators of s_x and s_y is not small, a lot more eigenvalues are scattered over the entire range; in practice the threshold (for the least common multiple) to achieve better convergence than with general $\vec{s}$ is quite low.

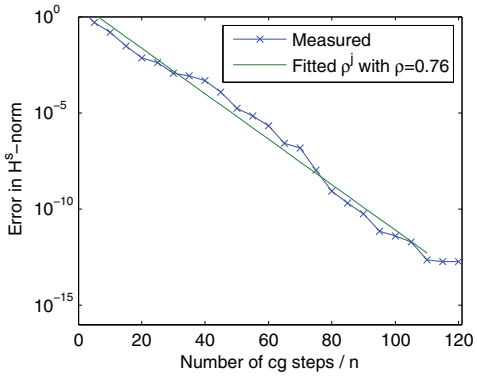


(a) $\varphi = 0$.

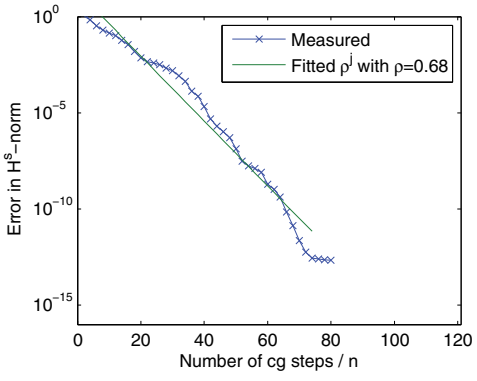
(b) $\varphi = \frac{\pi}{4}$.



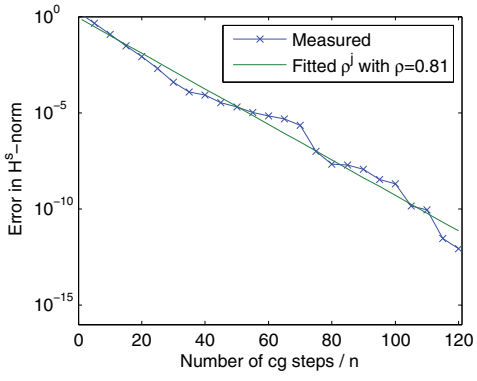
(c) $\varphi = \tan^{-1}(\frac{1}{2})$.



(d) $\varphi = \frac{\pi}{4} - 0.1$.



(e) $\varphi = \tan^{-1}(\frac{1}{4})$.



(f) $\varphi = 0.1$.

FIG. 6. *Convergence of CG applied to the ridgelet space RTE* (4.6) *using* $J = 4$, $\vec{s} = (\cos(\varphi), \sin(\varphi))^\top$, $\kappa = 1$, *and* $f(x, y) = \mathrm{e}^{-100((x-0.5)^2+(y-0.5)^2)}$. *The error was measured in the* $H^{\vec{s}}$*-norm; compare Definition* 4.4.

TABLE 1
Experimental convergence rates for FFRT solver for right-hand side f of given Sobolev order.

Power of sinc	r	2	3	4	5	6
Sobolev order of f	k	1.5	2.5	3.5	4.5	5.5
Base	ρ	0.3675	0.1721	0.0849	0.0470	0.0225
Corresp. power of 2	$\log_2(\rho)$	−1.4440	−2.5383	−3.5579	−4.4103	−5.4736

TABLE 2
Experimental convergence rates for FFRT solver for recovering solution u of given Sobolev order.

Power of sinc	r	2	3	4	5	6
Sobolev order of u	k	1.5	2.5	3.5	4.5	5.5
Base	ρ	0.6778	0.3552	0.1689	0.0926	0.0438
Corresp. power of 2	$\log_2(\rho)$	−0.5611	−1.4931	−2.5659	−3.4327	−4.5144

5.2. Convergence of basic RTE solver. We tested the convergence theory from subsection 4.2 in two ways. The first scheme follows Theorem 4.6:

(i) Fix κ and $\vec{s}$ to some constant value; in our case $\kappa = 8$ and $\vec{s} = (1, \frac{1}{2})^\top$ (normed to $\mathbb{S}^1$).
(ii) Take a right-hand side f with known smoothness H^k (taking r powers of an appropriately dilated sinc in each coordinate in Fourier space gives order $k = r - \frac{1}{2} - \varepsilon$ for any $\varepsilon > 0$).
(iii) Compute explicit reference solution u in Fourier space for the chosen values of κ, s, and u.
(iv) Compute u_j with ridgelet solver for $j = 1, \dots, 6$.
(v) Compute error in Fourier domain.
(vi) Estimate the convergence rate ρ such that the error of the ridgelet-based RTE solver is approximately $\mathcal{O}(\rho^j)$ for $j = 1, \dots, 6$.

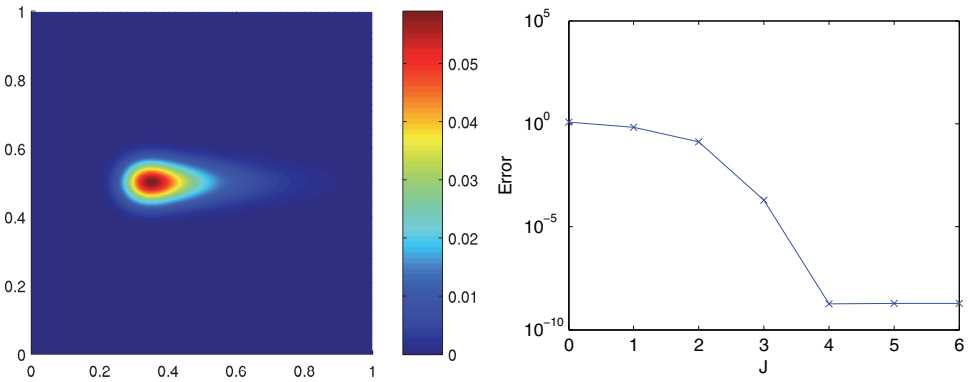
In this way, we obtained the data presented in Table 1, which agree with the predicted $\rho = 2^{-k}$ with good accuracy.

Where the first scheme doesn't follow Theorem 4.6 is in (v), the calculation of the error—since we don't have an explicit solution in physical space, the error is calculated in the Fourier domain. To underscore the claim that $\mathrm{err}_{\mathcal{F}}$ does not dominate the overall error, we proceeded with another test (which differs only in the following points from the first):

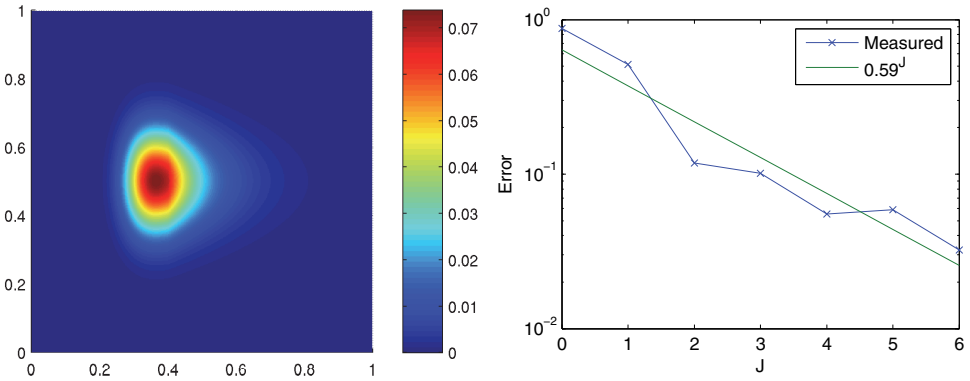
(ii) Take *solution* u with known smoothness H^k.
(iii) Compute *right-hand side* f for the chosen values of κ, s, and u.
(v) Compute the error in *physical space*.

Note that the loss of an order of convergence (i.e., $\rho = 2^{-(k-1)}$) in Table 2 is expected, because in general the $H^{\vec{s}}(\widehat{\Omega})$-norm of $\hat{u}$ restricted to $\widehat{\Omega} \setminus [-\frac{N}{2} : \frac{N}{2})$ will only be $\mathcal{O}(N^{-(k-1)})$ since we cannot compensate for the weight of the anisotropic derivative as in Lemma 4.5, where u had one additional order of smoothness along $\vec{s}$. Of course, f is then only in H^{k-1}; therefore the resulting rates substantiate the above claim that the power of N in the decay in $\mathrm{err}_{\mathcal{F}}$ corresponds to the Sobolev order of f.

Figures 7 and 8 give further examples of convergence rates. In particular, Figure 7(b) confirms that the existence of a jump in the parameters influences the convergence rate as well: If we introduce a jump in κ, the convergence rates approaches $\rho = 0.5$. Without the kink, however, the convergence rate is significantly better.



(a) $\vec{s} = (1,0)^\top$, $\kappa = 8$, $f(x,y) = \mathrm{e}^{-300\left((x-0.3)^2+(y-0.5)^2\right)}$.



(b) $\vec{s} = (1,0)^\top$, $\kappa = \begin{cases} 4, & x \le 0.4, \\ 10 & \text{otherwise} \end{cases}$, $f(x,y) = \mathrm{e}^{-300\,(x-0.3)^2 - 50\,(y-0.5)^2}$.

FIG. 7. *Convergence of ridgelet method in terms of J for smooth function, respectively, smooth and discontinuous absorption coefficient.*

For both figures, the error was computed as the difference in Fourier space norm to the "exact" solution obtained with $J = 6$, which is shown in physical space on the left. CG iterations were aborted once either the relative residual (measured in ridgelet coefficient space) dropped below 10^{-8} or 100 iteration steps were executed.

5.3. General Dirichlet boundary conditions. Assuming that κ and f are constant in the transport direction s, it is easily checked that the solution of the monodirectional RTE

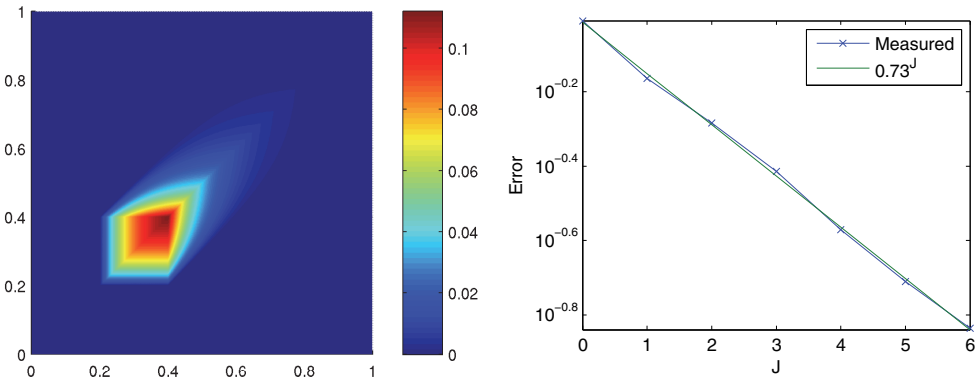
$$\vec{s} \cdot u + \kappa\, u = f$$

is given by

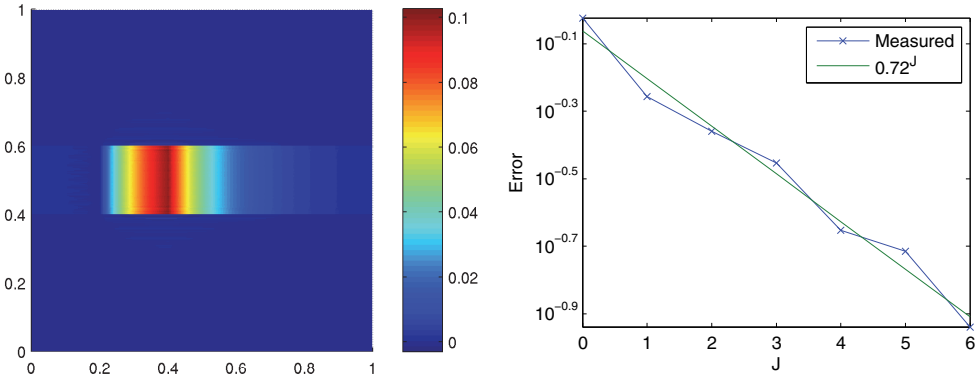
$$u(x,y) = C(x_-, y_-)\, \mathrm{e}^{-\kappa\, \vec{s}\cdot(x,y)^\top} + \frac{f}{\kappa},$$

where $C(x_-, y_-)$ is a function of the values on the inflow boundary

$$\Gamma_{\text{in}} := \left\{(x,y)^\top \in \partial\Omega \,\middle|\, \vec{s} \cdot n\big((x,y)^\top\big) < 0\right\},$$



(a) $\vec{s} = \frac{1}{\sqrt{2}}(1,1)^\top$, $\kappa = 8$, $f(x,y) = \begin{cases} 1, & |x-0.3| < 0.1 \wedge |y-0.3| < 0.1, \\ 0 & \text{otherwise.} \end{cases}$



(b) $\vec{s} = (1,0)^\top$, $\kappa = 8$, $f(x,y) = \begin{cases} 1, & |x-0.3| < 0.1 \wedge |y-0.5| < 0.1, \\ 0 & \text{otherwise.} \end{cases}$

FIG. 8. *Convergence of ridgelet method in terms of J for discontinuous function, with transport diagonally across, respectively, parallel to the singularity.*

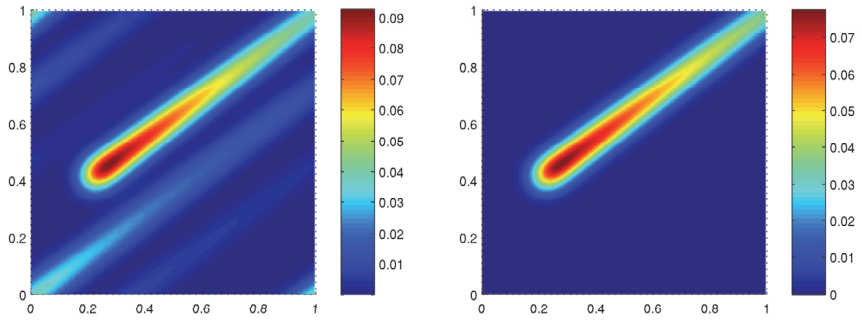
while $n\big((x,y)^\top\big)$ is the outward-facing normal vector, and x_- and y_- are the projections along $\vec{s}$ to Γ_{in}. This result can be used to generalize the basic RTE solvers from subsection 4.1—which require periodic boundaries—to arbitrary inflow boundary conditions.

The main problem is that, by construction, the domain is periodic, and anything leaving Ω reenters at the opposite point on Γ_{in}. One workaround to eliminate this is to enlarge the domain slightly by adding $\Omega_{\text{aux}} := [-\beta, 1]^2 \setminus \Omega$ and setting the absorption coefficient high enough there, such that (practically) nothing leaves Ω_{aux}; compare Figure 9. In both subfigures,

$$f(x,y) = \mathrm{e}^{-300\left((x-0.2)^2+(y-0.4)^2\right)}.$$

For Figure 9(a), $\kappa(x,y) = 1$, whereas for Figure 9(b), we add $\Omega_{\text{aux}} = [-0.2, 1]^2 \setminus \Omega$ and set

$$\kappa(x,y) = \begin{cases} 1, & (x,y) \in \Omega, \\ 30, & (x,y) \in \Omega_{\text{aux}}. \end{cases} \tag{5.1}$$



(a) Periodic boundaries on $[0,1]^2$. (b) Permitting outflow.

FIG. 9. *By extending $\Omega = [0,1]^2$ slightly and increasing κ in the enlarged region, we can fully absorb the otherwise periodized outflow.*

For constant κ, the reduction is at least $\mathrm{e}^{-\kappa_{\mathrm{aux}}\beta}$. Smoother choices of κ work similarly, but without limiting convergence due to a lack of smoothness.

Similarly to forcing the outflow to zero by enlarging the domain and tweaking κ, we can impose inflow boundary conditions

$$u(x,y) = u_{\mathrm{in}}(x,y) \quad \text{on } \Gamma_{\mathrm{in}}$$

by enlarging the domain before (in the direction of the transport s) Γ_{in} and setting an auxiliary forcing term f there. This is illustrated in Figure 10, where κ_{aux} is again set to (5.1)—high enough to prevent periodic contributions.

The forcing term (which has to scale with κ_{aux}) f is defined by extending the boundary values of Γ_{in} along $\vec{s}$ into Ω_{aux}, which is shown in Figure 10(b).

In Figure 10(c), we test the procedure, which yields good agreement between desired (dashed line) and effective (solid line) boundary values. By way of explanation, the line from $(0,1)$ to $(0,0)$ is folded to the negative half-axis, while the positive half-axis represents the line from $(0,0)$ to $(1,0)$.

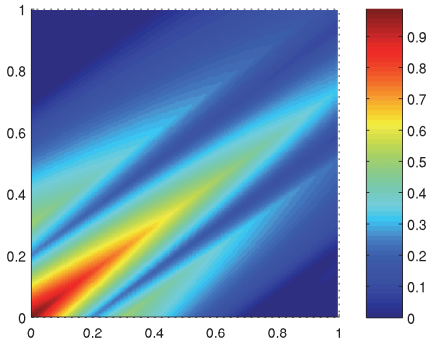
Note that in principle, it doesn't matter if the auxiliary domain enlargement is before or after Ω, in particular for eliminating the periodic pollution. However, for ease of imposing the inflow boundary condition, $\Omega_{\mathrm{aux}} = [-0.2,1]^2 \setminus \Omega$ is added before Γ_{in}.

5.4. Convergence of SDOM compared to DOM. When deriving the DOM and SDOM, we showed that their convergence estimates differ only in a logarithmic factor. This theoretical result was put to the test for smooth functions in Figure 11. Figures 11(a) and 11(b) show that the results hold true approximately.

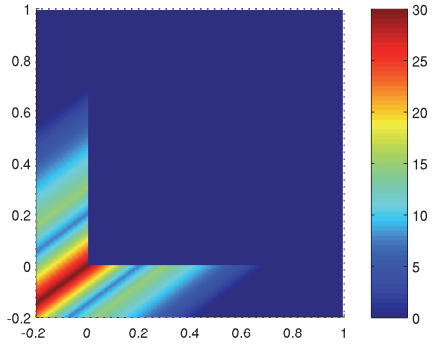
In all subfigures, $\kappa = 4$ and the source term is

$$f(x,y,\varphi) = \mathrm{e}^{-300\left((x-0.5)^2+(y-0.5)^2\right)}\,\mathrm{e}^{-2\min\{\varphi,2\pi-\varphi\}},$$

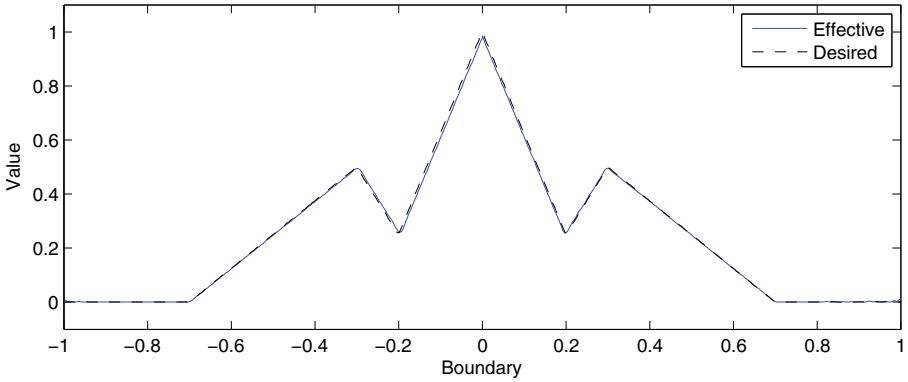
while the parameters are $N_s = 4^J$ for the DOM and $b = 4$ for the SDOM. The error was measured as $\|G[u](x,y) - G[\tilde{u}](x,y)\|_\Omega$, where the reference solution u was obtained by the DOM with $J = 6$, which allows us to calculate the SDOM convergence one scale further than the DOM. CG iterations were aborted once either the relative residual (measured in ridgelet coefficients) dropped below 10^{-8} or 100 iteration steps were executed.



(a) Inflow boundary condition on $[0,1]^2$.



(b) Plot of the f used in (a).



(c) Comparison between the effective boundary conditions and the desired ones.

FIG. 10. *Imposing nonhomogeneous boundary conditions by choosing suitable values for κ and f near the boundaries.*

In Figure 11(c), we plot the incident radiation $G[u]$ with the DOM using $J = 6$ and $N_s = 4^6$. Figure 11(d) shows the difference between this and the SDOM solution obtained with $J = 6$ and $b = 4$.

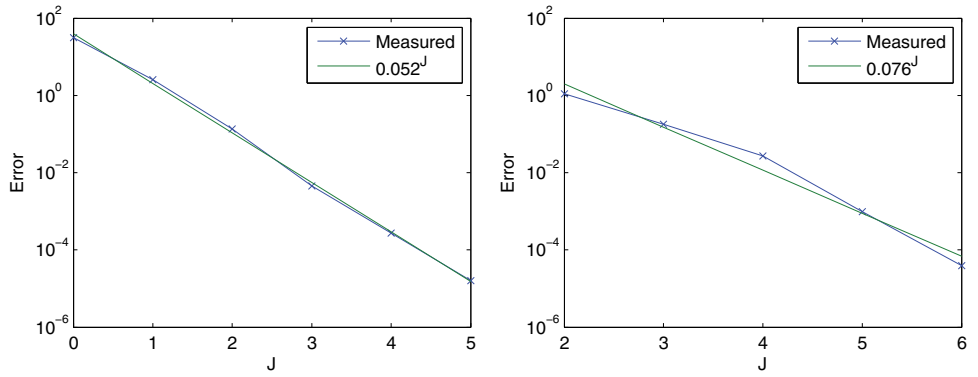
5.5. Source iterations. Figure 12 shows the solutions of the complete RTE—subfigures (a), (c), and (d) show the incident radiation for three different value of the scattering coefficient σ (0, 0.2, and 0.5, respectively). In Figure 12(b), the f and κ are illustrated: The red line on the left (for color images, see online version) shows the shape of the source term

$$f(x, y, \varphi) = \mathrm{e}^{-500\,(x-0.15)^2 - 10\,\min\{\varphi, 2\pi-\varphi\}^2}$$

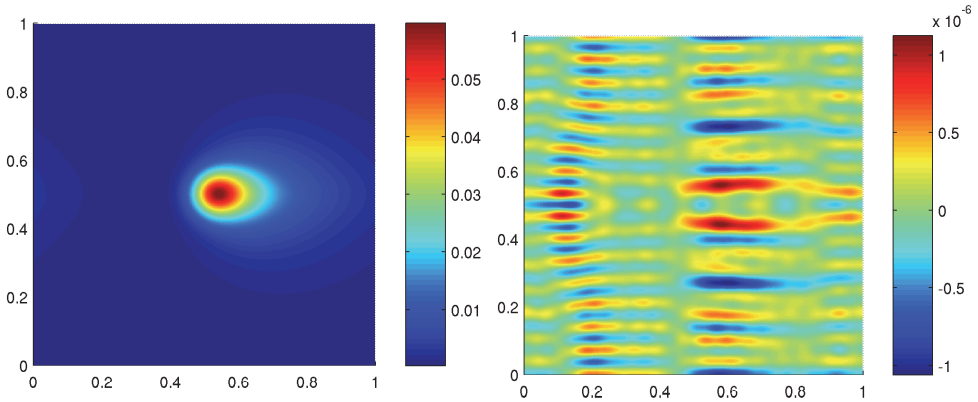
for some constant φ. The light blue area in the middle represents the obstacle which corresponds to the second term in

$$\kappa(x, y) = 2 + 18\,\mathrm{e}^{-2000\,(x-0.4)^4 - 1000\,(y-0.5)^4} + 98\,\mathrm{e}^{-900\,(x-0.9)^2}.$$

The last term in κ, shown as a dark blue bar on the right in Figure 12(b), was introduced in order to avoid that radiation flows across the y-boundary (compare also with Figure 9).



(a) DOM. (b) SDOM.

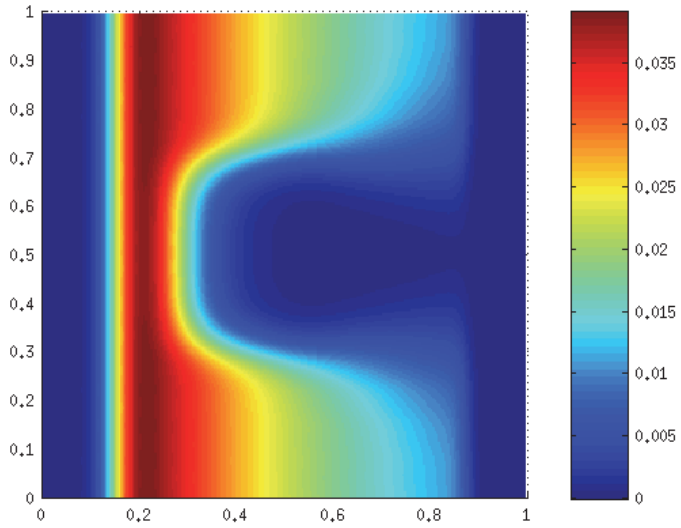


(c) DOM. (d) DOM - SDOM.

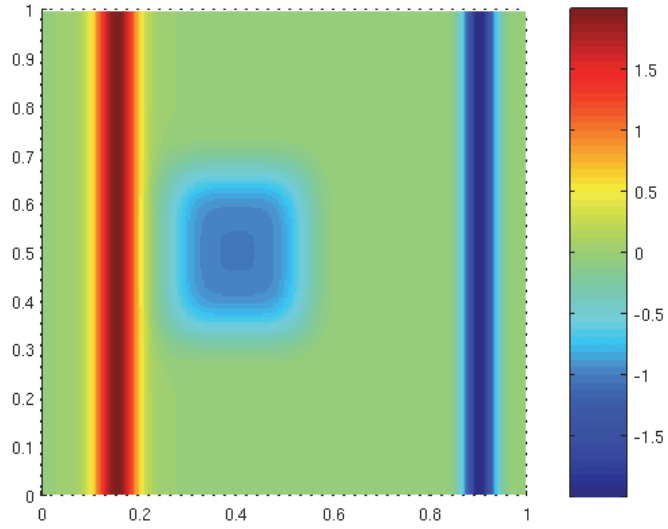
FIG. 11. *Comparison of the DOM and SDOM.*

Figures 13(a) and 13(b) show the convergence of the source iterations in the L^2-norm for the two nonzero values of σ. We observe exponential convergence in both cases, but the rate of convergence deteriorates with increasing σ. Since the source iterations are a fixed-point iteration and σ is proportional to the change in the right-hand sides between consecutive source iteration steps, both findings were to be expected.

The error was measured as the difference to the solution after 10 and 40 source iteration steps for Figures 13(a) and 13(b), respectively. We used the SDOM with $J = 5$ and $b = 4$ to solve the RTE for each source iteration step. CG iterations were aborted once either the relative residual (measured in ridgelet coefficients) dropped below 10^{-4} or 100 steps were executed.

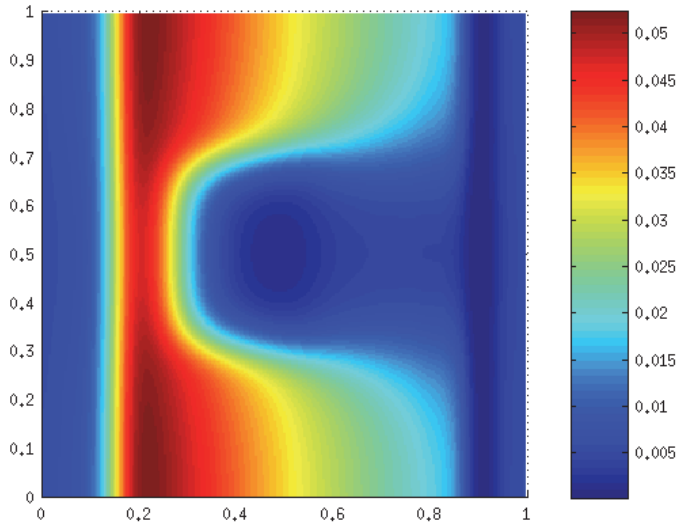


(a) Solution without scattering.

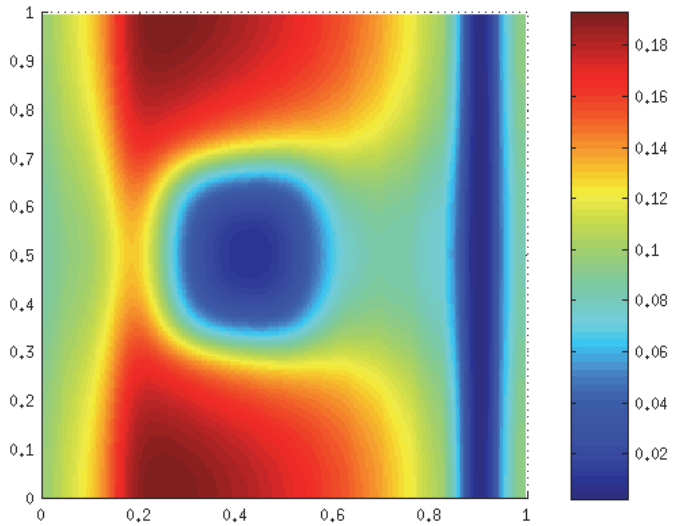


(b) The parameters of the problem.

FIG. 12. *Scattering of radiation around an obstacle for different values of the scattering coefficient σ. (Continued on next page.)*

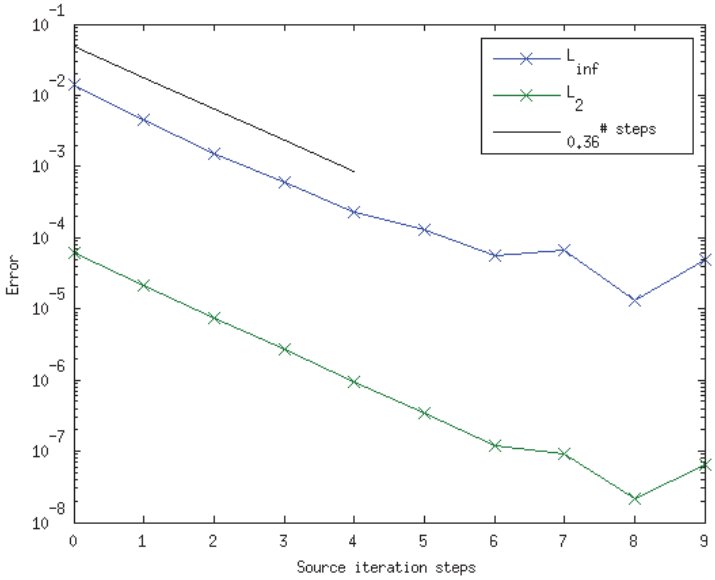


(c) Solution with $\sigma = 0.2$ after 10 source iteration steps.

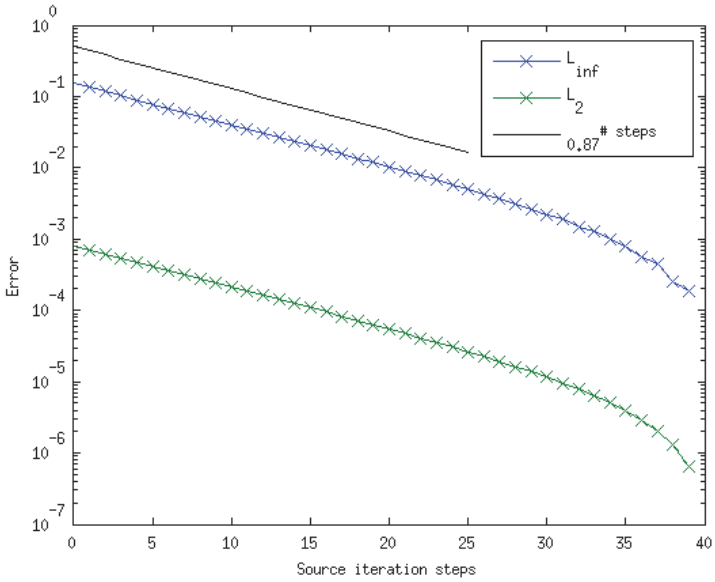


(d) Solution with $\sigma = 0.5$ after 40 source iteration steps.

FIG. 12. *Scattering of radiation around an obstacle for different values of the scattering coefficient* σ.



(a) Convergence for $\sigma = 0.2$.



(b) Convergence for $\sigma = 0.5$.

FIG. 13. *Convergence of the source iteration.*

REFERENCES

[1] O. Axelsson and G Lindskog, *On the rate of convergence of the preconditioned conjugate gradient method*, Numer. Math., 48 (1986), pp. 499–523.

[2] E. J. Candès, *Ridgelets: Theory and Applications*, Ph.D. thesis, Stanford University, 1998.

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

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

[5] E. J. Candès and D. L. Donoho, *Continuous curvelet transform:* I. *Resolution of the wavefront set*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 198–222.

[6] E. J. Candès and D. L. Donoho, *Continuous curvelet transform:* II. *Discretization and frames*, Appl. Comput. Harmon. Anal., 19 (2005), pp. 198–222.

[7] A. Cohen, *Numerical Analysis of Wavelet Methods*, Stud. Math. Appl. 32, North-Holland, Amsterdam, 2003.

[8] A. Cohen, W. Dahmen, and R. DeVore, *Adaptive wavelet methods for elliptic operator equations: Convergence rates*, Math. Comp., 70 (2001), pp. 27–75.

[9] W. Dahmen, C. Huang, G. Kutyniok, W.-Q. Lim, Ch. Schwab, and G. Welper, *Efficient resolution of anisotropic structures*, in Extraction of Quantifiable Information from Complex Systems, Lect. Notes Comput. Sci. Eng. 102, Springer, New York, 2014, pp. 25–51.

[10] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 61, SIAM, Philadelphia, 1992.

[11] M. N. Do and M. Vetterli, *The finite ridgelet transform for image representation*, IEEE Trans. Image Process., 12 (2003), pp. 16–28.

[12] M. N. Do and M. Vetterli, *The contourlet transform: An efficient directional multiresolution image representation*, IEEE Trans. Image Proc., 14 (2005), pp. 2091–2106.

[13] G. Easley, D. Labate, and W.-Q Lim, *Sparse directional image representations using the discrete shearlet transform*, Appl. Comput. Harmon. Anal., 25 (2008), pp. 25–46.

[14] A. G. Flesia, H. Hel-Or, A. Averbuch, E. J. Candès, R. R. Coifman, and D. L. Donoho, *Digital implementation of ridgelet packets*, in Beyond Wavelets, Stud. Comput. Math. 10, Academic Press/Elsevier, San Diego, CA, 2003, pp. 31–60.

[15] K. Grella and Ch. Schwab, *Sparse discrete ordinates method in radiative transfer*, Comput. Methods Appl. Math., 11 (2011), pp. 305–326.

[16] P. Grohs, *Ridgelet-type frame decompositions for Sobolev spaces related to linear transport*, J. Fourier Anal. Appl., 18 (2012), pp. 309–325.

[17] P. Grohs, S. Keiper, G. Kutyniok, and M. Schäfer, *α-Molecules*, submitted, 2014; preprint available as a SAM Report (2014), ETH Zürich, http://www.sam.math.ethz.ch/sam_reports/index.php?id=2014-16.

[18] P. Grohs and G. Kutyniok, *Parabolic molecules*, Found. Comput. Math., 14 (2014), pp. 299–337.

[19] P. Grohs and A. Obermeier, *Optimal Adaptive Ridgelet Schemes for Linear Transport Equations*, submitted, 2014; preprint available as a SAM Report (2014), ETH Zürich, http://www.sam.math.ethz.ch/sam_reports/index.php?id=2014-21.

[20] S. Häuser, *Fast Finite Shearlet Transform: A Tutorial*, University of Kaiserslautern, 2012.

[21] G. Kutyniok and D. Labate, *Introduction to shearlets*, in Shearlets: Multiscale Analysis for Multivariate Data, Birkhäuser, 2012, pp. 1–38.

[22] G. Kutyniok, D. Labate, W.-Q Lim, and G. Weiss, *Sparse multidimensional representation using shearlets*, in Wavelets XI (San Diego, CA), SPIE Proc., 5914 (2005), pp. 254–262.

[23] G. Kutyniok, M. Shahram, and X. Zhuang, *ShearLab: A rational design of a digital parabolic scaling algorithm*, SIAM J. Imaging Sci., 5 (2012), pp. 1291–1332.

[24] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures, University Lecture Series 22, AMS, Providence, RI, 2001.

[25] M. F. Modest, *Radiative Heat Transfer*, Academic Press, 2013.