Compressed wavefield extrapolation

Tim T.Y. Lin[*] **and Felix J. Herrmann**[*]

[*]*Seismic Laboratory for Imaging and Modeling,*

Department of Earth and Ocean Sciences,

University of British Columbia,

6339 Stores Road,

Vancouver, V6T 1Z4, BC, Canada

(March 18, 2007)

Running head: ***Compressed extrapolation with curvelets***

ABSTRACT

An *explicit* algorithm for the extrapolation of one-way wavefields is proposed which combines recent developments in information theory and theoretical signal processing with the physics of wave propagation. Because of excessive memory requirements, explicit formulations for wave propagation have proven to be a challenge in 3-D. By using ideas from "compressed sensing", we are able to formulate the (inverse) wavefield extrapolation problem on small subsets of the data volume ,thereby reducing the size of the operators. Compressed sensing entails a new paradigm for signal recovery that provides conditions under which signals can be recovered from incomplete samplings by *nonlinear* recovery methods that promote sparsity of the to-be-recovered signal. According to this theory, signals can successfully be recovered when the measurement basis is *incoherent* with the representation in which the wavefield is sparse. In this new approach, the eigenfunctions of the Helmholtz operator are recognized as a basis that is incoherent with curvelets that are known to compress seismic

1

wavefields. By casting the wavefield extrapolation problem in this framework, wavefields can successfully be extrapolated in the modal domain, despite evanescent wave modes. The degree to which the wavefield can be recovered depends on the number of missing (evanescent) wave modes and on the complexity of the wavefield. A proof of principle for the "compressed sensing" method is given for wavefield extrapolation in 2-D, together with a pathway to 3-D during which the multiscale and multiangular properties of curvelets in relation to the Helmholz operator are exploited. The results show that our method is stable, has reduced dip limitations and handles evanescent waves in inverse extrapolation.

INTRODUCTION

Our motivation

The main result from the relatively new field of "compressed sensing" (Candès et al., 2006b; Donoho, 2006; Tsaig and Donoho, 2006) states that an arbitrary k non-zero sparse spike train of length $N \gg k$ can exactly be recovered from m incoherent measurements with $m \sim k$ ($\sim$ means proportional to within $\log N$ constants). The term *incoherent* here refers to a quality between two bases. Qualitatively speaking, a basis is incoherent with respect to another basis if a sparse signal in one of the bases generally does not have a sparse representation in the other basis (see appendix B for a formal definition of this quality). A classical example of two bases that are incoherent with each other is the identity basis and the Fourier basis.

This result means that the unknown spike train can, for instance, exactly be recovered from m random Fourier measurements. These measurements correspond to taking inner products between random rows, selected from the Fourier matrix, and the vector containing the sparse spike train. This measurement could explicitly be stated as

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \tag{1}$$

with $\mathbf{A} := \mathbf{RMS}^T \in \mathbb{C}^{m \times N}$ the synthesis matrix, $\mathbf{M}$ the measurement matrix defined in terms of the Fourier matrix ($\mathbf{M} := \mathcal{F}$ with $\mathcal{F}$ denoting the discrete Fourier transform matrix), and $\mathbf{R}$ a restriction matrix randomly selecting m rows from $\mathbf{M}$. Here , $\mathbf{S}$ and $\mathbf{S}^T$ are the sparsity analysis and synthesis matrices for a domain that compresses the signal. The restriction matrix is defined such that the columns of $\mathbf{A}$ are 2-norm normalized to unity. The symbol := is used to denote definition. Simply speaking, Eq. 1 corresponds to

randomly selecting m Fourier coefficients from the Fourier transform of $\mathbf{x}_0$.

Spike trains are sparse in the Dirac/identity basis so we set $\mathbf{S}^T := \mathbf{I}$. The rows of $\mathbf{I}$ are incoherent with the rows of the Fourier measurement matrix that consists of complex exponentials. Because $\mathbf{x}_0$ has few non zeros, it is sparse and this sparse vector can exactly be recovered by solving the following nonlinear optimization problem

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{i=1}^{N} |x_i| \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

with the symbol $\widetilde{\ }$ hereby reserved for quantities obtained by solving an optimization problem. The $\arg\min_{\mathbf{x}}$ stands for the argument of the minimum, i.e., the value of the given argument for which the value of the expression attains its minimum value. This recovery is successful when the measurement and sparsity representations are incoherent and when m is large enough compared to the number of non-zero entries in $\mathbf{x}_0$. Since $m \ll N$ this recovery involves the inversion of an underdetermined system. As long as the vector $\mathbf{x}_0$ is sparse enough, recovery according to Eq. 2 is successful. Typically for Fourier measurements 5 coefficients per non-zero entry are sufficient for full recovery (Candès, 2007).

Instead of asking ourselves the question of how to recover $\mathbf{x}_0$ from incomplete data suppose now that we ask ourselves how to apply an integer shift by τ to an arbitrary but sparse vector $\mathbf{x}_0$, without having to shift each single entry. We all know that shifts translate to phase rotations in the Fourier domain and that the Fourier basis functions (rows of the Fourier matrix, $\mathcal{F}$) are incoherent with the Dirac basis, $\mathbf{I}$. More formally, consider the approximate shift operation, defined in terms of the exponentiation of the discrete difference matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$. In that case, the shift by τ can be written as

$$\mathbf{u} = e^{-\mathbf{D}\tau}\mathbf{v} = \mathbf{L}e^{j\mathbf{\Omega}\tau}\mathbf{L}^H\mathbf{v}, \tag{3}$$

where the decomposition matrix $\mathbf{L}^H$, with the symbol ^{H} denoting the Hermitian transpose,

is derived from the eigenvalue problem

$$\mathbf{D} = \mathbf{L}\boldsymbol{\Omega}\mathbf{L}^{H}. \tag{4}$$

In this expression, $\boldsymbol{\Omega}$ is a diagonal matrix with the eigenvalues $\omega = \mathrm{diag}(\boldsymbol{\Omega})$ on its diagonal. These eigenvalues correspond to the angular frequencies, while the orthonormal (de-)composition matrices $\mathbf{L}^{H}$, $\mathbf{L}$ correspond, when applying Neumann boundary conditions, to the forward and inverse discrete cosine transforms, respectively. The accuracy of this discrete approximation of the shift operator depends on the type and order of the finite-difference approximation in $\mathbf{D}$. Because the eigenvectors of the above shift operation correspond to the rows of Fourier-like (discrete cosine) measurement matrix of the previously posed recovery problem, we can define an alternative "compressed" procedure for applying the shift by solving the following nonlinear optimization program

$$\begin{cases} \mathbf{y}' = \mathbf{R}e^{j\boldsymbol{\Omega}\tau}\boldsymbol{\mathcal{F}}\mathbf{v} = \mathbf{R}\mathbf{M}'\mathbf{v} \\[2ex] \widetilde{\mathbf{u}} = \arg\min_{\mathbf{u}} \|\mathbf{u}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{u} = \mathbf{y}' \end{cases} \tag{5}$$

in which we took the liberty to overload the symbol $\boldsymbol{\mathcal{F}}$ with the discrete cosine transform. The input for this nonlinear program is given by the phase-rotated Fourier transform of $\mathbf{v}$, restricted to a (small) random set of m frequencies. The symbol $'$ is hereby reserved for phase rotated quantities. The shifted spike train is obtained by nonlinear recovery of the phase-rotated measurement vector $\mathbf{y}$. Instead of applying a full matrix-vector multiplication involving all temporal frequency components as in Eq. 3, the shift according to the above program involves the repeated evaluation of the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ and its transpose. In the extreme case of a vector with a single non-zero entry for $\mathbf{v}$, the matrix $\mathbf{A}$ will usually only need to be of size $5 \times N$, leading to a significant reduction for the size of the matrix. An example of the above procedure is included in Fig. 1, where 5 spikes with random

positions and amplitudes in a vector of length $N = 200$ are circular shifted by 20 samples. Comparison of the results of applying the *full* shift operator (cf. Eq. 3) and the compressed shift operator according to Eq. 5 shows that these results are identical. Only 15 random Fourier measurements were necessary for the recovery of the shifted spike train. Instead of applying a *full* 200×200 operator, application of the *compressed* operator of size 15×200 is sufficient. These results were calculated with the ℓ_1-solver of Basis Pursuit (Chen et al., 2001).

The idea of norm-one sparsity based recovery is not exactly new to the seismic imaging community. For instance, there exists a large body of literature on sparsity-promoting penalty functions. Since the seminal work of Claerbout and Muir (1973), norm-one regularized inversion problems have been prevalent in the formulation of geophysical inverse problems with applications ranging from deconvolution (Taylor et al., 1979; Oldenburg et al., 1981; Ulrych and Walker, 1982; Santosa and Symes, 1986; Levy et al., 1988; Sacchi et al., 1994) to filtering and seismic data regularization based on high-resolution Fourier (Sacchi and Ulrych, 1996; Zwartjes and Gisolf, 2006) and curvelet transforms (see e.g. Hennenfent and Herrmann, 2006a) and non-parametric seismic data recovery (F. J. Herrmann and G. Hennenfent, personal communication, 2007) , adaptive subtraction for multiple attenuation (Guitton and Verschuur, 2004; Herrmann et al., 2007) and Bayesian approaches with *priors* consisting of long-tailed Cauchy distributions (Sacchi and Ulrych, 1996).

What is new in compressed sensing is the insight into the criteria of successful recovery. For example, compressed sensing looks for the existence of a transform that compresses the (inverse) extrapolated wavefield and which is incoherent with the measurement basis. In that case, the wavefield can be recovered from a relative small subset of measurements. We leverage these new insights towards the formulation of the (inverse) wavefield extrapolation

6

problem by identifying the eigenfunctions of the modal transform (Grimbergen et al., 1998) as the measuring basis and curvelet frames (Candès and Donoho, 2004; Candès et al., 2006a) as the sparsity representation. The modal transform is an orthonormal transform defined by the eigenfunctions of the Helmholtz operator. Curvelets are known to be sparse on seismic wavefields and to preserve sparsity* after (inverse) extrapolation. Because curvelets are localized, they are incoherent with the radiating eigenfunctions. This property opens the way towards the formulation of a compressed version of the inverse wavefield extrapolation problem that only requires a small subsampling of the wavefield in the modal domain. In this way, not only could the size of the explicit extrapolation matrices (explicit matrices refer to actual matrices as opposed to implicit matrix-free extrapolation matrices where the matrix is not formed explicitly) be reduced but also the "missing" evanescent wave modes could correctly be handled in inverse extrapolation. However, this latter problem proves to be more difficult and involves the solution of an ill-posed inverse problem where the to-be-inverted matrix contains columns that are exponentially decaying, i.e., the matrix no longer adheres to the conditions imposed by compressed sensing.

Our work is in spirit of recent work by Mulder and Plessix (2004), who describe how seismic images can be obtained by using only a subset of the temporal frequencies. We extend their approach by exploiting the nonlinear recovery of compressed sensing in the context of (inverse) wavefield extrapolation. In this way, we are working towards the formulation of an explicit extrapolation scheme while meeting the computational challenges of

---

*Some care should be taken with this statement. For very smooth (constant) media curvelets are more or less invariant under wave propagation a property exploited by Douma and de Hoop (2006) and Chauris (2006). For more realistic velocity models curvelets diffuse and loose their strict sparsity. Empirical studies on seismic data have, however, shown that seismic data remains compressible in the curvelet domain (Candès et al., 2006a; Hennenfent and Herrmann, 2006b).

such schemes in 3-D.

Our approach

This paper introduces an explicit algorithm for the compressed evaluation of one-way inverse wavefield extrapolation. The main challenges in (inverse) extrapolation are

- the *size* of the eigenvectors in the modal domain (Grimbergen et al., 1998) that diagonalize the one-way wavefield extrapolation operators;

- the presence of *evanescent* wave modes that hamper a proper refocusing of the propagated wavefield which leads to poor resolution of steeply dipping events.

We provide a solution strategy that addresses both issues. Our method involves the rapid evaluation of monochromatic extrapolation operators that are solutions of the one-way wave equation in d dimensions (Claerbout, 1971; Berkhout, 1982; Hale et al., 1992; Grimbergen et al., 1998)

$$\partial_3 W^{\pm}(x_\nu, x_3; x'_\nu, x'_3) = \pm j\mathcal{H}_1 W^{\pm}(x_\nu, x_3; x'_\nu, x'_3) \tag{6}$$

with $j = \sqrt{-1}$, $\mathcal{H}_1$ the square-root of the Helmholz operator

$$\mathcal{H}_2 = \mathcal{H}_1 \mathcal{H}_1 \tag{7}$$

(see appendix A for detail), x_3 the vertical coordinate, x_ν with $\nu = 1 \cdots d-1$ the horizontal coordinate(s) and the initial condition

$$W^{\pm}(x_\nu, x_3 = x'_3; x'_\nu, x'_3) = \delta(x_\nu - x'_\nu). \tag{8}$$

In this expression, W^+ represents the downward wavefield extrapolation operator that carries the wavefield from depth level x_3 and lateral position $x_\nu \in \mathbb{R}^{d-1}$ to $x_3 > x'_3$ and

$x'_\nu \in \mathbb{R}^{d-1}$. Similarly, W^- represents the upward extrapolation operator carrying an upgoing one-way wavefield from depth level x_3 to $x_3 < x'_3$. The above expressions are monochromatic in the angular frequency ω. For notational simplicity, we drop the ω-dependence in the expressions. The examples in this paper are for two-dimensional ($d = 2$) wavefields but the theory also readily extends to $d = 3$.

After discretization along the space and time axes, the multi-frequency downward wavefield extrapolation of a downgoing time-domain wavefield vector $\mathbf{p}^+|_{x_3} \in \mathbb{R}^M$ at depthlevel x_3 can be written in the form of a matrix-vector product (see Appendix A for further details).

$$\mathbf{u} = e^{j\mathbf{H}_1 \Delta x_3} \mathbf{v} = \mathbf{W}\mathbf{v} \tag{9}$$

with $\mathbf{u} = \mathbf{p}^+|_{x'_3}$, $\mathbf{v} = \mathbf{p}^+|_{x_3}$ for $x_3 > x'_3$. The wavefield vectors contain the reordered entries of discretized wavefield . The propagation over the interval $\Delta x_3 = x_3 - x'_3$ is determined by the square-root of the block-diagonal multi-frequency $M \times M$ Helmholtz matrix $\mathbf{H}_2 = \mathbf{H}_1 \mathbf{H}_1$ with $M = n_\nu \times n_f$ the size of the discretized wavefield, n_ν the number of total spatial samples and n_f the number of angular frequencies. Similar expressions hold for upward extrapolation of an upgoing wavefield and for the remainder of the paper we will drop the symbols $\pm$ in cases where equivalent expressions hold for down- and upgoing wavefields. Moreover, we use lowercase symbols for continuous quantities, calligraphic symbols for operators on continuous functions and bold lower- and uppercase symbols for discretized matrix operators and vectors, respectively.

The evaluation of the matrix $\mathbf{W} \in \mathbb{C}^{M \times M}$ is expensive since it involves the solution of a sparse eigenvalue problem that diagonalizes the discretized Helmholtz matrix $\mathbf{H}_2$. The solution of this eigenproblem leads to the following factorization of the extrapolation operator

(see Appendix A for a more detailed treatment)

$$\mathbf{W} \;=\; \boldsymbol{\mathcal{F}}^H \mathbf{L} e^{j\boldsymbol{\Lambda}^{1/2}\Delta x_3} \mathbf{L}^H \boldsymbol{\mathcal{F}} \tag{10}$$

$$\;=\; \mathbf{M}^T \mathbf{M}' \tag{11}$$

with $\mathbf{L}^H$ the orthonormal modal decomposition matrix. The modal transform matrix $\mathbf{L}$ contains the 2-norm normalized eigenfunctions that solve

$$\mathbf{H}_2 = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}^H \tag{12}$$

with $\boldsymbol{\Lambda}$ a diagonal matrix with the eigenvalues on the diagonal. From this point on the system of eigenvectors and eigenvalues pertains to the Helmholtz equation. Commensurate with the language of compressed sensing, we will call the matrix $\mathbf{M} := \mathbf{L}^H \boldsymbol{\mathcal{F}}$ the "measurement" matrix and $\mathbf{M}' := e^{j\boldsymbol{\Lambda}^{1/2}\Delta x_3}\mathbf{M}$ the phase-rotated measurement matrix. Notice that the memory imprint is different for the Fourier and modal transforms. The Fourier transform is calculated *implicitly* with algorithmns which are trivial to store compared to the modal transform which consists of a *full* matrix-vector operator which must be explicitly constructed and kept in memory .

In practice, solving the above eigenvalue problem is computationally prohibitive. The common approach to limit the computational complexity of one-way wavefield extrapolation operators is to split the square-root operator (Collino and Joly, 1995). This type of approximation has led to the so-called paraxial, phase-screen (de Hoop et al., 2000) and split-step extrapolation operators (Stoffa et al., 1990). These approximations have the disadvantage that they are only valid up to certain dips and that they impose smoothness on the background velocity model defining the Helmholtz equation. See Bednar and Bednar (2006) and Mulder and Plessix (2004) for a comparison between one- and two-way 'wave equation' migration.

10

Aside from the above approximation errors, taking the square-root of $\mathbf{\Lambda}$ in Eq. 10 introduces additional difficulties related to the negative eigenvalues that may lead to an erroneous treatment of the evanescent wavefield associated with these negative eigenvalues. To negotiate problems related to the evanescent wavefield and possible aperture limitation, Berkhout (1982) proposes to define the inverse extrapolation operators by regularized least-squares inversion. The result of this inversion depends on a regularization parameter, which stabilizes the inversion at the expense of smoothing. This smoothing effect leads to dip limitations as illustrated in Fig. 2. The refocused wavefields according to matched filtering and regularized least-squares inversion, only recover the initial wavefield up to limited dips. These dip limitations are clearly visible in the frequency-wavenumber spectrum and eventually contribute to the dip limitation of migrated images. After 200 iterations of LSQR (Paige and Saunders, 1982), the spectrum is only partly filled in as can be seen from Fig. 2 (h). In this paper, a new approach is presented which aims to negate these smoothing effects while it also attempts to come up with a computationally feasible *explicit* method for wavefield extrapolation. A first result of this method is presented in Fig. 3, where the wavefield is completely refocused. We used the method of Basis Pursuit (BP) (Chen et al., 2001) to recover the broad-band initial wavefield. This recovery is successful because the initial wavefield is excited by point sources, thus it is strictly sparse. In practice, however, wavefields are not sparse in the spike basis, an observation that will be addressed in this paper.

Outline

First, we present one-way forward wavefield extrapolation in the modal domain based on the eigenfunction decomposition of the Helmholtz operator. Next, the compressed formula-

tion for the forward wavefield extrapolation is introduced exploiting the mutual coherence between the eigenfunctions and curvelets. Inverse wavefield extrapolation, including a discussion on the behavior of evanescent waves is discussed next, followed by its compressed formulation that accounts for the evanescent waves. Different restriction strategies that lead to a size reduction (compression) of the extrapolation operators are reviewed. The algorithm is tested on examples that range in complexity from a medium with a single velocity low or high to a physically hard velocity profile from an overthrust velocity model. The paper is concluded by a complexity analysis aimed to address size issues of wavefield extrapolation in 3-D. We show that a multiscale and angular formulation of the one-way wavefield extrapolation problem may lead to a feasible parallel algorithm in 3-D. Details on the modal transform, including the definition of the Helmholtz operator, its spectrum and forward and inverse wavefield extrapolation in the modal domain, are included in Appendix A: One-way wave-propagation in the modal domain while Appendix B: Stable recovery by curvelet frames details the ℓ_1-norm recovery with curvelets.

COMPRESSED WAVEFIELD EXTRAPOLATION IN THE MODAL DOMAIN

As the examples in the previous section show, the recovery of seismic wavefields is possible from measurements with large percentages of the data missing. This observation allows us to address the "incomplete measurement" problem associated with compressed wavefield extrapolation and the problem of the "missing" evanescent wavemodes in inverse extrapolation. By invoking the appropriate restrictions to the measurement matrices, $\mathbf{M}$ and $\mathbf{M}'$ (cf. Eq.10), the propagation operators can be reduced, leading to considerable savings on the computation time and memory use. First, the forward wavefield extrapolation oper-

ator is reviewed and the compressed forward wavefield extrapolation problem is discussed in terms of nonlinear recovery followed by a discussion on different restriction strategies. Next, the compressed inverse extrapolation problem is formulated. Since inverse extrapolation involves mediation of the evanescent waves, we will also focus on that aspect of wavefield extrapolation.

Forward wavefield extrapolation

In Appendix A it is shown that one-way wavefield extrapolation involves taking the square-root of the discretized Helmholtz equation. Using the orthonormality of the modal (de)composition matrices, this square-root can be written as

$$\mathbf{H}_1 = \mathbf{L}\mathbf{\Lambda}^{1/2}\mathbf{L}^H. \tag{13}$$

The square-root in this expression is chosen such that the imaginary part of the square-root is always negative avoiding exponential blow up for the propagation operator defined earlier. We choose the signs of the diagonal entries of $\mathbf{\Lambda^{1/2}}$ as follows

$$\Re(\lambda_i^{1/2}) \geq 0 \quad \text{for} \quad \lambda \geq 0 \quad i = 1 \cdots M \tag{14}$$

and

$$\Im(\lambda_i^{1/2}) < 0 \quad \text{for} \quad \lambda < 0 \quad i = 1 \cdots M. \tag{15}$$

With this choice, the forward extrapolation matrix defined in Eq. 10 is stable. The negative λ lead to an exponential decay of their corresponding propagated wavemodes. These attenuated wavemodes are called evanescent wavemodes.. The forward extrapolated wavefield can be written as

$$\mathbf{u} = \mathbf{W}\mathbf{v} = \mathbf{M}^H\mathbf{M}'\mathbf{v}. \tag{16}$$

Because of the evanescent wavemodes, $\mathbf{M}'$ is no longer unitary and

$$\mathbf{W}^H \mathbf{W} \neq \mathbf{I}. \tag{17}$$

Compressed forward extrapolation

Following the discussion in the introduction and in the Appendix: Stable recovery with curvelet frames, forward wavefield extrapolation can be cast into the following nonlinear optimization problem

$$\mathbf{W_1} : \begin{cases} \mathbf{y}' = \mathbf{R}\mathbf{M}'\mathbf{v} \\ \\ \mathbf{A} := \mathbf{R}\mathbf{M}\mathbf{C}^T \\ \\ \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{y}' \\ \\ \widetilde{\mathbf{u}} = \mathbf{C}^T\widetilde{\mathbf{x}}, \end{cases} \tag{18}$$

for an appropriately chosen restriction matrix and for the measurement matrix $\mathbf{M}$, defined as in Eq. (10), and with its rotation define by

$$\mathbf{M}' := e^{j\mathbf{\Lambda}^{1/2}\Delta x_3}\mathbf{M}. \tag{19}$$

In Eq. 18, $\mathbf{C}^T$ refers to curvelet synthesis by the fast inverse curvelet transform (Candès et al., 2006a; Ying et al., 2005; Hennenfent and Herrmann, 2006b, and Appendix B). During the compressed extrapolation, the wavefield is extrapolated with a compressed forward extrapolation operator $\widehat{\mathbf{W}'} = \mathbf{R}\mathbf{M}' \in \mathbb{C}^{m \times M}$. The forward extrapolated wavefield is subsequently recovered by the nonlinear inversion, promoting sparsity in the curvelet domain. This *compressed* formulation, which we write in short hand, as

$$\widetilde{\mathbf{u}} = \mathbf{W}_1[\mathbf{v}], \tag{20}$$

is designed to yield the same results as for the *full* forward extrapolation, $\mathbf{u} = \mathbf{W}\mathbf{v}$.

Inverse wavefield extrapolation

A key component of migration is the forward and inverse extrapolation of (one-way) wave-fields. The purpose of forward extrapolation is to propagate a source wavefield into the subsurface. Inverse extrapolation aims to revert this process by refocusing the wavefield. Ideally, inverse extrapolation maps a forward propagated impulsive wavefield back to a point. In the absence of evanescent waves, the inverse extrapolation operators are simply given by the adjoint of the forward extrapolation operator, i.e.

$$\mathbf{F} = \mathbf{W}^H. \tag{21}$$

In this case, we have

$$\mathbf{FW} = \mathbf{I} \tag{22}$$

by virtue of the Fourier and modal transform matrices being orthonormal. However, as we have seen in Fig. 2, the presence of the evanescent field causes problems. In that case, the inverse is no longer equal to the Hermitian transpose of the forward operator, i.e. $\mathbf{F} = \mathbf{W}^{-1} \neq \mathbf{W}^T$. The common practice of ignoring the evanescent field in the definition of the inverse extrapolation operator leads to a blurring as discussed earlier.

To show how this inverse extrapolation method may adversely affect results, we create a forward extrapolated wavefield according to

$$\mathbf{u} = \mathbf{Wv} \tag{23}$$

with an initial wavefield $\mathbf{v}$ given by the reflectivity of an overthrust model[†]. The forward propagated wavefield is computed with Eq. 23, where the reflectivity of the overthrust

[†]This model was developed by Sam Gray and is representative for the Canadian overthrust front. Velocity model courtesy BP.

model acts as an 'exploding' reflector. The propagation operator is defined for a medium that varies in the lateral direction only. The lateral variations are taken from the 'physically hard' overthrust model at depth $x_3 = 4\,\mathrm{km}$ and the wavefield is forward propagated over a distance of $600\,\mathrm{m}$. Next, the forward extrapolated wavefield $\mathbf{u}$ is inverse extrapolated by ignoring the evanescent waves with the matched-filter approach. The inverse extrapolated wavefields are explicitly given by

$$\widetilde{\mathbf{v}}_m = \mathbf{F}\mathbf{u} \quad \text{with} \quad \mathbf{F} := \mathbf{W}^H \tag{24}$$

or by   As the results in Figs. 2 and 4 indicate, this approach to the inverse extrapolation problem leads to an unsatisfactory refocusing of the original wavefield given by the 'exploding' overthrust model. Evanescent wavemodes, the regularization and the finite aperture all lead to undesirable artifacts.

Compressed inverse extrapolation

In summary, inverse extrapolation differs from forward extrapolation in the way the evanescent wavefield is handled. The choice for the square-root in Eq. 15 guarantees negative imaginary values that correspond to decaying exponentials for the forward extrapolation. Defining the inverse extrapolation as the inverse of the forward extrapolation leads to an exponential growth of the evanescent waves and people either ignore these evanescent modes (cf. Eq. (24)) or compute the inverse extrapolation operator as the regularized pseudo inverse of the forward extrapolation operator (cf. Eq. (24)).

We propose a method which addresses two main challenges of refocusing seismic wavefields namely,

- how to recover the initial wavefield from a random sampling of the modal plane

<div align="center">16</div>

amongst the propagating wavemodes $(\lambda \geq 0)$?

- how to mediate the dip-limitation due to decaying evanescent wave modes $(\lambda < 0)$?

These are distinct questions. The first question concerns the solution of a recovery problem, involving phase-rotated measurement matrices that are "*unitary*" before restriction (i.e., the exponent of the phase rotations is purely imaginary). In that case, the measurement basis adheres to the conditions that allow for recovery with compressed sensing as long as the restriction is carried out intelligently. The second question involves the solution of an ill-posed inverse problem, where the dip limitation attributed to the exponential decaying wavemodes is mediated. Both problems are solved using the same nonlinear optimization program albeit with a much slower convergence for the recovery of the evanescent wavefield.

According to our method the inverse extrapolated wavefield is calculated by solving the following nonlinear program

$$\mathbf{F_1}: \quad \begin{cases} \mathbf{y} = \mathbf{RMu} \\[2ex] \mathbf{A}' := \mathbf{RM}'\mathbf{C}^T \\[2ex] \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{A}'\mathbf{x} = \mathbf{y} \\[2ex] \widetilde{\mathbf{v}} = \mathbf{C}^T\widetilde{\mathbf{x}}. \end{cases} \tag{25}$$

In this formulation, the inverse extrapolated wavefield is calculated by first "measuring" the wavefield $\mathbf{u} = \mathbf{Wv}$, followed by inverting the phase rotated synthesis operator $\mathbf{A}'$. This definition of the synthesis operator is different from the definition for the forward problem because this specific choice allows us to curvelet-sparsity regularize the forward extrapolation operator that includes exponentially decaying evanescent waves. The inversion is stabilized by the ℓ_1-norm and involves the inversion of a propagation operator with decay-

ing exponentials for the evanescent wavemodes. Again we adapt the following short-hand notation for the nonlinear extrapolation procedure outlined in Eq. 25

$$\widetilde{\mathbf{v}} = \mathbf{F}_1[\mathbf{u}]. \tag{26}$$

Before comparing results of these nonlinear formulations for the forward and inverse extrapolation operators, we first discuss different restriction strategies to reduce the size of the operators.

Restriction strategies

The compressed formulation for the inverse wavefield extrapolation problem is (i) stable with respect to the "missing" evanescent wavemodes and (ii) allows for a reduction of the computational cost to synthesize the propagation matrices as well as of their storage requirements. While devising a strategy for the restriction (sampling of the modal plane), the following issues need to be taken into consideration for the evaluation of the synthesis matrix $\mathbf{A}'$

- the cost of solving the eigenvalue decomposition for each temporal frequency. For a wavefield of size M, i.e. $\mathbf{u} \in \mathbb{R}^M$ with $n_f = \mathcal{O}(M^{1/d})$ angular frequencies and $n_\nu = \mathcal{O}(M^{(d-1)/d})$ spatial samples, the computation of the eigenfunction decompositions for the sparse Helmholtz matrix requires $\mathcal{O}(M^{(2d-1)/d})$ operations for the *full* problem and $\mathcal{O}(m_\nu^2 \times m_f)$ for the *compressed* problem. The reduced number of spatial samples equals $m_\nu = p_\nu \cdot \mathcal{O}(M^{(d-1)/d})$ and the reduced number of frequencies $m_f = p_f \cdot n_f$ with p_ν, $p_f < 1$ the fractions of the size of the restriction over the total size of the wavefield's discretization;

- the cost of applying (repeated) matrix-vector products, which is $\mathcal{O}(M^{2(d-1)/d})$ for the *full* problem and $p_\nu \cdot p_f \cdot \mathcal{O}(M^{2(d-1)/d})$ for the *compressed* problem;

- the memory use for each frequency amounting to storing a $M^{(d-1)/d} \times M^{(d-1)/d}$ matrix for the *full* problem and a $m_\nu \times m$ with $m = m_\nu \times m_f$ for the *compressed problem*;

- an additional $\mathcal{O}(N \log N)$ for the computation of the curvelet transform;

For $d = 3$, the size of the wavefield vector grows cubicly ($M = n_f \times n_1 \times n_2$), which illustrates the formidable challenge we are faced with when designing *explicit* (inverse) wavefield extrapolation operators in higher dimensions. By formulating the application of the extrapolation as a "compressed sensing" problem, we can devise different strategies to limit the size of the wavefield extrapolation problem. The choice for the restriction depends on following main factors, namely

- the properties of the eigenvectors. Mutual coherency (see Appendix B for the definition of mutual coherency) between curvelets and 'Fourier' or 'non-Fourier like' eigenfunctions may differ. Since curvelets are strictly localized in the Fourier domain, eigenmodes that spread in the Fourier domain lead to a more favorable compressive sampling;

- the energy distribution of the wavefield in the frequency-modal domain. Recovery from a compressively sampled wavefield is more favorable for successful recovery if the random samples are collected from parts of the spectrum where the energy of the to-be-extrapolated wavefield resides.

- the presence of evanescent wavemodes. These modes decay exponentially and can be omitted when the purpose is to refocus the wavefield up to the dip of the evanescent

19

wave modes;

- the memory imprint. This imprint is directly related to the number of eigenfunctions that need to be stored to calculate the explicit matrix-vector multiplications. This storage is not an issue for the temporal frequency that is matrix free.

- the uniformity of the restriction. The more uniformly distributed the random restriction in the band in the frequency-modal domain where the energy is concentrated, the easier the recovery (D. L. Donoho et al., personal communication, 2006). For a fixed measurement size, this translates into a more accurate recovery or a faster convergence for the recovery. Alternatively, a more uniform (in both directions) restriction allows for a recovery from a smaller number of measurements.

Depending on the requirements (memory imprint *versus* number of flops), the restriction can be designed to

- select a subset of temporal frequencies that leads to a direct reduction of the number of "block diagonal" eigenproblems that need be solved and to a reduction of the number of eigenvectors that need to be stored;

- select a subset of eigenvalues In principle, discrete eigenfunctions can be calculated at a random subset of discrete eigenvalues by applying the appropriate shifts towards the eigenvalues. This random selection also leads to a reduction of the number of eigenfunctions that need to be stored in memory.

- a combination of temporal-frequency and eigenvalue restrictions;

Examples of the different restrictions are included in Fig. 5.

EXAMPLES

Extrapolation in Gaussian medium profiles

In this section, we apply our proposed method to two different lateral varying medium profiles, consisting of a single Gaussian shaped low- or high-velocity zone as shown in Fig 6. The background velocity is $2000\,\text{ms}^{-1}$. The minimum and maximum velocities are set to $1200\,\text{ms}^{-1}$ and $3500\,\text{ms}^{-1}$, respectively.

The Helmholtz operator is discretized with $\Delta x_1 = 4\,\text{m}$, $\Delta t = 0.004\,\text{s}$ with $n_t = n_\nu = 256$. During the experiments, a chain of horizontally oriented fine-scale curvelets is extrapolated (see Fig. 7(a)). In this experiment, the propagation direction is perpendicular to the direction in which the medium varies.

The recovery of evanescent wave modes in inverse extrapolation is studied by submitting a fan of fine-scale curvelets oriented with different angles (see Fig. 7(b)) to forward extrapolation in a homogenous medium, followed by non-linear inverse extrapolation.

The spectrum: Before examining the extrapolation results, let us first briefly study the behavior of the spectrum of the discretized Helmholtz matrix and its eigenvectors.

As described by Grimbergen et al. (1998), the eigenvalue spectrum for a medium with a low-velocity zone contains "continuous" and "discrete" parts, while the spectrum for the profile with the high-velocity zone is strictly continuous. The eigenfunctions corresponding to the continuous part of the spectrum are radiating. For the low-velocity profile, the eigenfunctions for the discrete part of the spectrum correspond to the guided wavemodes. These guided waves are evanescent outside the waveguide. For the high-velocity profile the situation is reversed. In that case, all wavemodes are radiating, but in the region of

the high-velocity perturbation they become evanescent. As Fig. 8 illustrates, there is little difference in the 'frequency spread' between radiating and non-radiation wave modes. This observation suggests that the restriction should not be preferential to either of these wavemodes based on the coherency argument.

Forward extrapolation: To demonstrate the performance of the compressed extrapolation algorithm, a chain of horizontally-oriented curvelets are vertically propagated over a distance of $\Delta x_3 = 600\,\mathrm{m}$ in a laterally varying medium. The results for the forward extrapolation are plotted in Figs 9 and 10, for the profiles depicted in Fig. 6(a) and 6(b). Both figures include the results with the *full* modal-domain forward extrapolation operator, $\mathbf{u} = \mathbf{W}\mathbf{v}$, and with the nonlinear compressed operator, $\widetilde{\mathbf{u}} = \mathbf{W}_1[\mathbf{v}]$ (cf. Eq. (18)). The initial wavefield $\mathbf{v}$ is plotted in Fig. 7(a). The discretized wavefields consist of $M = 2^{16}$ samples. The compressed extrapolation results are obtained with the cooling method outlined in Table 1 for restrictions with increasing numbers of uniform measurements. The examples show that a fraction $p = 0.16$ is sufficient to recover the forward propagated wavefields for the Gaussian low and high (with p denoting the total fractional reduction of the measurement matrix, calculated as the product of p_f and p_ν). The compressed formulation leads to a gain of in memory usage and more importantly a gain in the cost for computing the operators. For a given number of measurements, the compressed extrapolation is more accurate for the medium profile with the velocity high, which can be understood because the complexity of the extrapolated wavefield is simpler since there are no guided modes in that case.

Inverse extrapolation: Again horizontally-aligned curvelets (Fig. 7(a)) are used that are first forward extrapolated according $\mathbf{u} = \mathbf{W}\mathbf{v}$. This extrapolated wavefield is subsequently

inverse extrapolated according $\widetilde{\mathbf{v}} = \mathbf{F_1 u}$ (cf. (25)). The results of this procedure for the two medium profiles of Fig. 6 are included in Fig. 11. As the number of measurements increases, the wavefield is better recovered. For $p = 0.16$, the recovery is satisfactory. Again the results for the medium with the velocity high are slightly better, which can be explained by the absence of guided modes for this profile. Compared to compressed forward extrapolation, the inverse extrapolation requires less measurements leading to an improved compression of the operator. This behavior can be understood because the initial wavefield is sparser than the propagated wavefield and hence requires less measurements.

Inverse extrapolation of evanescent waves: Results for the more challenging inverse wavefield extrapolation for steep and hence evanescent wave modes are included in Fig. 12 for a constant velocity medium set to $2000\,\mathrm{ms}^{-1}$ with $\Delta x_1 = 40\,\mathrm{m}$ and a propagation distance of $\Delta x_3 = 50\,\mathrm{m}$. Comparisons are made between the original wavefield and the results after refocusing with the matched filter $\widetilde{\mathbf{v}} = \mathbf{F}_m \mathbf{u}$ and inverse extrapolation $\widetilde{\mathbf{v}} = \mathbf{F_1}[\mathbf{u}]$ according Eq. (25). As shown in Fig. 12(c) the evanescent curvelets are fully recovered.

Inverse extrapolation in the overthrust model

In this section, the proposed extrapolation methodology is applied to a model representative for the Canadian foothills. The lateral velocity profile used for the computation of the extrapolation operators is plotted in Fig. 13. To maintain the validity of the velocity profile while keeping our examples to a reproducible size, we chose $\Delta x_1 = 40\,\mathrm{m}$, with all other parameters the same as the previous examples in a Gaussian medium. This experiment involves compressed inverse extrapolation which aims to refocus a wavefield that propagated over a distance of $\Delta x_3 = 400\,\mathrm{m}$.

Aside from the mutual coherence, compressive sampling that emphasizes the regions of the frequency-modal domain where the to-be-recovered wavefield resides will lead to a more favorable recovery. Fig. 14 shows that most of the information of the wavefields resides at the large eigenvalues that correspond to the guided wavemodes. The spectra in Fig. 14 also confirm the loss of steep dipping events due to the evanescent waves.

The results for the compressed recovery are summarized in Fig. 15. First, the exploding reflector given by the reflectivity of the overthrust model is forward extrapolated. There are artifacts due to wrap around in the time axis. The results after applying the 'full' matched filter and ignoring the evanescent wave contribution are given in Fig. 4. Notice the slight reduction in bandwidth and dominant artifacts. The result obtained by computing the pseudo inverse, $\widetilde{\mathbf{v}}_{LS} = \mathbf{W}^\dagger \mathbf{u}$, leads to an improved result. However, some of the artifacts remain while steep events seem to suffer some deterioration which can be explained from the missing evanescent waves. The restricted matched filter does not recover the initial wavefield as we can see from Fig. 15(b). The compressed recovery result for $p_f = 0.35$ and $p_\nu = 0.7$, on the other hand recaptures most of the initial wavefield as can be observed Fig. 15(c).

COMPLEXITY ANALYSIS

The examples of the previous section gave us a proof of concept that compressed extrapolation is feasible. The question is whether *compressed* extrapolation yields benefits compared to *full* extrapolation in the modal domain. Since modal-domain extrapolation is considered expensive, we include in our discussion the computational costs associated with other "wave-equation" extrapolation schemes. We will also discuss additional measures that can be taken to gain computational efficiency by exploring the multiscale and angle decompo-

sition of the curvelet transform.

Computational cost of wavefield extrapolation

Wave-equation migration codes spend most of their time in (forward/inverse) wavefield extrapolation. There are two main categories of extrapolation operators. There are the *implicit* methods such as time-domain reversed-time migration (Bednar and Bednar, 2006) or spectral methods (Riyanti et al., 2006), both depending on right-hand sides, and *explicit* methods such as one-way extrapolation in the modal domain (Grimbergen et al., 1998) and frequency-domain direct solvers (Plessix and Mulder, 2004). Compared to the *implicit* methods, which require a recalculation of the operator for each "shot", explicit methods have the advantage that the forward and inverse extrapolation operators can be applied to each shot without the need of resynthesizing the operator. For the two-way frequency-domain methods, this step entails a LU factorization, whereas for the one-way wave extrapolation one needs to solve an eigenproblem (Grimbergen et al., 1998). When there are many shots, these explicit methods lead to a method that is an order of magnitude faster than the implicit methods. In addition, spectral methods have the advantage that only one (monochromatic) forward and one inverse extrapolated wavefield needs to be stored as opposed to time-domain techniques that require storage over time intervals (W. Symes, personal communication, 2006).

Following the analysis by Riyanti et al. (2006), the time-domain complexity is $n_s n_t \mathcal{O}(n^d)$ with n_s the number of shots, n_t the number of time steps and $M = n^d$ the size of a snapshot of the discretized wavefield for each shot. Since we are interested in obtaining the correct amplitudes, we have to take into account the cost of computing the regularized pseudo

inverse of inverse wavefield extrapolation. This brings the total computational complexity to $n_{LS}n_sn_t\mathcal{O}(n^d)$ with $n_{LS} = \mathcal{O}(1)$, the number of iterations for the LSQR solver (Paige and Saunders, 1982). This computational complexity is approximately the same for the explicit spectral method, amounting to a total complexity of $n_{LS}\mathcal{O}(n^{d+2})$, when assuming $n_s = n_t = n_f = \mathcal{O}(n)$ (Riyanti et al., 2006). The *compressed* modal transform requires $p \cdot n_{sol} \cdot \mathcal{O}(n^{2(d-1)+1})$ with $p = p_\nu \cdot p_f$ the fraction of the restriction and n_{sol} the computational cost of the ℓ_1 solver. These estimates show that as long as n_{sol} is of the same order as n_{LS}, we gain a factor of p. As the inverse extrapolation examples with the strictly sparse curvelets have shown, compressed operators can lead to a small p and hence to a substantial reduction of the computational cost for wavefields with low complexity.

Unfortunately, the memory requirements for 3-D more or less preclude explicit methods. The size of the full monochromatic modal decomposition matrix equals $n^{2(d-1)}$, which is clearly prohibitive and has been the major bottleneck for the construction of explicit methods for complex wavefields in 3-D. In the next section, we propose a new method that exploits multiscale and angular properties of the curvelet transform that allow us to further compress explicit propagation operators.

Multiscale and angular compressed extrapolation

Following the work by Tsaig and Donoho (2006), the unconstrained optimization problem for the recovery in $\mathbf{F_1}$ (cf. Eq. (25)) can be replaced by a sequence of smaller subproblems,

i.e.

$$
\mathbf{F_1^j} : \begin{cases}
\mathbf{y_j} = \mathbf{R_j M_j u} \\[2ex]
\mathbf{A_j'} := \mathbf{R_j M_j' C_j^T} \\[2ex]
\widetilde{\mathbf{x}}_\mathbf{j} = \arg\min_{\mathbf{x_j}} \|\mathbf{x_j}\|_1 \quad \text{s.t.} \quad \mathbf{A_j' x_j} = \mathbf{y_j} \\[2ex]
\widetilde{\mathbf{v}} = \sum_\mathbf{j} \mathbf{C_j^T} \widetilde{\mathbf{x}}_\mathbf{j},
\end{cases}
\tag{27}
$$

where $\mathbf{j} = \{j, l\}$ is the index set for the scale j and the angle l (remark that there are two angles for curvelets in $d = 3$). The $\mathbf{j}$ subscripted symbols refer to quantities at a particular scale and angle. For instance, $\mathbf{C_j^T}$ refers to the inverse curvelet transform of the coefficients at scale 2^j and angle $\theta_\mathbf{j} = 2\pi l 2^{\lfloor j/2 \rfloor}$. Since this inverse only reconstructs contributions to the total wavefield at this scale and angle, the modal (de)composition matrices can be constructed accordingly. The algorithm consists of the following steps

1. Discretize the Helmholtz operator with scale-dependent sample intervals $\Delta x_1 \sim 2^{-j}$ and $\Delta t \sim 2^{-j}$. This requires a definition for the "measurement" matrices $\mathbf{M_j}$ and $\mathbf{M_j'}$ that includes a coarsening (smoothing to scale 2^{-j} and downsampling to 2^j samples) operator $\mathbf{B_j}$. So explicitly, $\mathbf{M_j} = \mathbf{L_j^H \mathcal{F}_j B_j}$ and $\mathbf{M_j'} = e^{j\mathbf{\Lambda_j}^{1/2}\Delta x_3} \mathbf{M_j}$ where $\mathbf{L_j}$ is the modal decomposition and phase-rotation matrices adapted to the scale-angle pair $\mathbf{j}$;

2. Define a restriction operator $\mathbf{R_j}$ such that the angular frequencies and eigenfunctions are selected pertinent to the wedge defined by $\mathbf{j}$. Use the size of the wedge and the empirical histogram of the curvelet coefficients of the to-be-propagated wavefield to select the appropriate number of measurements $m_\mathbf{j}$ at each scale and angle.

The first step is an automatic consequence of the lack of high temporal and spatial frequencies at the coarse scales of the wavefield $\mathbf{u_j} = \mathbf{B_j u}$. Clearly, the adaptation in discretization leads to a dramatic reduction of the computational cost and size of the (de)composition

matrices at the coarser scales. The second step depends on the fact that we can limit our-selves to eigenfunctions that correspond to angles and temporal frequencies that pertain to a specific $\mathbf{j}$. Since the square-root of the eigenvalues $\lambda^{1/2} = \mathrm{diag}(\mathbf{\Lambda}^{1/2})$ can be associated with the vertical wave numbers ξ_3, we can establish a relationship between the angle θ and the eigenvalues,

$$\lambda^{1/2} = \xi_0 \cos\theta \quad \text{and hence} \quad \theta = \arccos\left(\xi_0^{-1}\lambda^{1/2}\right) \quad \lambda \geq 0. \tag{28}$$

In this expression, the negative angles were omitted for convenience. Curvelets at $\mathbf{j}$ are localized near the symmetric wedge

$$W_{j,l} = \{\pm\boldsymbol{\xi}, 2^j \leq |\boldsymbol{\xi}| \leq 2^{j+1}, |\theta - \theta_{\mathbf{j}}| \leq \pi.2^{-\lfloor j/2 \rfloor}\},$$

with $|\boldsymbol{\xi}| = \omega c_0^{-1}$. Combining these two equations leads to a multiscale and angular restriction, limiting the *angular* frequencies and *angles* to

$$2^j c_0|\boldsymbol{\xi}_0| \leq \omega \leq 2^{j+1} c_0|\boldsymbol{\xi}_0| \quad \text{and} \quad \lambda^{1/2} \leq \xi_0 \cos(\theta) \quad \text{with} \quad |\theta - \theta_{\mathbf{j}}| \leq \pi.2^{-\lfloor j/2 \rfloor}. \tag{29}$$

The 'random' restriction is done such that the regions of the modal plane are adaptively sampled according to where most of the information of the to-be-extrapolated wavefield re-sides. To accomplish this adaptive sampling, the forward curvelet transform of the wavefield $\mathbf{u}$ is calculated from which the fractions

$$p_{\mathbf{j}} := \sum_{\mu \in \mathcal{M}_{\mathbf{j}}} \mathrm{hist}\{\mathbf{c}_{\mu \in \mathcal{M}_{\mathbf{j}}}\} \tag{30}$$

are obtained, with $\mathbf{c} = \mathbf{C}\mathbf{u}$ the vector with curvelet coefficients, and $\mathrm{hist}\{\mathbf{c}_{\mathbf{j}}\}$ the normalized (by the total number of curvelet coefficients) histogram of the curvelet coefficients in the $\mathbf{j}^{\text{th}}$ wedge. The sum runs over the index set $\mathcal{M}_{\mathbf{j}}$ for the curvelets in each wedge. According to this definition, $\sum_{\mathbf{j}} p_{\mathbf{j}} = 1$ and the number of measurements for each subproblem can

28

be defined as $m_{\mathbf{j}} = p_{\mathbf{j}} \cdot m$ with m the total number of measurements. With this sampling adaptation, the number of measurements taken in each wedge is proportional with the number of significant coefficients in each wedge.

The above procedure sketches a truly multiscale and angular compressed extrapolation method that (i) significantly reduces the size of the compressed explicit operators; (ii) is adapted to where the information of the to-be-extrapolated wavefield resides; (iii) is amenable for parallel implementation. During the parallel implementation, the wavefield in each wedge can be extrapolated independently while the orthonormality of the Fourier and modal transforms also allows for straightforward parallelization.

The proposed multiscale method goes at the expense of a moderate overhead of no more than a factor of two, since $\log M_\mu$ eigenproblems are solved at scales 2^{-j} for $j = 0 \cdots \log M_\mu$. It is beyond the scope of this paper to prove whether the solution of $\mathbf{F_1}$ is equivalent to the solution of $\mathbf{F_1^j}$.

DISCUSSION

Initial findings

It was shown that compressed extrapolation can lead to a substantial reduction for the size of the compressive one-way extrapolation operators with extrapolation results that are equivalent to results obtained with "full" extrapolation in the modal domain. The performance of compressive extrapolation depends on three factors, namely the mutual coherence between curvelets and the monochromatic eigenfunctions; the shape of the frequency-modal amplitude spectrum and the complexity of the wavefield. Because curvelets are strictly localized in the Fourier domain, the mutual coherence decreases for eigenmodes that are

more spread in the Fourier domain. Comparison between the spectra of radiating and guided modes showed no distinctive difference in the amount of spreading. The amplitude spectrum in the Fourier-modal domain, however, showed large amplitudes for guided wave-modes with the largest eigenvalues and with temporal frequencies localized in the band of the extrapolated wavefield. The complexity of the forward extrapolated wavefield also has a noticeable effect. For complicated (physically "hard") velocity models, the forward extrapolated wavefield loses some of its sparsity in the curvelet domain and this means that more entries of the unknown curvelet coefficient vector need to be recovered. The larger the number of significant entries in the curvelet vector, the more compressive sampling of the wavefield is necessary in the modal domain.

For inverse extrapolation, our findings indicate a successful recovery from compressively sampled wavefields, where the evanescent wavefield is ignored and where the sampling is restricted to a small subset of the modal domain. For compressible initial wavefields, this approach can lead to a significant reduction for the size of the inverse extrapolation operator, which leads to a leaner explicit scheme. For example, inverse extrapolation towards a strictly sparse initial wavefield consisting of spikes only, leads to a dramatic reduction for the size of the extrapolation operators. The examples also showed that recovery is possible for more complicated wavefields that compress in the curvelet domain. For moderate propagation distances, our formalism also allowed for the recovery of the evanescent wavefield increasing the resolution of steep dips.

Of course the proposed method for the compressed evaluation of operators stands or fails with the compressibility of the signal after applying the extrapolation operators. As long as the initial wavefield is compressible, we are all set for inverse extrapolation. As we have seen, this is not necessary the case for forward extrapolation. Fortunately, this is not

really an issue for seismic applications where inversely extrapolated data are correlated with forward extrapolated sources during imaging. As long as the image is sparse, the inverse extrapolated wavefield will be sparse.

The computational cost of explicit (inverse) wavefield extrapolation consists of two parts namely the solution of the eigenproblem and the application of the matrix products of the explicit and full modal transform matrices. By reducing the number of rows both these costs are reduced. The cost for the solution of the eigenproblem scales quadratically with the size of the discretization so any reduction of the number will translate in reduced computation time. The restriction also reduces the size of the explicit matrices which is a prerequisite for a large-scale deployment of this explicit algorithm.

The potential benefits of our algorithm depend on the ability to (approximately) solve very large-scale ℓ_1 problems. As long as the cost of these solvers in conjunction with the gain in solving the eigenproblem are on the order of the cost of computing the least-squares inversion for the full problem our approach will be feasible. Additional benefits are the speedup due to the explicit nature of the operators and the reduction in memory imprint.

Multiscale and angular extensions

As with any three-dimensional explicit formulation of the extrapolation problem, the size of the matrices becomes prohibitively large. By casting the problem in terms of a multiscale and angular "compressed sensing" problem, it is possible to significantly reduce the size of the matrices involved. This size reduction can be accomplished by (i) adapting the spatio-temporal discretization of the Helmholz operator; (ii) adapting the restriction to temporal frequencies and eigenvalues pertinent to the curvelet partitioning (per curvelet wedge) and

finally (iii) by adapting the size of the restriction to the information content of the to-be-propagated wavefield. In particular, the reduced size for the coarse-scale discretization leads to a drastic reduction of the computational cost involved with the solution of the Helmholtz eigenproblem.

Future research

This paper is based on a new idea where the eigenfunctions that diagonalize a certain operator are used as a "measurement" basis in the formulation of a compressed sensing problem. This insight allows for a significant reduction for the cost of composing and applying large-scale explicit operators. The following issues need to be resolved before applying the presented methodology to large-scale migration:

Fast calculation of curvelet transforms: The recovery from compressed extrapolation requires multiple evaluations of the Fast Discrete Curvelet Transform. This transform is $\mathcal{O}(M \log M)$ with a constant that can be reduced by implementation in programmable gate arrays (He et al., 2004);

Fast ℓ_1 solvers: The success of all compressed sensing techniques depend on the ability to quickly solve ℓ_1 optimization problems. Currently, there is a surge in research activity addressing this important issue (see e.g. D. L. Donoho et al., personal communication, 2006 and Tibshirani (1996); Candès and Romberg (2005); Donoho et al. (2005); Figueiredo et al. (2007); Kol

Restricted modal-domain imaging conditions: In this paper, we focused on the extrapolation of wavefields. During migration, imaging conditions (and their adjoints) are used. For compressive imaging to be successful, restricted modal domain imaging conditions will have to be derived;

Smart sampling: The success of compressive sampling depends for a large part on sampling strategies. In our case, this means that an adaptive sampling strategy needs to be devised.

Fast computation of compressive sampled eigenvalues: The algorithm derives its reduction for the size of the operators from selecting a subset of eigenproblems. This means that fast eigenproblem solvers need to be developed that can efficiently compute random subsets of eigenvectors;

Construction of multiscale Helmholtz operators: The multiscale and angular extension opens the way to compute the eigenfunctions of the Helmholtz operator by smoothing the velocity model to a scale commensurate to the scale of the wavefield;

Theoretical compression criteria: The theory of compressed sensing provides results for measurement and sparsity matrices that are orthonormal. Theoretical performance estimates for extensions towards redundant frames (curvelets) is still an open problem. Since the idea of applying compressed sensing techniques to the evaluation of operators is new, theory is still mostly lacking on this exciting topic.

Other compressive approaches

Compression of imaging and hence extrapolation operators has been the "holy grail" of seismic imaging. So far, most of the attempts have been directed towards the diagonalization of the operators. Indeed recently reported results by Candès and Demanet (2005); Douma and de Hoop (2006); Chauris (2006) on curvelets and by other researchers on other generic expansions have shown promise. However, the difficulty with these approaches is that it is difficult to diagonalize extrapolation operators for "non-smooth" velocity models. Our

approach differs from these attempts in the sense that we do not try to replace the imaging operators by a transformed-domain diagonally dominant matrix. Instead, we use a velocity model-adaptive transform, the modal transform, which yields eigenfunctions that are incoherent with curvelets. So, we do not use the curvelets to diagonalize. Instead, we use this signal representation to compress the to-be-extrapolated wavefields. For complicated velocity models, our approach also loses its efficiency since the wavefield will lose some of its sparsity properties. However, since seismic imaging entails the collapse of seismic energy onto reflectors, we will continue to benefit from compression as long as the image permits curvelets as a sparse representation.

CONCLUSIONS

In this paper, we proposed a compressed (inverse) extrapolation algorithm that combines the diagonalization of the Helmholtz operator by the modal transform with recent insights from the field of compressed sensing. In our approach, the eigenfunctions spanning the modal transform are recognized as a suitable measurement basis that is incoherent with the curvelet frames that are known to compress seismic wavefields. This observation allows for a compression of operators by casting the action of the operator into a compressed sampling and subsequent recovery problem. The data is compressively sampled in the modal domain, where the operator is diagonalized, and subsequently recovered with the nonlinear sparsity promoting techniques from compressed sensing. This procedure leads to a reduction of the size of the operators with an overall computational gain that is contingent upon the cost for the recovery.

ACKNOWLEDGMENTS

APPENDIX A

ONE-WAY PROPAGATION IN THE MODAL DOMAIN

The one-way wave-extrapolation algorithm presented in this paper derives from the work of Grimbergen et al. (1998). In this appendix, we provide the basic equations for the Helmholtz operator, its discretization and diagonalization, leading to the definition of the one-way extrapolation operators that serve as the point of departure for our compressed extrapolation algorithm. For further details on the derivation of the one-wave field extrapolation operators in the modal domain, the reader is referred to Grimbergen et al. (1998); Dessing (1997).

Consider a medium with a constant velocity in the vertical direction over the interval $(x_3, x_3']$ and variations in the lateral direction(s).

Introduce a kernel associated with the Helmholtz operator

$$\mathcal{H}_2(x_\nu, x_3; x_\nu') = \left(\frac{\omega}{c(x_\nu, x_3)^2}\right)^2 \delta(x_\nu - x_\nu') + \partial_\nu \partial_\nu \delta(x_\nu - x_\nu'). \tag{A-1}$$

After discretization over the continuous variables (x_1, x_1') for $d = 2^{\ddagger}$, we have at angular frequency $\omega = n\Delta\omega, \quad n = 1 \cdots n_f$

$$\mathbf{H}_2 = \mathbf{C} + \mathbf{D}_2 \tag{A-2}$$

with

$$\mathbf{C} = \begin{bmatrix} \left(\frac{\omega}{c_1}\right)^2 & 0 & \cdots & 0 \\ 0 & \left(\frac{\omega}{c_2}\right)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{\omega}{c_{n_1}}\right)^2 \end{bmatrix}, \tag{A-3}$$

where $c_i = c(i\Delta x_1, x_3)$, $i = 1 \cdots n_1$ represent the discretized velocity in the x_1 direction with n_1 the number of discretization points. The matrix $\mathbf{D}_2$ is a band Toeplitz matrix with Neumann boundary conditions, representing the second-order derivative $\mathbf{D}_2 = \Delta x_1^{-2} \mathrm{Toep}\{\mathbf{d}_2^l\}$ with the row vector $\mathbf{d}_2^l$ the l-point stencil approximation of the second derivative operator and Δx_1 the spatial sample interval in the x_1 direction. In this paper, a 11-point stencil is used ($l = 11$), yielding 21 non-zero entries per row.

Spectrum of the discretized Helmholtz operator

Even though the setting of this paper is discrete, our insight in the behavior will benefit if we study the behavior of the spectrum of the continuous Helmholtz operator

$$\mathcal{H}_2 = \left(\frac{\omega}{c(x_\nu, x_3)}\right)^2 + \nabla_\nu^2 = \xi^2(x_\nu, x_3) + \partial_\nu \partial_\nu, \qquad \nu = 1 \cdots d - 1 \tag{A-4}$$

---

‡The discretization for $d = 3$ runs along similar lines and has been omitted.

with $\partial_\nu \partial_\nu$ the Laplacian in the horizontal coordinate(s). Notice that we use the convention of Einstein's subscript notation. The eigenfunctions of $\mathcal{H}_2$ for each temporal frequency satisfy

$$\mathcal{H}_2 \phi(x_\nu) = \lambda \phi(x_\nu) \tag{A-5}$$

with $\phi(x_\nu)$ the eigenfunctions and λ the eigenvalues. The eigenvalues of the self-adjoint operator $\mathcal{H}_2$ are real. For a medium that only varies laterally over a finite range and has constant velocity c_0 outside, the spectrum σ of $\mathcal{H}_2$ consists of a *continuous* and *discrete* part:

$$\sigma(\mathcal{H}_2) = \sigma_0(\mathcal{H}_2) \cup \sigma_1(\mathcal{H}_2), \tag{A-6}$$

where σ_0 refers to the continuous part of the spectrum and σ_1 to the discrete part. The transition from the continuous to the discrete part occurs at the wavenumber $\xi_0 = \omega/c_0$. The eigenfunctions of the continuous part to the left of ξ_0 are global and are called radiating while the eigenfunctions to the left of ξ_0 pertaining to the discrete spectrum are localized and correspond to so-called "bound states". These bound eigenfunctions are also known as non-radiating or guided modes. The eigenvalues λ can be interpreted as the square of the vertical wavenumber ξ_3 suggesting the following relationship

$$\lambda(\xi) = \xi_0^2 - \xi_\nu \xi_\nu \tag{A-7}$$

with ξ_0 the background wavenumber and ξ_ν the horizontal wavenumber(s).

For the discrete Helmholtz matrix, the eigenvalue spectrum is discrete. However, as the examples in Grimbergen et al.'s paper (their Fig. 4(d)) show the spectrum for lateral velocity profiles with a low-velocity well give rise to localized and radiating eigenfunctions. The corresponding eigenvalues look discrete for the latter and "continuous" for the former. The transition from the radiating to non-radiation eigenvalues occurs at ξ_0^2.

The modal transform

The eigenfunctions for the decomposition of the Helmholtz operator are complete and can be used to define a transform, the so-called modal transform, which decomposes a function with respect to the eigenfunctions. The forward transform of an arbitrary finite energy function f reads

$$\hat{f}(\xi_\nu) = \langle f, \psi_{\lambda(\xi_\nu)} \rangle = \int_{x_\nu \in \mathbb{R}^d} f(x_\nu) \overline{\psi_{\lambda(\xi_\nu)}(x_\nu)} \mathrm{d}x_\nu. \tag{A-8}$$

The symbol $\hat{}$ is hereby reserved for transformed domain quantities. In this expression, the hat symbol refers to functions in the temporal frequency-eigenvalue domain. The above modal transform corresponds to taking inner products with respect to the eigenfunctions of the $\mathcal{H}_2$ operator. The inverse modal transform composes the function f from the modal domain

$$f(x_\nu) = \int_{\xi_\nu \in \mathbb{R}^d} \hat{f}(\xi_\nu) \psi_{\lambda(\xi_\nu)}(x_\nu) \mathrm{d}\xi_\nu. \tag{A-9}$$

For simplicity, we ignore degeneracy and the existence of the discrete spectrum that would give rise to an additional summation over the discrete eigenvalues for the integral in Eq. A-9.

The results in the continuous setting largely carry over to the discrete setting in which we are primarily interested in this paper. These monochromatic expressions readily translate into a multi-frequency matrix-vector framework. The multi-frequency modal transform has a block-diagonal structure

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & & \\ & \ddots & \\ & & \mathbf{L}_{n_f} \end{bmatrix} \tag{A-10}$$

with $\mathbf{L}_1 \cdots \mathbf{L}_{n_f}$ the $n_\nu \times n_\nu$ monochromatic modal composition matrices with the eigenfunctions for each of the n_f temporal frequencies. The size of the discretization in the horizontal

coordinate direction(s) is n_ν. After the discretization, we have

$$\mathbf{L}\mathbf{L}^H = \mathbf{I} \tag{A-11}$$

with $\mathbf{I}$ the discrete identity matrix. This identity holds because $\mathbf{L}$ is orthornormal.

One-way extrapolation operator

As previously described, for each temporal frequency $\omega = n\Delta\omega$, the matrix $\mathbf{H}_2$ is self-adjoint, yielding the following decompositions

$$\mathbf{H}_{2,n} = \mathbf{L}_n \mathbf{\Lambda}_n \mathbf{L}_n^H, \quad n = 1 \cdots n_f \tag{A-12}$$

with real eigenvalues, i.e., $\mathrm{diag}(\mathbf{\Lambda}_n)$ is real. This equation is defines the discrete modal transform pair (cf. Eq. A-8)

$$\widehat{\mathbf{f}} = \mathbf{L}_n^H \mathbf{f} \quad \text{and} \quad \mathbf{f} = \mathbf{L}_n \widehat{\mathbf{f}}. \tag{A-13}$$

After the appropriate choices for the square-root of the eigenvalues (cf. Eq.15), the one-way extrapolation operator, carrying the wavefield from x_3 to x_3', can be written as

$$\widehat{\mathbf{W}}_n^{\pm} = \mathbf{L}_n \exp\{\mp j(x_3 - x_3')\mathbf{\Lambda}_n^{1/2}\}\mathbf{L}_n^H. \tag{A-14}$$

This operator propagates a monochromatic wavefield at angular frequency $\omega = n\Delta\omega$ by modal transformation it, followed by applying a diagonal phase rotation and the inverse modal transform. By combining the different monochromatic extrapolations, we arrive at the final expression for one-way wavefield extrapolation (cf. Eq. (10))

$$\mathbf{W}^{\pm} = \mathcal{F}^H \mathbf{L} e^{\mp j\mathbf{\Lambda}^{1/2}(x_3 - x_3')}\mathbf{L}^H \mathcal{F} \tag{A-15}$$

with $\mathcal{F}$ the discrete temporal Fourier transform and

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_1 & & \\ & \ddots & \\ & & \mathbf{L}_{n_f} \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_1 & & \\ & \ddots & \\ & & \mathbf{\Lambda}_{n_f} \end{bmatrix} \tag{A-16}$$

the block diagonal modal composition and eigenvalue matrices, respectively. The extrapolation of a wavefield in the $x_1 - t$ domain now entails taking a temporal Fourier transform, followed by application of the block-diagonal decomposition and phase rotation matrices. The propagated wavefield is recovered by applying the block-diagonal modal composition and the inverse temporal Fourier transform. The measurement and phase-rotated measurement matrices, which form the basis for our compressed extrapolation algorithm, can now be defined as

$$\mathbf{M} := \mathbf{L}^H \mathcal{F} \quad \text{and} \quad \mathbf{M}' := \exp\{j\mathbf{\Lambda}^{1/2}(x_3 - x_3')\}\mathbf{L}^H \mathcal{F}. \tag{A-17}$$

APPENDIX B

STABLE RECOVERY WITH CURVELET FRAMES

In this section, we will briefly introduce the framework of compressed sensing (Candès et al., 2006b; Donoho, 2006; Tsaig and Donoho, 2006). First, we introduce recovery by nonlinear optimization, followed by a discussion of curvelet frames as sparsity promoting representations for seismic wavefields.

Recovery by norm-one nonlinear optimization

The principle of compressed sensing is derived from the fact that an arbitrary discretized function, residing in the vector $\mathbf{f}$, can exactly be recovered from incomplete measurements

40

y given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \tag{B-1}$$

and with **A** a wide $m \times N$ matrix with $m \ll N$.

Nonlinear recovery: When the matrix **A** obeys certain conditions, this recovery is accomplished by solving the following nonlinear optimization problem

$$\mathbf{P_1}: \quad \begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y} \\[2mm] \widetilde{\mathbf{f}} = \mathbf{S}^T \widetilde{\mathbf{x}} \end{cases} \tag{B-2}$$

with the synthesis matrix defined as $\mathbf{A} := \mathbf{RMS}^T$. The success of the recovery depends on three main factors

- the *mutual coherence* between the rows of the measurement matrix **M** and sparsity synthesis matrix $\mathbf{S}^T$. This mutual coherence is given by

$$\mu(\mathbf{M}, \mathbf{S}) = \sqrt{N} \max_{(i,j)\in[1\cdots N]\times[1\cdots N]} |\langle \mathbf{m}_i, \mathbf{s}_j \rangle| \tag{B-3}$$

  with $\mathbf{m}_i$ and $\mathbf{s}_j$ the rows of **M** and **S**, respectively. The mutual coherence between the Dirac-Fourier pair is minimal ($\mu = 1$). The smaller the coherence the fewer observations are required for successful recovery (Candès et al., 2006b; Tsaig and Donoho, 2006; Hennenfent and Herrmann, 2006a).

- the *compression rate* of the to-be-recovered function **f** in the sparsity representation. This compression rate is quantified by a power-law decay rate for the magnitude sorted coefficients ($\mathbf{c} = \mathbf{Sf}$)

$$|c_{i\in I}| \leq C_r i^{-r} \quad \text{with} \quad r \geq 1, \tag{B-4}$$

41

with I the indices such that $c_{I(1)} \geq c_{I(2)} \geq \cdots \geq c_{I(N)}$ and r the compression rate

with C_r a constant depending on the signal's energy. The faster the decay the larger

r.

- the *randomness* of the restriction that is optimal for cases where $\mathbf{R}$ selects rows according to a uniform distribution.

Depending on the mutual coherence and the compression rate obtained by the sparsity representation, the program $\mathbf{P_1}$ is able to recover $\mathbf{f}$ for data $\mathbf{y}$ with large percentages (up to $80\,\%$ in 2-D) missing. This nonlinear recovery depends on the appropriate choice of the sparsity representation for seismic data.

Recovery by the cooling method: Following Daubechies et al. (2005) and Elad et al. (2005) and ideas dating back to Figueiredo and Nowak (2003), Eq. (B-2) is solved by an iterative thresholding technique that is derived from the Landweber descent method. Following Elad et al. (2005), the optimization problem $\mathbf{P}_1$ is replaced by a series of simpler optimization problems

$$
\mathbf{P}_\eta : \quad
\begin{cases}
\widetilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \eta \|\mathbf{x}\|_1 \\[2mm]
\widetilde{\mathbf{f}} = \mathbf{A}\widetilde{\mathbf{x}}.
\end{cases}
\tag{B-5}
$$

These optimization problems depend on a coupling parameter $\eta > 0$ that controls the emphasis on the ℓ_1 norm versus the ℓ_2 norm misfit. This parameter η is lowered during an outer loop.

Iterative soft thresholding: After m iterations of this outer cooling loop, estimations for the coefficient vector are computed for fixed η by the following inner loop

$$\mathbf{x}^{m+1} = T_\eta \left(\mathbf{x}^m + \mathbf{A}^T \left(\mathbf{y} - \mathbf{A}\mathbf{x}^m \right) \right), \tag{B-6}$$

with $\eta = \eta_m$ and the soft thresholding defined by

$$T_\eta(x) := \operatorname{sgn}(x) \cdot \max(0, |x| - \eta). \tag{B-7}$$

As shown by Daubechies et al. (2005), this iteration for fixed η converges to the solution of the subproblem in Eq. (B-5) for m large enough and $\|\mathbf{A}\| < 1$. The cost of each iteration is a synthesis and subsequent analysis. The details of the cooling algorithm are presented in Table. 1.

Initialize:

$m = 0;\ \mathbf{x}^0 = \mathbf{0};$

Choose: $L,\ \|\mathbf{A}^T\mathbf{y}\|_\infty > \eta_1 > \eta_2 > \cdots$

while $\|\mathbf{y} - \mathbf{A}\mathbf{x}^m\|_2 > \epsilon$ **do**

   for $l = 1$ to L **do**

      $\mathbf{x}^{m+1} = T_{\eta_m} \left(\mathbf{x}^m + \mathbf{A}^T \left(\mathbf{y} - \mathbf{A}\mathbf{x}^m \right) \right)$ {Iterative thresholding}

   end for

   $m = m + 1;$

end while

$\widetilde{\mathbf{f}} = \mathbf{A}\mathbf{x}^m.$

Table 1: The cooling method with $\eta_1 > \eta_2 > \cdots$ the series of decreasing thresholds. The inner loop is repeated L times.

Curvelet frames

The conditions for successful nonlinear recovery place constraints on the choice for the sparsity analysis and synthesis matrices. As described in the previous section, the analysis matrices must achieve a high compression rate for the wavefield while simultaneously possessing a sufficiently low mutual coherence with the measurement matrix. Taking these requirements as well as our goal of compressing seismic extrapolation operators into consideration, we will argue that the recently introduced curvelet transform provides the appropriate signal representation (Candès and Donoho, 2000a, 2004; Candès et al., 2006a; Ying et al., 2005).

Compression of seismic wavefields by curvelet frames

Curvelets achieve a rapid decay of magnitude-sorted coefficients (and hence compression) by expanding the wavefield in terms of localized, multiscale and multidirectional prototype waveforms that are anisotropically shaped. Without *prior* information, the location and direction of wavefronts are found though the process of alignment that leads to large inner products between curvelets (rows of the curvelet transform matrix) and wavefronts that locally have the same direction and frequency content. This principle of alignment is illustrated in Fig. B-1, where curvelets aligned with curved events are shown to have large coefficients, while curvelets that make an angle have small coefficients. As such curvelets are capable of representing seismic images with vectors that show a rapid decay for the magnitude-sorted coefficients. By comparison, the wavelet and Fourier transforms decay more slowly (Candès et al., 2006a; Hennenfent and Herrmann, 2006b). Refer to Candès and Donoho (2000b), for a discussion on the near optimal asymptotic decay rate for the nonlinear approximation error of curvelets.

The curvelet transform (see e.g. Candès and Donoho, 2004; Candès et al., 2006a) composes signals in terms of waveforms that are multiscale and multidirectional. Because the rows of the transform contain prototype waveforms that behave locally like 'little waves', the curvelet transform obtains near optimal sparsity on bandwidth-limited seismic data (Candès et al., 2006a; Hennenfent and Herrmann, 2006b). Because the number of rows with waveforms exceeds the number of samples in the image, the curvelet transform is overcomplete.

By using the fast discrete curvelet transform (FDCT by wrapping, see e.g. Candès et al., 2006a; Ying et al., 2005; Candes et al., 2005), data is perfectly reconstructed after decomposition by applying the adjoint of the curvelet transform, i.e., we have $\mathbf{f} = \mathbf{C}^T \mathbf{C} \mathbf{f}$ for an arbitrary finite-energy vector $\mathbf{f}$. In this expression, $\mathbf{C} \in \mathbb{R}^{N \times M}$ represents the curvelet decomposition, transforming a column vector $\mathbf{f} \in \mathbb{R}^M$ into the curvelet domain, $\mathbf{c} = \mathbf{C} \mathbf{f}$ with $\mathbf{c} \in \mathbb{R}^N$. The $\mathbf{C}^T$ represents the transpose of the curvelet transform. For the above choice of curvelet transform, the pseudo inverse equals the transpose, i.e., $\mathbf{C}^T \mathbf{c} = \mathbf{C}^\dagger \mathbf{c}$. The transform is a tight frame that preserves energy, i.e., $\|\mathbf{f}\| = \|\mathbf{C} \mathbf{f}\|$, so we have $\mathbf{C}^T \mathbf{C} \mathbf{f} = \mathbf{I} \mathbf{f}$. Since the discrete curvelet representation is overcomplete, with a moderate redundancy (a factor of roughly 8 for $d = 2$), the converse is not the identity, i.e., $\mathbf{C} \mathbf{C}^T \neq \mathbf{I}$, which makes it difficult to recover the sparsity vector $\mathbf{x}_0$ from $\mathbf{f} = \mathbf{C}^T \mathbf{x}_0$.

Curvelet frames were originally developed for studying high-frequency asymptotic solutions to the wave equation (Smith, 1997; Candès and Demanet, 2005). As such curvelets are known to display certain invariance properties under wave propagation. In our context, these invariance properties translate to a compression of wavefields by curvelets. This property is essential for the formulation of our compressed extrapolation algorithm.

Curvelet properties

Curvelets are redundant directional frames that represent a tiling of the two-dimensional frequency plane into multiscale and multi-angular wedges (see Fig. B-2). Because the directional sampling increases every-other scale doubling, curvelets become more and more anisotropic for finer and finer scales. They become 'needle-like' as illustrated in Fig. B-2. Curvelets are localized in both domains and are oscillatory in one and smooth in the other direction. Even though curvelets are not of compact support (non-zero over a finite interval) in the physical domain, they are of rapid decay with an effective support given by ellipsoids parameterized by a width $\propto 2^{j/2}$, length $\propto 2^j$ and angle $\theta = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale and l the angular index with the number of angles doubling every other scale doubling (see Fig. B-2). Curvelets are indexed by the multi-index $\mu := (j, l, \mathbf{k}) \in \mathcal{M}$ with $\mathcal{M}$ the multi-index set running over all scales, j, angles, l, and positions $\mathbf{k}$ (see for details Candès et al., 2006a; Ying et al., 2005). At each scale j the curvelet transform consists of a decomposition with respect to curvelets at each grid point and for $2^{\lfloor j/2 \rfloor}$ different angles. Therefore, conflicting angles are possible.

REFERENCES

Bednar, J. B. and C. J. Bednar, 2006, Two-way versus one-way: a case study style comparison: Presented at the 76th Ann. Internat. Mtg., SEG, Soc. Expl. Geophys., Expanded abstracts.

Berkhout, A. J., 1982, Seismic migration. Imaging of acoustic energy by wave field extrapolation: Elsevier, Amsterdam.

Candès, E., 2007, Compressive sensing: Presented at the 06/07 IAM Seminars, UBC.

Candès, E., L. Demanet, D. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: SIAM Multiscale Model. Simul., **5**, 861–899.

Candès, E., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Comm. Pure Appl. Math., **59**, 1207–1223.

Candès, E. J. and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Comm. Pure Appl. Math, **58**, 1472–1528.

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

Candes, E. J., D. Donoho, L. Demanet, and L. Ying, 2005, Curvelab: Fast discrete curvelet transform.

Candès, E. J. and D. L. Donoho, 2000a, Curvelets – a surprisingly effective nonadaptive representation for objects with edges, *in* L. L. Schumaker et al., eds., Curves and Surfaces. Vanderbilt University Press.

———, 2000b, Recovering Edges in Ill-posed Problems: Optimality of Curvelet Frames: Ann. Statist., **30**, 784–842.

Candès, E. J. and J. Romberg, 2005, ℓ_1-magic. Software, `http://www.acm.caltech.edu/l1magic/`.

Chauris, H., 2006, Seismic imaging in the curvelet domain and its implications for the curvelet design: Presented at the 76th Ann. Internat. Mtg., SEG, Soc. Expl. Geophys., Expanded abstracts.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Journal on Scientific Computing, **43**, 129–159.

Claerbout, J. and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **38**, 826–844.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Collino, F. and P. Joly, 1995, Splitting of operators, alternate directions, and paraxial approximations for the three-dimensional wave equation: SIAM J. Sci. Comput., **16**, 1019–1048.

Daubechies, I., M. Defrise, and C. de Mol, 2005, An iterative thresholding algorithm for linear inverse problems with a sparsity constrains: Comm. Pure Appl. Math., 1413–1457.

de Hoop, M., J. L. Rousseau, and R.-S. Wu, 2000, Generalization of the phase-screen approximation for the scattering of acoustic waves: Wave Motion, **31**, 43–70.

Dessing, F. J., 1997, A wavelet transform approach to seismic processing: PhD thesis, Delft University of Technology, Delft, the Netherlands.

Donoho, D. L., 2006, Compressed sensing: IEEE Trans. Inform. Theory, **52**, 1289–1306.

Donoho, D. L., I. Drori, V. Stodden, and Y. Tsaig, 2005, SparseLab. Software, `http://sparselab.stanford.edu/`.

Douma, H. and M. de Hoop, 2006, Leading-order seismic imaging using curvelets: Presented at the 76th Ann. Internat. Mtg., SEG, Soc. Expl. Geophys., Expanded abstracts.

Elad, M., J. Starck, P. Querre, and D. Donoho, 2005, Simulataneous Cartoon and Tex-

ture Image Inpainting using Morphological Component Analysis (MCA): Appl. Comput. Harmon. Anal., **19**, 340–358.

Figueiredo, M. and R. Nowak, 2003, An EM algorithm for wavelet-based image restoration: IEEE Trans. Image Processing, **12**, 906–916.

Figueiredo, M., R. D. Nowak, and S. J. Wright, 2007, Gradient Projection for Sparse Reconstruction. Software, `http://www.lx.it.pt/~mtf/GPSR/`.

Grimbergen, J., F. Dessing, and C. Wapenaar, 1998, Modal expansion of one-way operator on laterally varying media: Geophysics, **63**, 995–1005.

Guitton, A. and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the l^1-norm: Geophys Prospect, **52**, 27–27.

Hale, D., N. R. Hill, and J. Stefani, 1992, Imaging salt with turning seismic waves: Geophysics, **57**, 1453–1462. Discussion and reply by authors in GEO-58-8-1205-1206.

He, C., M. Lu, and C. Sun, 2004, Accelerating seismic migration using FPGA-based coprocessor platform: 12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines, 207–216.

Hennenfent, G. and F. J. Herrmann, 2006a, Application of stable signal recovery to seismic interpolation: Presented at the SEG International Exposition and 76th Annual Meeting.

——, 2006b, Seismic denoising with non-uniformly sampled curvelets: Comp. in Sci. and Eng., **8**, 16–25.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: Geoph. J. Int. (Accepted for publication).

Koh, K., S. J. Kim, and S. Boyd, 2007, Simple matlab solver for l1-regularized least squares problems. Software, `http://www-stat.stanford.edu/~tibs/lasso.html`.

Levy, S., D. Oldenburg, and J. Wang, 1988, Subsurface imaging using magnetotelluric data:

Geophysics, **53**, 104–117.

Mulder, W. and R. Plessix, 2004, How to choose a subset of frequencies in frequency-domain finite-difference migration: Geoph. J. Int., **158**, 801–812.

Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: Geophysics, **46**, 1528–1542.

Paige, C. C. and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM TOMS, **8**, 43–71.

Plessix, R. and W. Mulder, 2004, Frequency-domain finite difference amplitude-preserving migration: Geoph. J. Int., **157**, 975–987.

Riyanti, C., Y. Eriangga, R. Plessix, W. Mulder, C. Vulk, and C. Oosterlee, 2006, A new iterative solver for the time-harmonic wave equation: Geophysics, **71**.

Sacchi, M. and T. Ulrych, 1996, Estimation of the discrete fourier transform, a linear inversion approach: Geophysics, **61**, 1128–1136.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, 938–945.

Santosa, F. and W. Symes, 1986, Linear inversion of band-limited reflection seismogram: SIAM J. of Sci. Comput., **7**.

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

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

Taylor, H. L., S. Banks, and J. McCoy, 1979, Deconvolution with the ℓ_1 norm: Geophysics, **44**, 39.

Tibshirani, R., 1996, Least Absolute Shrinkage and Selection Operator. Software, `http://www-stat.stanford.edu/~tibs/lasso.html`.

Tsaig, Y. and D. Donoho, 2006, Extensions of compressed sensing: Signal processing, **86**, 549–571.

Ulrych, T. J. and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, 1295–1302.

Ying, L., L. Demanet, and E. Candés, 2005, 3d discrete curvelet transform: Presented at the Wavelets XI, SPIE.

Zwartjes, P. and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: Geophysics, **71**, V171–V186.

LIST OF FIGURES

the top left corner corresponds to negative eigenvalues with eigenmodes that are evanescent.

6 Lateral velocity profiles with background velocity $2000\,\mathrm{ms}^{-1}$. **(a)** Profile with velocity low of $1200\,\mathrm{ms}^{-1}$. **(b)** Profile with velocity high of $3500\,\mathrm{ms}^{-1}$. Spatial sampling interval of the profiles is set to $4\,\mathrm{m}$ with 256 samples, while the sigma of both Gaussian functions (which roughly correspond to its width at half maximum amplitude) are set to $80\,\mathrm{m}$.

7 Initial wavefields used for the extrapolation examples. **(a)** A chain of horizontally-oriented fine-scale curvelets playing the role of a "plane wave". **(b)** A fan of fine-scale curvelets with different angle.

8 Radiating and non-radiating wave modes for the velocity model with the Gaussian low (Fig. 6(a)) at 30Hz. **(a)** The non-radiating or guided wave mode. **(b)** the radiating wave mode. **(c)**, **(d)** the corresponding (f-k) spectra. The 'frequency spread' of these two wavemodes is not significantly different.

9 Compressed forward extrapolation according to $\mathbf{W_1}$ (cf. Eq. (18)) for different restrictions. The velocity model corresponds to the velocity low and is plotted in Fig. 6(a). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** The *full* extrapolated wavefield $\mathbf{u} = \mathbf{Wv}$ is included for reference; **(b)** The compressed forward propagated wavefield with $p_f = 0.2$ and $p_\nu = 0.0.2$; **(c)** The same as **(b)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(d)** The same as **(b)** but with $p_f = 0.6$ and $p_\nu = 0.4$. Observe that the forward propagated wavefield is largely recovered for the restriction in **(c)**.

10 Compressed forward extrapolation according to $\mathbf{W_1}$ (cf. Eq. (18)) for different restrictions. The velocity model corresponds to the velocity high and is plotted in Fig. 6(b). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** The *full* extrapolated wavefield $\mathbf{u} = \mathbf{Wv}$ is included for reference; **(b)** The compressed forward propagated wavefield with

$p_f = 0.2$ and $p_\nu = 0.2$; **(c)** The same as **(b)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(d)** The same as **(b)** but with $p_f = 0.6$ and $p_\nu = 0.4$. Observe that the forward propagated wavefield is largely recovered for the restriction in **(c)**.

11   Compressed inverse extrapolation according $\mathbf{F_1}$ (cf. Eq. (25)) for different restrictions. For **(a-c)** the velocity model is given by the Gaussian low (Fig. 6(a)) and in **(d-f)** by the Gaussian high (Fig. 6(b)). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** Inverse extrapolated wavefield for $p_f = 0.2$ and $p_\nu = 0.2$; **(b)** The same as **(a)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(c)** The same as **(a)** but with $p_f = 0.6$ and $p_\nu = 0.4$; **(d-f)** the same as **(a-c)** but for the velocity high. Observe that the recovery for the velocity high is slightly better.

12   Inversion of the evanescent wavemodes according $\widetilde{\mathbf{v}}_m = \mathbf{W}^T\mathbf{u}$ or $\widetilde{\mathbf{v}} = \mathbf{F_1}[\mathbf{u}]$ (cf. Eq. 25). The velocity model is constant at $2000\,\mathrm{ms}^{-1}$. The initial source wavefield, $\mathbf{v}$, is defined in terms of a the curvelet fan plotted in Fig. 7(b). **(a)** The *full* forward propagated wavefield $\mathbf{u} = \mathbf{W}\mathbf{v}$; **(b)** The matched filter; **(c)** The ℓ_1 recovery. Observe that the steep evanescent angles are fully recovered.

13   Lateral velocity profile for the overthrust examples.

14   Frequency-modal spectra for the exploding reflector model. **(a)** spectrum of the exploding reflector. **(b)** spectrum of the forward extrapolated exploding reflector. Notice the missing evanescent waves for **(b)** and the concentration of amplitudes in the area of the guided modes at the large eigenvalues.

15   Compressed inverse extrapolation according to $\mathbf{F_1}$ (cf. Eq. (25)). The data $\mathbf{u}$ is given by the exploding reflector data plotted in Fig. 4(a). **(a)** The *full* forward propagated wavefield $\mathbf{u} = \mathbf{W}\mathbf{v}$; **(b)** the inverse extrapolated wavefield obtained through matched filtering; **(c)** the compressed inverse extrapolated wavefield $\widetilde{\mathbf{v}} = \mathbf{F_1}[\mathbf{u}]$ with $p_f = 0.35$ and

$p_\nu = 0.7$. Compared to the matched filter , our approach recovers most of the initial wave-field.

B-1 Example of the alignment of curvelets with curved events.

B-2 Spatial and frequency representation of curvelets. **(a)** Four different curvelets in the spatial domain at three different scales. **(b)** Dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. The position of its wedge in k-space dictates the scale and angle of a curvelet, while the phase information in k-space corresponds to its physical translation. This figure illustrates the micro-local correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientation of curvelets in the two domains. The 90° rotation is a property of the Fourier transform.

Figure 1: Example of compressed shifting of length 200 with 5 arbitrary spikes. **Top:** the 5 spikes; **Middle:** shifted spikes by 20 samples according to Eq. 3. **Bottom:** the same but according to the compressed program of Eq. 5. Notice that there is virtually no difference.

Herrmann *et.al.* –
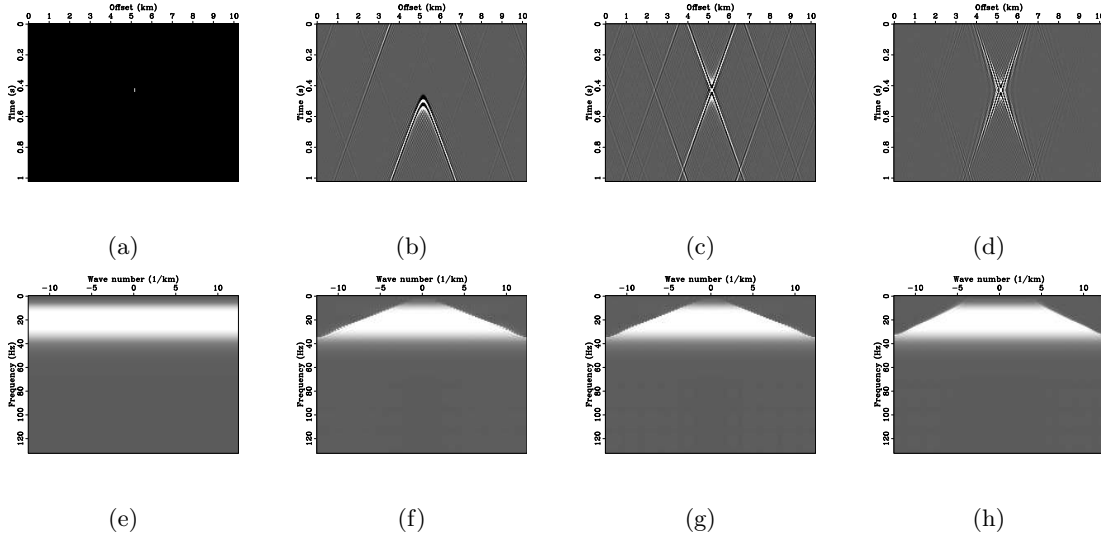
Figure 2: Illustration of the dip limitation during inverse extrapolation. **(a)** a bandwidth-limited impulsive source. **(b)** the forward extrapolated wavefield (cf. Eq. (16)). **(c)** The refocused pulse through inverse extrapolation with matched filtering (cf. Eq. (24)). **(d)** the same but with regularized least-squares inverse extrapolation; **(e-h)** The time-spatial spectra of **(a-d)**. Notice the lack of spatial frequencies corresponding to steep dips in **(f-h)**. The wrap-around effects in **(b-d)** are due to the periodicity of the temporal Fourier transform.

Herrmann *et.al.* –
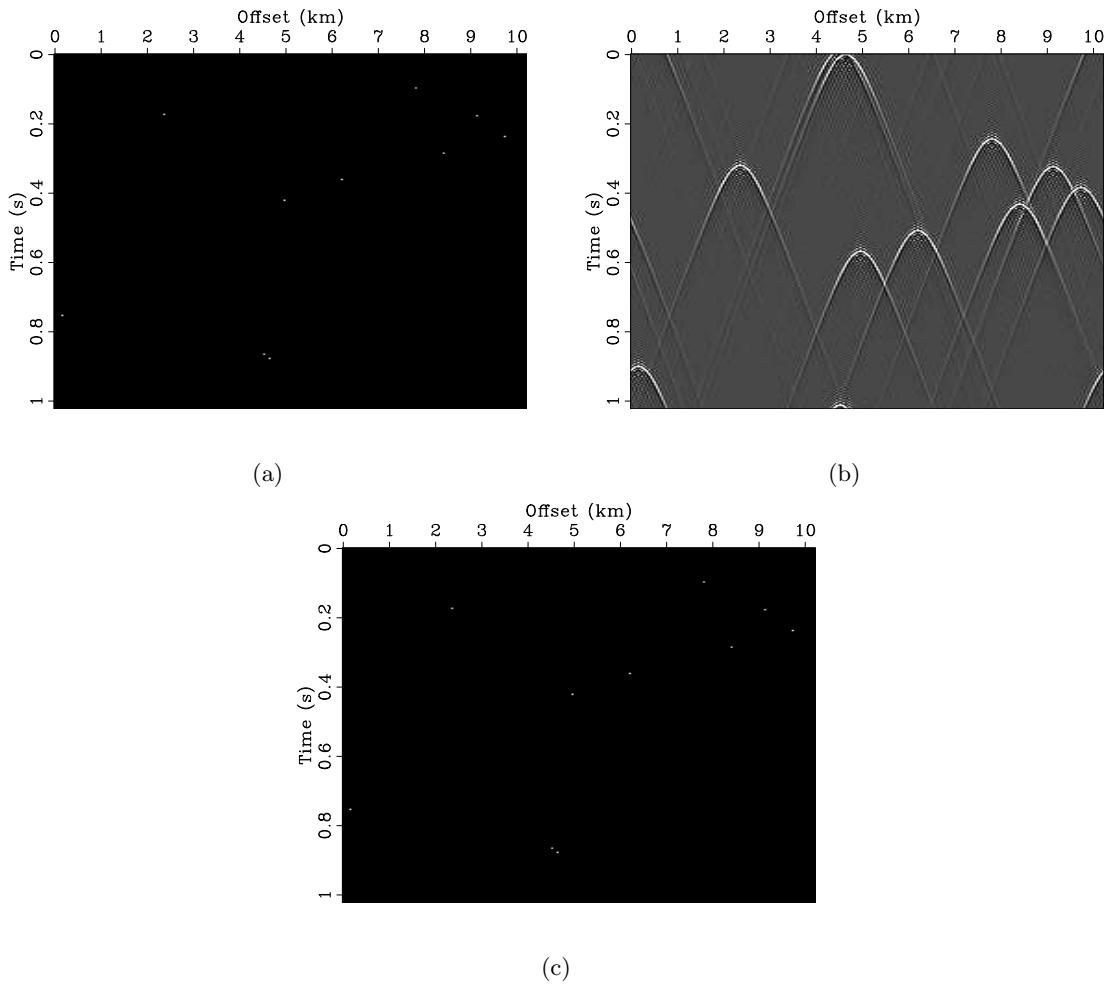
Figure 3: Compressed recovery from the forward extrapolated wavefield with a 100-fold reduction for the size of the extrapolation operators. **(a)** some broad-band spikes; **(b)** The forward extrapolated wavefield; **(c)** compressive nonlinear recovery of the initial wavefield.

Herrmann *et.al.* –
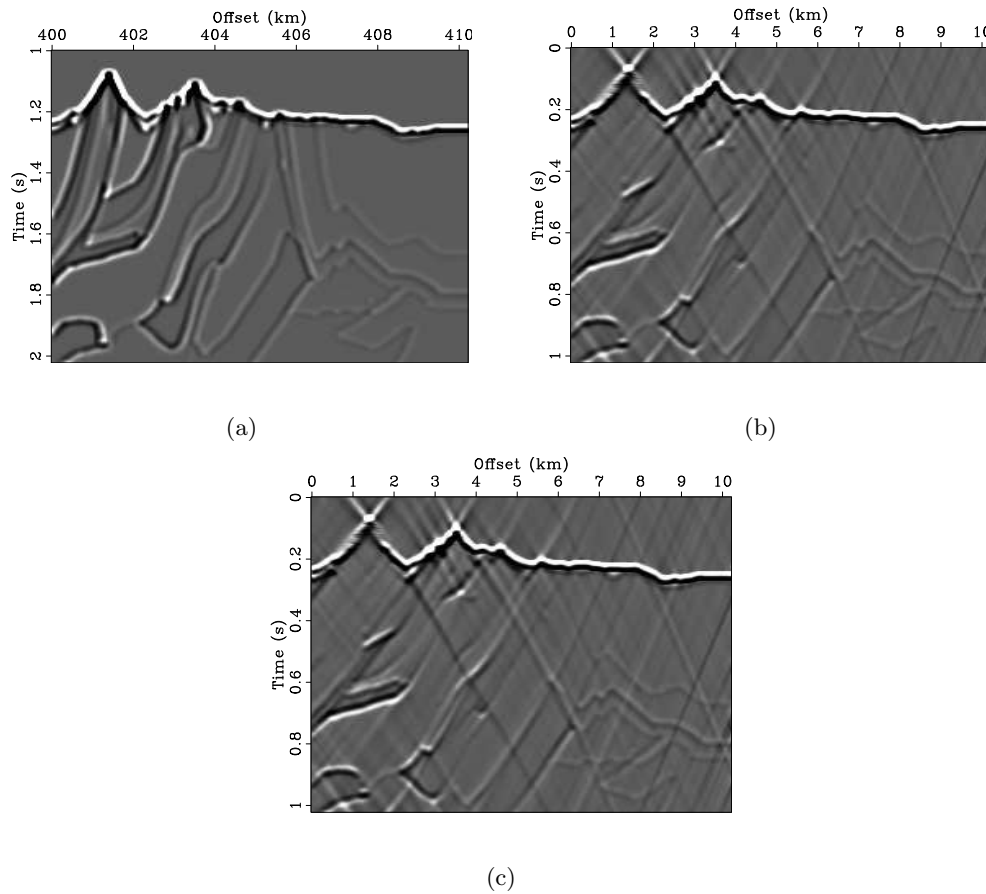
Figure 4: Inverse extraplation based on the matched filter (cf. Eq. (24)). **(a)** Exploding bandwidth-limited reflectivity representative of a Canadian overthrust front; **(b)** inverse extrapolation through matched filtering. Notice the smoothing, missing steep events and finite-aperture artifacts.

Herrmann *et.al.* –
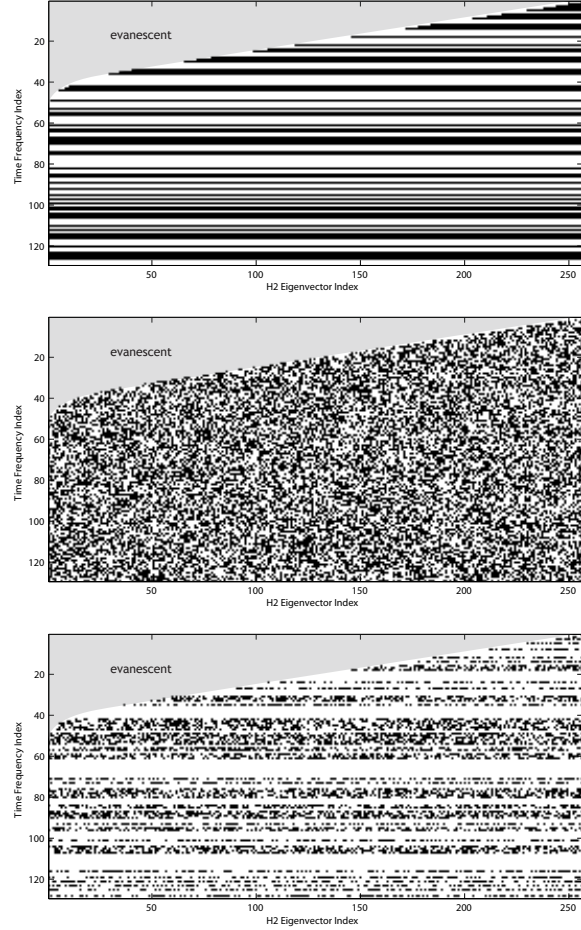
Figure 5: Different restriction strategies. **Top:** angular-frequency restriction with uniformly random-selected frequencies; **Middle:** Wavemode restriction with uniformly random selected wavemodes per frequency. Since the wavemodes for each frequency are obtained by individual eigenfunction problems, restriction of the wavemodes can be made independent across the frequency spectrum. **Bottom:** Mixed uniform restriction. The black areas correspond to individually selected angular frequencies and wave numbers. The grey area on the top left corner corresponds to negative eigenvalues with eigenmodes that are evanescent.

Herrmann *et.al.* –
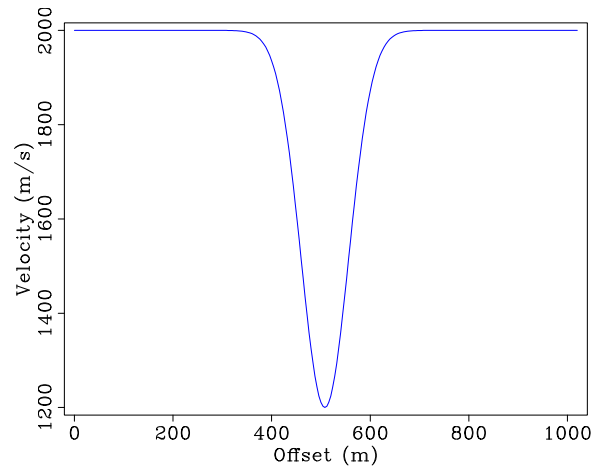
(a)
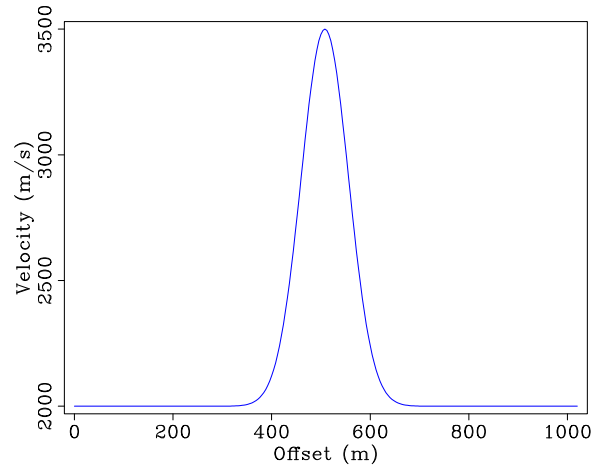
(b)

Figure 6: Lateral velocity profiles with background velocity $2000\,\mathrm{ms}^{-1}$. **(a)** Profile with velocity low of $1200\,\mathrm{ms}^{-1}$. **(b)** Profile with velocity high of $3500\,\mathrm{ms}^{-1}$. Spatial sampling interval of the profiles is set to $4\,\mathrm{m}$ with 256 samples, while the sigma of both Gaussian functions (which roughly correspond to its width at half maximum amplitude) are set to $80\,\mathrm{m}$.

Herrmann *et.al.* –
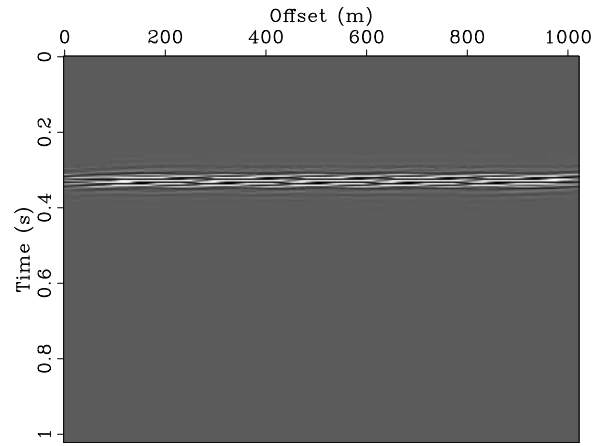
(a)
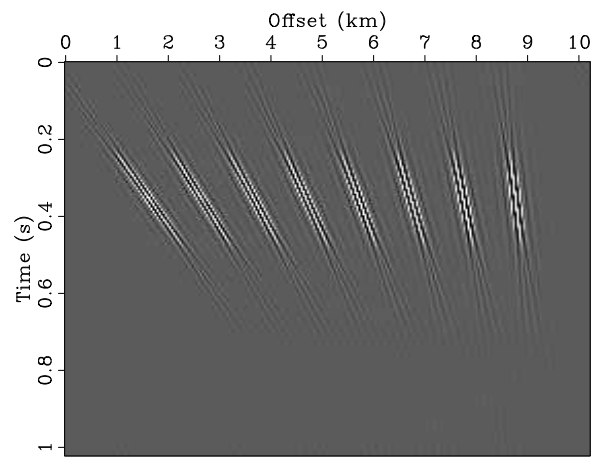
(b)

Figure 7: Initial wavefields used for the extrapolation examples. **(a)** A chain of horizontally-oriented fine-scale curvelets playing the role of a "plane wave". **(b)** A fan of fine-scale curvelets with different angle.

Herrmann *et.al.* –
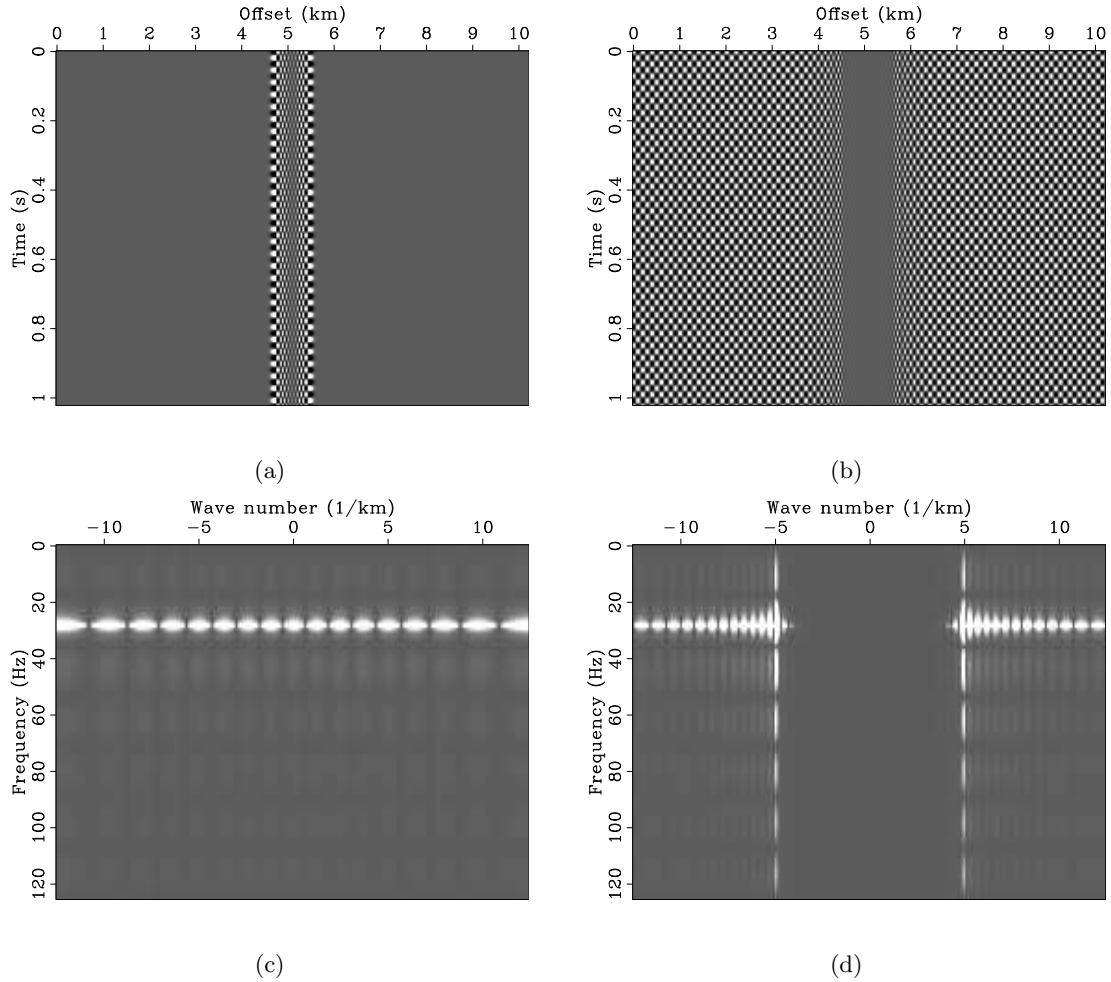
Figure 8: Radiating and non-radiating wave modes for the velocity model with the Gaussian low (Fig. 6(a)) at 30Hz. **(a)** The non-radiating or guided wave mode. **(b)** the radiating wave mode. **(c)**, **(d)** the corresponding (f-k) spectra. The 'frequency spread' of these two wavemodes is not significantly different.
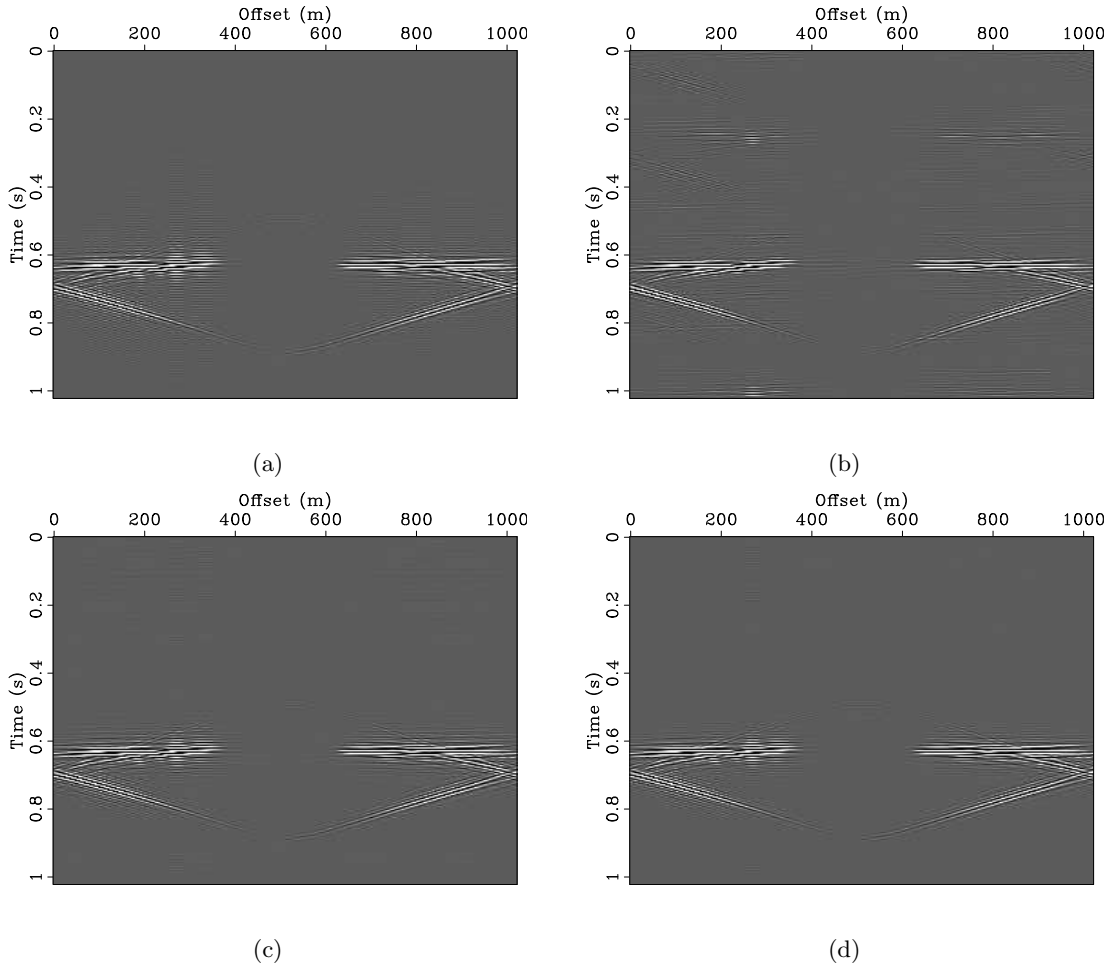
Herrmann *et.al.* –
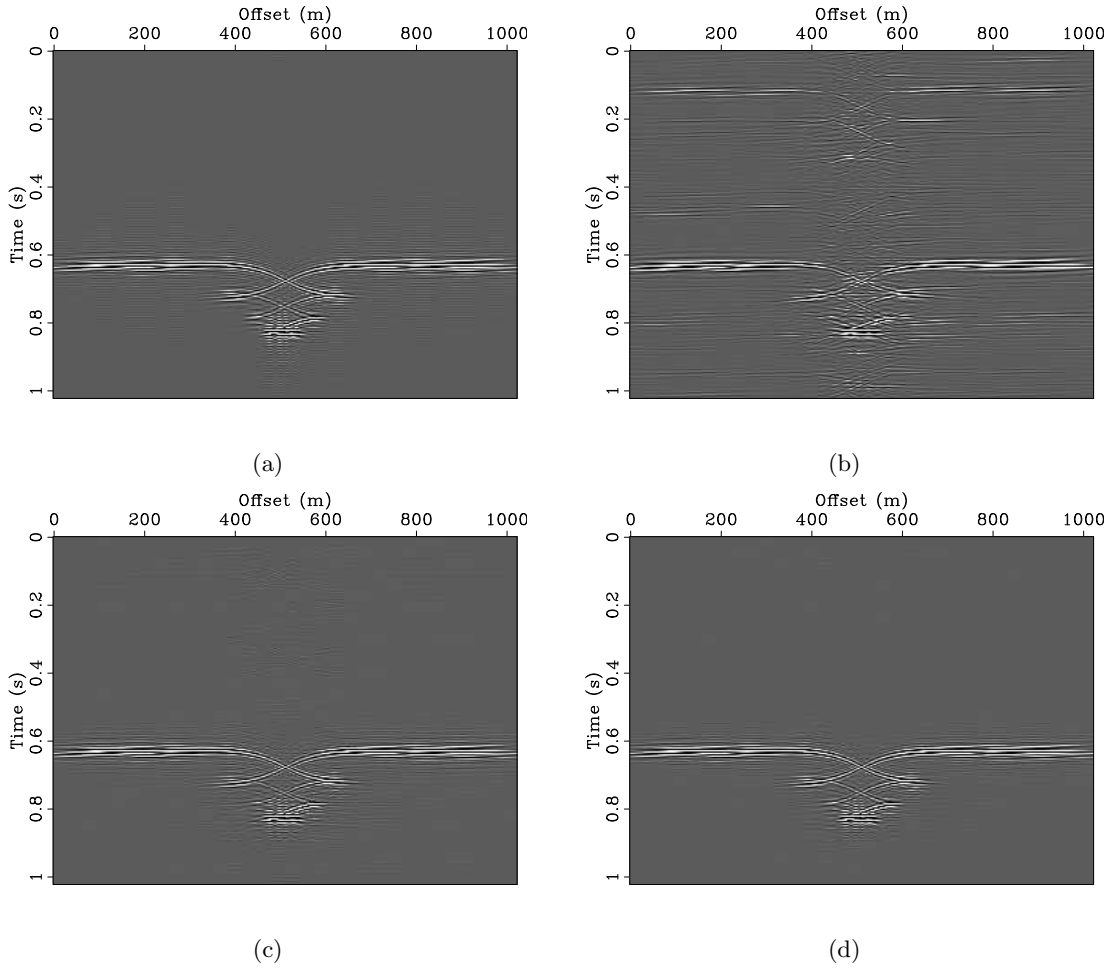
Figure 9: Compressed forward extrapolation according to $\mathbf{W_1}$ (cf. Eq. (18)) for different restrictions. The velocity model corresponds to the velocity low and is plotted in Fig. 6(a). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** The *full* extrapolated wavefield $\mathbf{u} = \mathbf{Wv}$ is included for reference; **(b)** The compressed forward propagated wavefield with $p_f = 0.2$ and $p_\nu = 0.0.2$; **(c)** The same as **(b)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(d)** The same as **(b)** but with $p_f = 0.6$ and $p_\nu = 0.4$. Observe that the forward propagated wavefield is largely recovered for the restriction in **(c)**.
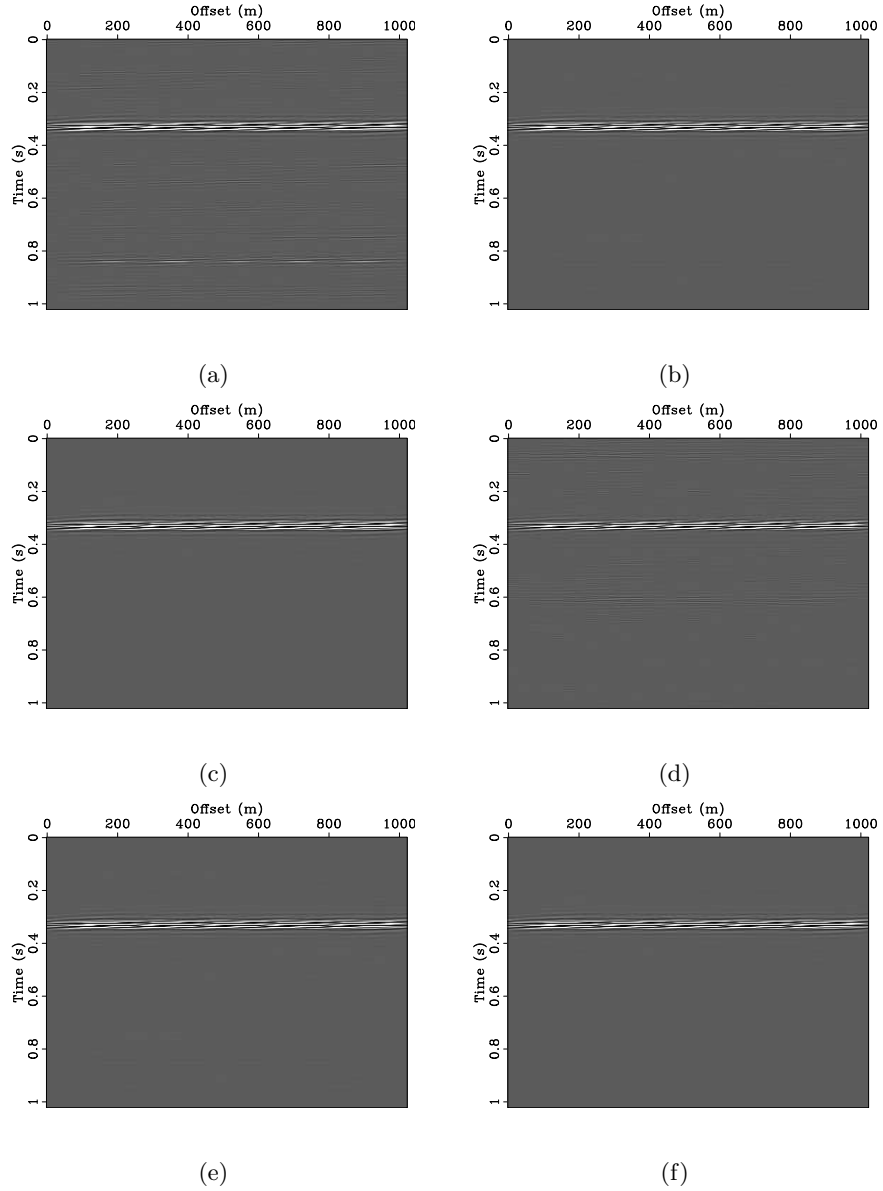
Herrmann *et.al.* –

Figure 10: Compressed forward extrapolation according to $\mathbf{W_1}$ (cf. Eq. (18)) for different restrictions. The velocity model corresponds to the velocity high and is plotted in Fig. 6(b). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** The *full* extrapolated wavefield $\mathbf{u} = \mathbf{Wv}$ is included for reference; **(b)** The compressed forward propagated wavefield with $p_f = 0.2$ and $p_\nu = 0.2$; **(c)** The same as **(b)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(d)** The same as **(b)** but with $p_f = 0.6$ and $p_\nu = 0.4$. Observe that the forward propagated wavefield is largely recovered for the restriction in **(c)**.

Herrmann *et.al.* –

Figure 11: Compressed inverse extrapolation according $\mathbf{F_1}$ (cf. Eq. (25)) for different restrictions. For **(a-c)** the velocity model is given by the Gaussian low (Fig. 6(a)) and in **(d-f)** by the Gaussian high (Fig. 6(b)). The initial source wavefield $\mathbf{v}$ is plotted in Fig. 7(a). **(a)** Inverse extrapolated wavefield for $p_f = 0.2$ and $p_\nu = 0.2$; **(b)** The same as **(a)** but with $p_f = 0.4$ and $p_\nu = 0.4$; **(c)** The same as **(a)** but with $p_f = 0.6$ and $p_\nu = 0.4$; **(d-f)** the same as **(a-c)** but for the velocity high. Observe that the recovery for the velocity high is slightly better.

Herrmann *et.al.* –

(a)

(b)

(c)
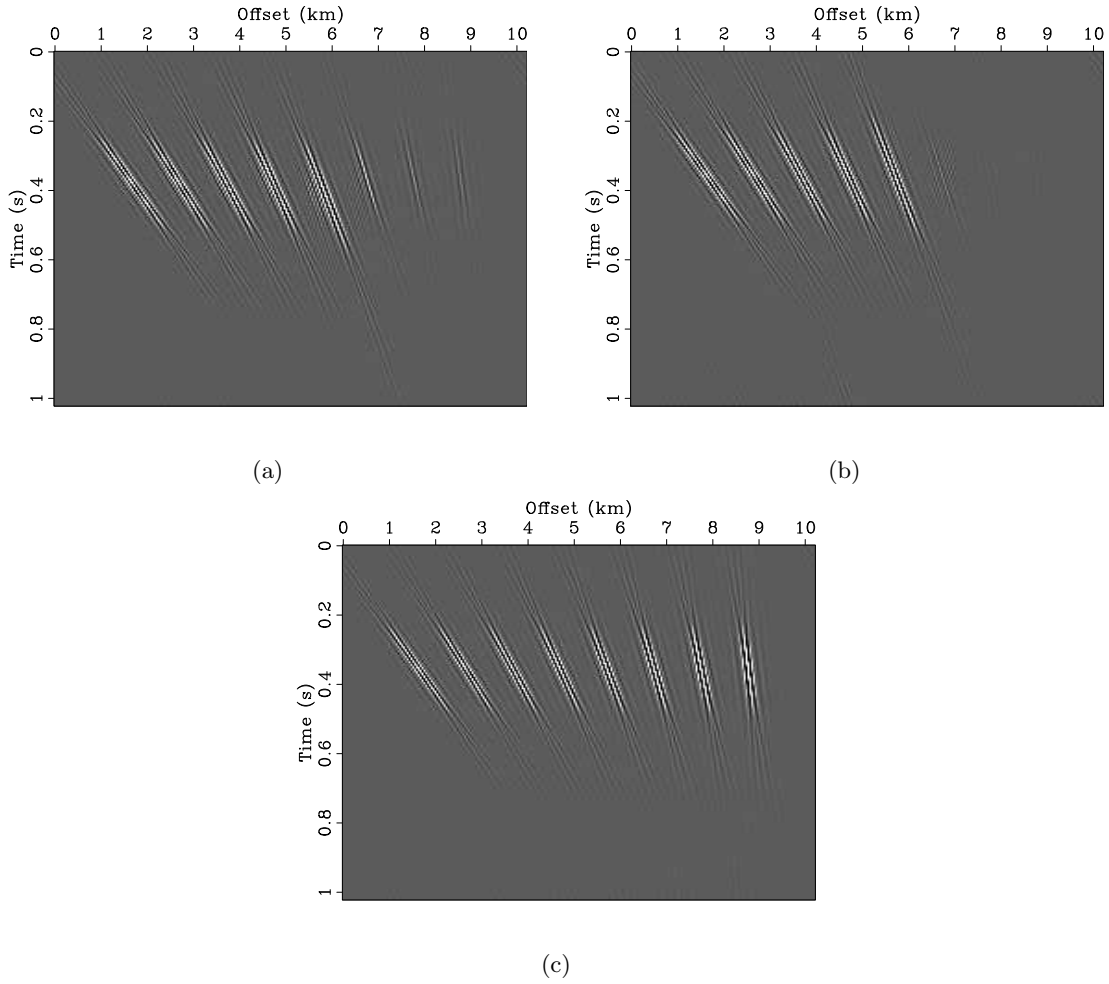
Figure 12: Inversion of the evanescent wavemodes according $\widetilde{\mathbf{v}}_m = \mathbf{W}^T\mathbf{u}$ or $\widetilde{\mathbf{v}} = \mathbf{F_1}[\mathbf{u}]$ (cf. Eq. 25). The velocity model is constant at $2000\,\mathrm{ms}^{-1}$. The initial source wavefield, $\mathbf{v}$, is defined in terms of a the curvelet fan plotted in Fig. 7(b). **(a)** The *full* forward propagated wavefield $\mathbf{u} = \mathbf{Wv}$; **(b)** The matched filter; **(c)** The ℓ_1 recovery. Observe that the steep evanescent angles are fully recovered.

Herrmann *et.al.* –
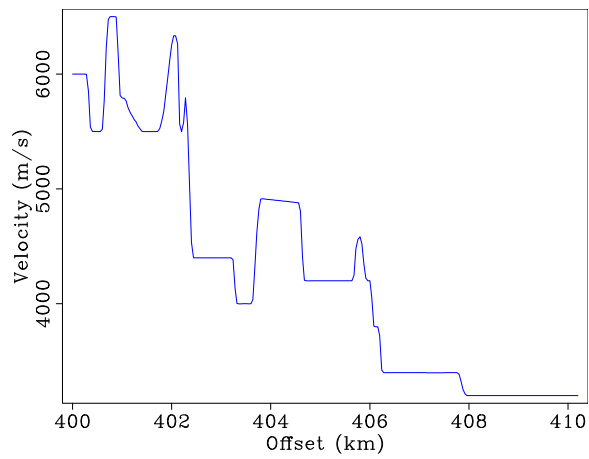
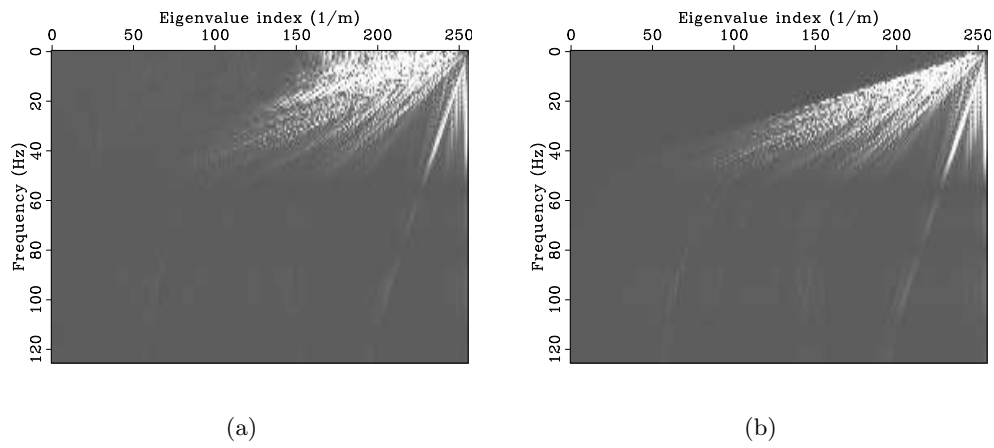Figure 13: Lateral velocity profile for the overthrust examples.**Herrmann** *et.al.* –

Figure 14: Frequency-modal spectra for the exploding reflector model. **(a)** spectrum of the exploding reflector. **(b)** spectrum of the forward extrapolated exploding reflector. Notice the missing evanescent waves for **(b)** and the concentration of amplitudes in the area of the guided modes at the large eigenvalues.
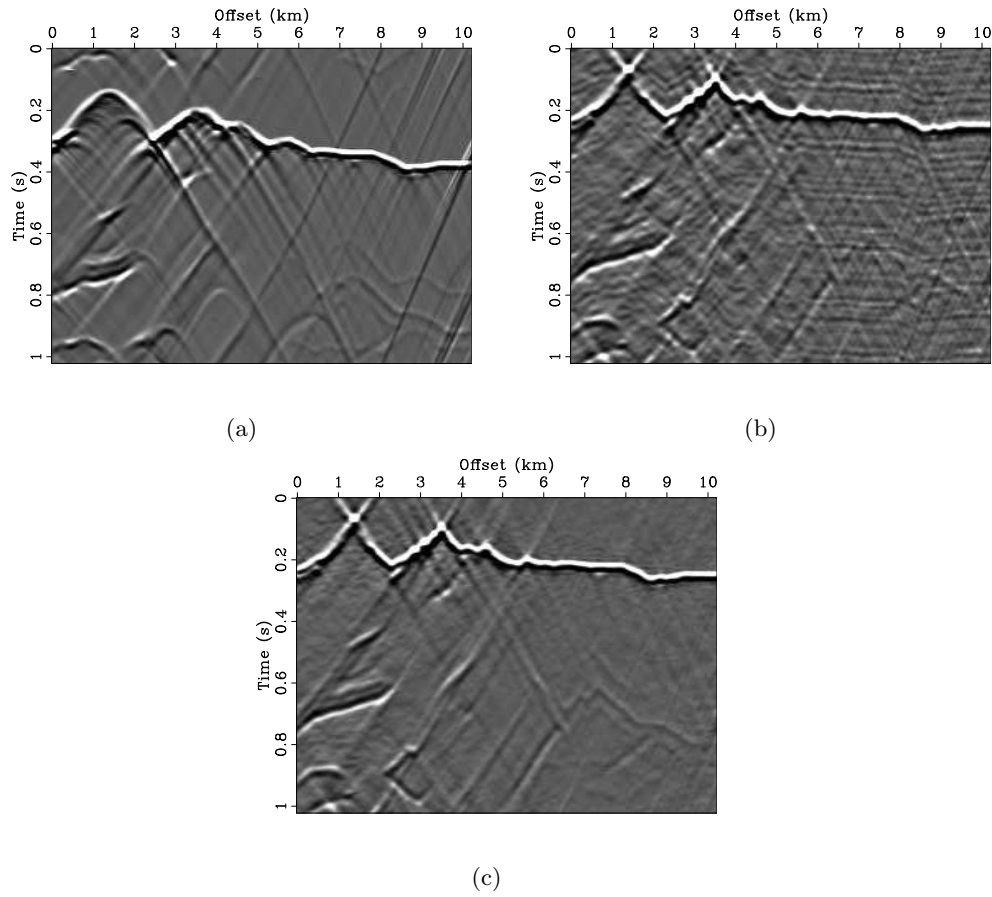
Herrmann *et.al.* –

(a)   (b)

(c)

Figure 15: Compressed inverse extrapolation according to $\mathbf{F_1}$ (cf. Eq. (25)). The data $\mathbf{u}$ is given by the exploding reflector data plotted in Fig. 4(a). **(a)** The *full* forward propagated wavefield $\mathbf{u} = \mathbf{W}\mathbf{v}$; **(b)** the inverse extrapolated wavefield obtained through matched filtering; **(c)** the compressed inverse extrapolated wavefield $\widetilde{\mathbf{v}} = \mathbf{F_1}[\mathbf{u}]$ with $p_f = 0.35$ and $p_\nu = 0.7$. Compared to the matched filter , our approach recovers most of the initial wavefield.
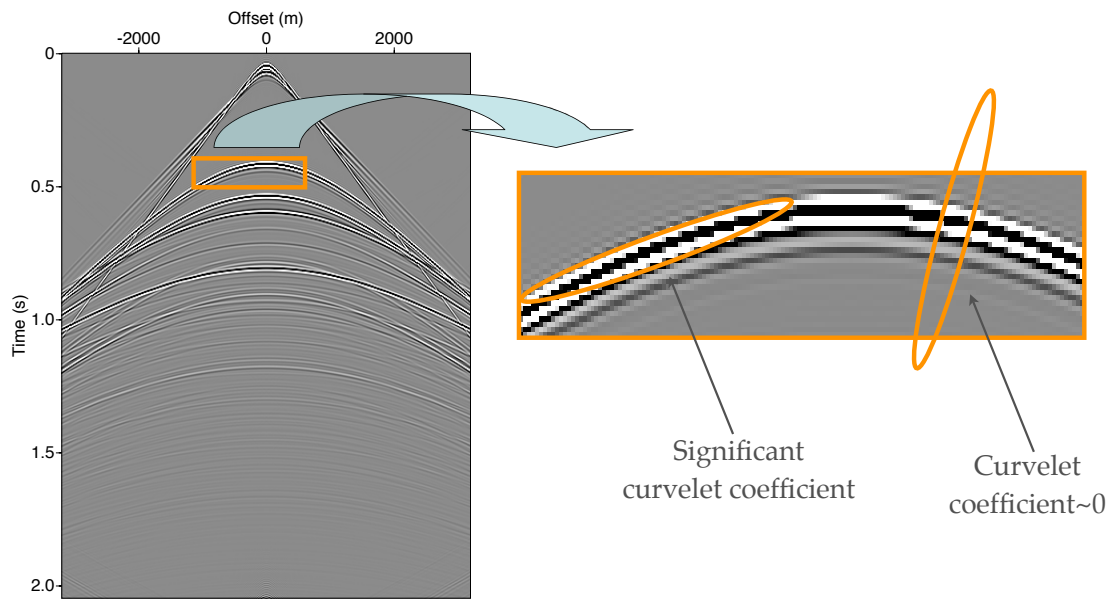
Herrmann *et.al.* –

70

Figure B-1: Example of the alignment of curvelets with curved events.
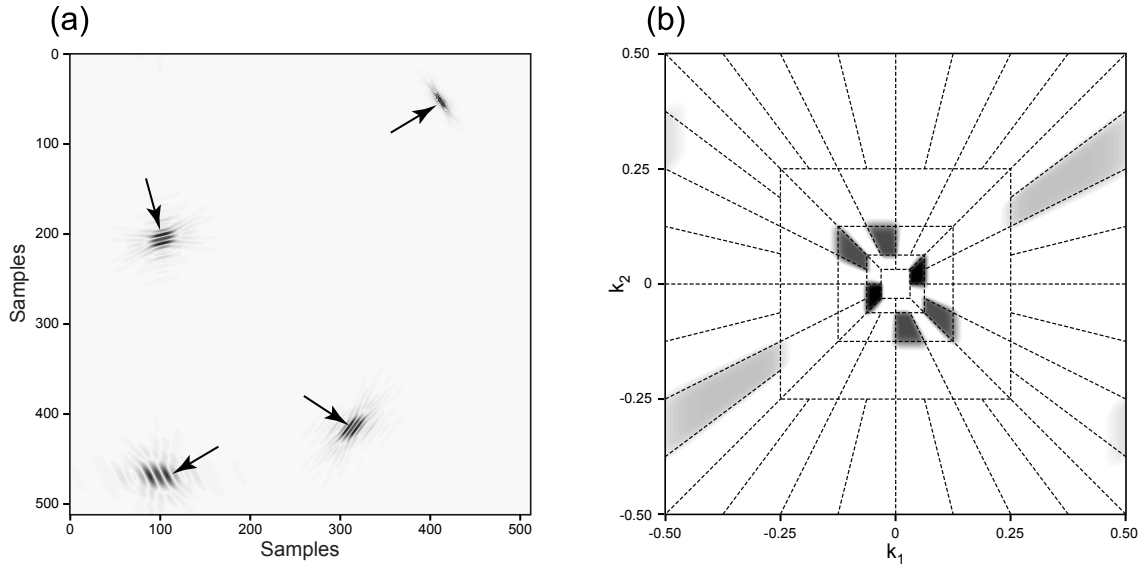
Herrmann *et.al.* –

Figure B-2: Spatial and frequency representation of curvelets. **(a)** Four different curvelets in the spatial domain at three different scales. **(b)** Dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. The position of its wedge in k-space dictates the scale and angle of a curvelet, while the phase information in k-space corresponds to its physical translation. This figure illustrates the micro-local correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientation of curvelets in the two domains. The 90° rotation is a property of the Fourier transform.

Herrmann *et.al.* –