On common-offset pre-stack time migration with curvelets

Huub Douma and Maarten V. de Hoop
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401-1887, USA

ABSTRACT

Recently, curvelets have been introduced in the field of applied harmonic analysis and shown to optimally sparsify smooth (C_2, i.e., twice continuously differentiable) functions away from singularities along smooth curves. In addition, it was shown that the curvelet representation of wave propagators is sparse. 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. In this paper, we study the use of curvelets in pre-stack time migration, and show that simply translating, rotating and dilating curvelets according to the pre-stack map time-migration equations we developed earlier, combined with amplitude scaling, provides a reasonably accurate approximation to time-migration. We demonstrate the principle in two dimensions but emphasize that extension to three dimensions is possible using 3D equivalents of curvelets. We treat time-migration in an attempt to learn the basic characteristics of seismic imaging with curvelets, as a preparation for future imaging in heterogeneous media with curvelets.

Key words: Pre-stack time-migration, common-offset, curvelets, map migration, translation, rotation, dilation

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 reflection occured. The determination of a reflector position *and* orientation from the location of a reflection in the data *and* the local slope, is generally referred to as map migration (Kleyn, 1977). For an overview of literature on this topic, and for an explanation of the applicability condition of map migration, we refer to Douma & de Hoop (2005).

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 (location, times and slopes) that persist in being identical under small perturbations of the reflectors; see Douma & de Hoop (2005) 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 is summed over diffraction surfaces [see, e.g., Bleistein *et al.* (2000)]; such mappings are many-to-one because all points along the diffraction surface are mapped to one output 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 cer-

tainly 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, 1992, p.186-201) or plane-wave destruction filters (Fomel, 2002; Claerbout, 1992, p.93-97).

Recently, in the field of harmonic analysis, Candès and Guo (2002) and Candès and Donoho (2000; 2004b) introduced a tight frame of curvelets (see appendix B for an explanation of tight frames), which provide 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 mainly 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). Of course, at points where the recorded wavefronts have caustic points or at point-like discontinuities in the subsurface (e.g. along faults), the level of sparsity achieved with a curvelet representation naturally will be somewhat less than the sparsity achieved for the smooth parts of the wavefronts or geologic interfaces. Since curvelets are anisotropic 2D extensions of wavelets and thus have a direction associated with them, using curvelets as building blocks of seismic data, the 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 gives the directions associated with the recorded wavefronts.

Smith (1998) and later Candès and Demanet (2002) have shown that curvelets sparsify a certain class of Fourier integral operators. Since the seismic imaging operator can be constructed from Fourier integral operators that belong to this class, and since reflections in seismic data lie mainly along smooth curves, it seems that curvelets are plausible candidates for simultaneous compression of seismic data and the imaging operator. Curvelets have a multiresolution character just like wavelets do. This means that curvelets of different scales have different dominant wavelengths. It is known that waves with a certain dominant wavelength are sensitive to variations in the medium with certain lengths scales only; 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. Therefore curvelets of different scales are sensitive to media with variations on different scales. This allows the possibility to filter 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 media. Smith (1998) has shown that the propagation of a curvelet through such a filtered medium is governed by the Hamiltonian flow associated with the center of the curvelet. Here the center of the curvelet is its center in phase-space, meaning the center location of the curvelet combined with the center direction. This means that a curvelet is treated as if it was a particle with an associated momentum (or direction). For each filtered medium, this observation reduces to the statement of Candès and Demanet (2004) that the propagation of a curvelet (through an infinitely smooth medium) is "well-approximated by simply translating the center of the curvelet along the corresponding Hamiltonian flow." 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.

For homogeneous media the above mentioned filtering is unnecessary. For such media, wave-equation based seismic imaging is really the same as Kirchhoff-style imaging. In this paper, we study the use of curvelets in homogeneous media (i.e., in time migration) and verify the statement that curvelets can be treated as particles with associated directions (or momenta) in an imaging context. We focus on the simple case of homogeneous media in an attempt to learn the basic characteristics of seismic imaging with curvelets, as a preparation for imaging in heterogeneous media with curvelets. This work is a follow-up on earlier work (Douma & de Hoop, 2004) that showed that (at least for time-migration) the kinematics of seismic imaging with curvelets are governed by map migration. This paper is a report on research in progress on pre-stack time imaging with curvelets.

In this paper we first present an intuitive description of curvelets, with examples of digital curvelets from the digital curvelet transform (Candès *et al.*, 2005). A detailed treatment of the construction of real-valued curvelets is included in appendix A. Subsequently, we show an example of the use of curvelets as building blocks of seismic data, and explain the relation between curvelets and map migration. We proceed to explain our current understanding of common-offset (CO) pre-stack time migration with curvelets, and introduce a transformation that consists of translations, rotations and dilations of curvelets to perform such migration. This transformation is largely based on map migration. Finally, we present numerical examples that show the use of this transformation for time-imaging with curvelets, and finish with a discussion and conclusion of the results.

Curvelets

In this section we explain intuitively what curvelets are, how they are constructed, and their main properties. Appendix A provides a detailed treatment of their construction in the frequency domain, which closely follows

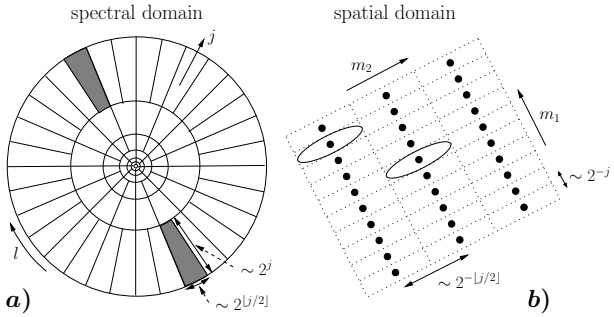

Figure 1. Tilings of the curvelet frame in the spectral domain (a) and the spatial domain (b). In the frequency domain a curvelet is supported 'near' a wedge on a polar grid (i.e. the actual support extends slightly further than the indicated wedge), where the width of the wedge is proportional to $2^{\lfloor j/2 \rfloor}$ 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).

the original treatment of the construction of real-valued curvelets by Candès and Donoho (2004b) but provides additional explanations and derivations to guide the non-specialist (i.e., not harmonic analysts). We include this extensive appendix because most of the literature on the construction of curvelets is rather dense and thus aim to make the construction of curvelets more accessible to a broader audience. For a short summary of (the more general) complex-valued curvelets, we refer the reader to Candès and Demanet (2004).

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 of course 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 direction associated with them. Just like wavelets are 'localized' in one variable *and* its Fourier dual, curvelets are 'localized' in two variables *and* their two Fourier duals. Analogously to wavelets, curvelets can be translated and dilated. The dilation is given also by a scale index j, and, since we are in 2D, the translation is indexed by two indices m_1 and m_2; we defer from the standard notation k_1 and k_2 to avoid confusion with the wave-vector components. A main difference between curvelets and wavelets is that curvelets can be rotated. 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 Figure 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 $\approx$ length^2 in the spatial domain (where we ignore the dimensional difference between width and length^2). This is generally refered to as the *parabolic scaling*. This anisotropic character of curvelets is the key to the proof from Candès and 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 the key difference between wavelets and curvelets.

Curvelets are constructed through the following sequence of operations. First, the spectral domain is bandpass filtered (i.e. 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 outer annulus having twice the radial width as the inner annulus. Each subband is subsequently subdivided into angular wedges (see Figure 1a), where the number of wedges in each subband is determined by the frequency content (or the scale index j) of the subband. 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. This is a consequence of the dyadic nature of the subband filtering done in the first step combined with the desired parabolic scaling. 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 direction associated with 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 multiresolution character (just like 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. Curvelets are in essence a tiling of phase-space; i.e., a tiling of two variables *and* its 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.

An angular wedge in the frequency domain has length proportional to 2^j (i.e., in the radial direction) and width proportional to $2^{\lfloor j/2 \rfloor}$ (see appendix A for the derivation). This means that in the spatial domain the curvelet is oscillatory in the direction of the main $\boldsymbol{k}$-vector (i.e., the $\boldsymbol{k}$-vector pointing to the middle of the angular wedge in the frequency domain), while it is smooth in the orthogonal direction. In some of the papers on curvelets, they are therefore referred to as (Candès & Demanet, 2004) "little needle(s) whose envelope is a specified 'ridge' ... which displays an oscillatory behavior across the main ridge". Intuitively, we can roughly 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 even before the first ever construction of curvelets [e.g. Smith (1997; 1998)], and dates back to the work of Córdoba and Fefferman (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) used in the study of Fourier Integral Operators [see e.g. Stein (1993)], that 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 B for a quick introduction to tight frames, and appendix A for the derivation of this tight frame). This means that, much like in the case of an orthonormal basis, we have a reconstruction formula

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

where c_μ denotes a curvelet with multi-index $\mu = (j, l, m_1, m_2)$, the superscript $*$ denotes taking the complex conjugate, M is an 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 term (f, c_μ) is the coefficient of curvelet c_μ given by the projection of the function f on curvelet c_μ with multi-index $\mu = (j, l, m_1, m_2)$. Of course, $(\cdot, \cdot)$ given in equation (1) is the familiar inner product on $L^2\left(\mathbb{R}^2\right)$.

Digital curvelets versus continuous curvelets

In the construction of continuous curvelets, the sampling of the spectral domain is done in polar coordinates, while the sampling of the spatial domain is Cartesian (see Figure 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. Also, in the field of image analysis [where the digital curvelet transform was 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, since this allows straightforward application of Fast Fourier Transform algorithms. Of course, this holds for seismic data too.

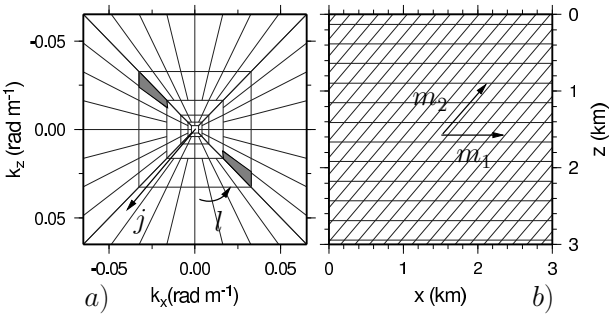

Figure 2. Tilings for digital curvelets in the spectral domain (a) and the spatial domain (b). For digital curvelets, the concentric dyadic circles in the spectral domain are replaced with concentric dyadic squares, and the Cartesian spatial grid is sheared.

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 no longer a rotated Cartesian grid, but is a sheared Cartesian grid (cf. Figure 1b and Figure 2b)*. This construction allows a rapidly computable digital curvelet transform. Whether this digital analog of the continuous curvelets introduces artefacts due to the loss of the rotational symmetry in the spectral domain (this is most severe near the corners of the concentric squares) is currently unclear to us. For more details on the implementation of digital curvelet transforms, we refer to Candès and Donoho (2004a) and Candès *et al.* (2005).

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, the spectral tiling of the digital curvelet transform is shown. The middle column shows

*Here the centers of the cells are the actual possible locations of the centers of the curvelets in space.

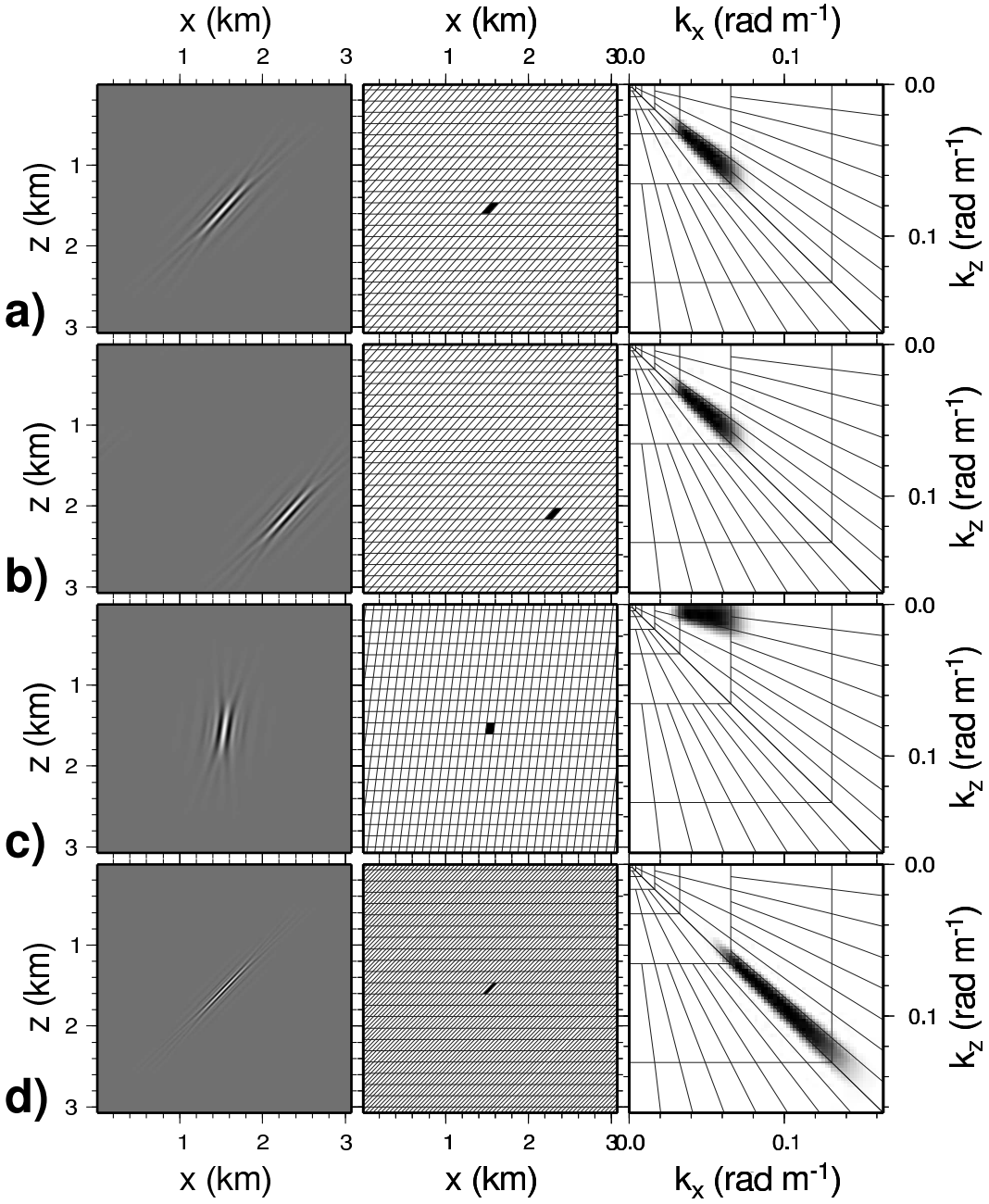

Figure 3. First column: curvelets in the spatial domain. Second column: associated spatial lattices, and spatial cells colored according to the value of the coefficient. Third column: amplitude spectra and frequency-domain tilings. First row: a curvelet. Second row: curvelet from Figure (a) with different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

the associated spatial lattice for each of the curvelets, where the centers of the cells are the actual possible locations of the centers of the curvelets in space. Here the spatial cells on the spatial lattice are colored according to the coefficient of the curvelet (here always one); 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 rotated version of the curvelet in Figure 3a; now the spatial location is the same, but the spectral

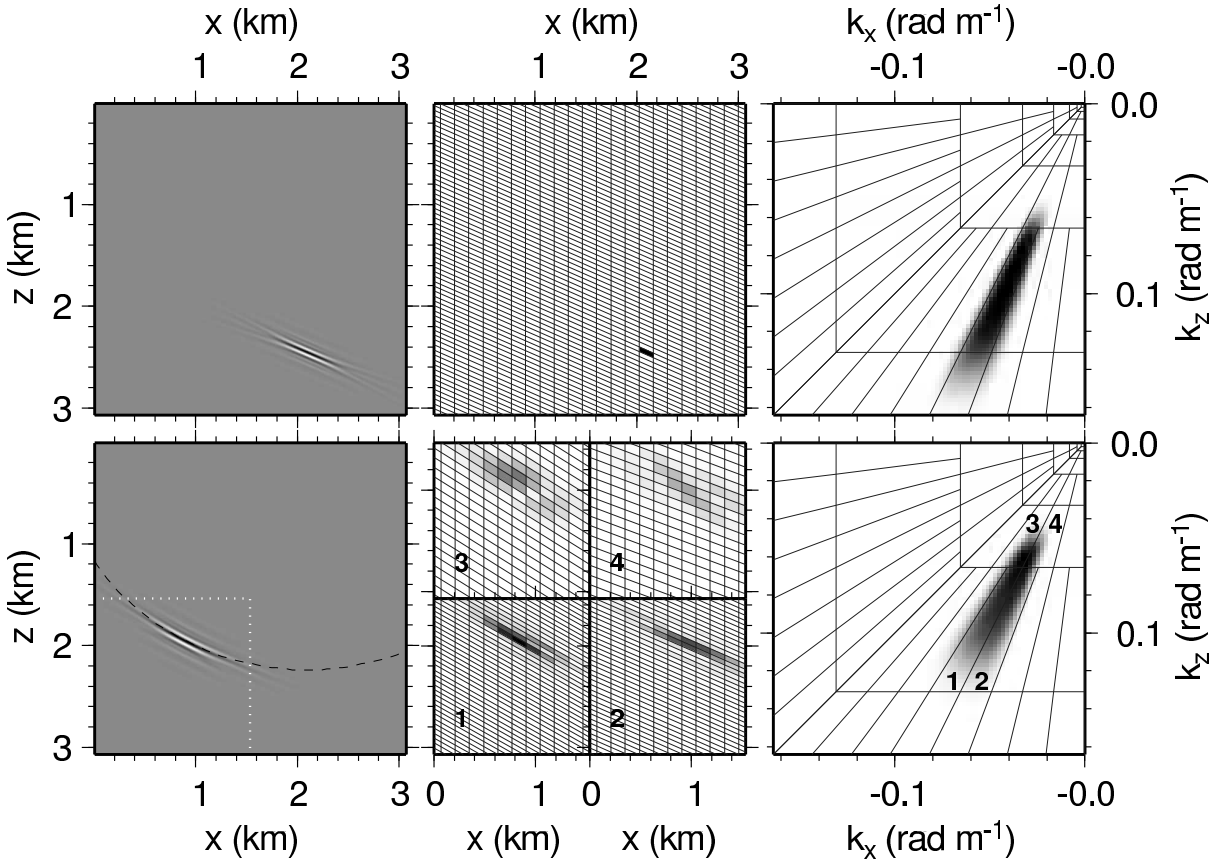

Figure 4. Top row: a curvelet with a dominant frequency of about 30 Hz (left, shown in depth $z = vt_u/2$, for consistency), the normalized absolute value of the coefficient on the spatial lattice (middle), and its amplitude spectrum (right). Bottom row: CO Kirchhoff migration of the curvelet in the 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). The Kirchhoff migration of a curvelet determines only *part of the isochron*, and shows that a curvelet is not mapped onto one other curvelet, but rather several other curvelets. The resulting curvelets are clustered together both in space and spectrum, at least for the constant media case shown here, indicating that a curvelet remains curvelet-like after CO Kirchhoff time migration.

tile has moved within the same concentric squares, i.e., indices j, m_1, and m_2 are the same, but index l has changed. Notice 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 is increased by one. Since the neighboring annulus is subdivided into more wedges, the angular index l has also changed, although the direction of the curvelet is essentially 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, while the curvelet location is essentially the same.

Curvelets remain curvelet-like when subjected to our class of operators

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. The remark in the introduction thus applies: The action on a curvelet of a particular scale can be approximated by flowing out the center of the curvelet in phase space in accordance with the Hamiltonian associated with the medium filtered for this scale. This means that, in the appropriately filtered media, curvelets remain fairly localized in both the spatial domain and the spectral domain. Hence, the propagated 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

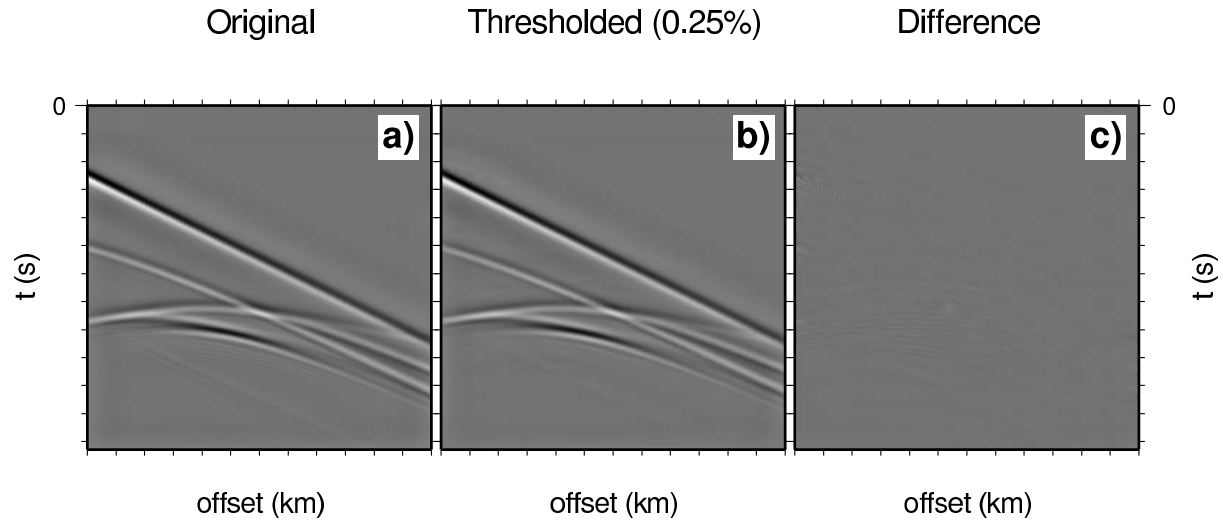

Figure 5. Synthetic common-shot gather with cusped wavefront: original (a), reconstructed using only the 0.25 % largest curvelet coefficients (b), and the difference (c). The reconstruction with 0.25 % of the curvelets is almost identical to the original common-shot gather. In this example, using only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

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.

For homogeneous media, filtering is unnecessary, and the above observation applies to the same medium for curvelets of all scales. To illustrate this, Figure 4 shows the result of CO Kirchhoff migration of a curvelet [taken from Douma and 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), and its associated amplitude spectrum. Again the coefficient of the curvelet 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 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 localized after the migration, and overlies four wedges in the curvelet tiling of the spectrum, indicating some leakage into neighboring curvelets in the spectral domain. 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'. Indeed there is also some leakage to neighboring curvelets in space, but again this can be considered small. This confirms that curvelets remain localized in both the spatial and spectral domain (i.e., they remain 'curvelet-like') after pre-stack time migration.

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, since 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 simply adopt this intuition and illustrate this with a simple synthetic example below.

Figure 5a shows part of a synthetic common-shot gather, where the wavefront has a cusp. This data relates 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 less curvelets than there are sample values in the original gather, to reconstruct the data. From Figure 5c it is clear that the difference between the original and reconstructed data is close to zero. This exemplifies that, using curvelets as building blocks of seismic data, the data can be sparsely represented with curvelets, with much fewer curvelets than the data has samples, and with essentially no residual, even in areas where the wavefront has cusps.

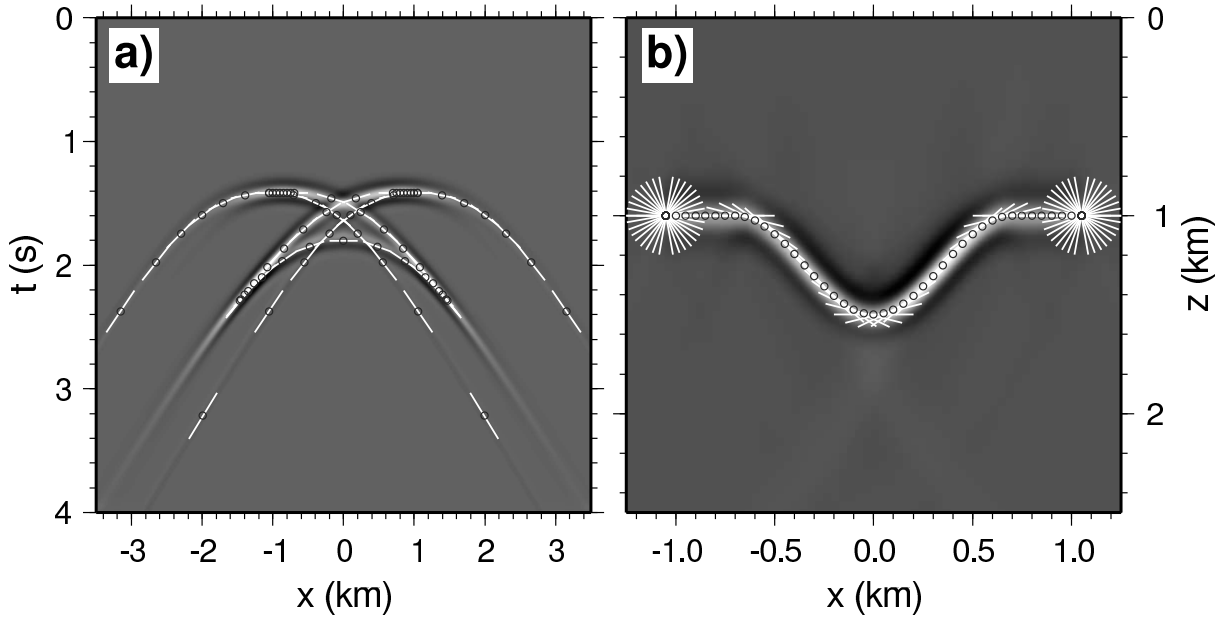

Figure 6. Common offset (h = 1000 m) data (a) and migrated data (b) from a syncline shaped reflector embedded in a constant velocity (v = 2000 m/s) medium, and demigrated and migrated line elements superposed on the data and migrated data, respectively. The excellent agreement between the demigrated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

In our example, we have applied a hard thresholding to the data; we simply did not use 99.75 % of the curvelets. At first sight one might think that therefore the compression ratio should be 400. However, the curvelet transform is redundant, 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). (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 used transform is about 3.) The hard thresholding that we used in our example, will in practice certainly not be ideal to determine the threshold level, especially in a practical situation where we have noise. Here, we refrain from any denoising issues, and focus on the imaging with curvelets. Hence, we assume that an intelligent thresholding of the data has already determined the significant curvelet coefficients in the data.

We emphasize 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 slopes from the data, such as local slant stacking in CDR (Zavalishin, 1981; Harlan & Burridge, 1983; Sword, 1987; Riabinkin, 1991), stereotomography (Billette & Lambaré, 1998; Billette *et al.*, 2003), and parsimonious migration (Hua & McMechan, 2001; Hua & McMechan, 2003), or multidimensional prediction-error filters (Claerbout, 1992, p.186-201) and plane-wave destruction filters (Fomel, 2002; Claerbout, 1992, p.93-97). Therefore, curvelets provide an appropriate reparameterization of seismic data, that have the wave-character of the data built into them.

2D Common-offset map time migration

Douma & de Hoop (2005) present explicit expressions for common-offset map time migration (i.e., migration in a medium with constant velocity), that use only the slope in a common-offset gather (and the velocity), rather than the slope in a common-offset gather *and* the slope in a common-midpoint gather (and the velocity), such as the equations presented by Sword (1987, p.22). The expressions in three dimensions from Douma & de

Hoop (2005) simplify in two dimensions to

$$x_m = x_u - \left(\frac{vt_u}{2}\right)^2 \frac{\Lambda_u}{h}, \quad (2)$$

$$t_m = \sqrt{\left[1-\left(\frac{vt_u\Lambda_u}{2h}\right)^2\right]\left\{t_u^2-\left(\frac{2h}{v}\right)^2\right\}}, \quad (3)$$

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

in which

$$\Lambda_u = \Lambda_u\left(p_u,\Theta_u,h\right) = \frac{1}{2\sqrt{2}p_uh}\sqrt{\Theta_u\left(1-\sqrt{1-\frac{64\left(p_uh\right)^4}{\Theta_u^2}}\right)}, \quad (5)$$

with

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

In these expressions x_u, t_u, and $p_u = \frac{1}{2}\partial t_u/\partial x_u$ are the midpoint location, two-way (unmigrated) traveltime, and the (unmigrated) slope in a CO section, respectively, while x_m, t_m and $p_m = \frac{1}{2}\partial t_m/\partial x_m$ are their migrated counterparts. Also, h denotes the half-offset and v is the medium velocity. Equations (2)-(4) are explicit expressions that determine the migrated reflector coordinates (x_m, t_m, p_m) from the specular reflection coordinates (x_u, t_u, p_u), given h and v. Equations (2)-(4) do not use the offset horizontal slowness $p_h = \frac{1}{2}\partial t_u/\partial h$, such 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 (Sergey Fomel, personal communication). In this way, map time-migration can be done *without knowledge of the medium velocity*. This idea dates back to the work of Ottolini (1983). In the context of pre-stack time-migration with curvelets, this would require 3D equivalents of curvelets for 2D imaging since both p_u and p_h would need to be known.

Figure 6a shows a common-offset gather ($h = 1000$ m) from synthetic data generated from a syncline model with constant velocity above the reflector. On top of 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 (x_m, z_m), with $z_m = vt_m/2$, and the local dip angle $\tan\phi = vp_m$ can be determined (ϕ is the angle with the horizontal measured clockwise positive). Figure 6b shows the migrated common-offset equivalent of the data shown in Figure 6a, with the migrated line elements, determined using x_m, t_m, and p_m, drawn on top of the image. The migrated line elements accurately follow the directions in the image, indicating that equations (2)-(4) accurately 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 identical to building a (band-limited) delta-function with plane waves from all directions; i.e., the Fourier transform of a delta function has all directions.

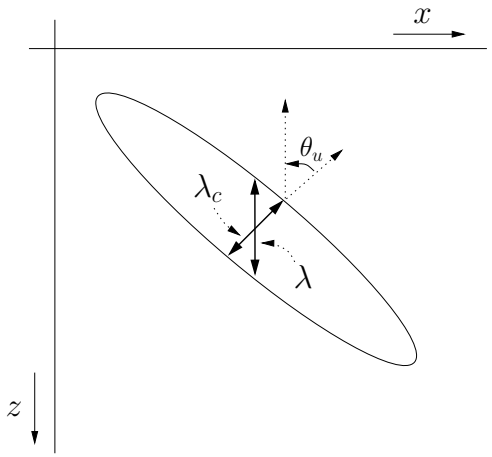

Figure 7. The width of the curvelet in the direction perpendicular to its 'main ridge' determines wavelength λ_c, while the vertical width of the curvelet determines the wavelength λ relevant to seismic migration. It follows that $\lambda_c = \lambda\cos\theta_u$ and hence $\omega = \omega_c\cos\theta_u$.

Knowing that each curvelet has a (few) direction(s) 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), gives us 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. This was also mentioned by Douma and de Hoop (2004).

2.5-D Common-offset time migration with curvelets

As mentioned in the previous section, the CO map time-migration equations (2)-(4) use the slope in the CO domain only. This implies that the flow from a curvelet in the data domain to the image domain is determined by one slope only. Therefore, the following scheme for time-imaging with curvelets emerges. Using curvelets as building blocks of seismic data, the directions (or local slopes) in the data follow from a straightforward projec-

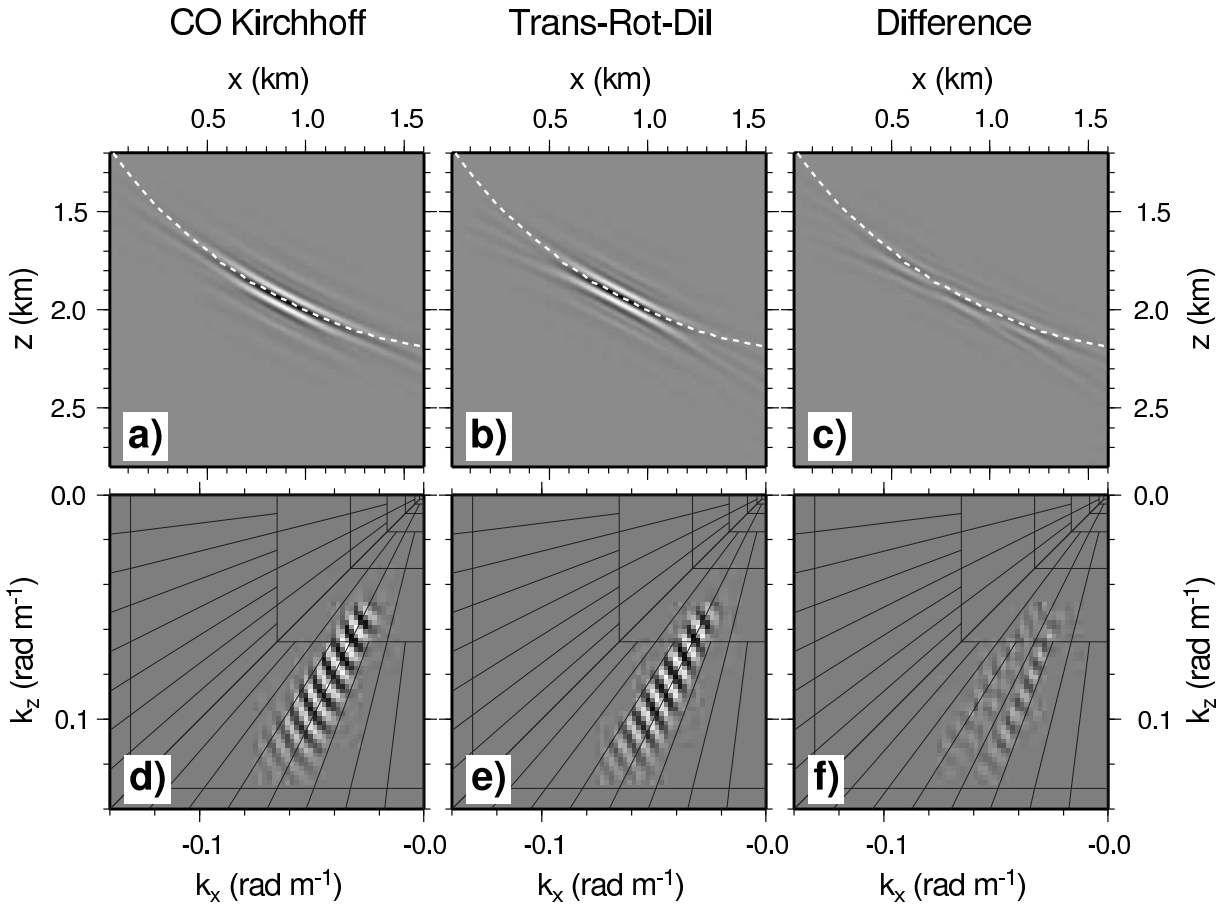

Figure 8. CO Kirchhoff migration of the curvelet shown in Figure 4 (a), TRD transformation of this curvelet (b), and the difference (c). The bottom row shows the real part of the spectrum for the CO Kirchhoff migrated curvelet (d), the TRD transform (e), and the difference (f). The TRD transform gives a good approximation to the Kirchhoff migrated curvelet.

tion of the data onto the curvelet frame. Thresholding of the curvelet coefficients then gives the curvelets associated with the main wavefronts in the data, and thus the directions associated with these wavefronts also. Subsequently, these directions can be used in our map migration equations (2)-(4) (together with an estimate of the velocity), to determine the migrated location and orientation of the curvelet. Hence, these equations determine a translation and a rotation of the curvelet.

It is known (e.g., Bleistein *et al.* (2000), p.223 and p.235) that after pre-stack migration, the length of the $\boldsymbol{k}$-vector changes according to

$$|\boldsymbol{k}| = \omega \nabla_y \phi(\boldsymbol{y}, \boldsymbol{x}_s, \boldsymbol{x}_r) = \frac{\omega}{v(\boldsymbol{y})} \cos\theta(\boldsymbol{y}, \boldsymbol{x}_s, \boldsymbol{x}_r) \ , \qquad (7)$$

where ω is the angular frequency, $\boldsymbol{x}_{s,r}$ are the source and receiver locations, $\boldsymbol{y}$ is the output location in the image, $\phi(\boldsymbol{y}, \boldsymbol{x}_s, \boldsymbol{x}_r)$ is the two-way traveltime from the source location $\boldsymbol{x}_s$ to the reflector at $\boldsymbol{y}$ to the receiver position $\boldsymbol{x}_r$, $v(\boldsymbol{y})$ is the velocity at output location $\boldsymbol{y}$, and $\theta(\boldsymbol{y}, \boldsymbol{x}_s, \boldsymbol{x}_r)$ is the half opening-angle between the ray from the source to the scattering point, and the ray from the receiver to the scattering point. The term $\cos\theta(\boldsymbol{y}, \boldsymbol{x}_s, \boldsymbol{x}_r)$ is usually referred to as the *obliquity* factor. In 2D, for a constant background velocity and a CO acquisition geometry, equation (7) simplifies to $|\boldsymbol{k}| = \frac{\omega}{v} \cos\theta(x_m, t_m, x_u, h)$, where h is the half offset, x_u the common midpoint location along the 2D acquisition profile, and the angle $\theta(x_m, t_m, x_u, h)$ is given by

$$\theta(x_m, t_m, x_u, h) = \frac{1}{2}\left\{\tan^{-1}\left(\frac{2\,(x_u + h - x_m)}{v t_m}\right) - \tan^{-1}\left(\frac{2\,(x_u - h - x_m)}{v t_m}\right)\right\} . \qquad (8)$$

A curvelet has dominant wavelength λ_c in the direction orthogonal to the curvelet (see Figure 7). However, the dominant wavelength λ in the vertical direction determines the frequency relevant to migration. It follows that $\lambda_c = \lambda \cos\theta_u$, which gives

$$\omega = \omega_c \cos\theta_u \ , \qquad (9)$$

where θ_u is the phase-angle (measured clockwise positive with the vertical) of the curvelet in the data domain.

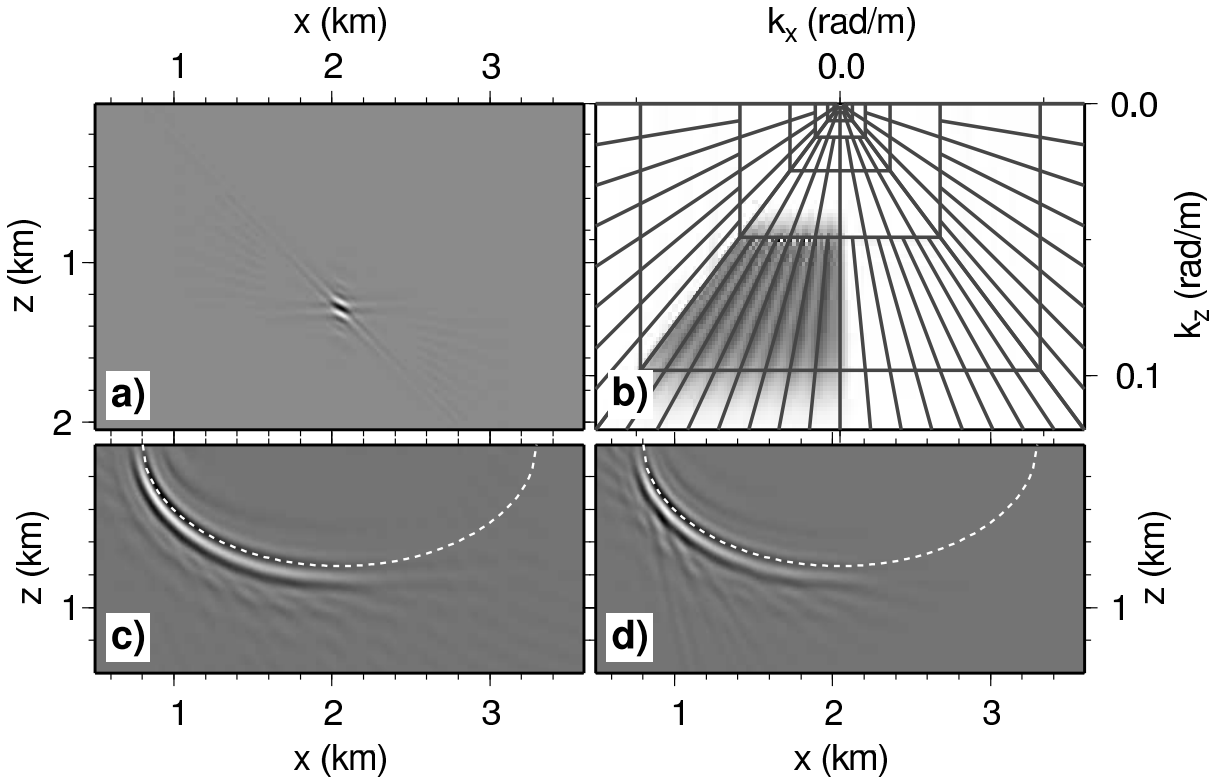

Figure 9. Superposition of 8 curvelets (a), the associated amplitude spectrum (b), the CO Kirchhoff migrated result (c), and the result from the amplitude corrected TRD transform (d). The amplitude corrected TRD transform gives a decent approximation to the Kirchhoff migrated result, and the interference between different curvelets results in a reasonably smooth isochron. Since we used only curvelets with $\boldsymbol{k}$-vectors pointig to the left, only the left part of the isochron is constructed.

Combining equations (7) and (9) it follows that, after migration, the length of the $\boldsymbol{k}$-vector can be written as

$$|\boldsymbol{k}| = \frac{\omega_c}{v(\boldsymbol{y})} \cos\theta(x_m, t_m, x_u, h) \sqrt{1 - (v(\boldsymbol{y})p_u)^2}\,, \qquad (10)$$

where we used $\cos\theta_u = \sqrt{1-(v(\boldsymbol{y})p_u)^2}$. This means that the curvelet needs to be dilated (or stretched) in the image domain with dilation factor D, given by

$$D = \frac{1}{\cos\theta(x_m, t_m, x_u, h)\sqrt{1-(vp_u)^2}}\,. \qquad (11)$$

where we specialized to the constant background velocity case. Therefore, CO pre-stack time migration with curvelets can be done by translating and rotating the curvelets in the data domain according to the map migration equations (2)-(4), and dilating the curvelets with a stretch factor given by equation (11). In these equations, the parameters x_u, t_u, and p_u are the position (x and t) and direction of the center of the curvelet respectively. Throughout this paper, we refer to the translation, rotation and dilation of a curvelet as the *TRD transformation* of a curvelet.

The TRD transformation of a curvelet provides the kinematics of imaging with curvelets, but ignores the dynamics, often referred to as the 'true amplitude' part of seismic imaging. For constant background velocities and a CO acquisition geometry, Bleistein *et al.* (2000) show that the equation for 2.5-D Kirchhoff inversion for reflectivity is given by [their equation (6.3.25); note the missing division with the velocity v in the phase function]

$$\begin{aligned}\beta(\boldsymbol{y}) \quad &= \quad \frac{1}{\sqrt{2\pi}} \int\int A(\boldsymbol{y}, x_u, \omega; h) e^{-i\omega[r_s+r_g]/v} \times \ldots \\ &\qquad u_S\left(x_g, x_s, \omega\right) dx_u d\omega, \end{aligned} \qquad (12)$$

with

$$\begin{aligned} A(\boldsymbol{y}, x_u, \omega; h) \quad &= \quad \frac{4y_3\sqrt{r_s+r_g}\left(r_s^2+r_g^2\right)}{\left(v\, r_s r_g\right)^{3/2}} \times \ldots \\ &\qquad \cos\theta(\boldsymbol{y}, x_u, h)\sqrt{|\omega|}e^{i\frac{\pi}{4}\mathrm{sgn}\omega}\,, \end{aligned} \qquad (13)$$

where $\boldsymbol{y} = (y_1, y_3)$ is the output location in the image domain, x_u is the midpoint location along the 2D acquisition profile, $r_{s,g}$ are the distance from the output location $\boldsymbol{y}$ to the source and receiver at $x_s = (x_u - h, 0)$ and $x_r = (x_u + h, 0)$, respectively, and u_S is the (singly) scattered data. To find the amplitude correction for the migration of a curvelet, we would therefore need to use a curvelet as scattered data u_S, and evaluate the

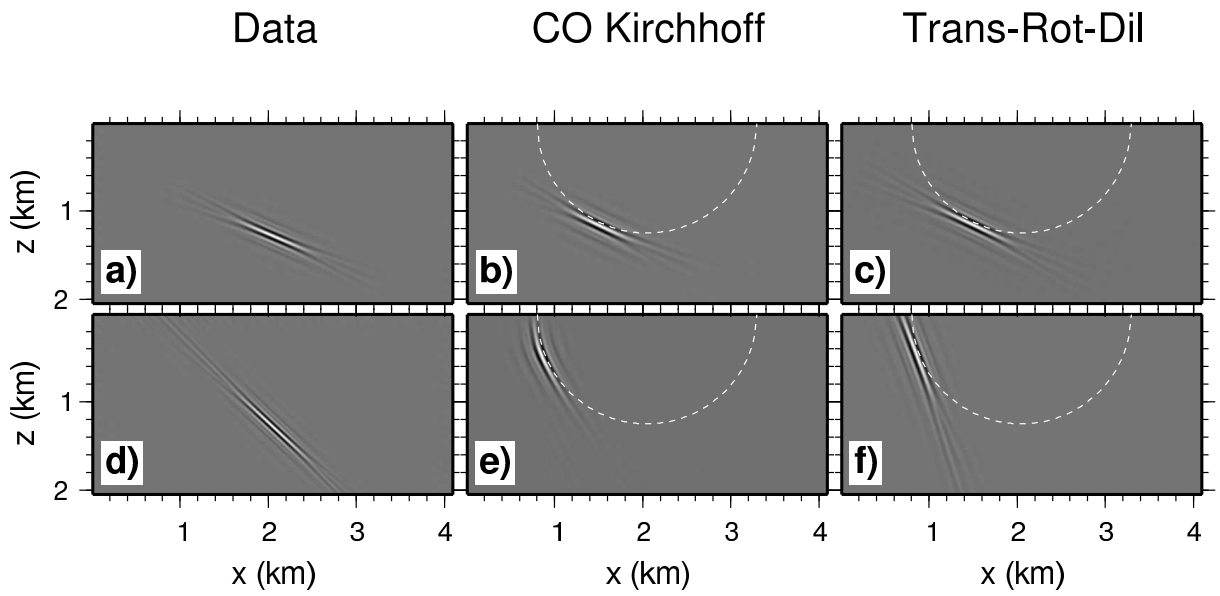

Figure 10. A shallow dipping curvelet (a) and a steep dipping curvelet (d), their CO Kirchhoff migrated counterparts [(b) and (e), respectively], and the TRD transformed counterparts [(c) and (f), respectively]. The steeper dipping curvelet is more strongly bended towards the isochron than is the shallower dipping curvelet.

integrals in equation (12). A natural way to evaluate the oscillatory integral in equation (12), is to use the method of stationary phase [e.g. Bleistein (1984, p.77-81) and Bleistein *et al.* (2000, p.127-135)]. Such an analysis should reveal the amplitude correction for the migrated curvelet. At the moment of writing this report, we have not done such an analysis. Therefore, for the purpose of generating the results in this report, we simply scale the curvelet with $A(\boldsymbol{y}, x_u, \omega; h)$ knowing that ultimately we want to do the stationary phase analysis.

Numerical examples

Figure 8a shows the 2.5-D CO Kirchhoff migrated curvelet from Figure 4 in more detail, while Figure 8d shows the real part of the associated spectrum. We plot the real part of the spectrum instead of the amplitude spectrum, to see the phase information and amplitude information combined in one plot. Figure 8b shows the result of the TRD transformation of the same curvelet, where the translation and rotation is determined by the map migration equations (2)-(4), and the dilation is determined by the stretch factor given in equation (11). Figure 8c shows the difference between the Kirchhoff migrated curvelet and the TRD transformed curvelet. Before subtraction, both images were normalized to have maximum amplitude one, such that the difference only shows relative amplitude differences between both images. Figures 8e and f show the real part of the amplitude spectra of the images shown in Figures 8b and c.

It is clear that for this particular curvelet, the TRD transformation of the curvelet gives a reasonable approximation to the Kirchhoff result. The maximum amplitude of the difference between both methods is about 20% of the maximum amplitude in the Kirchhoff image. From the lines of equal phase in the spectra, we see that the curvelet is slightly bent due to the migration, whereas the TRD transformation does not take this bending into account (cf. Figures 8d and e). The main difference in the spectrum occurs on the edges of the support of the curvelet in the frequency domain, where this bending is strongest. Overall, for this particular curvelet, the TRD transformation gives a satisfactory approximate image when compared to the Kirchhoff result.

Even though Figure 8 shows good results for one (particular) curvelet, this does not show the interference between different curvelets. Figure 9a shows a superposition of 8 curvelets, with the same central location in space, and the same scale index, but different directions (or angular indices). The vertical axis is here again converted from two-way traveltime to depth. Figure 9b shows the amplitude spectrum of all 8 curvelets combined, and it is clear that we used curvelets that have leftward pointing $\boldsymbol{k}$-vectors only. Figure 9c shows the 2.5-D CO Kirchhoff migrated result, while Figure 9d shows the result of our TRD transformation combined with an amplitude scaling given by equation (13). As expected, the Kirchhoff result gives the left part of the isochron only. Comparing Figures 9c and d, the TRD transformation combined with the amplitude correction of equation (13) gives a reasonable approximation to the

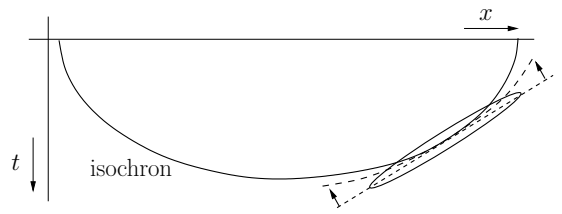

Figure 11. The observed bending of curvelets, mostly for the steeper dips in the data, could possibly be corrected for by bending the curvelet towards the isochron as depicted.

Kirchhoff migrated result. The interference of the different curvelets (of the same scale) is good, and compares favorably to the Kirchhoff result, although the comparison seems better for the shallower dipping part of the isochron. At the steeper parts of the isochron, the destructive interference between different curvelets away from the isochron, is somewhat less effective and leaves the tails of the curvelets somewhat visible.

In order to see the differences between the results for the steeper dipping curvelets and the shallower ones, Figure 10 shows the comparison between the Kirchhoff result and our TRD transformation for both a shallow and a steep dipping curvelet. Here we did zero-offset migration, to avoid the curvature of the isochron being different at different locations on the isochron; i.e., a semi-circle has constant curvature everywhere. Figure 10 shows that the steeper dipping curvelet is significantly 'bent' towards the isochron, while the shallower dipping curvelet is hardly bent at all. As a result, the TRD transform does better for the shallower dipping curvelets than for the steeper dipping ones, as it does not include any such bending.

It remains to be seen how severe the error is if we ignore this bending, and use our TRD transform on a synthetic data-set with many curvelets. In this case, there would be interaction between curvelets from several different scales, i.e., with different frequency content. We have not performed such a test yet, since this requires the proper frequency weighting for curvelets of different scales. As we mentioned earlier, this frequency weighting should follow from a stationary phase evaluation of equation (12) with a curvelet as scattered data u_S. At the moment of writing this report we have not done such a calculation. In case future tests using a combination of the amplitude correction, obtained in this manner, and the TRD transform would show that an omitted bending correction in the TRD transform causes large differences with the Kirchhoff result, we could introduce an extra bending correction to bend the curvelets towards the isochron as depicted in Figure 11.

Discussion

In this paper, the TRD transform is calculated using a brute-force approach in the spatial domain. For the significant curvelet coefficients, we apply an inverse curvelet transform, and transform the resulting curvelet in the spatial domain according to our TRD transform. This approach allows us to show the proof of concept of imaging with curvelets using our TRD transform, but does not provide an efficient way of such imaging with curvelets. Ultimately one would want to calculate the TRD transform in the curvelet frame, although an approach that makes use of the finite support of the curvelets in the spectral domain seems also worth investigating.

Even though the curvelets used in this paper are 2D, they can be extended to higher dimensions. Figure 12 shows an example of a 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[†]. Therefore, replacing the 2D map migration equations with their 3D equivalents, the TRD transform can be extended to 3D, hence allowing 3D CO pre-stack time-imaging with curvelets. Decomposing a 3D CO volume of data into 3D curvelets, the same procedure as outlined in this paper can be used to image 3D seismic data with 3D equivalents of curvelets. Of course, the proper 3D amplitude correction should be used in this case.

Even though in this paper we show results from migration only, modeling or demigration with curvelets works in the same way. In this case the TRD transform is defined according to the map demigration equations (Douma & de Hoop, 2005).

Conclusion

We have presented first examples of the use of curvelets in CO pre-stack time migration. A simple numerical example confirmed that curvelets can be used as building blocks for seismic data that allow for a sparse representation of such data. Since curvelets are roughly like pieces of bandlimited plane waves, the wave-character of the seismic data, i.e., that it is bandlimited and that the recorded wavefronts have directions associated with them, can be built into the representation of the data. Therefore, in essence, curvelets are an appropriate reparameterization of seismic data, that allow the wave-character of the data to be built into the representation of the data. Knowing that the local directions (or slopes) in the data (for a fixed scattering angle and azimuth, or offset and azimuth in homogeneous media) can be mapped one-to-one from the data to the image

[†]This rough description of course 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.

[b]

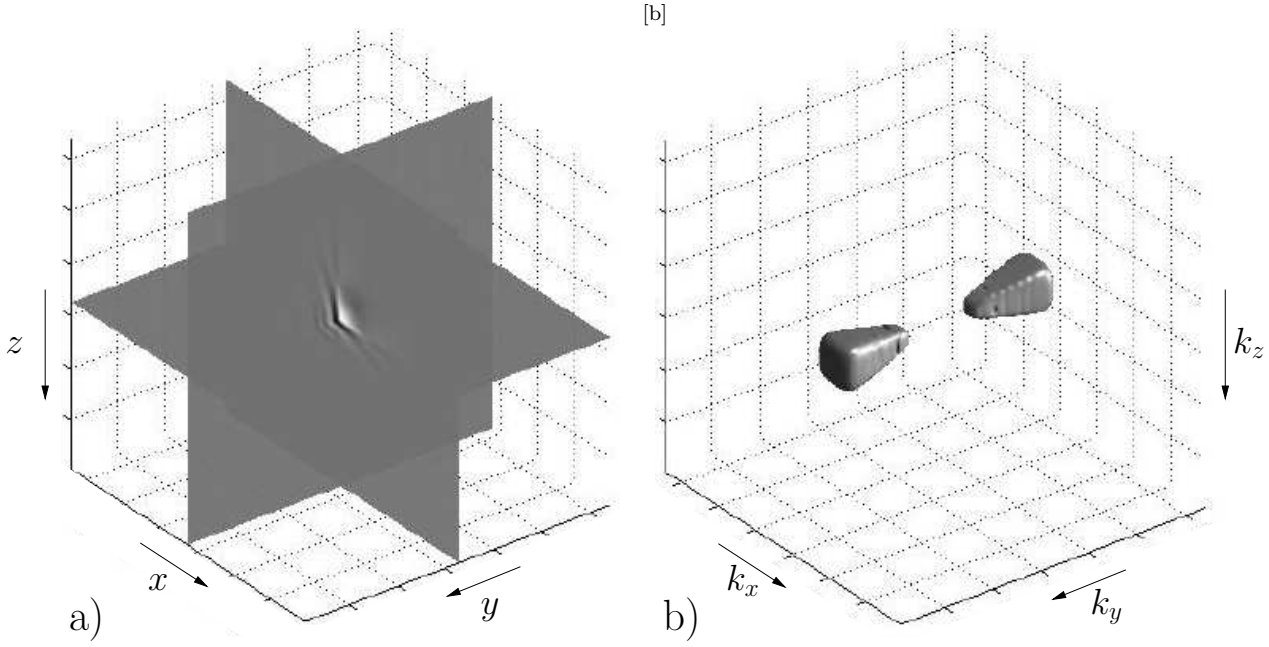

Figure 12. A 3D curvelet in space (a) and its associated amplitude spectrum (b). In the spatial domain, 3D equivalents of curvelets look like circular disks that are smooth along the disk and oscillatory orthogonal to the disk.

domain using map migration, we have studied the use of map migration to perform CO pre-stack time migration with curvelets.

The numerical tests of 2D CO pre-stack time imaging with curvelets we presented, showed that a simple translation, rotation, and dilation approach (which we refer to as the TRD transformation) of curvelets, largely determined by the 2D CO pre-stack map migration equations, provides a reasonably accurate approximation to the Kirchhoff migrated result, at least on a per scale basis, i.e., for curvelets with similar frequency content. Our results for CO pre-stack time migration with curvelets in 2D, through our TRD transformation, can be extended to 3D using 3D equivalents of curvelets.

Acknowledgments

The authors thank Roel Snieder for his critical review of the paper, and Emmanuel Candès for providing us with his Matlab implementation of the digital curvelet transform via unequally spaced fast Fourier transforms. This work was partly supported by Total E & P and the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena.

REFERENCES

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.

Bleistein, N. 1984. *Mathematical methods for wave phenomena*. Academic Press, Inc.

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

Candès, E. J., & Demanet, L. 2004. The curvelet representation of wave propagators is optimally sparse. *submitted, and can be downloaded from Candès website.*

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. 1992. *Earth soundings analysis: Processing versus inversion*. Blackwell Scientific Publications, Inc.
Córdoba, A., & Fefferman, C. 1978. Wave packets and Fourier Integral Operators. *Comm. PDE's*, **3**, 979–1005.
Douma, H., & de Hoop, M. V. 2005. Explicit expressions for pre-stack map time-migration in isotropic and VTI media and the applicability of map depth-migration in heterogeneous anisotropic media. Accepted for publication in *Geophysics*.
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*. Soc. Expl. Geophys.
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.
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*. Soc. Expl. Geophys.
Herrmann, Felix J. 2003b. *Multifractional splines: application to seismic imaging*. SPIE. 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.
Kleyn, A.H. 1977. On the migration of reflection time contour maps. *Geophysical Prospecting*, **25**, 125–140.
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.
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.
Stein, E. M. 1993. *Harmonic Analysis: Real-variable methods, orthogonality, and oscillatory integrals*. Second edn. Princeton University Press.
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.
Zavalishin, B.R. 1981. Perfection of methods for constructing seismic images using controlled directional reception. *Soviet Geology and Geophysics*, **22**, 98–104.

APPENDIX A: CURVELETS

In this appendix, we explain curvelets and their construction. This treatment closely follows the construction of real-valued curvelets from Candès and Donoho (2004b), except from 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 'splitting at every other scale' (that will become clear later).

Before treating the construction of curvelets, we mention that we want to be able to reconstruct a signal in a function space $\mathcal{X}$ with zero error; i.e., we want the 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 and Weiss (1996, pp.334-336) and appendix B in this paper] that if $\mathcal{X}$ is a Hilbert space, this is equivalent to requiring

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

to hold. 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 unit energy'. Condition (A2) is the definition of a (normalized) tight frame (see appendix B for a thorough explanation of tight frames). Hence, for the curvelets to satisfy the reconstruction formula (A1), we want them to constitute a partition of unit energy.

Curvelets can be thought of as 2D (anisotropic) extensions to wavelets. Just like wavelets are 'localized' in one variable *and* its Fourier dual, curvelets are 'localized' in two variables *and* their two Fourier duals. Such localization is of course understood within the limits imposed by the Heisenberg uncertainty principle. Choosing the variables to be x and z, such localization is obtained through 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) 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 the curvelets to constitute a tight frame (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 unit energy. The window $\chi_{j,l}(\boldsymbol{k})$ is constructed through multiplication of an angular window $\nu(\theta)$ and a radial window

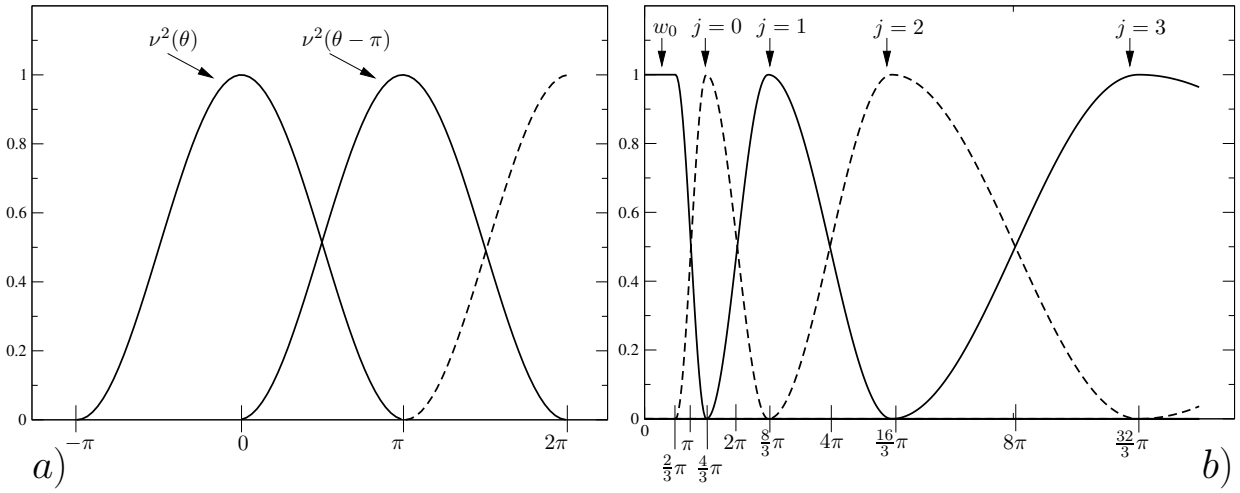

Figure A1. Example of a possible angular window $\nu(\theta)$ (a), and Lemarié-Meyer windows $|w(2^{-j}|\boldsymbol{k}|)|^2$ for the radial windows (b). The Lemarié-Meyer windows for j and $j+1$ are equal at $2^{j+1}\pi$.

$w(|\boldsymbol{k}|)$, 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 $[-\pi,\pi)$. We choose our function 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. 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

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

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

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

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, i.e., that the energy is conserved. There are many windows (with different decay properties) that satisfy this admissibility condition. Hence, curvelets can be constructed using different angular windows, much like in the construction of wavelets.

The angular windowing provides the curvelets with geometry. However, to make the curvelets multiscale, we need to subdivide the frequencies in subbands. This is done using a radial window w. A possible choice for such a window is the Lemarié-Meyer window [see (Hernández & Weiss, 1996, p.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' (i.e. a partition of unit energy, such that the sum of squares equals one), as

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

with w_0 a 'coarse scale' C^∞ window that equals one 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

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

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

that

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

where we used equations (A5) and (A6). Here, the notation $\lfloor j \rfloor$ denotes the integer part of j. From equation (A8), it is clear that $\chi_{j,l}(\boldsymbol{k})$ [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 would form such a partition of unity, just like for the angular windows; i.e., many different variants of curvelets can be constructed by using different radial window functions.

From the definition of $\chi_{j,l}(\boldsymbol{k})$, it is clear that $\chi_{j,l}(\boldsymbol{k})$ is obtained through a rotation of $\chi_{j,0}(\boldsymbol{k})$. With this observation, it becomes clear that 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 only increases every other scale (see Figure 1a). This 'splitting at every other scale' is a consequence of the choice of dyadic subbands; with the subband filtering being dyadic, the only way to have an integer number of angular windows per subband, is to increase 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 much 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 $O\left(2^j \times 2^{j/2}\right)$. Therefore, in the spatial domain, the curvelet is supported 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

$$\text{width} \approx \text{length}^2 \,, \qquad \text{(A8)}$$

where the approximation is used to indicate the omitted constants (and where we ignore the differences in dimensions between width and length squared). This scaling is referred to as the *parabolic scaling*. Note that the scaling relation (A8) holds in the spatial domain, while in the frequency domain we would have length $\approx$ width^2.

From the previous it follows that the volume of the angular wedge in the spectral domain is $O\left(2^j \times 2^{j/2}\right)$, while the volume of the rectangle in the spatial domain is $O\left(2^{-j/2} \times 2^{-j}\right)$. Therefore, curvelets are really Heisenberg tiles in phase-space, since their net volume in phase-space is

$$O\left(2^j \times 2^{j/2}\right) \times O\left(2^{-j/2} \times 2^{-j}\right) = O(1) \,,$$

which is in accordance with the Heisenberg uncertainty principle.

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 $t_j + L_j$ and l_j determine, respectively, the sampling intervals (or translation steps) in x and z direction in the spatial domain. It is known that $e^{iq\pi x/L}/\sqrt{2L}$, with $q \in \mathbb{Z}$, is an orthonormal basis for L^2 functions with period $2L$; this just follows from a Fourier series expansion for 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 and Weiss (1996, Chapter 1.4)]. Therefore, the sequence $(u_{j,m})_{m\in\mathbb{Z}^2}$ [with $m = (m_1, m_2)$] defined as

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

is an orthonormal basis for $L^2\left(\pm I_{1j} \times I_{2j}\right)$. 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

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

For normalization purposes, the term $2^{-(\lfloor j/2\rfloor + j)/2}$ is rewritten as $2^{-3j/4}$, i.e., $\lfloor j/2 \rfloor$ is replaced with $j/2$, giv-

ing

$$u_{j,m}(k_x,k_z) = \frac{2^{-3j/4}}{2\pi\sqrt{\delta_1\delta_2}} e^{i\pi(m_1+1/2)2^{-j}k_x/\delta_1} e^{im_2k_z2^{-\lfloor j/2\rfloor}/\delta_2},$$

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

$$\hat{c}_\mu(\boldsymbol{k}) = 2\pi\chi_{j,l}(\boldsymbol{k})u_{j,m}\left(R^*_{\theta_{j,l}}\boldsymbol{k}\right), \quad \mu=(j,l,m), \tag{A9}$$

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 (or better wavelets) are isotropic.

Using the definition of curvelets given in equation (A9), it can be shown that curvelets form a (normalized) tight frame. From equation (A9) it is clear that 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 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

$$\begin{aligned}
&\textstyle\sum_{j\geq0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}\sum_{m_1,m_2}\left|\left(\hat{f},\hat{c}_{(j,l,m_1,m_2)}\right)\right|^2 = \\
&\quad(2\pi)^2\textstyle\sum_{j\geq0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}\int\left|\hat{f}(\boldsymbol{k})\right|^2|\chi_{j,l}(\boldsymbol{k})|^2\,d\boldsymbol{k} \\
&=(2\pi)^2\int\left|\hat{f}(\boldsymbol{k})\right|^2\left[\textstyle\sum_{j\geq0}\sum_{l=0}^{2^{\lfloor j/2\rfloor}-1}|\chi_{j,l}(\boldsymbol{k})|^2\right]d\boldsymbol{k} \\
&\quad=(2\pi)^2\int\left|\hat{f}(\boldsymbol{k})\right|^2d\boldsymbol{k}=(2\pi)^2\left\|\hat{f}\right\|_2^2.
\end{aligned}$$

This tells us that $(\hat{c}_\mu)_{\mu\in M}$ (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 $f\in L^2(\mathbb{R}^2)$

$$\sum_{\mu\in M}|(f,c_\mu)|^2 = \|f\|_2^2\ . \tag{A10}$$

This shows that $(c_\mu)_{\mu\in M}$ is a normalized tight frame for $L^2(\mathbb{R}^2)$, giving us 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**. E.g., 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 $\sum_{i=1}^4|(f,\phi_i)|^2=2\|f\|^2$. Here the frame bounds measure the 'redundancy' of the frame in $\mathbb{R}^2$, i.e., four vectors in two dimensions have a redundancy of two.

It is known that the linear span of frame elements is dense in $\mathcal{H}$ [e.g., Hernández and Weiss (1996, p.399)]. This means, that any element $f\in\mathcal{H}$ can essentially 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 $\mathcal{H}$. It is clear that 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 (B1). Of course, adding any vector to a basis, would destroy the linear independence of the basis, meaning it no longer would be a basis. When both frame bounds equal one ($A=B=1$), the frame is called a **normalized** tight frame. Each orthonormal basis is obviously a normalized tight frame, but the converse is not generally true, since the frame elements need not be linearly independent.

Even though the frame elements do not need to be normal, it follows that 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 $\|e_k\|\leq1$. Hence, for a normalized tight frame, the linear independence of a basis is traded for the condition $\|e_k\|\leq1$, $\forall k\in J$.

If a normalized tight frame has $\|e_k\|=1$, $\forall k\in J$, it follows that $\{e_j\}_{j\in J}$ is an orthonormal basis for $\mathcal{H}$. To see this, observe that for fixed $k\in J$, $1=\|e_k\|^4=|(e_k,e_k)|^2\leq\sum_{j\in J}|(e_k,e_j)|^2\leq1$. 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 and Weiss (1996, pp.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

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

to hold[‡]; in other words, statements (B2) and (B3) are equivalent. This tells us that for a tight frame with frame bound A, we have reconstruction formula (B3), for if we have $\sum_{j\in J} |(f, e_j)|^2 = A\,\|f\|_{\mathcal{H}}^2$, simply defining $e'_j = e_j/\sqrt{A}$, gives $\sum_{j\in J} \left|\left(f, e'_j\right)\right|^2 = \|f\|_{\mathcal{H}}^2$. From the equivalence of statements (B2) and (B3), we then have $f = \sum_{j\in J} \left(f, e'_j\right) e'_j = \dfrac{1}{A}\sum_{j\in J} (f, e_j)\, e_j$; i.e., for a tight frame we have reconstruction formula (B3). For a normalized tight frame ($A = 1$), we simply have $f = \sum_{j\in J} (f, e_j)\, e_j$, and thus also the reconstruction formula (B3). Note that this reconstruction formula is identical to the reconstruction formula for an orthonormal basis, but that 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.

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

[‡]Here the equality means that the sum on the right-hand side of equation (B3) converges to an element $f \in \mathcal{H}$.