A hybrid formulation of map migration and wave-equation-based migration using curvelets

Huub Douma[1] and Maarten V. de Hoop[2]

[1] *Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401-1887, USA*
[2] *Center for Computational and Applied Mathematics and Department of Earth and Atmospheric Sciences Purdue University, West Lafayette, IN 47907, USA*

ABSTRACT

Recently, curvelets have been introduced in the field of applied harmonic analysis and shown to optimally sparsify smooth functions away from singularities along smooth curves. In addition, it has been shown that curvelets allow a sparse representation of wave propagators. Since the wavefronts in seismic data lie mainly along smooth surfaces (or curves in two dimensions) and since the imaging operator belongs to the class of operators that is sparsified by curvelets, curvelets are plausible candidates for simultaneous sparse representation of both the seismic data and the imaging operator. Here, the first curvelet-based seismic imaging method is presented, using common-offset time-migration as an example. It is shown that with curvelets the leading-order approximation (in angular frequency, horizontal wavenumber, and migrated location) to such migration becomes a simple transformation of the coordinates of the curvelets in the data, combined with amplitude scaling. This transformation is calculated using common-offset (CO) map time-migration, which uses the local slopes provided by the curvelet decomposition of the data. The accuracy of the method is verified using numerical examples that indicate that using the leading-order approximation only, provides a good approximation to CO time-migration. Even though the current work treats constant media only, the essence of the presented leading-order derivation applies to pre-stack depth-migration also. The combination of map migration with curvelets, as presented in this work, unifies the finite-frequency character of curvelets with the ability of map migration to handle caustics. Because wave-equation-based migration is not troubled by the presence of caustics and honors the finite-frequency character of the data, this work can be viewed as a hybrid formulation of map migration and wave-equation-based migration.

Author's note: Part of this work has previously been reported as Douma and de Hoop, "On common-offset pre-stack time migration with curvelets" (CWP-510), in the 2005 project review book. It is included here to present all material in a coherent and complete fashion.

Key words: Map migration, sparse representation, leading order, coordinate transformation, finite frequency, curvelets

Introduction

In the high-frequency approximation, seismic waves propagate along rays in the subsurface. The local slopes of reflections in seismic data measured at the surface determine (together with the velocity of the medium at the surface), the directions in which we need to look into the earth from the surface to find the location and orientation of the reflector in the subsurface where the reflectionw occurred. The determination of a reflector position *and* orientation from the location of a reflec-

tion in the data *and* the local slope, is generally referred to as map migration (Kleyn, 1977). For an overview of the literature on this topic, and for an explanation of the applicability condition of map migration, we refer to Douma & de Hoop (2006).

Given the slopes at the source and at the receiver locations, map migration provides a one-to-one mapping from the surface seismic measurements, i.e., locations, times, and slopes, to the reflector position and orientation in the image (provided the medium does not allow different reflectors to have identical surface seismic measurements that persist in being identical under small perturbations of the velocity model; see Douma & de Hoop (2006) for an explanation of this condition). This is in sharp contrast to migration techniques that do not make explicit use of the slopes in the data, such as Kirchhoff methods, where the data are summed over diffraction surfaces [see, e.g., Bleistein *et al.* (2000)]; such mappings are many-to-one in a way that all points along the diffraction surface are mapped to one diffractor location. The benefit of the explicit use of the local slopes in the data, is exploited in several seismic applications such as parsimonious migration (Hua & McMechan, 2001; Hua & McMechan, 2003), controlled directional reception (CDR) (Zavalishin, 1981; Harlan & Burridge, 1983; Sword, 1987; Riabinkin, 1991), and stereo tomography (Billette & Lambaré, 1998; Billette *et al.*, 2003). This list is not complete and many more applications exist. In all these methods, the slopes are estimated from the data using additional processing techniques such as local slant-stacking, multidimensional prediction-error filters (Claerbout, 1992p.186-201) or plane-wave destruction filters (Fomel, 2002; Claerbout, 1992p.93-97).

Recently, in the field of applied harmonic analysis (Candès & Guo, 2002; Candès & Donoho., 2000; Candès & Donoho, 2004b) a tight frame of curvelets has been introduced (see Appendix B for an explanation of tight frames) which provides an essentially optimal representation of objects that are twice continuously differentiable (C^2) away from discontinuities along C^2 edges. Due to the wave character of seismic data, the reflections recorded in seismic data lie predominantly along smooth surfaces (or curves in 2D), just as geologic interfaces in the subsurface lie primarily along smooth surfaces. Therefore, it is plausible to assume that seismic data and their images can be sparsely represented using curvelets. This was earlier also noted by Herrmann (2003a; 2003b). At points where the recorded wavefronts or the subsurface contain point-like discontinuities (e.g., at the edges of a fault in the subsurface), however, the level of sparsity that can be achieved with a curvelet representation naturally will be somewhat less than the sparsity achieved for the smooth parts of wavefronts or geologic interfaces. Because curvelets are anisotropic inseparable 2D extensions of wavelets that have main associated local directions, i.e., oriented wavepackets, using curvelets as building blocks of seismic data the local slopes in the data are built into the representation of the data; a simple projection of the data onto the curvelet frame (combined with an intelligent thresholding scheme to separate signal from noise) then provides the local directions associated with the recorded wavefronts.

Smith (1998) and Candès & Demanet (2002) have shown that curvelets sparsify a class of Fourier integral operators from which the seismic imaging operator can be constructed. In combination with the observation that reflections in seismic data lie mainly along smooth curves, it seems therefore that curvelets are plausible candidates for simultaneous compression of seismic data and the imaging operator. Waves with a given dominant wavelength are sensitive to variations in the medium with certain lengths scales only (Jannane *et al.*, 1989); i.e., a wave with a dominant wavelength of say 100 meters is hardly sensitive to variations in the medium on the scale of one centimeter (because the first Fresnel zone of a wave is proportional to the square-root of the wavelength). Just like wavelets, curvelets have a bandlimited character and thus have an associated dominant frequency. Because of this bandlimited nature, curvelets with different dominant frequencies are sensitive to variations in the medium at different scales. This allows the possibility to smooth the background velocity with filters related to the dominant wavelength of the curvelets (i.e., the scale of the curvelets), and propagate curvelets of different scales through different smoothed versions of the medium. Smith (1998) has shown that the propagation of a curvelet through such a smoothed medium is governed by the Hamiltonian flow in this smoothed medium associated with the center of the curvelet. Here the center of a curvelet is defined by its center in phase-space, meaning the central location of a curvelet combined with its central direction. This means that a curvelet is treated as if it was a particle with an associated momentum (or direction). For each smoothed medium, this observation reduces to the statement that the propagation of a curvelet (through an infinitely smooth medium) is well-approximated by translating the center of the curvelet along the corresponding Hamiltonian flow, i.e., translating it along the ray associated with the center of a curvelet (Smith, 1998; Candès & Demanet, 2005). In fact, the procedure just outlined yields a leading-order contribution to the solution of the wave equation (Smith, 1998). Hence this procedure admits wave-equation-based seismic imaging with curvelets.

In this work we focus on the simple case of homogeneous media to learn the basic characteristics of seismic imaging with curvelets and as preparation for imaging in heterogeneous media with curvelets. Because in homogeneous media the above-mentioned smoothing is unnecessary, we ignore such smoothing in this paper altogether. Instead, we focus our attention on verify-

ing, in the context of seismic imaging, the statement that curvelets can be treated as particles with associated directions (or momenta). We show that by treating curvelets as such, common-offset (CO) time-migration becomes, to leading order in angular frequency, horizontal wavenumber, and migrated location, a transformation of the coordinates (i.e., midpoint and time) of each curvelet in the data, combined with amplitude scaling. The coordinate transformation can be calculated with the aid of the explicit formulas for CO map time-migration developed by Douma & de Hoop (2006). Numerical verification of the proposed leading-order approximation based on this simple coordinate transformation confirms that it provides a good approximation to CO time-migration. We emphasize that the essence of the derivation is not limited to homogeneous media but mention that the leading-order approximation to pre-stack depth-migration is obtained also by a transformation of coordinates. In that case, however, the accuracy of the leading-order approximation is currently unknown.

The coordinate transformation presented below is calculated using map migration. Because map migration makes explicit use of the slopes in the data, it constitutes a mapping in phase-space (Douma & de Hoop, 2006). Moreover, since there are no caustics in phase-space, map migration naturally handles caustics. The combination of map migration with curvelets, as presented in this work, unifies the ability of map migration to handle caustics with the finite frequency character of curvelets. The ability to handle caustics while honouring the finite-frequency character of the data, is also found in wave-equation-based migration [e.g., Claerbout (1970; 1976), Stoffa *et al.* (1990), Biondi & Palacharla (1996), Le Rousseau & de Hoop (2001)]. Therefore, we view this work as a hybrid formulation of map migration and wave-equation-based migration.

The outline of this paper is as follows. First we explain what curvelets are, how they are constructed, and what are their main properties. Subsequently we present examples of digital curvelets from the digital curvelet transform (Candès *et al.*, 2005). A more detailed treatment of a particular construction of real-valued curvelets is included in Appendix A. Then, we show an example of the use of curvelets as building blocks of seismic data, and explain the relationship between curvelets and map migration. We proceed to derive that to leading order in angular frequency, horizontal wavenumber, and migrated location, CO time migration is equivalent to a transformation of the coordinates of the curvelets in the data combined with amplitude scaling, and analyze the components that constitute the coordinate transformation. Subsequently we explain how this transformation can be calculated from pre-stack map migration. Finally, we verify the accuracy of the proposed method using numerical examples, and finish with a discussion of the results.

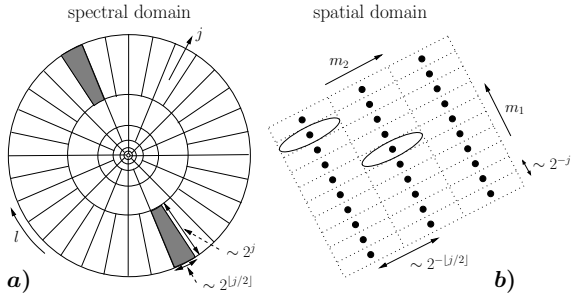

Figure 1. Tilings of the curvelet frame in the spectral domain (a) and the spatial domain (b). In the frequency domain a curvelet is supported near a wedge on a polar grid (i.e., the actual support extends slightly further than the indicated wedge), where the width of the wedge is proportional to $2^{\lfloor j/2 \rfloor}$ and its length is proportional to 2^j. On the support of such a wedge, a local Fourier basis provides a Cartesian tiling of the spatial domain (shown schematically in b). The essential support of a curvelet in the spatial domain is indicated by an ellipse (while again the actual support extends beyond this ellipse).

Curvelets

In this section we explain what curvelets are, how they are constructed, and what are their main properties. Appendix A, which provides a more detailed treatment of a particular construction of curvelets, closely follows the original treatment of the construction of real-valued curvelets by Candès & Donoho (2004b) but provides additional explanations and derivations to guide the non-specialist. For a short summary of the more general complex-valued curvelets, we refer the reader to Candès & Demanet (2005), while Candès *et al.* (2005) describe two different implementations of digital curvelet transforms. A frame similar to the curvelet frame was earlier introduced by Smith (1998).

In wavelet theory [e.g., Mallat (1998)], a 1D signal is decomposed into wavelets, where a wavelet is localized in both the independent variable *and* its Fourier dual, say time and frequency. Such localization is understood within the limits imposed by the Heisenberg uncertainty principle. These wavelets can be translated along the time axis through a translation index, and dilated in their frequency content through a scale index. They are uniquely determined by both indices: the translation index m determines their location along, say, the time axis, while the scale index j determines their location along, say, the frequency axis.

Curvelets are basically 2D *anisotropic* (see below) extensions to wavelets that have a main direction associated with them. Analogous to wavelets, curvelets can be translated and dilated. The dilation is given also by a scale index j that controls the frequency content of the curvelet, while the translation is indexed by m_1 and m_2

(in two dimensions)[⋆]. A curvelet has a main associated direction that can be changed through a rotation. This rotation is indexed by an angular index l. The relation between these indices and the location of the curvelet in the spatial and spectral domains is shown in Figures 1a and b. A curvelet is uniquely determined by all four indices (j, l, m_1, m_2).

As explained in Appendix A, curvelets satisfy the anisotropic scaling relation width $\propto$ squared length in the spatial domain[†]. This is generally referred to as *parabolic scaling*. This anisotropic scaling relation associated with curvelets is a key ingredient to the proof in Candès & Donoho (2004b) that curvelets provide the sparsest representations of C^2 (i.e., twice continuously differentiable) functions away from edges along piecewise C^2 curves. The search for sparse representations of such functions in the field of image analysis was the original motivation for their construction, as wavelets fail to sparsely represent such functions (Candès & Donoho, 2004b) due to their isotropic character. The anisotropic scaling relation is a key difference between wavelets and curvelets. The parabolic scaling is also the key to the proof that curvelets remain localized in phase-space (i.e., remain curvelet-like) under the action of the wave operator provided the medium is smoothed appropriately prior to propagation (Smith, 1998).

Curvelets are constructed through the following sequence of operations. First, the spectral domain is bandpass filtered in the radial direction into dyadic annuli (or subbands); this means that the radial widths of two neighboring annuli differs by a factor of two, the next outer annulus having twice the radial width of the inner one (see Figure 1a). Each subband is subsequently subdivided into angular wedges, where the number of wedges in each subband is determined by the frequency content (or the scale index j) of the subband (see Figure 1a). The number of wedges in a subband with scale j is $2^{\lfloor j/2 \rfloor}$, where the notation $\lfloor p \rfloor$ denotes the integer part of p. This means that the number of wedges in a subband increases only every other scale[‡]. Subsequently, each wedge is multiplied by a 2D orthonormal Fourier basis for the rectangle that just covers the support of the wedge. According to the discrete Fourier transform, this basis has the fewest members if the area of this rectangle is minimum, since then the product of both sampling intervals in space is largest. Therefore, the orientation of this rectangle rotates with the angular wedge, and the spatial tiling associated with the local Fourier basis is oriented along the central direction of the angular wedge (see Figure 1b); that is, the spatial tiling associated with each angular wedge depends on the particular orientation of the wedge. The subband filtering gives curvelets their bandlimited nature (just as with wavelets), whereas the subdivision of these subbands into angular wedges provides them with orientation. The local Fourier basis over the support of the angular wedge allows the curvelets to be translated in space. Therefore, curvelets are in essence a tiling of phase-space; i.e., a tiling of two variables and their two Fourier duals. The tiling is non-trivial in that the sampling of phase space is polar in the spectral domain, but Cartesian in the spatial domain. As explained in Appendix A, curvelets are essentially Heisenberg cells in phase-space.

Roughly speaking, we can think of curvelets as small pieces of bandlimited plane waves. The difference between this rough description and the actual interpretation lies, of course, in the fact that a bandlimited plane wave has associated with it one $\boldsymbol{k}$ direction only, whereas a curvelet is associated with a small range of $\boldsymbol{k}$ vectors. A better description is the term *coherent wave packet* which was around before curvelets were ever constructed [e.g. Smith (1997; 1998)] and dates back to the work of Córdoba & Fefferman (1978). The frequency-domain tiling of the curvelet frame is the same as the dyadic parabolic decomposition or second dyadic decomposition (Gunther Uhlmann, personal communication with Maarten de Hoop) used in the study of Fourier integral operators [see e.g. Stein (1993)], which was around long before the construction of the curvelet frame (Fefferman, 1973).

Curvelets form a *tight frame* for functions in $L^2\left(\mathbb{R}^2\right)$ — see Appendix A for the proof of this property and Appendix B for an introduction to tight frames. This means that, much as with an orthonormal basis, we have a reconstruction formula

$$f = \sum_{\mu \in \mathcal{M}} (f, c_\mu)\, c_\mu \ , \quad (f, c_\mu) = \int_{\mathbb{R}^2} f(\boldsymbol{x}) c_\mu^*(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \ , \quad (1)$$

where c_μ denotes a curvelet with multi-index $\mu = (j, l, m_1, m_2)$, the superscript $*$ denotes complex conjugation, $\mathcal{M}$ is a multi-index set, and $f(x_1, x_2) \in L^2\left(\mathbb{R}^2\right)$. Thus, we can express an arbitrary function in $L^2\left(\mathbb{R}^2\right)$ as a superposition of curvelets. The quantity (f, c_μ) is the

[⋆]We differ from the standard notation k_1 and k_2 to avoid confusion with the wave-vector components.

[†]Considering the difference in dimension between width and squared length, this scaling relation holds when both width and length are made dimensionless through division by a reference length, typically the sampling interval in numerical implementations. Otherwise this relation would involve a proportionality constant that has a dimension which adjusts for the difference in dimension between width and squared length^2.

[‡]This is a consequence of the dyadic nature of the subband filtering done in the first step combined with the desired parabolic scaling. Without this splitting at every other scale resulting from the parabolic scaling, the number of wedges would increase with a factor $\sqrt{2}$ if the scale index was increased by one. In this case we would not have an integer number of wedges, hence the doubling of the number of angular wedges every other scale only.

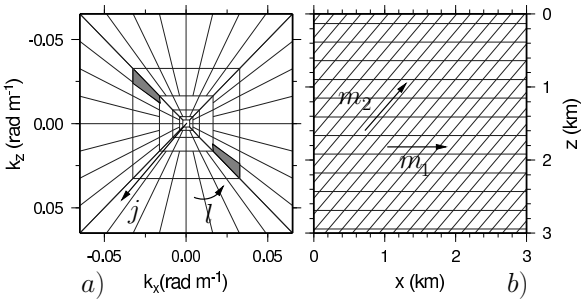

Figure 2. Tilings for digital curvelets in the spectral domain (a) and the spatial domain (b). For digital curvelets, the concentric circles in the spectral domain are replaced with concentric squares, and the Cartesian spatial grid is sheared.

coefficient of curvelet c_μ which denotes the projection of the function f on curvelet c_μ.

Digital curvelets

In the construction of curvelets treated so far, the sampling of the spectral domain is done in polar coordinates while the sampling of the spatial domain is Cartesian (see Figures 1a and b). From a computational point of view, this combination is not straightforward to implement. Combining Cartesian coordinates in both domains is straightforward and is standard in data processing. Therefore, for the purpose of digital curvelet transforms, the polar coordinates in the spectral domain are replaced with Cartesian coordinates (see Figures 2a). Also, in the field of image analysis [where the digital curvelet transform was originally developed (Candès & Donoho, 2004a; Candès *et al.*, 2005)], images usually have Cartesian spatial coordinates to begin with; hence it is natural to have Cartesian coordinates in the spectral domain also. This allows straightforward application of Fast Fourier Transform algorithms. This argument holds for seismic as well as image data.

To go from polar coordinates to Cartesian coordinates in the spectral domain, the concentric circles in Figure 1a are replaced with concentric squares (see Figure 2a); hence the rotational symmetry is replaced with a sheared symmetry. As a consequence, the Cartesian sampling in the spatial domain is a Cartesian grid that is sheared rather than rotated (cf., Figures 1b and 2b)[§]. This construction allows a rapidly computable digital curvelet transform (Candès *et al.*, 2005). For more details on the implementation of digital curvelet transforms, we refer to Candès & Donoho (2004a) and Candès *et al.* (2005).

[§]Here the centers of the cells are the locations of the centers of the curvelets in space.

Examples of digital curvelets

Figure 3 shows examples of digital curvelets. The left column shows curvelets in the spatial domain, while the right column shows their associated spectra. Superimposed on the spectra is the spectral tiling of the digital curvelet transform. The middle column shows the associated spatial lattice for each of the curvelets, where the centers of the cells are the locations of the centers of the curvelets in space[¶]. Here the spatial cells on the spatial lattice are colored according to the magnitude of the curvelet coefficient (here always unity); black equals one and white equals zero. Figure 3b shows a translated version of the curvelet in Figure 3a; the spectral tile is the same, but the spatial tile has changed, i.e., indices j and l are held constant, but the translation indices m_1 and m_2 are different. Figure 3c shows a curvelet with the same spatial location and the same scale index as that in Figure 3a, but with a different angular index l. That is, the spectral tile has moved within the same filter band, i.e., within the same concentric squares. Note how the spatial lattice changes as we change the angular index l. Finally, Figure 3d shows a dilated version of the curvelet shown in Figure 3a; the spatial location is the same, but the spectral tile has moved outward into a neighboring annulus (or subband), i.e., the scale index $j \rightarrow j + 1$. Since the neighboring annulus is subdivided into more wedges, the angular index l has also changed, but in such a way that the direction of the curvelet is basically the same. Similarly, since the larger scale has a finer spatial sampling associated with it, the translation indices m_1 and m_2 have also changed, but in such a way that the curvelet location is the same.

Curvelets remain curvelet-like when subjected to the class of operators relevant for seismic imaging

The action of operators belonging to the class of Fourier integral operators that can be sparsely represented using curvelets, which includes the CO time-migration operator, can be described in terms of propagation of singularities along a Hamiltonian flow (Smith, 1998; Candès & Guo, 2002). The action on a curvelet of a particular scale can be approximated by translating the curvelet along the ray associated with the center of the curvelet through the medium smoothed for that particular scale. This means that, in the appropriately smoothed media, curvelets remain fairly localized in both the spatial domain and the spectral domain. Hence, the propa-

[¶]Throughout this work, we used the nonequispaced fast Fourier transform-based curvelet transform (Candès *et al.*, 2005). Because this transform induces a sheared spatial grid, as opposed to the wrapping-based transform (Candès *et al.*, 2005), which induces a rectangular grid, all spatial lattices shown in this work are sheared grids.

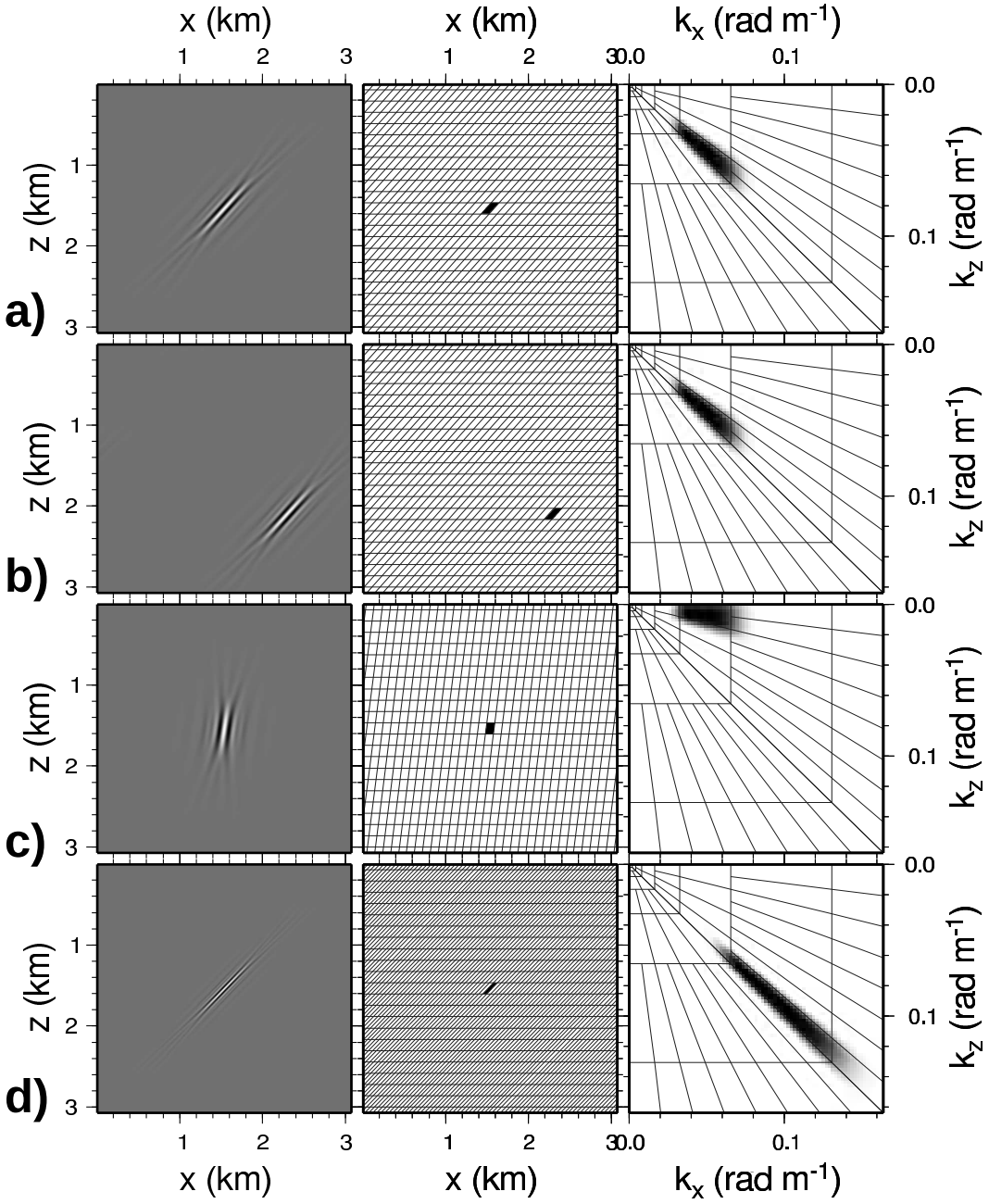

Figure 3. First column: curvelets in the spatial domain. Second column: associated spatial lattices, and spatial cells colored according to the value of the coefficient (black is one, white is zero). Third column: amplitude spectra and frequency-domain tilings. First row: a curvelet. Second row: curvelet from (a) with different translation indices. Third row: curvelet from (a) with a different angular index. Fourth row: curvelet from (a) with a different (higher) scale index (here the translation indices and the angular index are also different because they both depend on the scale index).

gated curvelet can be constructed by using neighboring curvelets only, where neighboring is understood in the context of phase-space; i.e., a neighboring curvelet is a curvelet that is close in the spatial domain and has orientation close to the orientation of the curvelet that is propagated along the central ray, i.e., the ray associated with the center of the curvelet in phase-space. Note that propagation of a curvelet in accordance with the Hamiltoninan is an operation in phase-space since it makes explicit use of the slope of the curvelet. Since in

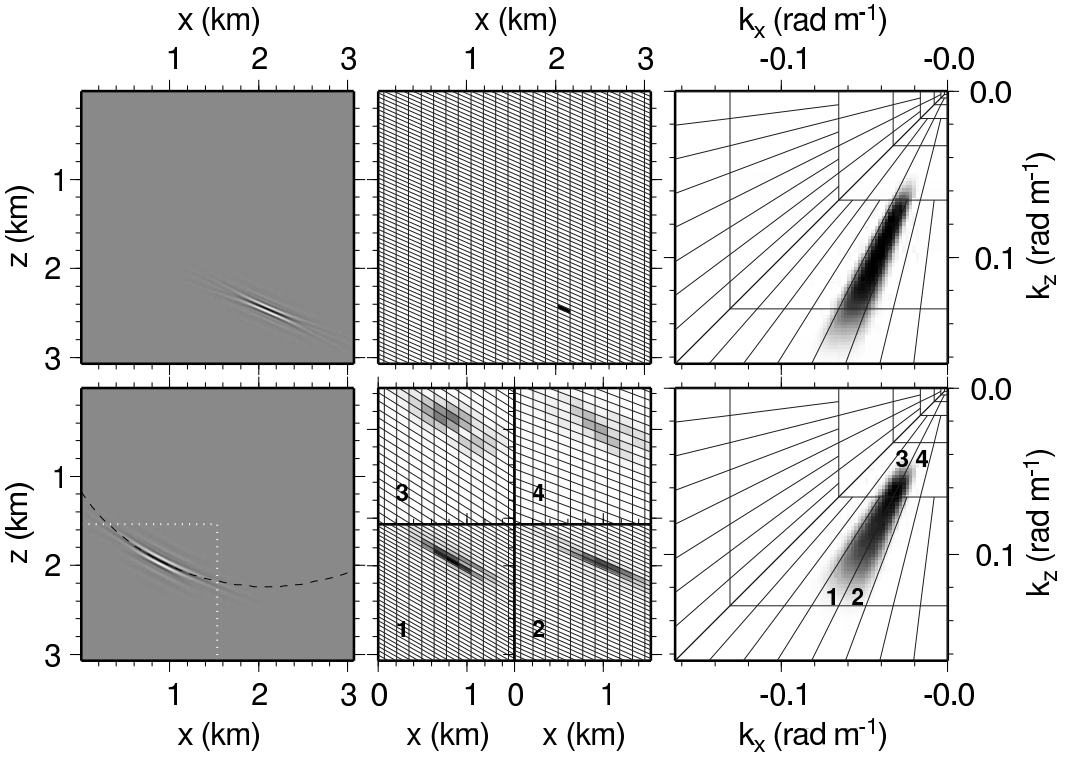

Figure 4. Top row: a curvelet with a dominant frequency of about 30 Hz (left, shown in depth $z = vt_u/2$, for consistency), the normalized absolute value of the coefficient on the spatial lattice (middle), and its amplitude spectrum (right). Bottom row: CO Kirchhoff migration of the curvelet in the top row. The middle panel in this row shows the coefficients on the spatial lattice in the lower left quadrant of the leftmost panel (indicated with the dotted lines in the leftmost panel) for each of the numbered wedges (labeled '1' to '4') in the spectrum (right).

phase-space no caustics can occur, curvelet-based imaging does not even notice the formation of caustics.

For homogeneous media smoothing is unnecessary and curvelets of all scales can be propagated along the central ray through the same medium. To illustrate this, Figure 4 shows the result of CO Kirchhoff migration of a curvelet [taken from Douma & de Hoop (2004)]. The top row shows the input curvelet in space (the vertical axis was converted to depth using $z = vt_u/2$ for convenience, with z denoting depth, and v and t_u denoting velocity and the two-way traveltime, respectively) and its associated amplitude spectrum. Again the spatial distribution of the coefficients is shown in the middle panel, just as in Figure 3. The left-most panel of the bottom row shows the CO Kirchhoff-migrated curvelet. Notice how the migrated curvelet is clearly localized in space and determines *only part of the isochron*, in sharp contrast to the whole isochron if a single sample (or a 'spike') would be used as input to the migration. This confirms that in the context of migration (at least for CO time-migration), curvelets are indeed a more appropriate choice for building blocks of seismic data than are spikes (that are currently used to represent seismic data). The spectrum of the migrated curvelet (bottom right) is clearly also localized after the migration, and overlies only four wedges in the curvelet tiling of the spectrum. The middle panel shows the coefficients for the spatial area in the lower left quadrant of the leftmost figure (outlined by the dotted lines), for the wedges labeled '1' through '4'; since curvelets of different orientations induce different spatial tilings[‖], and since the migrated curvelet overlies four wedges, the spatial distribution for the coefficients related to all four wedges are shown separately in the lower middle pannel of Figure 4. Indeed this reveals that the migrated curvelet is constructed from several curvelets, but only few of them having large coefficients. This confirms that curvelets remain fairly localized in both the spatial and spectral domain (i.e., they remain curvelet-like), at least after CO time-migration.

[‖]The wrapping-based curvelet transform (Candès *et al.*, 2005) does not induce different tilings for spectral wedges with the same scale index but different angular indices.

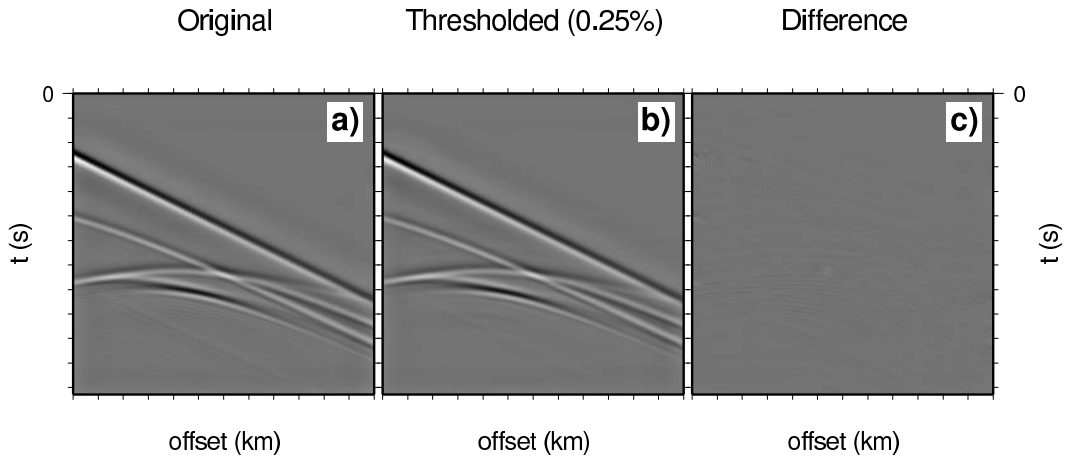

Figure 5. Synthetic common-shot gather with cusped wavefront: original (a), reconstructed using only the 0.25% largest curvelet coefficients (b), and the difference (c). The reconstruction with 0.25% of the curvelets is almost identical to the original common-shot gather. In this example, using only 0.25% of the curvelets results in about 30 times fewer curvelets than input samples in the gather.

Curvelets as building blocks of seismic data

Seismic reflections in seismic data lie primarily along smooth surfaces (or curves in two dimensions). Even diffractions from discontinuities in the earth's subsurface, such as edges of geologic interfaces caused by faulting, lie along smooth surfaces. This is a direct consequence of the wave-character of seismic data. As mentioned in the introduction, it is intuitive that curvelets can be used to sparsely represent seismic data because curvelets provide the sparsest representations of smooth (C^2) functions away from edges along piecewise C^2 curves (Candès & Donoho, 2004b). Throughout this work, we adopt this intuition and illustrate this with a simple synthetic example. Hence, we focus on the sparse representation of the imaging operator rather than the data.

Figure 5a shows part of a synthetic common-shot gather, where the wavefront has a cusp. These data relate to a model with a syncline-shaped reflector. Figure 5b shows the reconstructed gather where only the 0.25% largest curvelet coefficients were used. For the particular example shown, this relates to a compression ratio of about 30; i.e., we used 30 times fewer curvelets than the number of sample values in the original gather to reconstruct the data. From Figure 5c the difference between the original and reconstructed data is negligible. Note that this large compression ratio is partly a result of the synthetic data having many zero sample values to start with; for a field data example this compression ratio would likely be smaller. This example, however, shows that with curvelets as building blocks of seismic data, the data can be represented with fewer curvelets than samples in the data, and with essentially no residual, even in areas where the wavefront has cusps.

In this example, we have applied a hard thresholding to the data; we discarded 99.75% of the curvelets. At first sight one might therefore think that the compression ratio should be 400. The curvelet transform, however, is a redundant transform, meaning that if all curvelets are used to reconstruct the data, there are more curvelets than sample points in the data. Different digital implementations of the curvelet transform have different redundancies** (Candès *et al.*, 2005). The hard thresholding used in this numerical example will likely not be suitable for determining the threshold level with field data, especially where the data are contaminated with noise. In that case, we would need more sophisticated methods to separate signal from noise and obtain a sparse representation of the data. In this work we do not concern ourselves with any such denoising and compression methods but instead focus on the sparse representation of the imaging operator. We reiterate that by using curvelets as building blocks of seismic data, the local slopes (or directions) in the data are built into the data representation. Other than a straightforward projection of the data onto the curvelet frame (combined with an intelligent thresholding procedure), no additional processing steps are required to extract the local (central) slopes from the data. Therefore, curvelets provide an appropriate reparametrization of seismic data that has the wave-character of the data built into it.

**In this particular example, the apparent compression ratio (400) and the associated implied redundancy of $400/30 \approx 13$ is so large only because a lot of zero-padding was necessary to make the number of samples in the gather both horizontally and vertically equal to an equal power of 2; the actual redundancy of the transform used is about 3.

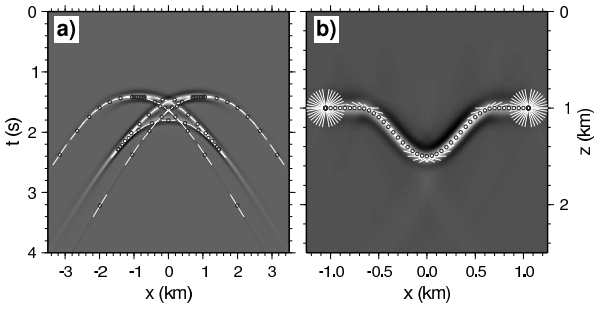

Figure 6. Common-offset (h = 1000 m) data (a) and migrated data (b) from a syncline shaped reflector embedded in a constant velocity (v = 2000 m/s) medium, and demigrated and migrated line elements superposed on the data and migrated data, respectively. The excellent agreement between the demigrated line elements and the seismic data (a), and the migrated line elements and the migrated data (b), indicates the validity of the common-offset map time-demigration and migration equations, respectively.

2D Common-offset map time migration

Douma & de Hoop (2006) present explicit expressions for common-offset map time-migration (i.e., migration in a medium with (effectively) constant velocity) that use only the slope in a common-offset gather (and the velocity), rather than the slopes in both a common-offset gather and a common-midpoint gather (and the velocity), such as the equations presented by Sword (1987p.22). In essence this means that CO time-migration can be written in terms of one Hamiltonian flow only, instead of two. The expressions in three dimensions from Douma & de Hoop (2006) simplify[††] in two dimensions to

$$y_1^m = x_u - \left(\frac{vt_u}{2}\right)^2 \frac{\Lambda_u}{h}\,, \tag{2}$$

$$y_2^m = \sqrt{\left[1-\left(\frac{vt_u\Lambda_u}{2h}\right)^2\right]\left\{\left(\frac{vt_u}{2}\right)^2 - h^2\right\}}\,, \tag{3}$$

$$p_m = \frac{p_u t_u\,|\Lambda_u - 1|\,|\Lambda_u + 1|}{\sqrt{\left[1-\left(\frac{vt_u\Lambda_u}{2h}\right)^2\right]\left\{t_u^2 - \left(\frac{2h}{v}\right)^2\right\}}}\,, \tag{4}$$

[††]Note that Douma & de Hoop (2006) used $p_u := \frac{1}{2}\frac{\partial t_u}{\partial x_u}$ and $p_m := \frac{1}{2}\frac{\partial t_m}{\partial y_1^m}$ instead of $p_u := \frac{\partial t_u}{\partial x_u}$ and $p_m := \frac{\partial t_m}{\partial y_1^m}$, as is done throughout this work.

in which

$$\Lambda_u = \Lambda_u\left(p_u, \Theta_u, h\right) := \frac{1}{\sqrt{2}p_u h} \times \sqrt{\Theta_u\left(1-\sqrt{1-\frac{4\left(p_u h\right)^4}{\Theta_u^2}}\right)}\,, \tag{5}$$

with

$$\Theta_u = \Theta_u\left(t_u, p_u, h\right) := t_u^2 + \left(\frac{2h}{v}\right)^4 \frac{1}{t_u^2} - 2\left(\frac{2h}{v}\right)^2\left(1-\left(\frac{vp_u}{2}\right)^2\right)\,. \tag{6}$$

In these expressions x_u, t_u, and $p_u := \partial t_u/\partial x_u$ are the midpoint location, two-way (unmigrated) traveltime, and the (unmigrated) slope in a CO section, respectively, while y_1^m, $y_2^m = vt_m/2$ and $p_m := \partial t_m/\partial y_1^m$ are the migrated horizontal and vertical location and the migrated slope, respectively, with t_m the migrated two-way traveltime. Furthermore, h denotes the half-offset and v is the medium velocity. Equations (2)-(4) are explicit expressions that determine the migrated reflector coordinates (y_1^m, y_2^m, p_m) from the specular reflection coordinates (x_u, t_u, p_u), given h and v.

At first sight one might expect these map migration equations to use both the local slopes at the source and the receiver, or equivalently the local slopes in a CO gather and in a CMP gather, because these slopes after all determine the directions in which to look into the earth for the reflector location. It turns out, however, that in homogeneous media only the slope in the CO gather needs to be known (Douma & Calvert, 2006). Equations (2)-(4) indeed do not use the offset horizontal slowness $p_h := \partial t_u/\partial h$, so that in practice only p_u needs to be estimated, and the slope in a common-midpoint gather can be ignored. It is possible to derive map-migration equations that use the offset horizontal slowness p_h instead of the velocity. In this way, the need to estimate the velocity from CMP gathers is replaced with the need to estimate the local slopes in a CMP gather (Fomel, 2005). This idea dates back to the work of Ottolini (1983). In the context of 2D pre-stack time migration with curvelets, the additional slopes p_h can be estimated by decomposition of the full data volume (i.e., time, midpoint, and offset) with 3D extensions of curvelets. In this way both p_u and p_h can be obtained. Hence, 2D pre-stack time migration without velocity picking would require 3D curvelets.

Figure 6a shows a common-offset gather (h = 1000 m) from synthetic data generated from a syncline model with constant velocity above the reflector. Superimposed on the reflections, line elements are drawn tangent to the reflections. Each of these line elements determines a local slope, p_u, while the center of the line determines the two-way traveltime t_u and the common-midpoint location x_u. Using equations (2)-(4) and the velocity v = 2000 m/s, the migrated location (y_1^m, y_2^m) and the

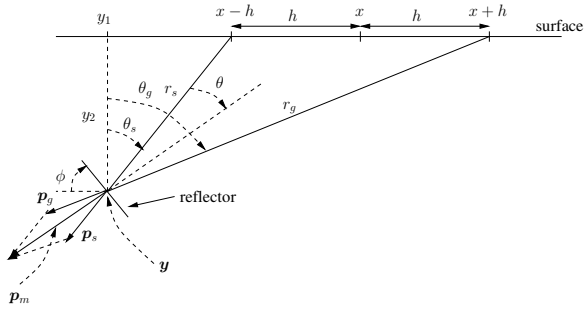

Figure 7. Geometry associated with 2-D CO time migration.

local dip $\tan\phi = vp_m/2$ can be determined (ϕ is the angle with the horizontal measured clockwise positive). Figure 6b shows the migrated common-offset counterpart of the data shown in Figure 6a, with the migrated line elements determined using y_1^m, y_2^m, and p_m, superimposed on the image. The migrated line elements accurately follow the directions in the image, indicating that equations (2)-(4) indeed capture the kinematics of CO time migration in a high-frequency approximate sense. Note that the line elements from diffractions of the edges of the syncline in Figure 6a, are all mapped to the same location but with different orientations. This is analogous to building a delta function with plane waves from all directions; i.e., a Fourier transformed delta function has all directions.

Knowing that each curvelet has a central direction associated with it, we aim to replace the line elements in Figure 6 with curvelets and see to what extent moving curvelets around according to the map migration equations (2)-(4) results in a good time-migrated image. We are thus aiming to lift the applicability of map migration beyond velocity model building, and show its use for pre-stack (here CO time) imaging with curvelets (Douma & de Hoop, 2004).

Leading-order approximation to 2.5-D CO time-migration using curvelets

The CO 2.5-D time-migration equation is given by (Bleistein *et al.*, 2000p.309, equation (6.3.25))

$$\beta(\boldsymbol{y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sqrt{|\omega|}\,a\,(\boldsymbol{y},x;h) \times e^{-i\omega(r_s+r_g)/v+i(\pi/4)\mathrm{sgn}\omega}\,U_s(x,\omega)\mathrm{d}x\mathrm{d}\omega, \quad (7)$$

where $\beta(\boldsymbol{y})$ is the reflectivity function at migrated location $\boldsymbol{y} = (y_1\ y_2)^T$, ω is the angular frequency, and $U_s(x,\omega)$ denotes the scattered field at midpoint location x, Fourier transformed with respect to time t. Figure 7 shows the geometry associated with 2-D CO time-migration. The amplitude $a(\boldsymbol{y},x;h)$ is given by

$$a\,(\boldsymbol{y},x;h) = \frac{4y_2}{\sqrt{2\pi v^3}}\frac{\sqrt{r_s+r_g}\left(r_s^2+r_g^2\right)}{\left(r_sr_g\right)^{3/2}}\cos\theta(x,\boldsymbol{y};h), \quad (8)$$

with $\theta(x,\boldsymbol{y};h)$ half the opening angle between the slowness vectors at the source and the receiver locations, and the distances r_s and r_g from the source and the receiver to the reflector location, respectively, given by

$$r_s = r_s(\boldsymbol{y},x;h) = \sqrt{(x-h-y_1)^2+y_2^2}\,, \quad (9)$$

$$r_g = r_g(\boldsymbol{y},x;h) = \sqrt{(x+h-y_1)^2+y_2^2}\,. \quad (10)$$

Anticipating that in the spectral domain a curvelet is localized near a wedge, we Fourier transform U_s with respect to x [using the convention given in Bleistein *et al.* (2000p.29-30)] to get

$$\beta(\boldsymbol{y}) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sqrt{|\omega|} \times\left[\int_{-\infty}^{\infty}a\,(\boldsymbol{y},x;h)\,e^{-i\{\omega(r_s+r_g)/v-k_xx\}}\mathrm{d}x\right] \times\hat{U}_s(k_x,\omega)e^{i(\pi/4)\mathrm{sgn}\omega}\mathrm{d}k_x\mathrm{d}\omega\,. \quad (11)$$

The quantity between the square brackets is an oscillatory integral that can be approximated using the method of stationary phase [e.g., Bleistein *et al.* (2000p.129)]. Treating $-\omega$ as the formal large parameter and defining the phase Φ as

$$\Phi(\boldsymbol{y},x,k_x,\omega;h) := \frac{r_s(\boldsymbol{y},x;h)+r_g(\boldsymbol{y},x;h)}{v} - \frac{k_x}{\omega}x\,, \quad (12)$$

we get

$$\beta(\boldsymbol{y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{a(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h);h)}{\sqrt{2\pi|\Phi''(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h)|}} \times e^{-i\omega\Phi(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h)}\hat{U}_s(k_x,\omega) \times e^{i(\pi/4)\mathrm{sgn}\omega[1-\mathrm{sgn}(\Phi''(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h))]}\mathrm{d}k_x\mathrm{d}\omega, \quad (13)$$

with

$$\Phi''(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h) := \left.\frac{\partial^2\Phi}{\partial x^2}\right|_{x(\boldsymbol{y},k_x/\omega;h)} \quad (14)$$

the Hessian evaluated at the stationary midpoint location $x(\boldsymbol{y},k_x/\omega;h)$. Note that the factor $\sqrt{|\omega|}$ in equation (11) has canceled out through the stationary phase approximation. Calculating the Hessian explicitly gives

$$\Phi'' = \frac{y_2^2}{v}\left(\frac{1}{r_s^3(\boldsymbol{y},x;h)}+\frac{1}{r_g^3(\boldsymbol{y},x;h)}\right)\,. \quad (15)$$

Since $r_s(\boldsymbol{y},x;h)$ and $r_g(\boldsymbol{y},x;h)$ represent distances (that are always positive) and because we have $y_2 \geq 0$, it follows that the Hessian is positive definite, whence $\mathrm{sgn}(\Phi'') = 1$. Using this in equation (13) gives

$$\beta(\boldsymbol{y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{a(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h);h)}{\sqrt{2\pi|\Phi''(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h)|}} \times e^{-i\omega\Phi(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h)}\hat{U}_s(k_x,\omega)\mathrm{d}k_x\mathrm{d}\omega\,. \quad (16)$$

The stationary midpoint location $x(\boldsymbol{y}, k_x/\omega; h)$ satisfies

$$\left.\frac{\partial \Phi(\boldsymbol{y}, x, k_x, \omega; h)}{\partial x}\right|_{x=x(\boldsymbol{y},k_x/\omega;h)} = 0\ . \tag{17}$$

This condition leads to a quartic equation in x with $x(\boldsymbol{y}, k_x/\omega; h)$ as its solution. This solution gives the midpoint location x as a function of the slope $p = k_x/\omega$ in the CO data and the migrated location $\boldsymbol{y}$. The resulting expression for $x(\boldsymbol{y}, k_x/\omega; h)$ is thus a map demigration as a function of unmigrated (or input) as well as migrated (output) quantities; i.e., the midpoint location is determined as a function of the unmigrated slope and the migrated location. Note that to find the stationary midpoint location $x(\boldsymbol{y}, k_x/\omega; h)$ the quartic equation does to be solved if we decompose the data into curvelets. In that case we can use the resulting central midpoint location (x), central traveltime (t) and central slope (p) to determine the migrated location ($\boldsymbol{y}$) through the map migration equations (2)–(4). In this way, for each input curvelet, we know the relation between the midpoint location x on one side, and the migrated location $\boldsymbol{y}$ and the slope (p) in the data on the other side. Hence, we need not solve the quartic equation if we setup the migration in the data-domain (i.e., an input-based algorithm, see below) rather than in the image-domain (i.e., an output-based algorithm). The explicit form of the quartic equation is given in Appendix C.

Let $\beta_\mu(\boldsymbol{y})$ be the image of one input curvelet c_μ, with multi-index $\mu = (j, l, m_1, m_2)$. That is, we replace the input data $\hat{U}_s(k_x, \omega)$ in equation (16) with one curvelet $\hat{c}_\mu(k_x, \omega)$, where the notation $\hat{c}_\mu$ denotes the Fourier transform of the curvelet $c_\mu(x, t)$ in the $x-t$ domain. Therefore we have

$$\begin{aligned}\beta_\mu(\boldsymbol{y}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} a'(\boldsymbol{y}, k_x, \omega; h) \\ &\quad \times e^{-iP(\boldsymbol{y},x(\boldsymbol{y},k_x/\omega;h),k_x,\omega;h)}\hat{c}_\mu(k_x,\omega)\mathrm{d}k_x\mathrm{d}\omega\ ,\end{aligned} \tag{18}$$

where we have defined

$$a'(\boldsymbol{y}, k_x, \omega; h) := \frac{a(\boldsymbol{y}, x(\boldsymbol{y}, k_x/\omega; h); h)}{\sqrt{2\pi|\Phi''(\boldsymbol{y}, x(\boldsymbol{y}, k_x/\omega; h), k_x, \omega; h)|}} \tag{19}$$

and

$$\begin{aligned}P(\boldsymbol{y}, k_x, \omega; h) &:= \omega\,\Phi(\boldsymbol{y}, x(\boldsymbol{y}, k_x/\omega; h), k_x, \omega; h) \\ &= \omega\, t(\boldsymbol{y}, x(\boldsymbol{y}, k_x/\omega; h); h) \\ &\quad -k_x\, x(\boldsymbol{y}, k_x/\omega; h)\ ,\end{aligned} \tag{20}$$

with

$$t(\boldsymbol{y}, x(\boldsymbol{y}, k_x/\omega; h); h) := (r_s(\boldsymbol{y}, x; h) + r_g(\boldsymbol{y}, x; h))/v \tag{21}$$

the traveltime from the source to the reflector and back to the receiver. This is reminiscent of plane-wave migration (Akbar *et al.*, 1996).

Noticing that $\hat{c}_\mu(k_x, \omega)$ is localized near a wedge with center angular frequency ω_u and center horizontal wavenumber k_x^u, we linearize the dependence of P on ω and k_x around ω_u and k_x^u. This means that, apart from the amplitude a', the oscillatory integral on the right-hand side of equation (18) becomes simply an inverse Fourier transform. Doing this, we have

$$\begin{aligned}P(\boldsymbol{y}, k_x, \omega; h) &\approx P(\boldsymbol{y}, k_x^u, \omega_u; h) \\ &\quad +(\omega-\omega_u)\left.\frac{dP}{d\omega}\right|_{(k_x^u,\omega_u)} \\ &\quad +(k_x-k_x^u)\left.\frac{dP}{dk_x}\right|_{(k_x^u,\omega_u)}\ ,\end{aligned} \tag{22}$$

where the derivatives $dP/d\omega$ and dP/dk_x denote total derivatives. From equation (20) it is immediate that

$$\begin{aligned}P(\boldsymbol{y}, k_x^u, \omega_u; h) &= \omega_u\, t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h) \\ &\quad -k_x^u\, x(\boldsymbol{y}, p_u; h)\ ,\end{aligned} \tag{23}$$

where $p_u := k_x^u/\omega_u$. Then, calculating the total derivative, it follows from equation (20) that

$$\begin{aligned}\left.\frac{dP}{d\omega}\right|_{(k_x^u,\omega_u)} &= t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h) \\ &\quad + \left.\left(\omega\frac{\partial t}{\partial x}\frac{\partial x}{\partial \omega}\right)\right|_{(k_x^u,\omega_u)} \\ &\quad - \left.\left(k_x\frac{\partial x}{\partial \omega}\right)\right|_{(k_x^u,\omega_u)}\ .\end{aligned} \tag{24}$$

Since $p = \partial t/\partial x$ and that $\omega p = k_x$, it follows that

$$\left.\frac{dP}{d\omega}\right|_{(k_x^u,\omega_u)} = t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)\ . \tag{25}$$

Similarly, it follows that

$$\begin{aligned}\left.\frac{dP}{dk_x}\right|_{(k_x^u,\omega_u)} &= \left.\left(\omega\frac{\partial t}{\partial x}\frac{\partial x}{\partial k_x}\right)\right|_{(k_x^u,\omega_u)} \\ &\quad -x(\boldsymbol{y}, p_u; h) - \left.\left(k_x\frac{\partial x}{\partial k_x}\right)\right|_{(k_x^u,\omega_u)} \\ &= -x(\boldsymbol{y}, p_u; h)\ .\end{aligned} \tag{26}$$

Using equations (23), (25), and (26) in (22), we have

$$P(\boldsymbol{y}, k_x, \omega; h) \approx \omega\, t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h) - k_x\, x(\boldsymbol{y}, p_u; h). \tag{27}$$

In addition, using that the amplitude $a'(\boldsymbol{y}, k_x, \omega; h)$ varies slowly over the spectral support (i.e., a spectral wedge) of a curvelet, we approximate

$$a'(\boldsymbol{y}, k_x, \omega; h) \approx a'(\boldsymbol{y}, k_x^u, \omega_u; h)\ . \tag{28}$$

Then, using equations (27) and (28) in (18), it follows that

$$\beta_\mu(\boldsymbol{y}) = a'(\boldsymbol{y},k_x^u,\omega_u;h)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{c}_\mu(k_x,\omega) \times e^{-i\{\omega\, t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)-k_x\, x(\boldsymbol{y},p_u;h)\}}\mathrm{d}k_x\mathrm{d}\omega\ , \tag{29}$$

which is indeed simply an inverse Fourier transform of a curvelet.

Furthermore, we make use of the fact that curvelets remain (fairly) localized in the spatial domain under the action of the migration operator (see Figure 4 for an illustration in the context of time migration). That is, given a curvelet in the data $c_\mu(x,t)$ with center location (x_u,t_u) and main slope p_u, this curvelet is localized near the map-migrated location $\boldsymbol{y}_m$. Therefore, we linearize the dependence of $t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)$ and $x(\boldsymbol{y},p_u;h)$ on $\boldsymbol{y}$ around the map-migrated location $\boldsymbol{y}_m$. Linearizing $x(\boldsymbol{y},p_u;h)$ gives

$$x(\boldsymbol{y},p_u;h) \approx x_u + (\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h))|_{\boldsymbol{y}_m}\cdot(\boldsymbol{y}-\boldsymbol{y}_m)\ , \tag{30}$$

where we used $x_u = x(\boldsymbol{y}_m,p_u;h)$ and where the notation $\nabla_{\boldsymbol{y}}$ denotes the gradient with respect to $\boldsymbol{y}$. Linearizing $t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)$ in $\boldsymbol{y}$ again involves total derivatives instead of partial derivatives since both t and x depend on $\boldsymbol{y}$. Hence, we have

$$\begin{aligned} t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h) &\approx t(\boldsymbol{y}_m,x(\boldsymbol{y}_m,p_u;h);h) \\ &\quad + \left.\frac{dt(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)}{d\boldsymbol{y}}\right|_{\boldsymbol{y}}\cdot(\boldsymbol{y}-\boldsymbol{y}_m) \\ &= t_u + (\boldsymbol{y}-\boldsymbol{y}_m)\cdot\left\{\nabla_{\boldsymbol{y}}t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)\right. \\ &\quad \left. + \frac{\partial t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)}{\partial x}\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h);h)\right\}\Big|_{\boldsymbol{y}_m} \\ &= t_u + (\boldsymbol{y}-\boldsymbol{y}_m)\cdot\left\{\nabla_{\boldsymbol{y}}t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)\Big|_{\boldsymbol{y}_m}\right. \\ &\quad \left. + p_u\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h);h)|_{\boldsymbol{y}_m}\right\}\ . \end{aligned} \tag{31}$$

Using equations (30) and (31) in (29) it follows that

$$\begin{aligned} \beta_\mu(\boldsymbol{y}) &= a'(\boldsymbol{y},k_x^u,\omega_u;h)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{c}_\mu(k_x,\omega) \\ &\times e^{-i(\omega t_u-k_x x_u)} \\ &\times e^{-i\omega\left\{\nabla_{\boldsymbol{y}}t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)+p_u\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h);h)\right\}|_{\boldsymbol{y}_m}\cdot(\boldsymbol{y}-\boldsymbol{y}_m)} \\ &\times e^{-k_x\,\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h)|_{\boldsymbol{y}_m}\cdot(\boldsymbol{y}-\boldsymbol{y}_m)}\mathrm{d}k_x\mathrm{d}\omega\ . \end{aligned} \tag{32}$$

Realizing that the integral is peaked at $\boldsymbol{y}=\boldsymbol{y}_m$, we can replace $a'(\boldsymbol{y},k_x^u,\omega_u;h)$ in equation (32) with $a'(\boldsymbol{y}_m,k_x^u,\omega_u;h)$. Doing this, while recognizing the inverse Fourier transform, we finally have

$$\beta_\mu(\boldsymbol{y}) = a'(\boldsymbol{y}_m,k_x^u,\omega_u;h)\ c_\mu(L\cdot(\boldsymbol{y}-\boldsymbol{y}_m)+\boldsymbol{x}_u)\ , \tag{33}$$

where we have defined

$$L := \begin{pmatrix} \nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h)|_{\boldsymbol{y}_m} \\ \{\nabla_{\boldsymbol{y}}t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)+p_u\nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h)\}|_{\boldsymbol{y}_m}\end{pmatrix} \tag{34}$$

and

$$\boldsymbol{x}_u := \begin{pmatrix} x_u \\ t_u\end{pmatrix}\ . \tag{35}$$

Therefore, it follows that *the leading-order approximation (in angular frequency, horizontal wavenumber, and migrated location) to CO time-migration, consists of a coordinate transformation applied to a curvelet in the data*, given by

$$\boldsymbol{y} = L^{-1}\cdot(\boldsymbol{x}-\boldsymbol{x}_u)+\boldsymbol{y}_m\ , \tag{36}$$

combined with an amplitude scaling, where both the matrix L and the amplitude a' are evaluated with the use of pre-stack map migration.

The matrix L can be written as

$$L = S_{p_u}\cdot T\ , \tag{37}$$

with

$$S_{p_u} := \begin{pmatrix} 1 & 0 \\ p_u & 1\end{pmatrix}\ , \tag{38}$$

$$T := \begin{pmatrix} \nabla_{\boldsymbol{y}}x(\boldsymbol{y},p_u;h)|_{\boldsymbol{y}_m} \\ \nabla_{\boldsymbol{y}}t(\boldsymbol{y},x(\boldsymbol{y},p_u;h);h)|_{\boldsymbol{y}_m}\end{pmatrix}\ . \tag{39}$$

The explicit form of the matrix T for CO time-migration in (39) is derived in Appendix D. Note that the matrix S_{p_u} defines a unilateral shear along the t-axis; i.e., the matrix S_{p_u} shears the input curvelet with slope p_u along the t-axis to have zero slope.

Now that we know the leading-order approximation to one CO time-migrated curvelet c_μ through equation (33), we can determine the total image of the (scattered) data u_s using the reconstruction formula (1). Letting M denote the 2.5-D CO time-migration operator, we have

$$\begin{aligned} \beta(\boldsymbol{y}) &= [Mu_s](\boldsymbol{y}) = \sum_{\mu\in\mathcal{M}}(u_s,c_\mu)\,[Mc_\mu](\boldsymbol{y}) \\ &= \sum_{\mu\in\mathcal{M}}(u_s,c_\mu)\,\beta_\mu(\boldsymbol{y})\ , \end{aligned} \tag{40}$$

where $\mathcal{M}$ is the index set containing all multi-indices μ represented in the data. Finally, because seismic data can be sparsely represented with curvelets, we can write the total image $\beta(\boldsymbol{y})$ as

$$\beta(\boldsymbol{y}) = \sum_{\mu\in\tilde{\mathcal{M}}}(u_s,c_\mu)\,\beta_\mu(\boldsymbol{y})\ , \tag{41}$$

with $\beta_\mu(\boldsymbol{y})$ given by equation (33), and $\tilde{\mathcal{M}}$ the index set that holds the multi-indices of curvelets that survive a certain threshold. Note that the index set $\tilde{\mathcal{M}}$ is determined through thresholding the projection of the seismic data u_s onto the curvelet frame.

Input- versus output-based imaging

Equation (41) is based on a decomposition of the data $u_s(\boldsymbol{x})$ with curvelets, followed by thresholding of the resulting coefficients, and subsequent imaging of each curvelet that survived the thresholding. Therefore, this is an *input-based* seismic migration. This is analogous to convolution of each sample in the data with its associated isochron in accordance with classical Kirchhoff time-migration. Alternatively, *output-based* Kirchhoff migration can be calculated based on diffraction stacks, as is done in practical imaging algorithms; i.e., for each position in the image, a diffraction surface is calculated and the data are stacked over this surface. The resulting stack then gives the image at that particular position in the image.

There is an important difference between input- and output-based migration algorithms. In output-based algorithms the image points are specified and the needed diffraction surfaces in depth migration are calculated by ray-tracing in heterogeneous (anisotropic) media from the image point to the surface. In input-based algorithms, for a particular point in the data, all locations in the image that have the same traveltime as the data point need to be calculated. This is a substantially more complex procedure than ray-tracing from the image point to the subsurface. For this reason many practical migration implementations are output-based algorithms. For the homogeneous media case treated in this work, this difference is irrelevant. Formulating the migration as an output-based imaging algorithm, we have

$$\begin{aligned}\beta(\boldsymbol{y}) &= [Mu_s](\boldsymbol{y}) = \sum_{\mu'\in\mathcal{M}'} ([Mu_s], c_{\mu'})\, c_{\mu'}(\boldsymbol{y}) \\ &= \sum_{\mu'\in\mathcal{M}'} (u_s, [M^* c_{\mu'}])\, c_{\mu'}(\boldsymbol{y}) \,, \end{aligned} \tag{42}$$

where M^* is the adjoint of the migration operator, i.e., the modeling (or demigration) operator. In this case, a curvelet in the image domain is demigrated and its projection onto the data is calculated. Subsequent thresholding then determines if the projection of the demigrated curvelet onto the data exceeds a specified threshold. This procedure would allow *target-oriented* wave-equation-based migration by demigrating curvelets contributing to a particular part of the image only. This requires calculating the leading-order approximation to the demigration operator acting on a single curvelet, i.e. $[M^* c_{\mu'}](\boldsymbol{x})$ in equation (42).

Analysis of the linear transformation L

In Appendix D we derive an explicit expression, equation (D13), for the linear transformation T in 2D CO time-migration. It is given by

$$T = \begin{pmatrix} 1 & \dfrac{\tan\theta_s r_g^3 + \tan\theta_g r_s^3}{r_s^3 + r_g^3} \\ \dfrac{-2\cos\theta\sin\phi}{v} & \dfrac{2\cos\theta\cos\phi}{v} \end{pmatrix} , \tag{43}$$

where θ is the half opening-angle (see Figure 7). Inspection of the second row of this matrix, reveals that this transformation contains a rotation with angle $\phi = (\theta_s + \theta_g)/2$, which is the migrated dip. Making this rotation explicit, we rewrite T as

$$T = T' \cdot R_{-\phi} \,, \tag{44}$$

with

$$T' := \begin{pmatrix} \cos\phi + X\sin\phi & X\cos\phi - \sin\phi \\ 0 & \dfrac{2\cos\theta}{v} \end{pmatrix} , \tag{45}$$

$$R_{-\phi} := \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} . \tag{46}$$

The matrix $R_{-\phi}$ describes a rotation with angle ϕ. Because ϕ is clockwise positive and since t increases downward while x increases to the right, $R_{-\phi}$ describes an anti-clockwise rotation. In the definition of T' we introduced

$$X := \frac{\tan\theta_s r_g^3 + \tan\theta_g r_s^3}{r_s^3 + r_g^3} \tag{47}$$

for notational convenience. Further, inspecting equations (45) it follows that T' can be written as the matrix product of a dilation matrix D and a unilateral shear matrix S. That is, we can write T' as

$$T' = D \cdot S \,, \tag{48}$$

with

$$D := \begin{pmatrix} \cos\phi + X\sin\phi & 0 \\ 0 & \dfrac{2\cos\theta}{v} \end{pmatrix} , \tag{49}$$

$$S := \begin{pmatrix} 1 & \dfrac{X\cos\phi - \sin\phi}{\cos\phi + X\sin\phi} \\ 0 & 1 \end{pmatrix} . \tag{50}$$

Therefore, using equations (44) and (48) in equation (37), it follows that

$$L = S_{p_u} \cdot D \cdot S \cdot R_{-\phi} \,. \tag{51}$$

Finally, using this representation of L in the coordinate transformation in equation (36), which describes the leading-order contribution to 2D CO time-migration, we have

$$\boldsymbol{y} = R_\phi \cdot S^{-1} \cdot D^{-1} \cdot S_{-p_u} \cdot (\boldsymbol{x} - \boldsymbol{x}_u) + \boldsymbol{y}_m \,, \tag{52}$$

where we used that $S_{-p_u} = S_{p_u}^{-1}$.

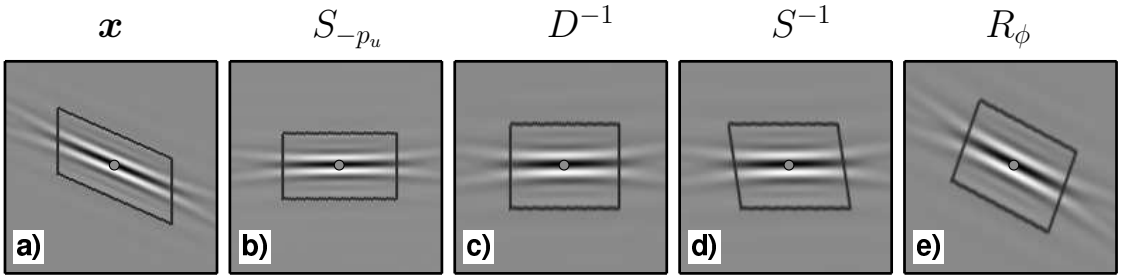

Figure 8. Illustration of the individual components of the linear transformation $L^{-1} = R_\phi \cdot S^{-1} \cdot D^{-1} \cdot S_{-p_u}$ [equation (51)]: input curvelet (a) after S_{-p_u} (b), D^{-1} (c), S^{-1} (d), and R_ϕ (e). The parallellogram superimposed on a curvelet in subfigure a and its transformed versions in b-e allow easy identification of the action of each individual component of L^{-1}.

From equation (52) it follows that the leading-order approximation to CO time-migration can be described by the following sequence of linear[‡‡] transformations. First, a curvelet in the data with center location $\boldsymbol{x}_u$ (i.e., center midpoint location x_u and center two-way traveltime t_u) and center slope p_u, is sheared along the time-axis to have zero slope. Subsequently, the curvelet is dilated in both the vertical and horizontal direction. Then, the curvelet is sheared along the horizontal direction and is rotated to have output dip ϕ. All these linear transformations have the center location $\boldsymbol{x}_u$ of the curvelet as their origin. Finally, the resulting transformed curvelet is translated to the migrated output location $\boldsymbol{y}_m$. This sequence of transformations (except the translation) is depicted for one curvelet in Figure 8, where a parallellogram is added to illustrate the geometric character of these transformations.

Based on our knowledge of map migration, the rotation and translation part of the sequence of transformations is intuitive. The dilation in the vertical direction with $v/(2\cos\theta)$ (after shearing the curvelet along the time axis to have zero slope) takes care of both the time-depth conversion and of the lowering of the frequency content resulting from the imaging condition; i.e., the frequency content is lowered by a factor $\cos\theta$ because the length of the migrated slowness $\boldsymbol{p}_m$ is smaller than the sum of the lengths of the input slownesses $\boldsymbol{p}_{s,r}$ (see Figure 7). Therefore, the translation, rotation and dilation part of the transformation conform to our intuition. However, the unilateral shears along both the vertical and horizontal axes, as well as the dilation of the horizontal axis, are somewhat surprising, and indicate that the leading-order approximation to CO time-migration incorporates a certain amount of linear deformation — this linearity results from the linearization of $t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)$ and $x(\boldsymbol{y}, p_u; h)$ about $\boldsymbol{y}_m$. To understand this, consider the special case of zero-offset migration. In that case the matrices S_{-p_u} and R_ϕ remain unchanged, while the shear matrix S becomes the identity matrix and the dilation matrix D is given by

$$D|_{h=0} = \begin{pmatrix} 1/\cos\phi & 0 \\ 0 & 2/v \end{pmatrix} . \tag{53}$$

Observe that this matrix includes no $\cos\theta$ term because for zero-offset $\theta = 0$ (i.e., $|\boldsymbol{p}_m| = 2/v$). Note also that this matrix includes a squeeze in the horizontal direction equal to $\cos\phi$. Because this squeeze is applied after S_{-p_u}, which shears the input curvelet to have zero slope, the squeeze shortens the long axis of a curvelet. If $\phi = \pi/2$ the long axis of the curvelet is mapped onto a point. This can be understood by noticing that, assuming a constant velocity, we can record a reflection from a dipping reflector with dip $\pi/2$ in the subsurface only if this reflector is at the surface. Since the reflection of this reflector will have a slope $2/v$ in the data, a straight line with this slope will be mapped onto a point.

Calculation of L

All quantities necessary for the calculation of L, can be found from the curvelet decomposition of the data, combined with map migration. First, the central location $\boldsymbol{x}_u = (x_u\ t_u)^T$ of the curvelet in the data is found from the translation indices m_1 and m_2, while the slope p_u is found from the scale index j and the angular index l. This slope determines the matrix S_{-p_u}. Then, using x_u, t_u, and p_u in the 2D map time-migration equations (2)–(4), we find the migrated location $\boldsymbol{y}_m$, and the migrated slope p_m. The migrated dip ϕ can then be calculated from $\phi = \tan^{-1}(vp_m/2)$, which determines the rotation matrix R_ϕ. From $\boldsymbol{y}_m$ and x_u the distances $r_{s,g}(\boldsymbol{y}_m, x_u; h)$ can be calculated using equations (9) and (10) for a given half-offset h, and the angles $\theta_{s,g}$ then follow from $\theta_{s,g} = \cos^{-1}(y_2^m/r_{s,g})$, which determine the half opening-angle $\theta = (\theta_g - \theta_s)/2$. Using the calculated values of ϕ, $r_{s,g}(\boldsymbol{y}_m, x_u; h)$, $\theta_{s,g}$, and θ, we can calculate the matrices D and S using equations (47), (49), and (50). Note that in digital implementations the matrix L needs to be applied on a grid. For this purpose, we derive the discrete form of L in Appendix E.

[‡‡]The translation part of the coordinate transformation is of course not linear.

Extension to pre-stack depth-migration

Although here we have concerned ourselves with time-migration only (i.e., straight rays), the main ingredients of the presented derivation apply to pre-stack depth-migration as well. That is, for pre-stack depth-migration the derivation of the leading-order approximation also contains a stationary phase evaluation of the integral with respect to the non-constant (or non-common) variable[§§], and subsequent linearization of the phase of the resulting oscillatory integral followed by linearization around the (map) migrated location $\boldsymbol{y}_m$. Therefore, for pre-stack depth-migration, a derivation similar to that presented here also leads to the leading-order approximation to depth-migration being a simple transformation of coordinates. For depth migration, however, the transformation is calculated for a smoothed medium, as pointed out by Smith (1998) in the context of solving the wave-equation; i.e., a curvelet with scale index j (i.e., a wavelength proportional to 2^{-j}) is migrated through a smoothed medium that has no heterogeneity beyond a scale proportional to $\sqrt{2^{-j}} = 2^{-j/2}$, i.e., beyond the width of a curvelet. This is reminiscent of the first Fresnel zone of a wave being proportional to $\sqrt{\lambda}$, with λ the wavelength.

In the context of one-way wave-equation-based migration we mention that deformation of a curvelet beyond the leading-order approximation could possibly be calculated using an approach similar to the one given by Smith (1998p.799), who presents a weak solution to the wave equation as the solution of a Volterra equation. The solution of this equation is evaluated based on an initial estimate of the solution constructed from roughly translating the centers of curvelets along the Hamiltonian flow associated with the smoothed media. In the context of one-way wave-equation-based imaging, the wave operator is to be replaced by a one-way operator, while time is replaced by depth. Then the extension of the presented leading-order approximation for CO time-migration to pre-stack depth-migration, would allow the calculation of the initial estimate of the solution of the Volterra equation. We emphasize that this idea remains to be verified.

Numerical examples

Figure 9a shows the 2.5-D CO Kirchhoff migrated curvelet from Figure 4 in more detail, while Figure 9d shows the real part of the associated spectrum. Figure 9b shows the result of using the leading-order approximation (33) to image the same curvelet, while Figure 9c shows the difference between the Kirchhoff result and that of the leading-order approximation. Before subtraction, both images were normalized to the same maximum amplitude so that the difference shows only relative amplitude differences between the two images. Figures 9e and f show the real part of the amplitude spectra of the images shown in Figures 9b and c.

The leading-order approximation (33) based on a simple transformation of coordinates of curvelets provides a good approximation to the Kirchhoff result. The maximum amplitude of the difference between both methods is 24% of the maximum amplitude in the Kirchhoff image. Note from the patterns in the real parts of the spectra, that the curvelet is slightly curved due to the migration, whereas the linear coordinate transformation does not take such bending into account (cf., Figures 9d and e). As a consequence, the main difference in the spectrum occurs on the edges of the support of the curvelet in the frequency domain (see Figure 9f). This difference, attributable to the nonlinear deformation of the curvelet which is absent in the presented leading-order approximation, can again be considered small.

Although Figure 9 shows a favorable comparison for one particular curvelet, of importance is the interference among several curvelets after migration. Figure 10a shows a superposition of eight curvelets, with the same central location in space, and the same scale index, but different directions (or angular indices). Figure 10b shows the amplitude spectrum of the superposition of all eight curvelets, revealing that we used right-ward sloping curvelets only. Figure 10c shows the 2.5-D CO Kirchhoff migrated result (the offset is here 2 km), while Figure 10d shows the result of the leading-order approximation [i.e., equation (40), with $\beta_\mu(\boldsymbol{y})$ calculated using (33)]. Since the data in Figure 10a contain right-ward sloping curvelets only, the Kirchhoff result contains the left part of the bandlimited isochron only, as expected. Comparing Figures 10c and d, the leading-order approximation again provides a good approximation to the Kirchhoff migrated result; the interference of the different curvelets leads to an overal smooth left-part of the bandlimited isochron that compares favorably with the Kirchhoff result. Away from the isochron, both results show the weak tails of the curvelets. For the Kirchhoff result these tails are again somewhat curved towards the isochron, while the leading-order approximation does not include such nonlinear deformation. The difference between both results, however, is small.

Figure 11a shows a CO section (with offset equal 1 km) of synthetic data generated from a syncline model in a constant-velocity background. We use these data to further test the curvelet-based migration, and to highlight that triplications in the data are not a problem; crossing events in the data are simply constructed from curvelets that have the same location but different orientations. In Figure 11a we used only the 0.75% largest curvelet coefficients of the data to reconstruct the data.

[§§]In this work the non-common variable is the midpoint, but in, for example, common-shot depth-migration, it would be the receiver location.

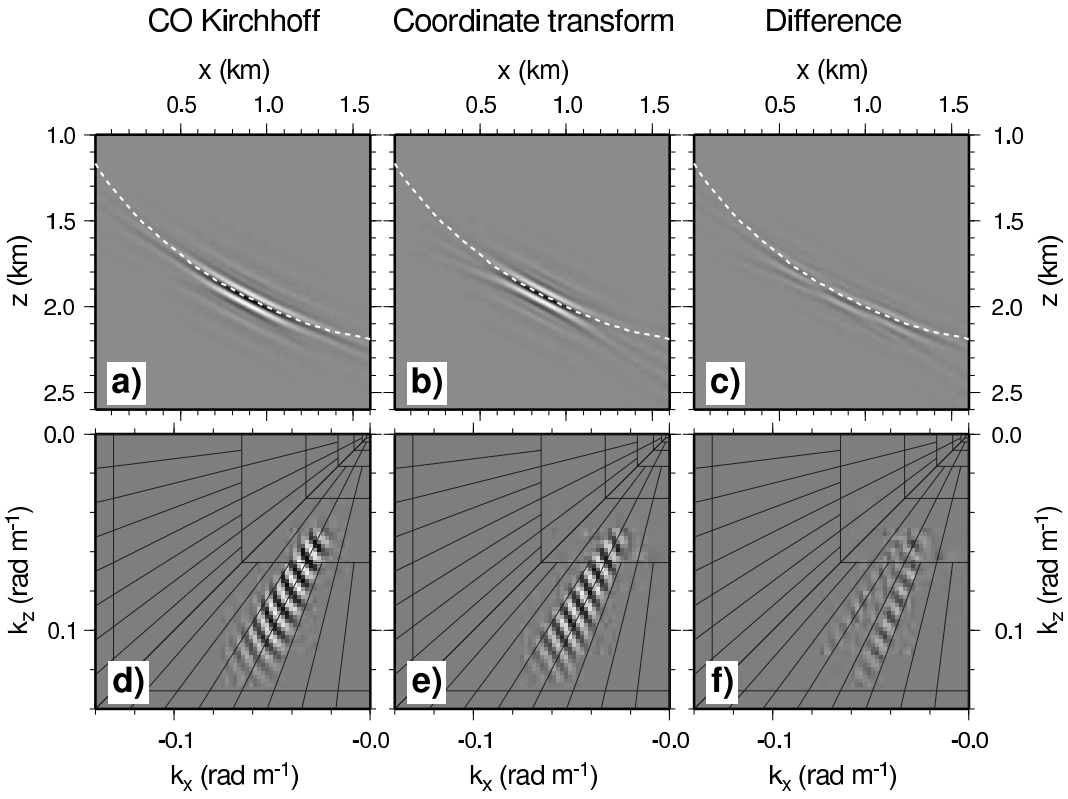

Figure 9. CO Kirchhoff migration of the curvelet shown in Figure 4 (a), coordinate transformation of this curvelet according to equation (36) (b), and the difference (c). The bottom row shows the real part of the spectrum for the CO Kirchhoff migrated curvelet (d), the coordinate transform of equation (36) (e), and the difference (f). The coordinate transformation provides a good first approximation to the Kirchhoff-migrated curvelet.

This thresholding percentage was determined in the same manner as presented earlier for the data shown in Figure 5, i.e., through visually verifying that the difference between the original and reconstructed data is negligible. The associated amplitude spectrum of the reconstructed data is shown in Figure 11b. Observe that almost all energy is contained in frequency bands associated with scale indices $4 \leq j \leq 6$. Figure 11c shows the 2.5-D CO Kirchhoff migrated result, while Figure 11d shows the curvelet-based migration calculated with the leading-order approximation [i.e, using equation (41)]. The leading-order curvelet-based migration provides a good approximation to the Kirchhoff result, with some small artefacts related to the absence of the nonlinear deformation in the leading-order approach.

The bandlimited nature of curvelets allows the possibility to do migration velocity analysis as a function of frequency. Such analysis has physical relevance since a wave averages the medium properties upon propagation over the first Fresnel zone, which is proportional to $\sqrt{\lambda}$, with λ the wavelength, and thus to $1/\sqrt{\omega}$ [e.g., Kravtsov (1988) and Spetzler & Snieder (2004)]. Hence, waves of different frequencies indeed are sentitive to the earth's structure at different scales. Using curvelets in seismic imaging allows naturally for such analysis by simply chosing to image curvelets with selected frequency content (i.e., scale index) only. Although for constant velocity media the medium is the same at every scale, Figure 12 illustrates the principle by showing the resulting images as a function of scale-index j (Figures 12a, b and c). Figures 12d, e, and f, show the associated cumulative images; the images for $j < 4$ are not shown because the data have essentially no energy for $j < 4$ (see Figure 11). Following such per-scale imaging, the associated image gathers can then be submitted to a velocity analysis. The opportunity for such frequency-dependent migration velocity analysis is a straightforward consequence of the curvelets frame being an appropriate representation of both the seismic data and the imaging operator. Note that per-scale imaging also allows the study of the finite-frequency interaction with the interfaces, i.e., the frequency-dependence of the reflection coefficient.

The locally-associated main directions of curvelets allow controlled illumination (Rietveld & Berkhout, 1992) of the subsurface. Figure 13a shows the amplitude spectrum of the data in Figure 11a, but reconstructed with rightward sloping curvelets only. Hence, a dip-filtering is here achieved by simply reconstruct-

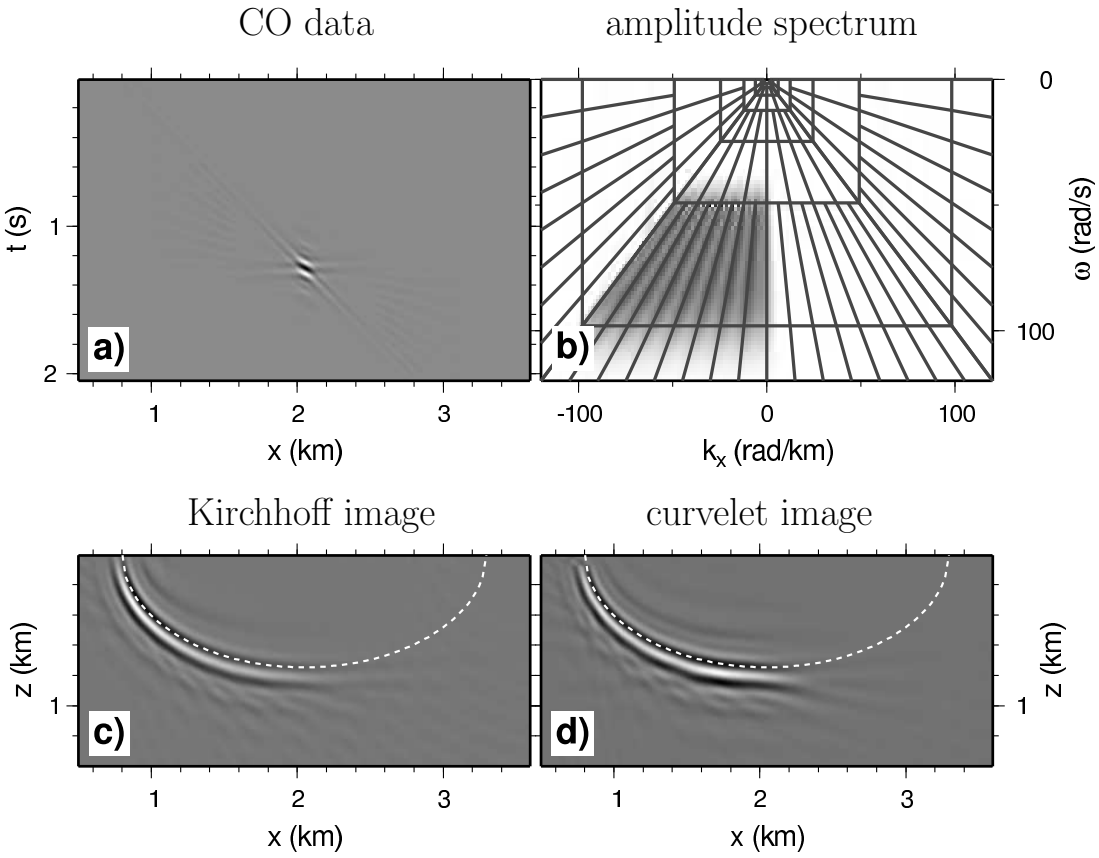

Figure 10. Superposition of eight curvelets (a), the associated amplitude spectrum (b), the CO Kirchhoff-migrated result (c), and the result from the coordinate transformation of equation (36), combined with amplitude scaling [equation (33] (d). The amplitude-corrected and coordinate-transformed curvelet gives a good approximation to the Kirchhoff-migrated result, and the interference between different curvelets results in a smooth isochron. Since we used only rightward-sloping curvelets, the left part of the isochron is constructed only.

ing the data using rightward sloping curvelets only. Figure 11b shows the resulting image, which now mainly shows the left part of the syncline only. Note that a partial image of the discontinuity on the right edge of the syncline is also visible, but now imaged with rightward sloping curvelets only. Here, the controlled illumination is achieved by dip-filtering the *data* with curvelets. Dip-filtering the *image* with curvelets, in contrast, leads naturally to the focusing-in-dip procedure of Brandsberg-Dahl *et al.* (2003). The output-based imaging algorithm based on equation (42) would then simply demigrate only curvelets in the image with certain dips in certain locations, and check if the resulting demigrated curvelet had a projection on the data that exceeded a chosen threshold. Again, these opportunities are in essence a straightforward consequence of curvelets being appropriate building blocks of both seismic data and the imaging operator; no additional preprocessing, such as slant-stacking, is needed.

Discussion

Here, we have calculated the coordinate transformation in equation (36) using a brute-force approach in the spatial domain. For each significant curvelet coefficient, we apply an inverse curvelet transform, and transform the resulting curvelet in the spatial domain according to equation (36). This approach has allowed us to generate the numerical examples and to show the proof of concept of imaging with curvelets using the derived leading-order approximation. This approach, however, does not provide an efficient algorithm to perform seismic imaging with curvelets. Ultimately one would want to calculate the transformation of coordinates in the curvelet frame through a mapping of curvelet indices and coefficients, although an approach that makes use of the compact support of the curvelets in the spectral domain, seems worth investigating also.

Curvelet-based seismic imaging can have several potential benefits over existing imaging algorithms. As mentioned in the previous section, the fact that

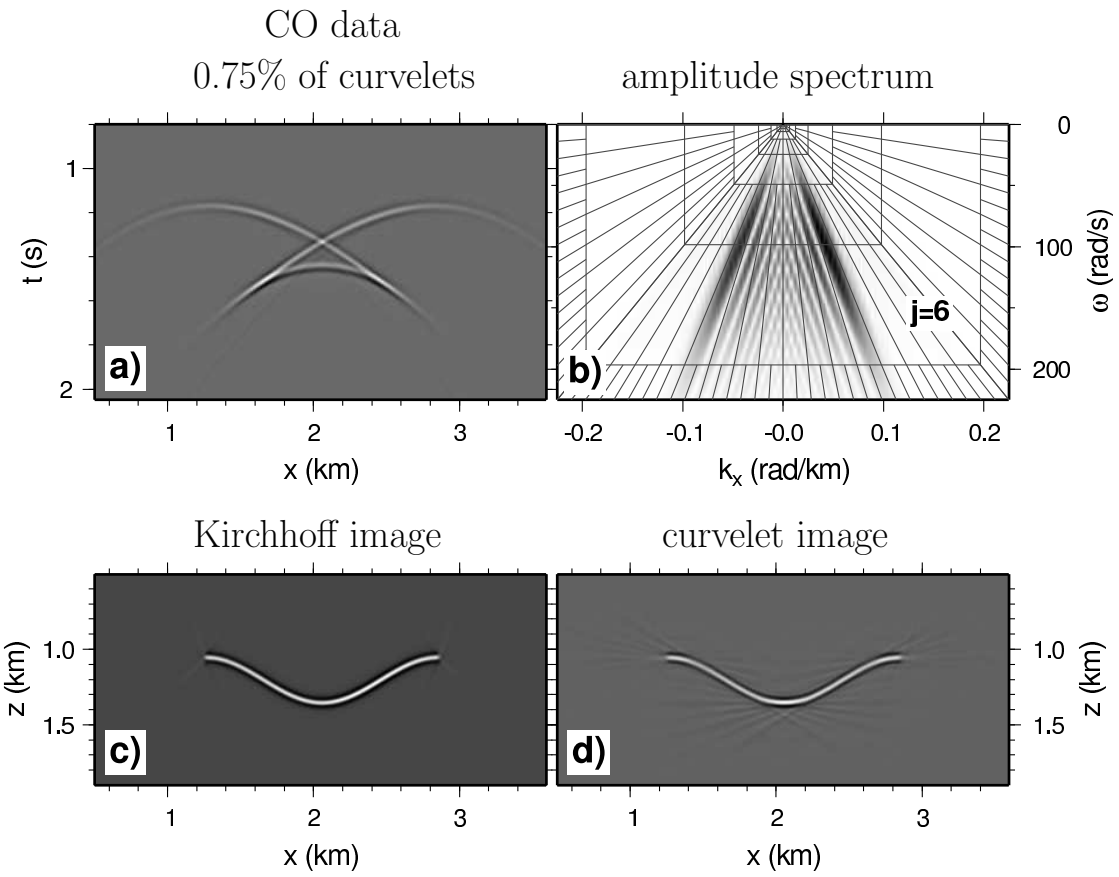

Figure 11. Sparse representation of CO (h = 1000 m) data from a syncline model using only the 0.75% largest curvelet coefficients (a) and its associated amplitude spectrum (b). Each of the curvelets used to reconstruct the data shown in subfigure a is migrated using equation (33), resulting in the image shown in (d).

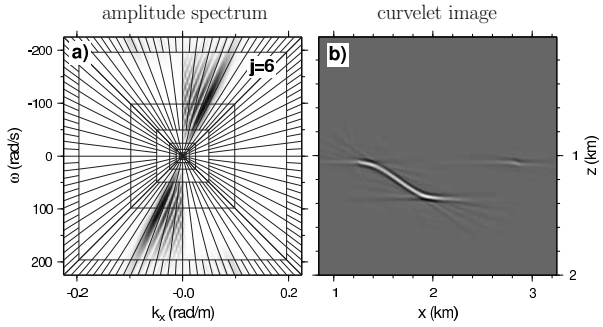

Figure 13. Amplitude spectrum of the thresholded data in Figure 11a, but now using only rightward-sloping curvelets (a), and the resulting image of these curvelets using equation (41) (b). By using curvelets with certain slopes (i.e., certain angular indices) only, we achieve a controlled illumination of the subsurface.

curvelets incorporate the character of the seismic data combined with the fact that they are elements of phase-space, turns extensions to seismic imaging such as controlled illumination, focussing in dip, and frequency-dependent migration velocity analysis, into natural consequences of an appropriate reparameterization of seismic data and the imaging operator. Moreover, the fact that the leading-order approximation to curvelet-based seismic imaging is calculated using map migration, suppresses the need for integration over diffraction surfaces. Combined with the intuition that curvelets allow a sparse representation of the data, which we exemplified here using simple examples, this could therefore lead to a gain in computation time over existing imaging algorithms. However, we currently do not know how much additional computational overhead is needed to calculate the deformation of curvelets beyond the leading-order approximation. Therefore, the question whether curvelet-based imaging will lead to a gain in computational efficiency over existing algorithms remains currently open. Finally, based on the work of Smith (1998), curvelet-based imaging allows imaging through media with limited smoothness, i.e., rougher media.

Even though the curvelets used in this paper are

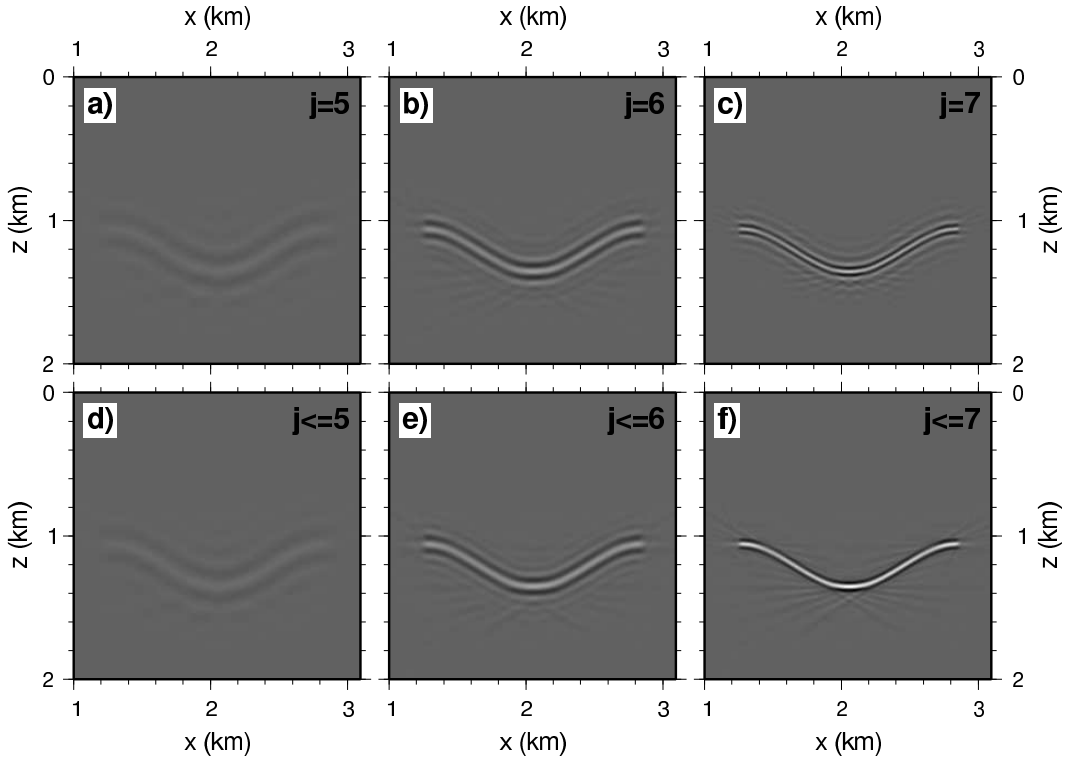

Figure 12. Images obtained using equation (33) from the thresholded CO data shown in Figure 11a, for scales $j = 5$ (a), $j = 6$ (b), $j = 7$ (c), and for the cumulative scales $j \leq 5$ (d), $j \leq 6$ (e), and $j \leq 7$ (f). For $j \leq 4$ there is hardly any energy in the data (see Figure 11b), and hence the resulting images are not shown. This figure illustrates the potential to image the subsurface at different frequencies by simply choosing to use curvelets with certain frequencies (i.e., scale indices) only.

two-dimensional, they can be extended to higher dimensions (Ying *et al.*, 2005). Figure 14 shows an example of a three-dimensional (3D) curvelet in both the spatial domain (a) and the spectral domain (b). In the spatial domain, 3D equivalents of curvelets look like circular disks that are smooth along the disk and oscillatory orthogonal to the disk. Roughly speaking they are smoothed circular pieces of a bandlimited plane wave in 3D[¶¶]. Hence, provided the leading-order approximation to seismic imaging is worked out in three dimensions in, e.g., a common-source setting (or common-offset as in this work), 3D curvelets allow seismic imaging in 3D.

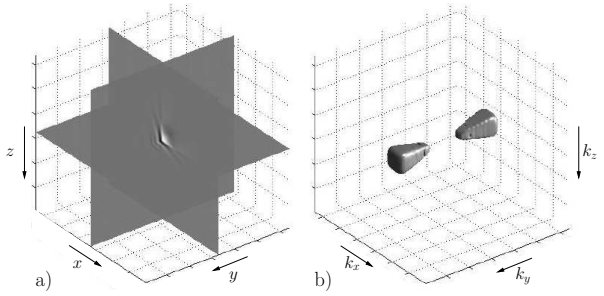

Figure 14. A 3D curvelet in space (a) and its associated (contoured) amplitude spectrum (b). In the spatial domain, 3D equivalents of curvelets look like circular disks that are smooth along the disk and oscillatory orthogonal to the disk.

Conclusion

With curvelets as building blocks of seismic data, the character of the data, i.e., bandlimitation and locally associated directions, can be built into the representation of the data. A simple projection of the data onto the curvelet frame, combined with an intelligent thresholding algorithm, then provides the local slopes in the data. We have presented simple numerical examples to illustrate this. These examples also illustrate the potential of curvelets to sparsely represent the data, even though this needs to be verified for field data in combination with a more sophisticated denoising algorithm.

[¶¶]This rough description ignores that each curvelet has a small range of $\boldsymbol{k}$-vectors associated with it, rather than only a single $\boldsymbol{k}$ direction, as has a plane wave.

We have shown that using curvelets, the leading-order approximation to CO time-migration becomes a simple transformation of the input coordinates of the curvelets in the data, combined with amplitude scaling. This transformation is calculated using CO map time-migration, which uses the local slopes provided by the curvelet decomposition of the data. Considering that the data can be sparsely represented with relatively few curvelets, CO time-migration becomes to leading order the same as a transformation of coordinates of only few curvelets combined with amplitude scaling. We have presented numerical examples indicating that this leading-order approximation provides a good approximation to such time migration.

Although the current work treats constant media only, we emphasize that the essence of the presented derivation applies to pre-stack depth-migration also. Specifically, the derivation of the leading-order approximation to pre-stack depth-migration also contains a stationary phase evaluation of the integral with respect to the non-constant variable, and subsequent linearization of the phase of the resulting oscillatory integral followed by linearization around the map-migrated location of the center of the curvelet. In addition, we briefly sketched a possible way to account for deformation of the curvelet beyond the leading-order approximation while emphasizing that this idea remains to be verified in the future. Because curvelets are basically elements of phase-space, curvelet-based seismic imaging naturally allows the formation of caustics in heterogeneous anisotropic media. Given the bandlimited nature of curvelets, curvelet-based migration combines high-frequency-asymptotics-based Kirchhoff imaging with finite-frequency-based wave-equation migration. This work is the first of a series of papers that will lead to a new generation of seismic imaging algorithms.

ACKNOWLEDGMENTS

The authors thank Emmanuel Candès for providing us with an early version of his Matlab implementation of the digital curvelet transform via unequally spaced fast Fourier transforms. H.D. thanks Roel Snieder for several valuable discussions.

REFERENCES

Akbar, F.E., Sen, M.E., & Stoffa, P.L. 1996. Prestack plane-wave Kirchhoff migration in laterally varying media. *Geophysics*, **61**, 1068–1079.

Billette, F., & Lambaré, G. 1998. Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International*, **135**, 671–690.

Billette, F., Le Bégat, S., Podvin, P., & Lambaré, G. 2003. Practical aspects and applications of 2D stereotomography. *Geophysics*, **68**, 1008–1021.

Biondi, B., & Palacharla, G. 1996. 3-D prestack migration of common-azimuth data. *Geophysics*, **61**(6), 1822–1832.

Bleistein, N., Cohen, J.K., & Stockwell, J.W.Jr. 2000. *Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion*. Springer.

Brandsberg-Dahl, S., de Hoop, M.V., & Ursin, B. 2003. Focusing in dip and AVA compensation on scattering-angle/azimuth common image gathers. *Geophysics*, **68**(1), 232–254.

Candès, E. J., & Demanet, L. 2005. The curvelet representation of wave propagators is optimally sparse. *Communications on Pure and Applied Mathematics*, **58**, 1472–1528.

Candès, E. J., & Donoho, D. L. 2004a. DCTvUSFFT: Digital Curvelet Transform via Unequispaced Fast Fourier Transforms. *Technical Report, California Institute of Technology*.

Candès, E. J., & Donoho, D. L. 2004b. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. on Pure and Appl. Math.*, **57**, 219–266.

Candès, E. J., & Guo, F. 2002. New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Processing*, **82**, 1519–1543.

Candès, E. J., Demanet, L., Donoho, D., & Ying, L. 2005. Fast discrete curvelet transforms. *preprint*.

Candès, E.J., & Donoho., D.L. 2000. *Curvelets - a surprisingly effective nonadaptive representation for objects with edges*. Vanderbilt University Press. Pages 105–120.

Claerbout, J. F. 1976. *Fundamentals of geophysical data processing with applications to petroleum prospecting*. McGraw Hill Company.

Claerbout, J. F. 1992. *Earth soundings analysis: Processing versus inversion*. Blackwell Scientific Publications, Inc.

Claerbout, J.F. 1970. Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure. *Geophysics*, **35**(3), 407–418.

Córdoba, A., & Fefferman, C. 1978. Wave packets and Fourier Integral Operators. *Comm. PDE's*, **3**, 979–1005.

Douma, H., & Calvert, A. 2006. Nonhyperbolic moveout analysis in VTI media using rational interpolation. Accepted for publication in *Geophysics*.

Douma, H., & de Hoop, M. V. 2006. Explicit expressions for pre-stack map time-migration in isotropic and VTI media and the applicability of map depth-migration in heterogeneous anisotropic media. *Geophysics*, **71**(1), S13–S28.

Douma, H., & de Hoop, M.V. 2004. Wave-character preserving pre-stack map migration using curvelets. *In: Expanded Abstracts 74th annual international meeting of the SEG*.

Fefferman, C. 1973. A note on spherical summation multipliers. *Israel J. Math.*, **15**, 44–52.

Fomel, S. 2002. Applications of plane-wave destruction filters. *Geophysics*, **67**(6), 1946–1960.

Fomel, S. 2005. Velocity-independent time-domain seismic imaging using local event slopes. *In: Expanded abstracts of the 75th annual international meeting of the SEG*.

Harlan, W., & Burridge, R. 1983. A tomographic velocity inversion for unstacked data. *Stanford Exploration Project report SEP-37, 1-7*.

Hernández, E., & Weiss, G. 1996. *A first course on wavelets*.

CRC Press, Inc.

Herrmann, F. 2003a. Optimal imaging with curvelets. *In: Expanded Abstracts 73th annual international meeting of the SEG*.

Herrmann, Felix J. 2003b. *Multifractional splines: application to seismic imaging*. Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X, Volume 5207. Pages 240–258.

Hua, B., & McMechan, G. A. 2001. Parsimonious 2D post-stack Kirchhoff depth migration. *Geophysics*, **66**, 1497–1503.

Hua, B., & McMechan, G. A. 2003. Parsimonious 2D prestack Kirchhoff depth migration. *Geophysics*, **68**, 1043–1051.

Jannane, H., Beydoen, W., Crase, E., Cao, D., Koren, Z., Landa, E., Menses, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., Trezeguet, D., & Xie, M. 1989. Wavelengths of Earth Structures that Can Be Resolved From Seismic Reflection Data. *Geophysics*, **54**, 906–910.

Kleyn, A.H. 1977. On the migration of reflection time contour maps. *Geophysical Prospecting*, **25**, 125–140.

Kravtsov, Y.A. 1988. Rays and Caustics As Physical Objects. *Pages 227–348 of:* Wolf, E. (ed), *Prog. in Optics, XXVI*. Amsterdam: Elsevier.

Le Rousseau, J.H, & de Hoop, M.V. 2001. Modeling and imaging with the scalar generalized-screen algorithms in isotropic media. *Geophysics*, **66**(5), 1551–1568.

Mallat, S. G. 1998. *A wavelet tour of signal processing*. First edn. Academic Press, USA.

Ottolini, R. 1983. Velocity independent seismic imaging. *Stanford Exploration Project report SEP-37, 1-7*.

Riabinkin, L.A. 1991. *Fundamentals of resolving power of controlled directional reception (CDR) of seismic waves*. Society of Exploration Geophysicists (Tulsa), Geophysics Reprint Series. Pages 36–60.

Richardson, M. 1958. *College Algebra*. Prentice Hall, Inc.

Rietveld, W.E.A., & Berkhout, A.J. 1992. Depth migration combined with controlled illumination. *SEG Technical Program Expanded Abstracts*, **11**, 931–934.

Smith, H. 1997. A Hardy space for Fourier integral operators. *J. Geom. Anal.*, **7**.

Smith, H. 1998. A parametrix construction for wave equations with $C^{1,1}$ coefficients. *Ann. Inst. Fourier*, **48**, 797–835.

Spetzler, J., & Snieder, R. 2004. The Fresnel volume and transmitted waves. *Geophysics*, **69**, 653–663.

Stein, E. M. 1993. *Harmonic Analysis: Real-variable methods, orthogonality, and oscillatory integrals*. Second edn. Princeton University Press.

Stoffa, P.L., Fokkema, J.T., de Luna Freire, R.M., & Kessinger, W.P. 1990. Split-step Fourier migration. *Geophysics*, **55**(4), 410–421.

Strang, G., & Nguyen, T. 1997. *Wavelets and filter banks*. Revised edn. Wellesley-Cambridge Press.

Sword, C. H. 1987. *Tomographic determination of interval velocities from reflection seismic data: the method of controlled directional reception*. Ph.D. thesis, Stanford Exploration Project, Stanford University.

Ying, L, Demanet, L., & Candés, E. 2005. 3D discrete curvelet transform. *preprint, available at http://www.curvelet.org*.

Zavalishin, B.R. 1981. Perfection of methods for constructing seismic images using controlled directional reception. *Soviet Geology and Geophysics*, **22**, 98–104.

APPENDIX A: A PARTICULAR CONSTRUCTION OF A TIGHT FRAME OF CURVELETS

In this appendix, we give a detailed explanation of the construction of curvelets and show that they constitute a tight-frame for functions in $L^2(\mathbb{R}^2)$. This treatment closely follows the construction of real-valued curvelets by Candès & Donoho (2004b) except for some added derivations and explanations to clarify the construction of curvelets for the non-specialist. We deviate in some places from the original treatment to clarify certain details. For example, we immediately use the notion of *splitting at every other scale* (explained later). A frame similar to the tight frame of curvelets was earlier introduced by Smith (1998).

Before treating the construction of curvelets, we mention that we want to be able to reconstruct a signal in a function space $\mathcal{X}$; i.e., we want curvelets to satisfy the reconstruction formula

$$f = \sum_{\mu \in M} (f, c_\mu)\, c_\mu \quad , \tag{A1}$$

where c_μ denotes a curvelet with multi-index μ (and M some index-set), and $f \in \mathcal{X}$. It is known [e.g., Hernández & Weiss (1996pp.334-336) and Appendix B in this paper] that if $\mathcal{X}$ is a Hilbert space and c_μ are elements in this space, equation (A1) is satisfied if and only if

$$\|f\|_{\mathcal{X}}^2 = \sum_{\mu \in M} |(f, c_\mu)|^2 \ , \quad \forall f \in \mathcal{X} \ , \tag{A2}$$

where $\|\cdot\|_{\mathcal{X}}$ denotes the norm on $\mathcal{X}$. The latter expression implies that the energy of the signal f is conserved through the decomposition (A1); in other words, the curvelets c_μ should be a *partition of unity* (i.e., unit energy) for the reconstruction formula (A1) to hold. Condition (A2) is the definition of a normalized tight frame (see Appendix B for an explanation of tight frames).

Curvelets can be thought of as 2D (anisotropic) extensions to wavelets. Just as in 1D wavelets are localized in one variable and its Fourier dual, curvelets are localized in two variables *and* their two Fourier duals. Such localization is understood within the limits imposed by the Heisenberg uncertainty principle, i.e., $\Delta t \Delta f \geq 1$, with t and f denoting time and frequency, respectively. This principle thus determines a *tile* in the space with t along one axis and f along the other. For curvelets in 2D, choosing the variables to be x and z, such localization is obtained through (hyperdimensional) tiling of the spatial domain (x, z) and the spectral domain (k_x, k_z). We treat the spectral localization first followed by the spatial one.

Let $\chi_{j,l}(\boldsymbol{k})$ be a window (or tile) in the spectral domain, with j an index related to the radial (i.e., scale)

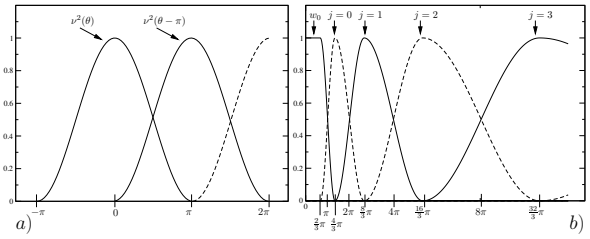

Figure A1. Example of a possible angular window function $\nu(\theta)$ (a), and Lemarié-Meyer windows $|w(2^{-j}|\boldsymbol{k}|)|^2$ for the radial windows (b). The Lemarié-Meyer windows for j and $j+1$ are equal at $2^{j+1}\pi$.

direction, and l an index related to the angular direction; i.e. the localization in the spectrum is dealt with in polar coordinates r and θ. In order for curvelets to constitute a tight frame and thus satisfy the reconstruction formula (A1) (see Appendix B), the windows $\chi_{j,l}(\boldsymbol{k})$ must satisfy

$$\sum_j \sum_l |\chi_{j,l}(\boldsymbol{k})|^2 = 1,\ j \in J, l \in L, \tag{A3}$$

such that $\chi_{j,l}(\boldsymbol{k})$ indeed is a partition of unity. The window $\chi_{j,l}(\boldsymbol{k})$ is constructed through multiplication of an angular window $\nu(\theta)$ and a radial window $w(|\boldsymbol{k}|)$, where $\boldsymbol{k}$ is the wave-vector. We treat the angular window first, followed by the radial window.

Let $\nu(\theta)$ be an even, real-valued, non-negative, angular window function that is infinitely continuously differentiable (i.e., C^∞) and compactly supported on $[-\pi, \pi)$. This function is chosen to be 2π-periodic, such that $\nu(\theta)$ and $\nu(\theta+\pi)$ are exactly out of phase; Figure A1a shows a possible choice of such a window function. Then the angular window satisfies

$$|\nu(\theta)|^2 + |\nu(\theta+\pi)|^2 = 1,\ \theta \in [0, 2\pi).$$

Defining the windows $\nu_{j,l}(\theta) = \nu(2^j\theta - l\pi)$, with $j \geq 0$ and $l = 0, 1, 2, \ldots, 2^j - 1$, it follows that these windows cover the interval $[-\pi/2^j, \pi)$. The windows $\nu_{j,l}(\theta+\pi)$ then cover $[-\pi - \pi/2^j, 0)$. Taking into account the 2π-periodicity of $\nu(\theta)$, it follows that $\nu_{j,l}(\theta)$ and $\nu_{j,l}(\theta+\pi)$ together cover $[0, 2\pi)$. It is important to note that each window $\nu_{j,l}(\theta)$ has width $2\pi/2^j$ and that the width of the window thus depends on the scale j.

Due to the 2π-periodicity of $\nu(\theta)$, we have

$$\begin{aligned}\nu_{j,l}(\theta+\pi) &= \nu(2^j(\theta+\pi) - l\pi) \\ &= \nu\left(2^j\theta - \left(l + 2^j\right)\pi\right) = \nu_{j,l+2^j}(\theta).\end{aligned} \tag{A4}$$

Using this, it follows that for each $j \geq 0$

$$\begin{aligned}\sum_{l=0}^{2^{j+1}-1} |\nu_{j,l}(\theta)|^2 &= \sum_{l=0}^{2^j-1} |\nu_{j,l}(\theta)|^2 + \sum_{l=2^j}^{2^{j+1}-1} |\nu_{j,l}(\theta)|^2 \\ &= \sum_{l=0}^{2^j-1} |\nu_{j,l}(\theta)|^2 + \sum_{l=0}^{2^j-1} \left|\nu_{j,l+2^j}(\theta)\right|^2 \\ &= \sum_{l=0}^{2^j-1} \left(|\nu_{j,l}(\theta)|^2 + |\nu_{j,l}(\theta+\pi)|^2\right) \\ &= 1\,.\end{aligned} \tag{A5}$$

Therefore, energy is indeed conserved in the angular direction using the window functions $\nu_{j,l}(\theta)$. Another way of saying this is that the windows $\nu_{j,l}(\theta)$ satisfy the *admissibility condition* (Candès *et al.*, 2005). There are many different window functions that satisfy this admissibility condition. Hence, curvelets can be constructed using different choices of angular window functions, much as in the construction of wavelets.

The angular windowing provides the curvelets with geometry, i.e., with their main associated directions. To give curvelets their bandlimited nature, howver, the (radial) frequency axis is subdivided into bandpass filters (i.e., subbands). This is done using a radial window w. A possible choice for such a window is the Lemarié-Meyer window [see Hernández & Weiss (1996p.27-28) and Figure A1b]. Let w be a C^∞ real-valued, non-negative, function with support included in $[2\pi/3, 8\pi/3]$. Then Meyer introduces a partition of unity as

$$|w_0(r)|^2 + \sum_{j\geq 0} \left|w(2^{-j}r)\right|^2 = 1, \quad \forall r \geq 0, \tag{A6}$$

with w_0 a *coarse scale* C^∞ window that equals 1 on $[0, 2\pi/3]$ and vanishes beyond $4\pi/3$.

Let $\boldsymbol{k}$ denote the (angular) frequency vector, i.e., $\boldsymbol{k} = (k_x, k_z)$. Then, combining the angular and radial windows into one window $\chi_{j,l}(\boldsymbol{k})$ as

$$\chi_{j,l}(\boldsymbol{k}) = w\left(2^{-j}\,|\boldsymbol{k}|\right)\left[\nu_{\lfloor j/2\rfloor,l}(\theta) + \nu_{\lfloor j/2\rfloor,l}(\theta+\pi)\right], \tag{A7}$$

and setting $|\chi_0(\boldsymbol{k})|^2 = |w_0(|\boldsymbol{k}|)|^2 + |w(|\boldsymbol{k}|)|^2$, it follows that

$$\begin{aligned}&|\chi_0(\boldsymbol{k})|^2 + \sum_{l=0,j\geq 1}^{2^{\lfloor j/2\rfloor}-1} |\chi_{j,l}(\boldsymbol{k})|^2 \\ &= |w_0(|\boldsymbol{k}|)|^2 + |w(|\boldsymbol{k}|)|^2 + \sum_{l=0,j\geq 1}^{2^{\lfloor j/2\rfloor}-1} \left|w\left(2^{-j}\,|\boldsymbol{k}|\right)\right|^2 \\ &\qquad \times \left|\nu_{\lfloor j/2\rfloor,l}(\theta) + \nu_{\lfloor j/2\rfloor,l}(\theta+\pi)\right|^2 \\ &= |w_0(|\boldsymbol{k}|)|^2 + |w(|\boldsymbol{k}|)|^2 + \sum_{j\geq 1}\left(\left|w\left(2^{-j}\,|\boldsymbol{k}|\right)\right|^2\right. \\ &\qquad \left. \times \sum_{l=0}^{2^{\lfloor j/2\rfloor}-1} \left[\nu_{\lfloor j/2\rfloor,l}(\theta) + \nu_{\lfloor j/2\rfloor,l}(\theta+\pi)\right]\right)\end{aligned}$$

$$= |w_0(|\boldsymbol{k}|)|^2 + \sum_{j\geq 0} \left| w\left(2^{-j}\,|\boldsymbol{k}|\right)\right|^2 ,$$

and thus

$$|\chi_0(\boldsymbol{k})|^2 + \sum_{l=0,j\geq 1}^{2^{\lfloor j/2\rfloor}-1} |\chi_{j,l}(\boldsymbol{k})|^2 = 1\,. \tag{A8}$$

where we used equations (A5) and (A6). Here, the notation $\lfloor j\rfloor$ denotes the integer part of j. From equation (A8), $\chi_{j,l}(\boldsymbol{k})$ [together with $\chi_0(\boldsymbol{k})$] form a partition of unity of the whole frequency plane. Note that by construction, the $\chi_{j,l}(\boldsymbol{k})$ have compact support, i.e., curvelets are compactly supported in the spectral domain. Note that there are many different radial windows that could form such a partition of unity, just as for the angular windows; i.e., many different variants of curvelets can be constructed by using different radial window functions.

From the definition of $\chi_{j,l}(\boldsymbol{k})$, $\chi_{j,l}(\boldsymbol{k})$ is obtained through a rotation of $\chi_{j,0}(\boldsymbol{k})$. With this observation, the tiling $\chi_{j,l}(\boldsymbol{k})$ of the frequency plane (k_x, k_z) can be depicted as seen in Figure 1a. Note that the definition of $\chi_{j,l}(\boldsymbol{k})$ using $\nu_{\lfloor j/2\rfloor,l}$ in equation (A7), means that the number of angular windows increases only every other scale (see Figure 1a). This *splitting at every other scale* is a consequence of the choice of dyadic subbands in combination with the parabolic scaling. With the subband filtering being dyadic (i.e., the length of the support of the radial window being multiplied by 2 each time the scale index increases by 1), the width of the curvelet is multiplied by $\sqrt{2}$ each time the scale index increases by 1. Hence, the only way to have an integer number of angular windows per subband with the windows satisfying the parabolic sclaling, is to double the number of angular windows at every other scale only.

Since the radial (or subband) windows $w(2^{-j}|\boldsymbol{k}|)$ for j and $j+1$ are equal at $2^{j+1}\pi$ (see Figure A1b), the frequencies of $\chi_{j,l}(\boldsymbol{k})$ are supported near the *dyadic coronae* $\{2^j\pi \leq |\boldsymbol{k}| \leq 2^{j+1}\pi\}$, depicted in Figure 1a. These dyadic coronae are directly related to the familiar dyadic subband filtering in wavelet theory. This decomposition of the frequency plane is the same as the second dyadic decomposition treated by Stein (1993) in relation to oscillatory integrals of the second kind (or Fourier Integral Operators).

Ignoring the splitting at every other scale, we see that the length of an angular wedge (i.e., in the radial direction) is proportional to $2^{j+1} - 2^j = 2^j$, while the width of the wedge is proportional to $2^j \times 2^{-j/2} = 2^{j/2}$. This means that the volume of the angular wedge is $O\left(2^j \times 2^{j/2}\right)$. Considering the Heisenberg uncertainty principle, a curvelet is therefore supported in the spatial domain near a rectangle of width proportional to 2^{-j} and length $2^{-j/2}$; i.e., the width of this rectangle is the inverse of the length of the angular wedge, while the length of the rectangle is the inverse of the width of the wedge. Therefore, curvelets obey the anisotropic scaling relation

$$\text{width} \propto \text{length}^2\,, \tag{A9}$$

where proportionality is used to indicate the omitted constants[|||]. This scaling is referred to as the *parabolic scaling*. Note that the scaling relation (A9) holds in the spatial domain, while in the frequency domain we have length $\propto$ width^2.

So far we have shown the tiling of the frequency plane with curvelets. The windows $\chi_{j,l}(\boldsymbol{k})$ allow us to rotate and dilate curvelets. To be able to translate curvelets in the spatial domain, the windows $\chi_{j,l}(\boldsymbol{k})$ are multiplied with a local (i.e., over the spectral support of a curvelet) L^2 orthonormal basis. Since all windows $\chi_{j,l}(\boldsymbol{k})$ are simply rotated versions of $\chi_{j,0}(\boldsymbol{k})$, we first consider the window $\chi_{j,0}(\boldsymbol{k})$. The support of this window is contained in the rectangle $R = I_{1j} \times I_{2j}$, with

$$I_{1j} = \{k_x, t_j \leq k_x \leq t_j + L_j\}, \quad I_{2j} = \{k_z, |k_z| \leq l_j/2\}\,,$$

with L_j, and l_j the smallest possible bounds, and t_j the largest possible bound, such that $\chi_{j,0}(\boldsymbol{k}) = 0$ outside this rectangle R. Here L_j and l_j determine, respectively, the sampling intervals (or translation steps) in x- and z-direction in the spatial domain. From a Fourier series expansion for functions with period $2L$, $e^{iq\pi x/L}/\sqrt{2L}$, with $q \in \mathbb{Z}$, is an orthonormal basis for L^2 functions with period $2L$. This means that $e^{i2\pi m_2 k_z/l_j}/\sqrt{l_j}$, with $m_2 \in \mathbb{Z}$, is an orthonormal basis for $L^2(I_{2j})$. Also, $e^{i\pi(m_1+1/2)k_x/L_j}/\sqrt{2L_j}$ with $m_1 \in \mathbb{Z}$, is an orthonormal basis for $L^2(\pm I_{1j})$ [see e.g. Hernández & Weiss (1996Chapter 1.4)]. Therefore, the sequence $(u_{j,m})_{m\in\mathbb{Z}^2}$ [with $m = (m_1, m_2)$] defined as

$$u_{j,m}\,(k_x, k_z) = \frac{e^{i\pi(m_1+1/2)k_x/L_j}\, e^{i2\pi m_2 k_z/l_j}}{\sqrt{2L_j l_j}}$$

is an orthonormal basis for $L^2\,(\pm I_{1j} \times I_{2j})$. Defining δ_1 and δ_2 by $L_j = \delta_1 \pi 2^j$ and $l_j = \delta_2 2\pi 2^{\lfloor j/2\rfloor}$, we have

$$\begin{aligned} u_{j,m}\,(k_x, k_z) \quad &= \quad \frac{2^{-3j/4}}{2\pi\sqrt{\delta_1\delta_2}} e^{i\pi(m_1+1/2)2^{-j}k_x/\delta_1} \\ &\quad \times e^{im_2 2^{-\lfloor j/2\rfloor}k_z/\delta_2}\,, \end{aligned}$$

where the multiplicative term $2^{-3j/4}$ follows from ignoring the splitting at every other scale in the term $2^{-(\lfloor j/2\rfloor + j)/2}$, i.e., replacing $\lfloor j/2\rfloor$ with $j/2$ in this term.

Using this definition, and denoting by $R_{\theta_{j,l}}$ the rotation by angle $\theta_{j,l}$, the curvelet in the frequency domain is then defined as

$$\hat{c}_\mu(\boldsymbol{k}) = 2\pi\chi_{j,l}(\boldsymbol{k}) u_{j,m}\left(R^*_{\theta_{j,l}}\boldsymbol{k}\right), \quad \mu = (j,l,m), \tag{A10}$$

[|||]This scaling relation holds when both width and length are made dimensionless through division by a reference length, typically the sampling interval in numerical implementations. Otherwise this relation would involve a proportionality constant that has a dimension which adjusts for the difference in dimension between width and squared length^2.

where $\hat{c}_\mu$ denotes the Fourier transform of the curvelet $c_\mu \in L^2(\mathbb{R}^2)$. These are the *fine scale* curvelets. The coarse scale curvelets are then defined as $\hat{c}_{\mu_0}(\boldsymbol{k}) = 2\pi\chi_0(\boldsymbol{k})u_m(\boldsymbol{k})$, with $u_m(\boldsymbol{k}) = (2\pi\delta_0)^{-1}e^{i(m_1k_x/\delta_0+m_2k_z/\delta_0)}$. Here, δ_0 is again some appropriate constant determining the sampling rate (or translations steps) in the spatial domain. Note that here δ_0 determines the sampling rate in both the x- and z-direction, since the coarse scale curvelets are isotropic (and thus really are wavelets).

Using the definition of curvelets given in equation (A10), curvelets form a normalized tight frame. From equation (A10), in L^2 we have

$$\sum_{m_1,m_2}\left|\left(\hat{f},\hat{c_\mu}\right)\right|^2 = (2\pi)^2\cdot\int\left|\hat{f}(\boldsymbol{k})\right|^2|\chi_{j,l}(\boldsymbol{k})|^2\,d\boldsymbol{k},$$

where we have used that $\left(u_{j,m}\left(R^*_{\theta_{j,l}}\boldsymbol{k}\right)\right)_{m\in\mathbb{Z}^2}$ is by construction an orthonormal basis for L^2 over the support of $\chi_{j,l}(\boldsymbol{k})$, i.e., $\left|u_{j,m}\left(R^*_{\theta_{j,l}}\boldsymbol{k}\right)\right|^2 = 1$. Using that the $\chi_{j,l}(\boldsymbol{k})$ constitute a partition of unity by equation (A8), it then follows that

$$\begin{aligned}
&\textstyle\sum_{j\geq 0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}\sum_{m_1,m_2}\left|\left(\hat{f},\hat{c}_{(j,l,m_1,m_2)}\right)\right|^2\\
&\quad= (2\pi)^2\sum_{j\geq 0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}\int\left|\hat{f}(\boldsymbol{k})\right|^2|\chi_{j,l}(\boldsymbol{k})|^2\,d\boldsymbol{k}\\
&\quad= (2\pi)^2\int\left|\hat{f}(\boldsymbol{k})\right|^2\left[\sum_{j\geq 0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}|\chi_{j,l}(\boldsymbol{k})|^2\right]d\boldsymbol{k}\\
&\quad= (2\pi)^2\int\left|\hat{f}(\boldsymbol{k})\right|^2 d\boldsymbol{k}\\
&\quad= (2\pi)^2\left\|\hat{f}\right\|_2^2.
\end{aligned}$$

Therefore $(\hat{c}_\mu)_{\mu\in M}$ (with $\mu = (j,l,m_1,m_2)$ a multi-index) is a tight frame for $L^2(\mathbb{R}^2)$. Here, M is a multi-index set that has the appropriate ranges for j, l, m_1 and m_2. Finally, using the Parseval formula and the Plancherel formula for $L^2(\mathbb{R}^2)$, i.e., $\left(\hat{f},\hat{c}_\mu\right) = (2\pi)^2(f,c_\mu)$ and $\left\|\hat{f}\right\|_2^2 = (2\pi)^2\|f\|_2^2$, respectively, it follows that for $f\in L^2(\mathbb{R}^2)$

$$\sum_{\mu\in M}|(f,c_\mu)|^2 = \|f\|_2^2\ . \tag{A11}$$

This shows that $(c_\mu)_{\mu\in M}$ is a normalized tight frame for $L^2(\mathbb{R}^2)$, giving the reconstruction formula (A1).

APPENDIX B: TIGHT FRAMES IN HILBERT SPACES

Let $\mathcal{H}$ be a Hilbert space, and let $(\cdot,\cdot)$ denote the inner product on $\mathcal{H}$. A collection of elements $\{e_j\}_{j\in J}$ in $\mathcal{H}$, with J an index set, is called a **frame** if there exist constants A and B, $0 < A \leq B < \infty$, such that

$$A\|f\|_{\mathcal{H}}^2 \leq \sum_{j\in J}|(f,e_j)|^2 \leq B\|f\|_{\mathcal{H}}^2 \quad \forall f\in\mathcal{H}, \tag{B1}$$

where A and B are called **frame bounds**, and where $\|\cdot\|_{\mathcal{H}}$ denotes the norm on $\mathcal{H}$ induced by the inner product. If the frame bounds are equal ($A = B$), the frame is called **tight**. For example, the four vectors $\phi_1 = (0,1)$, $\phi_2 = (1,0)$, $\phi_3 = \left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$, and $\phi_4 = \left(-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$, form a tight frame for $\mathbb{R}^2$ with frame bounds $A = B = 2$, since it follows that $\sum_{i=1}^4|(f,\phi_i)|^2 = 2\|f\|^2$.

The linear span of frame elements is *dense* in $\mathcal{H}$ [e.g., Hernández & Weiss (1996p.399)], meaning that any element $f\in\mathcal{H}$ can be written as a linear combination of the frame elements, and that the difference between this linear combination and f has a measure zero. If the frame elements were also linearly independent, they would form a basis for $\mathcal{H}$. The frame elements e_j are not a basis since adding the zero vector to $\{e_j\}_{j\in J}$ does not alter the inequalities in expression (B1). Adding any vector, even a zero vector, to a basis, destroys the linear independence of the basis, meaning it no longer would be a basis. When both frame bounds equal unity ($A = B = 1$), the frame is called a *normalized* tight frame. Each orthonormal basis is a normalized tight frame, but the converse is not generally true, since the frame elements need not be linearly independent.

Although the frame elements do not need to be normal, they must satisfy $\|e_j\| \leq \sqrt{B}$, $\forall j\in J$. To see this, let $f = e_k$ for some $k\in J$, and use frame condition (B1) to see that $\|e_k\|^4 = |(e_k,e_k)|^2 \leq \sum_{j\in J}|(e_k,e_j)|^2 \leq B\|e_k\|^2$. Therefore we have $\|e_k\| \leq \sqrt{B}$, $\forall k\in J$. If the frame is a normalized tight frame (i.e., $A = B = 1$), we must have $\|e_k\| \leq 1$. Hence, for a normalized tight frame, the linear independence of a basis is traded for the condition $\|e_k\| \leq 1$, $\forall k\in J$.

If a normalized tight frame has $\|e_k\| = 1$, $\forall k\in J$, $\{e_j\}_{j\in J}$ is an orthonormal basis for $\mathcal{H}$. To see this, observe that for fixed $k\in J$, $1 = \|e_k\|^4 = |(e_k,e_k)|^2 \leq \sum_{j\in J}|(e_k,e_j)|^2 \leq 1$. Therefore, we have $\sum_{j\in J}|(e_k,e_j)|^2 = |(e_k,e_k)|^2 + \sum_{j\neq k,j\in J}|(e_k,e_j)|^2 = \|e_k\|^2 + \sum_{j\neq k,j\in J}|(e_k,e_j)|^2 = 1$, and thus $\sum_{j\neq k,j\in J}|(e_k,e_j)|^2 = 0$. This implies $(e_k,e_j) = 0$, $\forall k\in J$ and $k\neq j$. Since k is arbitrary, and since $\|e_k\| = 1$, it follows that $\{e_j\}_{j\in J}$ is an orthonormal basis for $\mathcal{H}$.

It is known [e.g., Hernández & Weiss (1996pp.334-336)] that for a Hilbert space $\mathcal{H}$ and a family of elements $\{e_j\}_{j\in J}$ in $\mathcal{H}$, the condition

$$\|f\|_{\mathcal{H}}^2 = \sum_{j\in J}|(f,e_j)|^2 \quad \forall f\in\mathcal{H}, \tag{B2}$$

is necessary and sufficient for

$$f = \sum_{j\in J} (f, e_j)\, e_j \tag{B3}$$

to hold. That is, statements (B2) and (B3) are equivalent. This tells us that for a tight frame with frame bound A, reconstruction formula (B3) holds, for if $\sum_{j\in J} |(f,e_j)|^2 = A\,\|f\|^2_{\mathcal{H}}$, simply defining $e'_j = e_j/\sqrt{A}$, gives $\sum_{j\in J} \left|\left(f, e'_j\right)\right|^2 = \|f\|^2_{\mathcal{H}}$. From the equivalence of statements (B2) and (B3), then $f = \sum_{j\in J} \left(f, e'_j\right) e'_j = \frac{1}{A}\sum_{j\in J} (f, e_j)\, e_j$; i.e., for a tight frame we have reconstruction formula (B3). For a normalized tight frame ($A = 1$), $f = \sum_{j\in J} (f, e_j)\, e_j$, and thus also the reconstruction formula (B3). Note that this reconstruction formula is identical to the reconstruction formula for an orthonormal basis, but here the frame elements are not orthogonal; i.e., you can have reconstruction formula (B3) with linearly dependent elements of $\mathcal{H}$, provided (B2) holds.

Although we showed that the reconstruction formula (B3) holds for (normalized) tight frames, a similar reconstruction formula can be found if the frame is not tight. We refer the reader to Mallat (1998Chapter 5), Hernández & Weiss (1996Chapter 8), or Strang (1997Section 2.5) for treatments of non-tight frames.

APPENDIX C: THE QUARTIC EQUATION FOR THE STATIONARY MIDPOINT LOCATION AS A FUNCTION OF Y AND P

The stationary midpoint location $x(\boldsymbol{y}, p; h)$ satisfies the condition that the derivative of the phase Φ with respect to midpoint location is zero. That is, using equation (12), $x(\boldsymbol{y}, p; h)$ satisfies

$$\left(\frac{\partial}{\partial x}\left[\frac{(r_s + r_g)}{v} - px\right]\right) = 0\,, \tag{C1}$$

with r_s and r_g given in equations (9) and (10), respectively. Solving this equation for $x(\boldsymbol{y}, p; h)$ leads to a quartic equation in $\sigma := (x(\boldsymbol{y}, p; h) - y_1)^2$ given by

$$\sigma^4 + a_3\sigma^3 + a_2\sigma^2 + a_1\sigma + a_0 = 0\,, \tag{C2}$$

with

$$a_3 := \frac{y_2^2\left(3 - (vp)^2\right)}{(1 - (vp/2)^2)} - 4h^2\,, \tag{C3}$$

$$a_2 := 6h^4 - \frac{h^2y_2^2\left(3 - (vp)^2\right) - \frac{6}{4}y_2^4\left(2 - (vp)^2\right)}{1 - (vp/2)^2}, \tag{C4}$$

$$\begin{aligned} a_1 := {} & -4h^6 + \left(1 - (vp/2)^2\right)^{-1}\left[y_2^6\left(1 - (vp)^2\right)\right. \\ & -h^4y_2^2\left(3 - (vp)^2\right) \\ & \left.-h^2v^2y_2^4\left(4/p^2 - 2v^2 + p^2v^4\right)\right]\,, \end{aligned} \tag{C5}$$

$$a_0 := \left(h^2 + y_2^2\right)^3\left[h^2 - \frac{(vp)^2y_2^2}{4 - (vp)^2}\right]\,. \tag{C6}$$

Note that for $p > 0$ we have $y_1 < x$, and that for $p < 0$ we have $y_1 > x$. Therefore, in general we have

$$x(\boldsymbol{y}, p; h) = \mathrm{sgn}(p)\sqrt{\sigma} + y_1\,. \tag{C7}$$

The general solution to a quartic equation can be found in, for example, Richardson (1958p.329-331).

APPENDIX D: EXPLICIT EXPRESSION FOR THE LINEAR TRANSFORMATION T FOR COMMON-OFFSET TIME-MIGRATION WITH CURVELETS

To find an explicit expression for the linear transformation T given in equation (39), first consider the term $\nabla_{\boldsymbol{y}} t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)|_{\boldsymbol{y}_m}$. For the subsequent derivation it is important to realize that in calculating the gradient of t with respect to $\boldsymbol{y}$, the midpoint location x is fixed. Using equation (21) together with the definitions of r_s and r_g in equations (9) and (10), it follows that

$$\begin{aligned} &\nabla_{\boldsymbol{y}} t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)|_{\boldsymbol{y}_m} \\ &\qquad = \frac{1}{v}\begin{pmatrix} -\left(\dfrac{x_u - h - y_1^m}{r_s} + \dfrac{x_u + h - y_1^m}{r_g}\right) \\ \\ \dfrac{y_2^m}{r_s} + \dfrac{y_2^m}{r_g} \end{pmatrix} \\ &\qquad = \frac{1}{v}\begin{pmatrix} -(\sin\theta_s + \sin\theta_g) \\ \\ \cos\theta_s + \cos\theta_g \end{pmatrix}\,, \end{aligned} \tag{D1}$$

where $\boldsymbol{y}_m = (y_1^m\ y_2^m)^T$, and where the angles θ_s and θ_g are defined in Figure 7. Then, defining the slowness vectors $\boldsymbol{p}_s$ and $\boldsymbol{p}_g$ at the source and receiver locations, respectively, as

$$\boldsymbol{p}_{s,g} := \frac{1}{v}\begin{pmatrix} -\sin\theta_{s,g} \\ \cos\theta_{s,g} \end{pmatrix}\,, \tag{D2}$$

we can rewrite equation (D1) as

$$\nabla_{\boldsymbol{y}} t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)|_{\boldsymbol{y}_m} = \boldsymbol{p}_s + \boldsymbol{p}_g := \boldsymbol{p}_m\,, \tag{D3}$$

Here, $\boldsymbol{p}_m$ is the slowness vector associated with the dip covector $\boldsymbol{\xi}_m := \omega\boldsymbol{p}_m$ (i.e., the wave vector associated with the reflector). Defining $\phi := (\theta_s + \theta_g)/2$ as the dip (i.e., the angle with the horizontal measured clockwise positive), it follows that

$$\nabla_{\boldsymbol{y}} t(\boldsymbol{y}, x(\boldsymbol{y}, p_u; h); h)|_{\boldsymbol{y}_m} = \frac{2\cos\theta}{v}\begin{pmatrix} -\sin\phi \\ \cos\phi \end{pmatrix}\,, \tag{D4}$$

where $\theta := (\theta_g - \theta_s)/2$ is the half opening-angle.

Next, consider the term $\nabla_{\boldsymbol{y}} x(\boldsymbol{y}, p_u; h)|_{\boldsymbol{y}_m}$. From Appendix C, the stationary midpoint location satisfies equation (C1). Explicitly writing out this equation gives

$$\frac{x - h - y_1}{r_s} + \frac{x + h - y_1}{r_g} + pv = 0\,. \tag{D5}$$

Now, treating x as the dependent variable and $\boldsymbol{y}$ and $p = k_x/\omega$ as the independent variables, we can take the

partial derivative with respect to y_1 on both sides of equation (D5) to get

$$\left(\frac{\partial x}{\partial y_1} - 1\right) \times \left\{ \left(\frac{1}{r_s} + \frac{1}{r_g}\right) + \frac{(x-h-y_1)^2}{r_s^3} + \frac{(x+h-y_1)^2}{r_g^3} \right\} = 0 \,. \tag{D6}$$

Therefore, we have

$$\left.\frac{\partial x(\boldsymbol{y}, p; h)}{\partial y_1}\right|_{\boldsymbol{y}_m} = 1 \,. \tag{D7}$$

This result is intuitive, since changing the horizontal component y_1 of the image location $\boldsymbol{y}$, while keeping the slope p constant, should result in a simple translation of the whole geometry shown in Figure 7 along the horizontal axis. Similarly, taking the partial derivative with respect to y_2 on both sides of equation (D5), we get

$$\frac{1}{r_s}\frac{\partial x}{\partial y_2} - \frac{x-h-y_1}{r_s^2}\frac{\partial r_s}{\partial y_2} + \frac{1}{r_g}\frac{\partial x}{\partial y_2} - \frac{x+h-y_1}{r_g^2}\frac{\partial r_g}{\partial y_2} = 0 \,. \tag{D8}$$

Using that

$$\frac{\partial r_s}{\partial y_2} = \frac{x-h-y_1}{r_s^2}\frac{\partial x}{\partial y_2} + \frac{y_2}{r_s} \,, \tag{D9}$$

$$\frac{\partial r_g}{\partial y_2} = \frac{x+h-y_1}{r_g^2}\frac{\partial x}{\partial y_2} + \frac{y_2}{r_g} \,, \tag{D10}$$

and substituting these expressions in equation (D8), it follows that

$$\left.\frac{\partial x(\boldsymbol{y}, p)}{\partial y_2}\right|_{\boldsymbol{y}_m} = \frac{\tan\theta_s r_g^3 + \tan\theta_g r_s^3}{r_g^3 + r_s^3} \,. \tag{D11}$$

Using this together with equation (D7), it follows that

$$\nabla_{\boldsymbol{y}} x(\boldsymbol{y}, p_u; h)|_{\boldsymbol{y}_m} = \begin{pmatrix} 1 & \dfrac{\tan\theta_s r_g^3 + \tan\theta_g r_s^3}{r_g^3 + r_s^3} \end{pmatrix}^T . \tag{D12}$$

Finally, using the results from equations (D4) and (D12) in equation (39), we find the resulting explicit expression for the linear transformation T to be

$$T = \begin{pmatrix} 1 & \dfrac{\tan\theta_s r_g^3 + \tan\theta_g r_s^3}{r_s^3 + r_g^3} \\ \dfrac{-2\cos\theta\sin\phi}{v} & \dfrac{2\cos\theta\cos\phi}{v} \end{pmatrix} . \tag{D13}$$

APPENDIX E: DISCRETE FORM OF THE LINEAR TRANSFORMATION L^{-1}

For the purpose of digitally implementing the linear transformation L^{-1}, we express the shear, rotation, and dilation matrices that together form L as matrices on a grid with sampling intervals Δt and Δx in the time and midpoint axes, respectively. Starting with the shear matrix S_{-p_u} [see equation (38)], we observe that

$$S_{-p_u} \cdot \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} x \\ t - p_u x \end{pmatrix} \,. \tag{E1}$$

Writing the vector $(x \quad t)^T$ as $(n_x\Delta x \quad n_t\Delta t)^T$, it follows that

$$S_{-p_u} \cdot \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} n_x\Delta x \\ (n_t - p_u\dfrac{\Delta x}{\Delta t}n_x)\Delta t \end{pmatrix} \,. \tag{E2}$$

Therefore, the matrix S_{-pu} expressed on the discrete grid, with sampling intervals Δt and Δx, is given by

$$S^{\Delta}_{-p_u} = \begin{pmatrix} 1 & 0 \\ -p_u\dfrac{\Delta x}{\Delta t} & 1 \end{pmatrix} \,, \tag{E3}$$

where we use the superscript Δ to denote that the matrix is defined on a discrete grid. Hence, defining the vector $\boldsymbol{x}^{\Delta} := (n_x \quad n_t)^T$, we have

$$S^{\Delta}_{-p_u} \cdot \boldsymbol{x}^{\Delta} = \begin{pmatrix} n_x \\ n_t - p_u\dfrac{\Delta x}{\Delta t}n_x \end{pmatrix} \,. \tag{E4}$$

Going through the same analysis for the matrices D^{-1} and S^{-1} [equations (49) and (50)], we find

$$\left(D^{\Delta}\right)^{-1} = \begin{pmatrix} \dfrac{1}{\cos\phi + X\sin\phi} & 0 \\ 0 & \dfrac{1}{\cos\theta} \end{pmatrix} \,, \tag{E5}$$

$$\left(S^{\Delta}\right)^{-1} = \begin{pmatrix} 1 & -\dfrac{X\cos\phi - \sin\phi}{\cos\phi + X\sin\phi}\dfrac{v\Delta t}{2\Delta x} \\ 0 & 1 \end{pmatrix} \,, \tag{E6}$$

while for the rotation matrix $R_{-\phi}$ we get

$$R^{\Delta}_{-\phi} = \begin{pmatrix} \cos\phi & -\sin\phi\dfrac{v\Delta t}{2\Delta x} \\ \sin\phi\dfrac{2\Delta x}{v\Delta t} & \cos\phi \end{pmatrix} \,. \tag{E7}$$

In finding the expressions for $\left(D^{\Delta}\right)^{-1}$, $\left(S^{\Delta}\right)^{-1}$, and $R^{\Delta}_{-\phi}$, we used that the sampling interval Δz in the image satisfies $\Delta z = v\Delta t/2$.

Therefore, using equations (E4)–(E7), we have

$$\begin{pmatrix} n'_x \\ n_z \end{pmatrix} = \left(L^{\Delta}\right)^{-1} \cdot \begin{pmatrix} n_x \\ n_t \end{pmatrix} = R^{\Delta}_{-\phi} \cdot \left(S^{\Delta}\right)^{-1} \cdot \left(D^{\Delta}\right)^{-1} \cdot S^{\Delta}_{-p_u} \cdot \begin{pmatrix} n_x \\ n_t \end{pmatrix} . \tag{E8}$$

In general, the new grid coordinates n'_x and n_z are not integers. Therefore we need to interpolate the obtained grid coordinates to integer values of n'_x and n_z. We do this using bilinear interpolation.