Sparsity- and continuity-promoting seismic image recovery with curvelet frames

Felix J. Herrmann [a,*], Peyman Moghaddam [a], Christiaan C. Stolk [b]

[a] *Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada*

[b] *Department of Applied Mathematics, University of Twente, The Netherlands*

Abstract

A nonlinear singularity-preserving solution to seismic image recovery with sparseness and continuity constraints is proposed. We observe that curvelets, as a directional frame expansion, lead to sparsity of seismic images and exhibit invariance under the normal operator of the linearized imaging problem. Based on this observation we derive a method for stable recovery of the migration amplitudes from noisy data. The method corrects the amplitudes during a post-processing step after migration, such that the main additional cost is one application of the normal operator, i.e., a modeling followed by a migration. Asymptotically this normal operator corresponds to a pseudodifferential operator, for which a convenient diagonal approximation in the curvelet domain is derived, including a bound for its error and a method for the estimation of the diagonal from a compound operator consisting of discrete implementations for the scattering operator and its adjoint the migration operator. The solution is formulated as a nonlinear optimization problem where sparsity in the curvelet domain as well as continuity along the imaged reflectors are jointly promoted. To enhance sparsity, the ℓ_1-norm on the curvelet coefficients is minimized, while continuity is promoted by minimizing an anisotropic diffusion norm on the image. The performance of the recovery scheme is evaluated with a reverse-time 'wave-equation' migration code on synthetic datasets, including the complex SEG/EAGE AA$'$ salt model.

1. Introduction

This paper introduces a general-purpose algorithm for the stable recovery of the amplitudes in seismic images. The method involves the inversion in dimension d of a zero-order pseudodifferential operator (PsDO) of the type

$$(\Psi f)(x) = \int_{\mathbb{R}^d} e^{-ix \cdot \xi} a(x, \xi) \hat{f}(\xi) \, d\xi \tag{1}$$

---

* Corresponding author.
   E-mail address: fherrmann@eos.ubc.ca (F.J. Herrmann).

with

$$\hat{f}(\xi) = \frac{1}{(2\pi)^d} \int\limits_{\mathbb{R}^d} f(x) e^{ix\cdot\xi}\,\mathrm{d}x \tag{2}$$

the Fourier transform with $x, \xi \in \mathbb{R}^d$ the spatial coordinate and wavenumber vectors and $a(x,\xi)$ the symbol of the pseudodifferential operator. Under certain conditions that include no grazing rays and high-frequency asymptotics, the above linear operator describes seismic data after imaging [20,41,45]. A seismic image is derived from noisy data given by

$$d(x_s, x_r, t) = (Km)(x_s, x_r, t) + n(x_s, x_r, t) \tag{3}$$

with K the linearized Born scattering operator, $m(x)$ the model with the reflectivity and n zero-mean standard deviation σ white Gaussian noise. The seismic data volume is a function of the source and receiver positions, $x_s \in \mathbb{R}^{d-1}$ and $x_r \in \mathbb{R}^{d-1}$ and of time, $t \in \mathbb{R}_0^+$. After applying the migration operator K^T (the symbol T is reserved for the transpose) to the data volume an "image" is obtained according to

$$\begin{aligned}
(K^T d)(x) &= (K^T K m)(x) + (K^T)n(x), \\
y(x) &= (\Psi m)(x) + e(x).
\end{aligned} \tag{4}$$

This equation for the image y corresponds to the normal equation and contains contributions from the normal operator, Ψ, as well as from imaged noise e. The operator K preserves the singularities in y, i.e., $K^T K m \approx \mathrm{Id}\,m$ with Id the identity matrix and the symbol $\approx$ indicating equality in the locations of the singularities. With the increased demand for high-quality images, the above approximation of $\Psi \approx \mathrm{Id}$ is no longer justifiable because it may lead to errors in the relative amplitudes of the imaged reflectors. Our main goal is to derive a method that recovers these imaged amplitudes, i.e., estimate models m from "images" y with clutter.

1.1. Existing approaches

With the increasing demand for hydrocarbons and the increasingly hard to image (subsalt) targets, the recovery of correct seismic amplitudes has become more challenging. Approaches in the extensive literature on this topic range from least-squares migration, where the normal or Gramm matrix is inverted with Krylov subspace methods (see, e.g., [14,20,33]), to high-frequency (microlocal) methods (see, e.g., [20,41,45]), where the normal operator is inverted based on ray-asymptotic arguments, to methods that apply a diagonal scaling (see, e.g., [26,36,37] and more recently [42]) to approximately invert the Gramm operator of 'wave-equation' migration. These scaling methods either calculate the weighting analytically [36,45], based on certain assumptions regarding the acquisition and velocity model, or estimate the diagonal weighting from the action of the normal operator on some reference vector, an idea first suggest by Symes and reported in [17]. In this paper, a similar approach is followed, where the normal operator is replaced by a diagonal matrix acting on the curvelet coefficients of the image [7,27].

1.2. Our approach

The computational cost of evaluating the migration operator is $\mathcal{O}(n^{d+2})$ for a d-dimensional image with n samples in each direction. This large cost makes it a major challenge to conduct least-squares migration for $d > 2$ or large n [14,20,30,31,33,36]. We address this issue by exploiting recently developed curvelet frames. These frame expansions compress seismic images (see, e.g., [7,27] for the compression of seismic data and Fig. 1 for the compression of a typical migrated seismic image) and consist of a collection of frame elements 'curvelets' that are approximately invariant under PsDO's. These two properties allow for the development of an approach similar to the so-called wavelet–vaguelette method (WVD) (see, e.g., [10,21,32]). In this approach, scale-invariant homogeneous operators are nonlinearly inverted, using the eigenfunction-like behavior of wavelets and curvelets.

Our main contribution to earlier ideas on stable seismic image recovery (see, e.g., [28]) is threefold: First, the WVD method is extended to expanding PsDO's with respect to *redundant* curvelet frames. Second, it is observed that order-zero pseudodifferential operators with homogeneous principal symbols act approximately as a multiplicative scaling on curvelets. This property implies that such operators can be approximated by a diagonal matrix acting on the curvelet
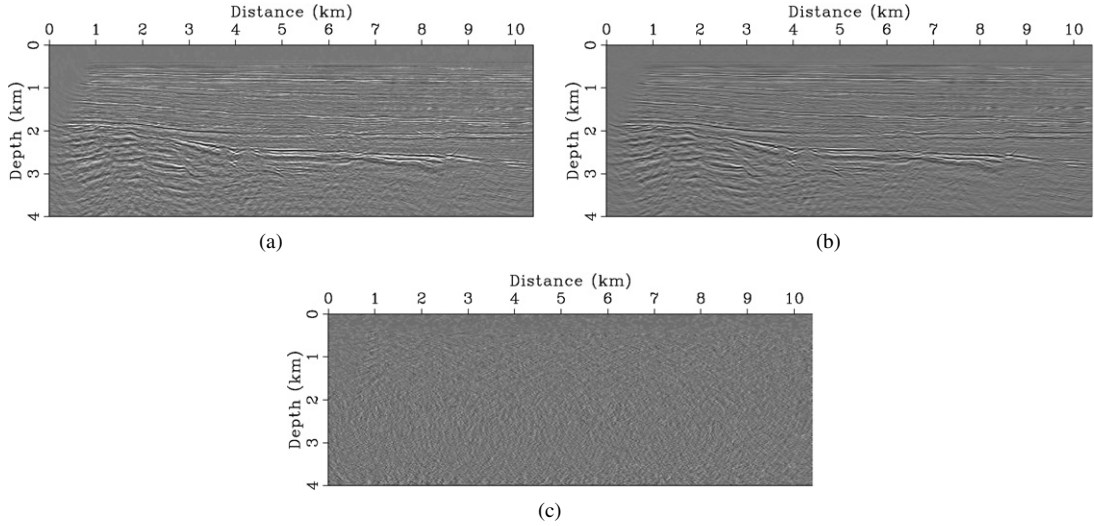
Fig. 1. Example of the recovery from different percentages of curvelet coefficients. The data set concerns a migrated image derived from the Mobil data set. (a) Near perfect recovery from $p = 99\%$ of the coefficients. (b) Recovery from only $p = 3\%$ of the coefficients. (c) The difference between near perfect recovery (a) and approximate recovery (b). This difference does not contain substantial coherent energy.

coefficients of an image, followed by an inverse curvelet transform. We refer to this procedure as a curvelet-domain approximate diagonal scaling (or weighting) that becomes more accurate for smaller scales (higher frequencies). Third, a formulation for the amplitude recovery is presented in terms of a nonlinear sparsity- and continuity-enhancing optimization problem.

After discretization, the nonlinear optimization problem for the seismic amplitude recovery (see, e.g., [13,18,48]) has the following form:

$$\mathbf{P}_\epsilon: \quad \begin{cases} \tilde{\mathbf{x}} = \min_{\mathbf{x}} J(\mathbf{x}) & \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leqslant \epsilon, \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \tilde{\mathbf{x}}. \end{cases} \tag{5}$$

During the optimization, the vector $\mathbf{x}$ is optimized with respect to the penalty functional $J(\mathbf{x})$ and the ℓ_2-data misfit. The term sparsity vector is used for $\mathbf{x}$ to stress the point that the magnitude-sorted coefficients of this vector are of rapid decay, by virtue of compression by curvelet frames.

The above optimization problem is solved for the model by finding a coefficient vector $\tilde{\mathbf{x}}$ that minimizes the penalty term subject to fitting the data to within a noise-level dependent tolerance level ϵ. The 'tilde' symbol (˜) is reserved for vectors that solve a nonlinear optimization problem. The penalty functional $J(\mathbf{x})$ is designed to promote the *sparsity* of the image in the curvelet domain as well as *continuity* along the imaged reflectors. Bold lowercase symbols refer to discretized vectors and bold uppercase symbols to discretized operators. Nonboldface symbols refer to continuous infinite-dimensional functions and operators.

Estimates for the model $\tilde{\mathbf{m}}$ are calculated by applying the pseudoinverse (denoted by the symbol †) of the analysis matrix to $\tilde{\mathbf{x}}$. This analysis matrix is given by the adjoint of the synthesis matrix ($\mathbf{A}^T$). This synthesis is given by the diagonally-weighted curvelet *synthesis* matrix with a weighting designed such that

$$\mathbf{A}\mathbf{A}^T \mathbf{r} \simeq \boldsymbol{\Psi} \mathbf{r}. \tag{6}$$

In this expression, $\mathbf{r}$ represents an appropriately chosen discrete reference vector and $\boldsymbol{\Psi}$ the discrete normal operator, formed by compounding the discrete scattering and its transpose the migration operator, i.e., $\boldsymbol{\Psi} := \mathbf{K}^T \mathbf{K}$ with the symbol := denoting 'defined as.' The symbol $\simeq$ is used to denote that this expression is approximate, a statement we will make more precise with a bound on the error.

After forming the normal operator with one's favorite numerical implementation for the migration operator and its adjoint, our amplitude recovery method consists of the following steps:
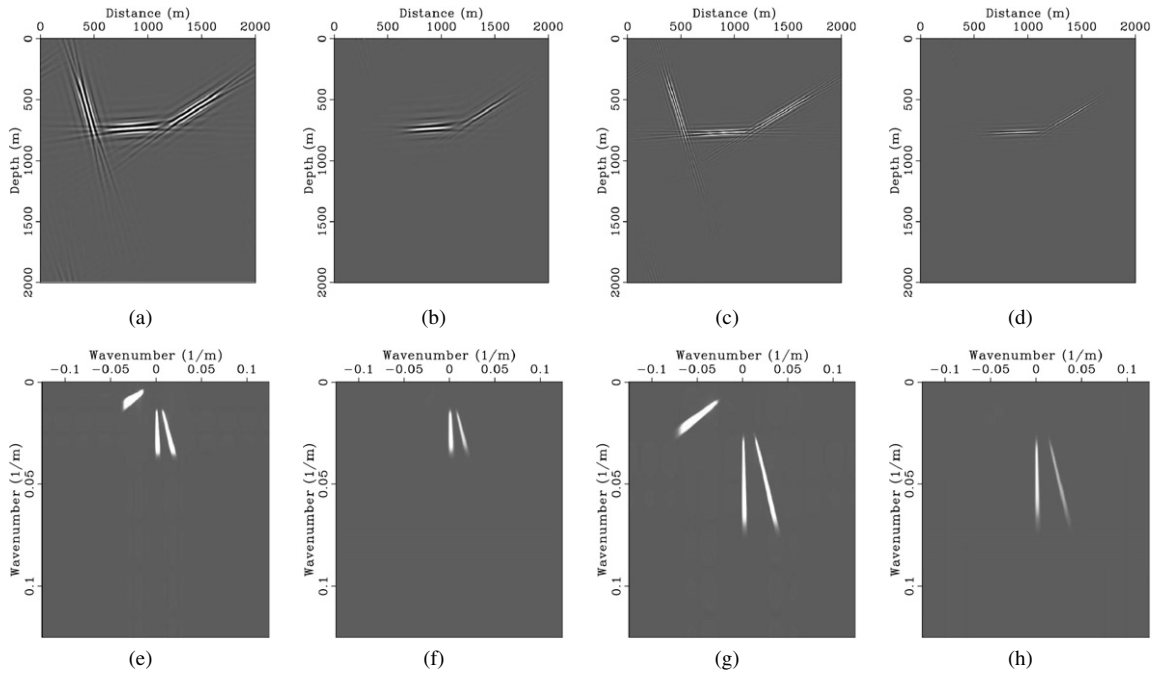
Fig. 2. Invariance of curvelets under the discretized normal operator Ψ for a smoothly varying background model (a so-called lens model see Fig. 4a). Three coarse-scale curvelets in the physical domain before (a) and after application of the normal operator (b) in the physical (a–b) and Fourier domain (e–f). The results for three fine-scale curvelets are plotted in (c–d) for the physical domain and in (g–h) for the Fourier domain. Remark: The curvelets remain close to invariant under the normal operator, a statement which becomes more accurate for finer scale which is consistent with Theorem 1. The example also shows that this statement only holds for curvelets that are in the support of the imaging operator excluding steeply dipping curvelets.

(1) Calculate an appropriate reference vector, $\mathbf{r}$, from the data that is close enough to the unknown image. Typically, this reference vector is defined by a conventional image to which a simple amplitude correction is applied valid for a constant velocity background medium;

(2) Estimate a diagonal curvelet-domain weighting that approximates the normal operator on the reference vector. This diagonal defines the synthesis matrix $\mathbf{A}$;

(3) Estimate $\tilde{\mathbf{x}}$ by solving the nonlinear optimization problem $\mathbf{P}_\epsilon$. This program inverts the synthesis matrix;

(4) Calculate the model $\tilde{\mathbf{m}}$ from the recovered coefficient vector $\tilde{\mathbf{x}}$ through the pseudoinverse of $\mathbf{A}^T$.

The proposed recovery method derives from two essential properties of curvelet frames, namely the ability of this signal representation to compress seismic images and the invariance of curvelets under the normal operator. Figures 1 and 2 are included to stress the importance of these properties by confirming that a real migrated image can accurately be recovered from only 3% of the curvelet coefficients, while curvelets remain approximately invariant under the normal operator defined for a smooth background velocity model.

1.3. Outline

First, a diagonal approximation of the normal operator in the curvelet domain is derived. The error of this approximation is bounded, using properties of the curvelet tiling of phase space. In this derivation, use is made of the property that the normal operator can be described as a zero-order pseudodifferential operator. Next, a method is proposed to estimate the diagonal weighting matrix that uses the property that the symbol of the normal operator is smooth in phase space and positive. This diagonal approximation leads to an approximate seismic image representation by a diagonally-weighted curvelet transform. With this representation, the seismic image amplitudes can be recovered by solving a nonlinear optimization problem. During this optimization problem, the image is recovered by minimizing the data mismatch while promoting sparsity in the curvelet domain and continuity along imaged reflectors. This paper is concluded by applying the algorithm to a synthetic data set derived from the SEG AA$'$ salt model.

2. Approximation of the normal operator

The primary focus of seismic imaging is to locate singularities in the Earth's elastic properties from seismic data recorded at the surface [2,4,5,19,40,45]. A seismic survey consists of multiple seismic experiments in which both the location of the sources and receivers are varied along the surface. The acquired data is subsequently used to create images of the singularities in the subsurface. The main purpose of the recovery method presented in this paper is to recover the relative amplitudes of seismic images from data that is possibly contaminated with noise. For this purpose, a diagonal approximation of the normal operator in the curvelet domain is presented. This approximation derives from the approximate invariance of curvelets under the normal operator (see Fig. 2). The approximation leads to a SVD-like decomposition for the normal operator and makes the large-scale seismic image recovery problem amenable to a solution by nonlinear optimization.

2.1. Zero-order imaging operators

In the high-frequency limit, the scattering operator and the normal operator can, under certain conditions on the medium and ray-geometry, be considered as Fourier integral operators (FIO's) [45]. In dimension two ($d = 2$), the scattering operator, K, and its adjoint the migration operator, K^T, can both be considered as FIO's of order $\frac{1}{4}$, while the leading behavior for their composition, the normal operator Ψ, corresponds to an order-one invertible elliptic PsDO [20,41,45]. To make this PsDO amenable to an approximation by curvelets, the following substitutions are made for the scattering operator and the model: $K \mapsto K(-\Delta)^{-1/2}$ and $m \mapsto (-\Delta)^{1/2}m$ with $((-\Delta)^\alpha f)^\wedge(\xi) = |\xi|^{2\alpha} \cdot \hat{f}(\xi)$. Alternatively, these operators can be made zero-order by composing the data side with a 1/2-order fractional integration along the time coordinate, i.e., $K \mapsto \partial_t^{-1/2} K$ (see, e.g., [4]). After these substitutions, the normal operator Ψ becomes zero-order. Similar substitution are made in the WVD methods where vaguelettes are introduced according to the same mappings.

Our seismic image amplitude problem is different from the recovery of images in ill-posed problems such as inverting the Radon transform [10]. In our case, the ill-posedness comes from small values of the symbol at certain positions and wave numbers and is not related to the ill-posedness stemming from the inversion of a normal operator that acts as a (fractional) inverse Laplacian. Our normal operator, without the above substitutions, acts as a (fractional) Laplacian. This behavior makes the problem ill-posed for the low-frequencies. Since seismic data and the approximations of the operators are both high-frequency, this ill-posedness is negated. The ill-posedness that is a concern are the entries in the symbol of the normal operator that correspond to regions in the model space that are badly insonified.

2.2. Curvelet frames

Curvelets are directional frames that represent a specific tiling of the two-/three-dimensional frequency plane into multiscale and multi-angular wedges (see Fig. 3). Because the directional sampling increases every-other scale doubling, curvelets become more and more anisotropic at finer and finer scales. They become 'needle-like' as illustrated in Fig. 3. Curvelets are localized in Fourier space and their smoothness in this domain leads to a rapid decay in the physical domain. Their effective support is given by ellipsoids with a width $\propto 2^{j/2}$ and length $\propto 2^j$ and an angle $\theta_{jl} = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale and l the angular index. Curvelets are indexed by the multi-index $\mu := (j, l, k) \in \mathcal{M}$ with $\mathcal{M}$ the multi-index set running over all scales, j, angles, l, and positions k (see, for details, [7,49]). Following [7], define the continuous curvelet family for $x \in \mathbb{R}^2$ as

$$\varphi_\mu(x) = 2^{3j/4}\varphi(D_j R_{\theta_{jl}} x - k), \tag{7}$$

where

- $\varphi(x)$ is a smooth bell-shaped function in the horizontal direction and oscillatory in the vertical direction;
- $D_j = \text{diag}(2^j, 2^{\lfloor j/2 \rfloor})$ is the parabolic scaling matrix;
- $R_{\theta_{jl}}$ is the rotation matrix over angles $\theta_{jl} = 2\pi l 2^{-\lfloor j/2 \rfloor}$ with $0 \leqslant \theta_{jl} < 2\pi$;
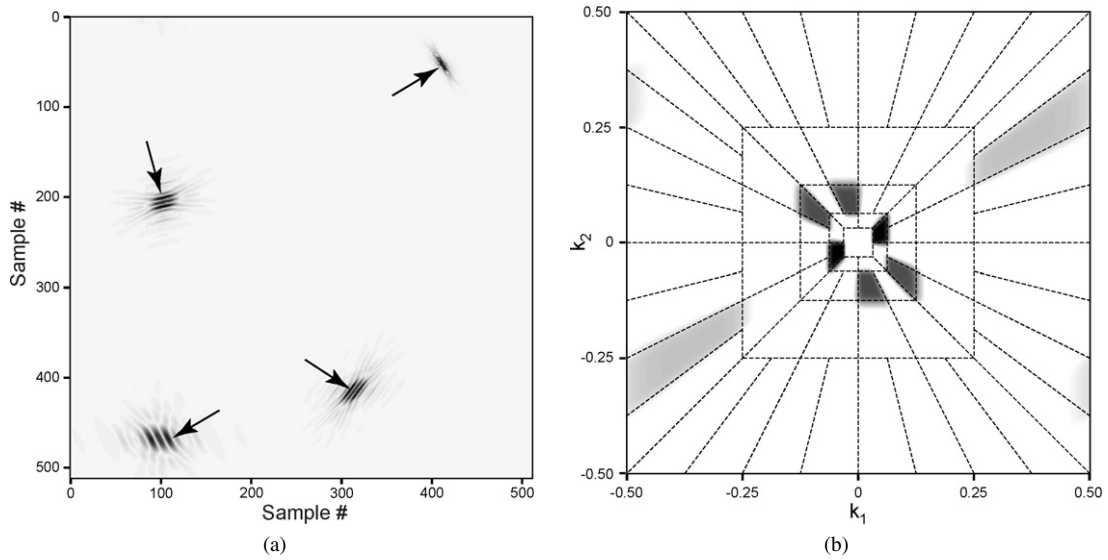- $k = (k_1, k_2) \in \mathbb{Z}^2$ the translation parameters.

Fig. 3. 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. This figure illustrates the microlocal 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.

In the frequency domain, curvelets are compactly supported and each element $\hat{\varphi}_\mu(\xi)$ is localized near the symmetric wedge

$$W_{jl} = \left\{ \pm\xi,\ 2^j \leqslant |\xi| \leqslant 2^{j+1},\ |\theta - \theta_{jl}| \leqslant \pi \cdot 2^{-\lfloor j/2 \rfloor} \right\}.$$

The number of wedges doubles every other scale doubling and hence the directional selectivity increases for finer scales (see Fig. 3). For each $\mu \in \mathcal{M}$, the curvelet coefficients are given by the inner product

$$c_\mu = \langle f, \varphi_\mu \rangle := \int_{\mathbb{R}^2} f(x)\overline{\varphi_\mu(x)}\,dx \tag{8}$$

of a function $f \in L^2(\mathbb{R}^2)$ with a curvelet $\varphi_\mu \in L^2(\mathbb{R}^2)$. Curvelets are tight frames, so we have the following reconstruction formula:

$$f = \sum_{\mu \in \mathcal{M}} c_\mu \varphi_\mu = C^T C f, \tag{9}$$

by virtue of the energy conservation

$$\sum_{\mu \in \mathcal{M}} |c_\mu|^2 = \|f\|_{L^2(\mathbb{R}^2)}, \quad \forall f \in L^2(\mathbb{R}^2), \tag{10}$$

with C and C^T the curvelet decomposition and composition, respectively. Refer to [7], for details on the construction of the fast discrete curvelet transform (FDCT) in dimension two.

2.3. Diagonal approximation of PsDO's

Data, and the scattering operator K can be so defined that the normal operator is a PsDO of order 0, which is real and selfadjoint, and has a homogeneous principal symbol $a(x,\xi)$. We will show that we can approximate Ψf in the curvelet domain by application of a diagonal matrix, if the wavenumbers in f are sufficiently high, relative to the smoothness of the symbol of Ψ.

So let $\Psi = \Psi(x, D)$ be a pseudodifferential operator of order 0, with homogeneous principal symbol $a(x,\xi)$. Assuming the operator is selfadjoint and real, the principal symbol a is also real and has an even symmetry under

the transformation $\xi \leftrightarrow -\xi$. In addition, we make the technical assumption that Ψ is compact, i.e., if f has compact support, then Ψf also has compact support.

Curvelets in $\mathbb{R}^2$ are denoted by φ_μ and we define $|\mu| = j$. The scale j ranges from some positive number $j_{\min}$, to infinity (or $j_{\max}$ in an implementation). The central position and wave vector for the curvelet will be denoted by (x_μ, ξ_μ), we let $\theta_\mu = \xi_\mu / \|\xi_\mu\|$ be the angle of the curvelet.

Next, we give some consequences of the localization of the curvelet in the space and Fourier domains. The support of a curvelet in the Fourier domain is contained in one of the domains $\pm\xi_2 > \epsilon|\xi_1|$, or $\pm\xi_1 > \epsilon|\xi_2|$, for some small ε, say we are in the first case, then ξ_2 can be used as the radial coordinate, with ξ_1/ξ_2 the angular coordinate in the Fourier domain. From the localization properties w.r.t. the angular coordinate, it follows that $(\xi_1/\xi_2 - \xi_{\mu,1}/\xi_{\mu,2})$ is bounded by $C_1 2^{-\lfloor j/2 \rfloor}$ on the support of the curvelet and with C_1 some finite positive constant (from here on we assume that any constant C_i is finite and positive), so that

$$\left\| (\xi_1/\xi_2 - \xi_{\mu,1}/\xi_{\mu,2}) \hat{\varphi}_\mu \right\|_{L^2(\mathbb{R}^2)} \leqslant C_1 2^{-\lfloor j/2 \rfloor}. \tag{11}$$

Localization in the space domain follows from the smoothness of the defining functions in the Fourier domain, in the radial direction we have

$$\left\| \theta_\mu \cdot (x - x_\mu) \varphi_\mu \right\|_{L^2(\mathbb{R}^2)} \leqslant C_2 2^{-j}$$

with C_2 another finite constant. For the other directions we have a weaker estimate, as the curvelet is in the Fourier domain narrower, and therefore in the space domain wider in the directions normal to θ_μ, i.e.,

$$\left\| (x - x_\mu) \varphi_\mu \right\|_{L^2(\mathbb{R}^2)} \leqslant C_3 2^{-\lfloor j/2 \rfloor}.$$

In fact parallel results hold for curvelets in $\mathbb{R}^d$.

We now compare the application of a PsDO to a curvelet with simply multiplying by the value of the symbol at (x_μ, ξ_μ). The result will be proved for x in $\mathbb{R}^d$.

Lemma 1. *Suppose a is in the symbol class $S_{1,0}^0$, then, with C' some constant, the following holds*:

$$\left\| \left(\Psi(x, D) - a(x_\nu, \xi_\nu) \right) \varphi_\nu \right\|_{L^2(\mathbb{R}^n)} \leqslant C' 2^{-|\nu|/2}. \tag{12}$$

To approximate Ψ, we define the sequence $\mathbf{u} := (u_\mu)_{\mu \in \mathcal{M}} = a(x_\mu, \xi_\mu)$. Let $\mathbf{D}_\Psi$ be the diagonal matrix with entries given by $\mathbf{u}$. Next we state our result on the approximation of Ψ by $C^T \mathbf{D}_\Psi C$.

Theorem 1. *The following estimate for the error holds*:

$$\left\| \left(\Psi(x, D) - C^T \mathbf{D}_\Psi C \right) \varphi_\mu \right\|_{L^2(\mathbb{R}^n)} \leqslant C'' 2^{-|\mu|/2}, \tag{13}$$

where C'' is a constant depending on Ψ.

This main result is proved in Appendix A and it shows that the approximation error for the diagonal approximation decreases for increasingly finer scales, j. The approximation derives from the property that the symbol is slowly varying over the support of a curvelet when the velocity model is sufficiently smooth.

2.4. Decomposition of the normal operator

By virtue of Theorem 1, the normal operator can approximately be factorized into

$$(\Psi \varphi_\mu)(x) \simeq \left(C^T \mathbf{D}_\Psi C \varphi_\mu \right)(x) = \left(A A^T \varphi_\mu \right)(x) \tag{14}$$

with $A := C^T \sqrt{\mathbf{D}_\Psi}$ and $A^T := \sqrt{\mathbf{D}_\Psi} C$. Because the seismic reflectivity can be written as a superposition of curvelets, the φ_μ can be replaced with the model m. The normal equation (cf. Eq. (4)) can now be rewritten in the following approximate form:

$$y(x) = (\Psi m)(x) + e(x) \simeq \left(A A^T m \right)(x) + e(x) = A x_0 + e. \tag{15}$$

The symbol $\simeq$ in these expressions is used to indicate that this identity is approximate with an error bounded by Eq. (13) for some constant. This approximation for the forward model forms the point of departure for our seismic amplitude recovery algorithm.

To illustrate the above results, a numerical approximation of a PsDO is applied to a number of curvelets at different location, angles and scales. This PsDO corresponds to the normal operator associated with a seismic experiment for a velocity model that contains a low-velocity lens (see Fig. 4a). The results of these experiments are summarized in Fig. 2. Figures 2b and 2d display the results of applying the PsDO to three different coarse- and fine-scale curvelets (Figs. 2a and 2c). The Fourier-domain images are plotted in Figs. 2f and 2h. Comparing the imaged curvelets with their originals, shows that curvelets remain invariant as long as they lie in the support of the normal operator. Steep dipping curvelets are outside the support of the operator and this explains the deterioration of their amplitudes for steep dips. As predicted by Theorem 1, the normal operator becomes more diagonal for increasingly finer scales, an observation reflected in the behavior of the coarse- *versus* the fine-scale curvelets. Not only is this observation consistent with the microlocal correspondence of curvelets reported by [10] but it is also consistent with the fact that the PsDO's correspond to high-frequency approximations of the normal operator.

2.5. Estimation of the diagonal

2.5.1. The discrete normal operator

Before proceeding with the formulation of the seismic image recovery problem, a method is introduced to estimate the diagonal entries of $\mathbf{D}_\Psi$. These entries are calculated from applying a discrete implementation of the normal operator to some appropriately chosen reference vector. This discrete normal operator is given by the composition of the discretized scattering operator that relates the discretized reflectivity to the discretized data and its adjoint the migration operator. This compound operator corresponds to a matrix-free implementation for $\boldsymbol{\Psi} := \mathbf{K}^T \mathbf{K} \in \mathbb{R}^{M \times M}$ with $\mathbf{K} \in \mathbb{R}^{L \times M}$ the scattering matrix and L and M the length of the data and image vectors.

2.5.2. The discrete curvelet transform

For the numerical implementation of the curvelet transform, the fast discrete curvelet transform via wrapping [7] is used, which corresponds to a matrix-free implementation of the tall curvelet decomposition matrix $\mathbf{C} \in \mathbb{R}^{N \times M}$ with $N = \#\{\mathcal{M}\} \gg M$ with the symbol $\#$ denoting 'number of.' This transform yields a redundant coefficient vector $\mathbf{c} = \mathbf{Cr}$ with $\mathbf{c} \in \mathbb{R}^N$. For this choice of curvelet transform, the pseudoinverse equals the transpose, i.e., $\mathbf{C}^T \mathbf{c} = \mathbf{C}^\dagger \mathbf{c}$. The transform is a numerical isometry that preserves energy, i.e., $\|\mathbf{r}\| = \|\mathbf{Cr}\|$, so we have $\mathbf{C}^T \mathbf{Cr} = \mathrm{Id}\,\mathbf{r}$ in the ℓ_2 sense. 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., $\mathbf{CC}^T \mathbf{r}$ is a projection. Because $\mathbf{CC}^T$ is a projection, not every curvelet vector is the forward transform of some discrete vector $\mathbf{f}$.

2.5.3. Nonuniqueness

The estimation of the diagonal $\mathbf{D}_\Psi$ involves the calculation of a diagonal curvelet-domain weighting vector that approximates the action of the normal operator on an appropriately chosen reference vector $\mathbf{r}$, i.e., $\mathbf{C}^T \mathbf{D}_\Psi \mathbf{Cr} \simeq \boldsymbol{\Psi}\mathbf{r}$ (cf. Eq. (13)). This estimation problem can be formulated in terms of a least-squares minimization problem. Define the 'data' vector as $\mathbf{b} = \boldsymbol{\Psi}\mathbf{r}$, i.e., the migrated demigrated reference vector. Let $\mathbf{D}_\Psi = \mathrm{diag}(\mathbf{u})$, with $\mathbf{u} \in \mathbb{R}^N$, be the diagonal curvelet-domain weighting matrix with coefficients $(u_\mu)_{\mu \in \mathcal{M}}$ on its diagonal and calculate the curvelet transform of the reference vector $\mathbf{v} = \mathbf{Cr}$. The diagonal estimation problem can now be formulated as finding the vector $\mathbf{u}$ that minimizes $\|\mathbf{b} - \mathbf{Pu}\|_2$ with $\mathbf{P} := \mathbf{C}^T \mathrm{diag}(\mathbf{v})$ and $\mathbf{b} \in \mathbb{R}^M$ the known demigrated-migrated reference vector. Because of the redundancy of the curvelet transform ($M \ll N$), there are unfortunately many possible solutions to this minimization problem, which gives rise to an undesired nonuniqueness.

2.5.4. Regularization via smoothness in phase space

One of the consequences of the nonuniqueness is the existence of negative entries in the estimated diagonal. These entries are inconsistent with the fact that $\boldsymbol{\Psi}$ is positive definite. Therefore, a regularization is needed that leads to an estimate for the diagonal with positive entries only. We argue that this can be accomplished by imposing smoothness in phase space. This smoothness is a property of the symbol that describes the behavior of the normal operator. In the curvelet domain, this smoothness translates to a smoothness in the amplitudes amongst neighboring points on

Table 1
Algorithm for the estimation of the diagonal via regularized least-squares inversion. The Lagrange multiplier, η, is increased up to the point that all entries in the vector for the diagonal are positive

---

Calculate: $\mathbf{b} = \boldsymbol{\Psi}\mathbf{r}$ and $\mathbf{v} = \mathbf{C}\mathbf{r}$
Set: $\eta = \eta_{\min}$
while $\exists(\tilde{u}_\mu)_{\mu\in\mathcal{M}} < 0$ *do*
   Solve
   $\tilde{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{b} - \mathbf{P}\mathbf{u}\|_2^2 + \eta^2\|\mathbf{L}\mathbf{u}\|_2^2$
   Increase the Lagrange multiplier
   $\lambda = \eta + \Delta\eta$
end while

---

the curvelet grid. For each entry in the diagonal, this smoothness can be imposed by including an additional penalty term in the least-squares formulation for the estimation of the diagonal. This damped least-squares problem can be formulated as

$$\tilde{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{b} - \mathbf{P}\mathbf{u}\|_2^2 + \eta^2\|\mathbf{L}\mathbf{u}\|_2^2, \tag{16}$$

where $\mathbf{L} = [\mathbf{D}_1^T \ \mathbf{D}_2^T \ \mathbf{D}_\theta^T]^T$ is a so-called sharpening operator, penalizing fluctuations amongst neighboring coefficients in $\mathbf{u}$. $\mathbf{D}_{1,2}$ contain first-order differences at each scale in the $x_{1,2}$ directions, i.e., $\mathbf{D}_{1,2} = [\mathbf{D}_{1,2}^{j_{\min}} \ \cdots \ \mathbf{D}_{1,2}^{j_{\max}}]$ with $\mathbf{D}_{1,2}^j$ the differences at the jth scale in the spatial directions, and $\mathbf{D}_\theta$ the first-order differences in the θ direction, i.e., $\mathbf{D}_\theta = [\mathbf{D}_\theta^{j_{\min}} \ \cdots \ \mathbf{D}_\theta^{j_{\max}}]$ with $\mathbf{D}_\theta^j$ the difference in the angle direction at the jth scale. These differences are scale dependent because the curvelet grid changes for each scale. The emphasis on the penalty term with respect to the misfit is controlled by η.

2.5.5. Estimation with positivity constraints

To ensure a correct balance between the misfit and the penalty functional, Eq. (16) is solved for a series of increasing Lagrange multipliers η. For each η, the following system of equations is inverted:

$$\begin{bmatrix} \mathbf{P} \\ \eta\mathbf{L} \end{bmatrix} \mathbf{u} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \tag{17}$$

with $\mathbf{0}$ a column vector with N zero entries. For a given reference vector, the diagonal estimation procedure consists of the following steps. Apply the normal operator to the reference vector and calculate its curvelet coefficients. Set the Lagrange multiplier to $\eta_{\min}$ and solve Eq. (16) by inverting the system of equations in Eq. (17). Subsequently, increase η by $\Delta\eta$ and repeat this procedure until all diagonal elements of $\tilde{\mathbf{u}}$ are nonnegative. Even though we do not have a proof of convergence for this method, increasing the η leads to positive entries for large enough η. The steps of this estimation procedure are summarized in Table 1.

Example 1. The procedure outlined above is tested on a stylized example with a velocity model given by a low-velocity lens (Fig. 4a) and a bandwidth-limited reflectivity (Fig. 4b), consisting of three events. The low-velocity zone is representative of a gas lens in the overburden. Data is generated from this reflectivity with a zero-order linearized Born modeling and consists of 32 shot records with 500 receivers each. For more details on the migration code, refer to [43] and to the numerical experiment section at the end of this paper. The data is subsequently imaged (Fig. 4c). This image and the bandwidth-limited reflectivity (Fig. 4d) define the 'data' vector $\mathbf{b}$ (cf. Eq. (17)) and the reference vector $\mathbf{r}$ used to estimate $\tilde{\mathbf{u}}$ according to the algorithm presented in Table 1. Fig. 5 displays the curvelet vectors for increasing $\eta = \{0.01, 0.1, 1, 10\}$. These plots clearly illustrate that $\tilde{\mathbf{u}}$ becomes positive for η large enough ($\eta \geqslant 1$ in this case). Figure 4d contains the diagonal approximation of the normal operator for the diagonal estimated with $\eta = 1$. Comparison between the results of applying the normal operator (Fig. 4c) and its diagonal approximation (Fig. 4d), shows a nice correspondence. The relative ℓ_2-error between the actual image and its approximation is 6.1%. This value corresponds to a moderate size for the constant C'' in the bound for the approximation error (Theorem 1).
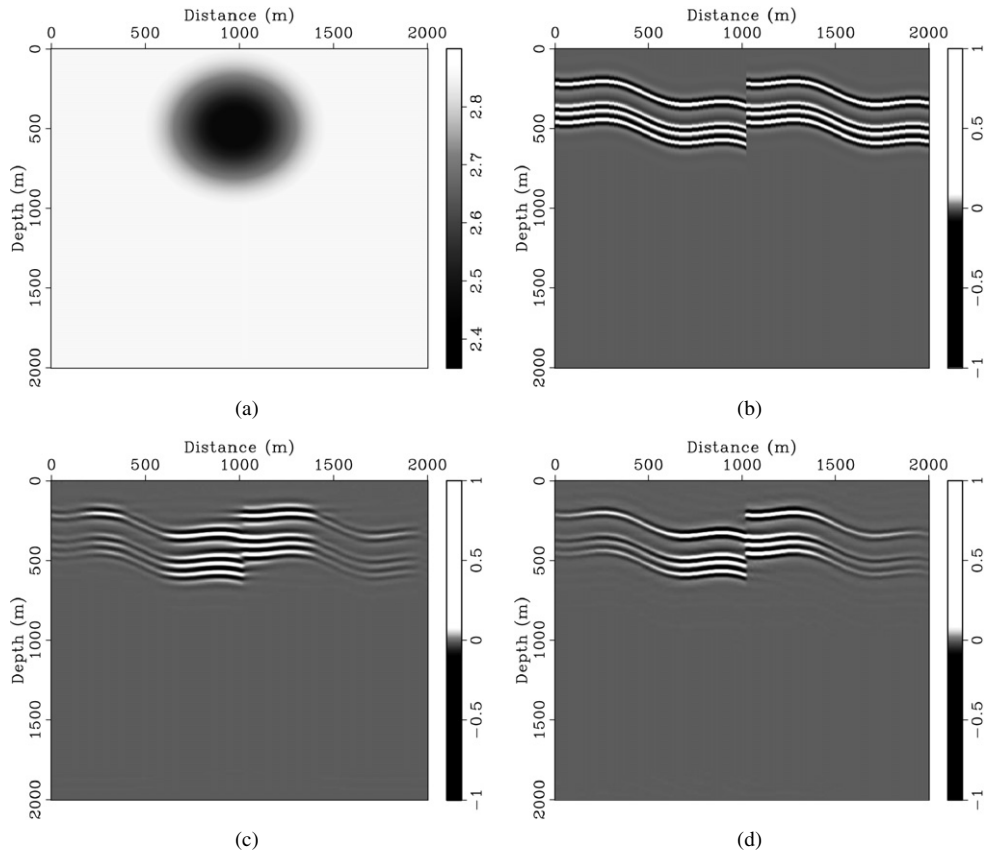
Fig. 4. Stylized example of the diagonal estimation for a smooth velocity model and a reflectivity consisting of three reflection events including a fault. (a) The smooth background model with the low-velocity lens. This velocity model is used to calculate the linearized scattering and migration operators. (b) The bandwidth limited reflectivity. (c) The imaged reflectivity from data consisting of 32 shot records with 500 traces each. (d) The approximated normal operator for $\mathrm{diag}(\tilde{\mathbf{u}})$ estimated with $\eta = 1$. The bandwidth limited reflectivity in (b) and the image in (c) served as input for the reference and 'data' vectors for the diagonal estimation procedure outlined in Table 1. Remarks: the main dimming of the normal operator is captured by the diagonal approximation. The relative ℓ_2-error between the actual and the approximate normal operator is 6.1%.

3. Stable seismic image recovery

Our formulation for the seismic amplitude recovery problem banks on two fundamental properties of curvelets, namely their ability to *detect wave-fronts* and their *invariance* under the normal operator. The combination of these two properties allow for a robust solution of the recovery problem in terms of a nonlinear optimization problem. In this section, the different steps that lead to this formulation are discussed starting with an argumentation why curvelets compress seismic images, followed by a quick review of ℓ_1-promoted image recovery, its extension to the seismic situation and its solution by a cooling method.

3.1. Curvelet frames for seismic images

The sedimentary crust consists of sheet-like layers that correspond to functions with singularities along piece-wise smooth curves. These singularities are associated with rapid variations in the medium fluctuations at which seismic waves reflect. During seismic imaging, these singularities are recovered. Examples of singularities imaged from real seismic data can be found in Fig. 1a. This image can be considered as a bandwidth limited version of a function with singularities on piece-wise C^2 curves that may include faults and pinch outs (see, e.g., Figs. 4b and 8a that are examples of synthetic reflectivity models). Fourier (and also wavelet) transforms are known to perform poorly on this type of functions [9,12]. This poor performance can be explained by the inability of the Fourier transform to localize and the lack of directivity of wavelets. Only when oriented perpendicular to the interfaces, wavelets decay rapidly. Since
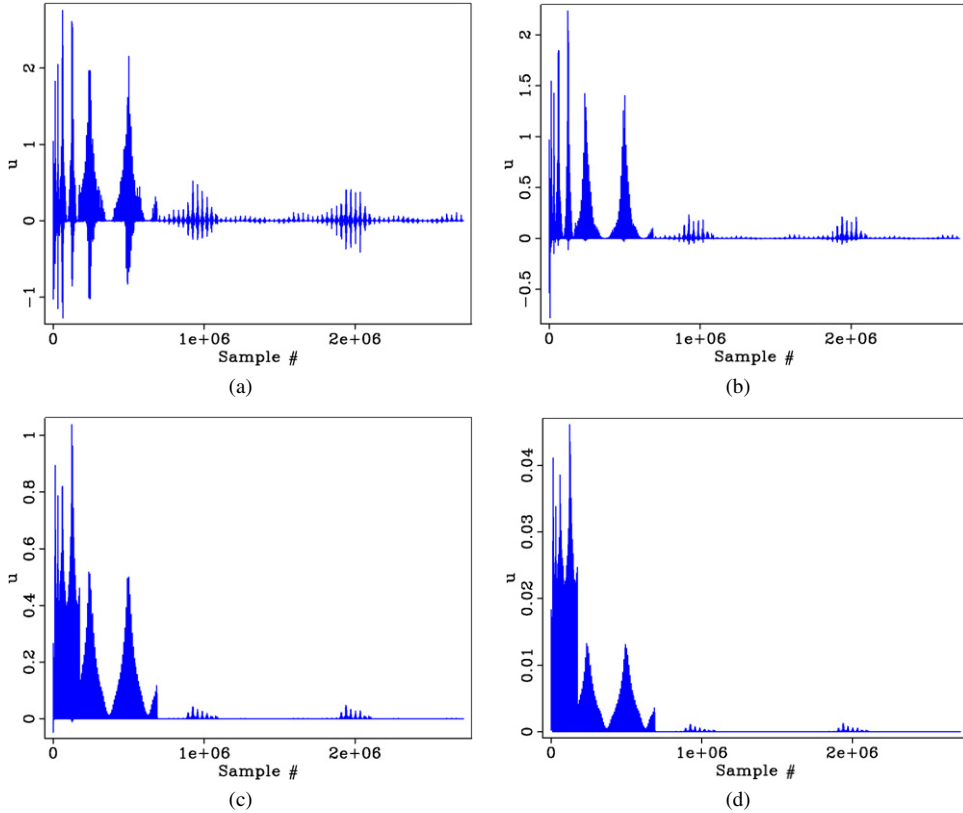
(a)

(b)

(c)

(d)

Fig. 5. Estimates for the diagonal $\tilde{\mathbf{u}}$ are plotted in (a–d) for increasing $\eta = \{0.01, 0.1, 1, 10\}$. The diagonal is estimated according the procedure outlined in Table 1 with the reference and 'data' vectors, $\mathbf{v}$ and $\mathbf{b}$, plotted in Figs. 4b and 4c. As expected the diagonal becomes more positive for increasing η.

wavelets lack directionality, they cannot resolve these directions, known as the wavefront set. This inability explains why wavelets do not significantly improve the compression rate for seismic images. Curvelet frames, on the other hand, are designed to detect the wavefront set [6,8,9,11] and lead, as shown in Fig. 6, to better empirical compression rates for real and synthetic seismic images. This improved compression rate explains the accurate reconstruction in Fig. 1a of a real seismic image from only 3% of the curvelet coefficients and is consistent with the theoretical non-linear approximations rates reported in the literature (see, e.g., [9]). These rates correspond to a rate of $\mathcal{O}(P^{-2})$ for curvelets, with P the number of largest entries used in the approximation, while Fourier and wavelets only attain $\mathcal{O}(P^{-1/2})$ and $\mathcal{O}(P^{-1})$, respectively. The empirical rates plotted in Fig. 6 seem to confirm these theoretical findings.

3.2. Stable recovery

Because of the redundancy of the curvelet transform, $\mathbf{C}\mathbf{C}^T$ is a projection and this means that not every curvelet vector is a forward transform of a vector $\mathbf{f}$. Therefore, the vector $\mathbf{x}_0$ cannot readily be calculated from $\mathbf{f} = \mathbf{C}^T\mathbf{x}_0$, because there exist infinitely many coefficient vectors whose inverse transform equals $\mathbf{f}$. Recent work on in the field of compressive sensing has shown that rectangular matrices such as the curvelet matrix, $\mathbf{C} \in \mathbb{R}^{N \times M}$ with $N \gg M$, can stably be inverted through a nonlinear sparsity promoting optimization program. These nonlinear recovery methods require fast decay for the magnitude sorted curvelet coefficients (for detail refer to the extensive literature on stable recovery also known as compressed sensing; see, e.g., [13,23,39]).

Following these results, the vector $\mathbf{x}_0$ can be recovered from noise-corrupted data

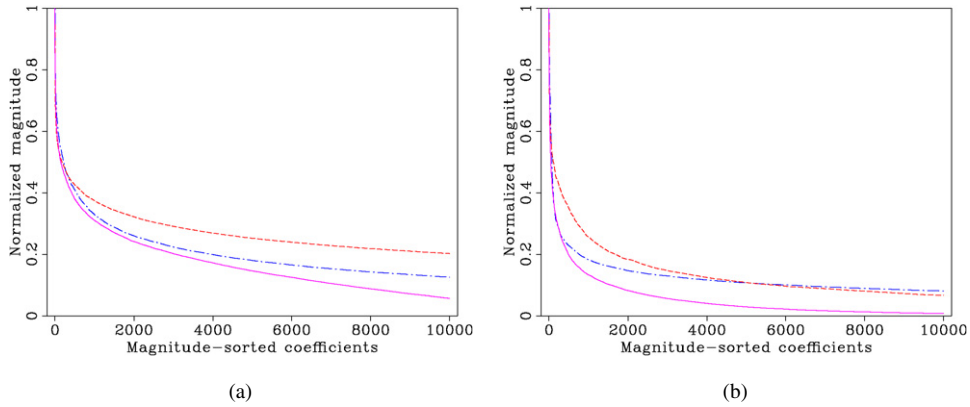$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \tag{18}$$

Fig. 6. Decay rate for the curvelet coefficients compared to the decay for the coefficients of Fourier (dashed line) and discrete wavelet transforms (dot-dashed line). (a) Decay rate for the sorted curvelet coefficients (solid line) of the real migrated image included in Fig. 1a. (b) The same as (a) but now for the synthetic reflectivity of the SEG/EAGE AA′ salt model (cf. Fig. 8a). The decay for the curvelet transform clearly compares favorably to these other two transforms.

with $\mathbf{A} := \mathbf{C}^T$. As long as the signal is sufficiently compressible $\mathbf{x}_0$ can successfully be recovered to within the noise level. This recovery involves the solution of a constrained convex optimization problem

$$\mathbf{P}_1: \quad \begin{cases} \tilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 := \sum_{j=1}^{N} |x_j| & \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leqslant \epsilon, \\ \tilde{\mathbf{m}} = \mathbf{A}\mathbf{x}. \end{cases} \tag{19}$$

For certain matrices $\mathbf{A}$, the solution of this unconstrained optimization problem lies to within the noise level (see, e.g., [13,23]). The optimization problem $\mathbf{P}_1$ is known as the constrained variation of the LASSO [46] and the basis-pursuit denoising (BPDN) [15] algorithms.

As part of the optimization, the sparsity vector is fitted within the tolerance ϵ. This tolerance depends on the noise level given by the standard deviation of the noise vector. Since $n_{1...M} \in N(0, \sigma^2)$, the probability of $\|\mathbf{n}\|_2^2$ exceeding its mean by plus or minus two standard deviations is small. The $\|\mathbf{n}\|_2^2$ is distributed according the χ^2-distribution with mean $M \cdot \sigma^2$ and standard deviation $\sqrt{2M} \cdot \sigma^2$. By choosing $\epsilon^2 = \sigma^2(M + \nu\sqrt{2M})$ with $\nu = 2$, we remain within the mean plus or minus two standard deviations. Following [23], the above constrained optimization problem ($\mathbf{P}_1$), is replaced by a series of simpler unconstrained optimization problems

$$\mathbf{P}_\lambda: \quad \begin{cases} \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \\ \tilde{\mathbf{m}} = \mathbf{A}\tilde{\mathbf{x}}. \end{cases} \tag{20}$$

These optimization problems depend on the Lagrange multiplier λ. A cooling method is used, where $\mathbf{P}_\lambda$ is solved for a Lagrange multiplier λ that is slowly decreased from a large starting value $\lambda_1 < \|\mathbf{A}^T\mathbf{y}\|_\infty$. The optimal $\tilde{\mathbf{x}}$ is found for the largest λ for which $\|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2 \leqslant \epsilon$. During the optimization, the underdetermined frame matrix $\mathbf{A}$ is inverted by imposing the sparsity promoting ℓ_1-norm. This norm regularizes the inverse problem of finding the unknown coefficient vector (see also [18]). We refer to [22,47] for the recovery conditions for Eqs. (19) and (20).

3.3. Solution by iterative thresholding

Following [18,23] and ideas dating back to [25], Eq. (19) is solved by an iterative thresholding technique that derives from the Landweber descent method. The method consist of an outer loop during which the Lagrange multiplier is lowered and an inner loop during which $\mathbf{P}_\lambda$ is solved. The inner loop is initialized by the solution obtained from the previous outer loop starting with zero vector for the first iteration. After m iterations of the outer cooling loop, the estimated coefficient vector is computed for fixed λ by the following inner loop:

$$\mathbf{x}^{m+1} = T_\lambda\big(\mathbf{x}^m + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^m)\big) \tag{21}$$

with $\lambda = \lambda_m$ and

$$T_\lambda(x) := \text{sgn}(x) \cdot \max\big(0, |x| - |\lambda|\big). \tag{22}$$

As shown by [18], this iteration for fixed λ converges to the solution of Eq. (20) for m large enough and $\|\mathbf{A}\| < 1$. The cost for each iteration is a curvelet synthesis and subsequent analysis.

3.4. Stable seismic recovery

The main purpose is to estimate the relative amplitudes of seismic images from data that is possibly contaminated with noise. This data is represented by the discretized linearized forward model

$$\mathbf{d} = \mathbf{Km} + \mathbf{n} \tag{23}$$

with $\mathbf{K}$ the discrete linearized Born modeling operator. Applying the adjoint of this operator to the data vector creates a noisy image (cf. Eq. (4))

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{m} + \mathbf{e}. \tag{24}$$

With the curvelet decomposition of the normal operator (cf. Eq. (15)) in discrete form, i.e.,

$$\boldsymbol{\Psi}\mathbf{m} \simeq \mathbf{A}\mathbf{A}^T\mathbf{m}, \tag{25}$$

this expression can be rewritten into the following approximate form:

$$\mathbf{y} \simeq \mathbf{A}\mathbf{x}_0 + \mathbf{e} \tag{26}$$

with $\mathbf{A} := \mathbf{C}^T\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma} := \sqrt{\mathbf{D}_\Psi}$. To avoid instabilities related to small entries in the estimated diagonal $\mathbf{u}$, we set $\mathbf{D}_\Psi = \mathrm{diag}\,(\tilde{\mathbf{u}} + \delta)/\delta$ with δ some small parameter. Comparing this image representation with the one in Eq. (18) shows that these expressions are equivalent, aside from the noise term. While the noise in Eq. (18) is white and Gaussian, the noise term in Eq. (26) is colored by the migration operator. The expectation for the covariance of this colored noise term equals $\mathbf{E}\{\mathbf{e}\mathbf{e}^T\} = \sigma^2\boldsymbol{\Psi} \simeq \sigma^2\mathbf{A}\mathbf{A}^T$. Because the nonlinear recovery of the vector $\mathbf{x}_0$ involves the inversion of the matrix $\mathbf{A}$ in which the covariance is factored, the noise term is approximately whitened. To understand this whitening, consider the pseudoinverse $\mathbf{A}^\dagger = \boldsymbol{\Gamma}^{-1}\mathbf{C}$ which, when applied to $\mathbf{e}$, approximately whitens the noise in the curvelet domain. This whitening is a well-known property of WVD techniques.

The nonlinear recovery of seismic images now involves solving

$$\mathbf{P}_1': \quad \begin{cases} \tilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 := \sum_{j=1}^N |x_j| & \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leqslant \epsilon, \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger\tilde{\mathbf{x}}. \end{cases} \tag{27}$$

The main assumption underlying this formulation is that the curvelet vector of the model remains (approximately) sparse after application of the normal operator. This assumption is valid when the curvelets remain sufficiently invariant under this operator. In that case, the nonlinear program $(\mathbf{P}_1')$ approximately inverts the normal operator in two stages, namely during the ℓ_1-norm regularized inversion of the synthesis matrix, yielding a sparse estimate for the curvelet coefficient vector, and during the calculation of the model vector *via* the pseudoinverse of $\mathbf{A}^T$. During each stage, the 'square-root' of the normal operator is approximately inverted.

4. Sparsity- and continuity-enhanced seismic image recovery

The above formulation provides a stable recovery for the amplitudes of seismic images, the relative strengths of the fluctuations in $\mathbf{m}$, without the necessity of multiple evaluations of the normal operator. This recovery is only accurate when the proposed diagonal approximation (cf. Eq. (25)) provides an adequate approximation to the normal operator. The accuracy of this approximation depends on the available scales in the seismic image (cf. Theorem 1), on the complexity of the background velocity model, the acquisition geometry and the closeness of the reference $\mathbf{r}$ to the actual unknown model $\mathbf{m}$. For the remainder of this paper, we assume the acquisition to be ideal and the reference vector to be close to the actual model.

4.1. Anisotropic diffusion

Aside from the aforementioned factors that enter into the seismic recovery problem, there is the issue of how to limit spurious curvelet artifacts. These artifacts are either related to the so-called "pseudo-Gibb's" phenomenon (or

better side-band effects (see [12])) inherent to curvelet or other harmonic expansions. To reduce these spurious artifacts, the sparsity enhancing penalty functional in Eq. (27) is complemented with an additional continuity-enhancing penalty functional. This functional enhances the continuity along the imaged reflectors, which tend to be smooth in the tangential directions. By applying an anisotropic smoothing technique, the 'wavefront' set of the imaged reflectivity is preserved. This technique differs from commonly used edge-preserving penalty functionals such as total variation (TV) (see, e.g., [16,38]) that tend to remove the oscillatory behavior of the reflectors in the normal direction.

The anisotropic-diffusion penalty term (see, e.g., [24]) is given by

$$J_c(\mathbf{m}) = \left\| \boldsymbol{\Lambda}^{1/2} \boldsymbol{\nabla} \mathbf{m} \right\|_2^2, \tag{28}$$

with $\boldsymbol{\nabla}$ the discretized gradient matrix defined as $\boldsymbol{\nabla} = [\mathbf{D}_1^T \ \mathbf{D}_2^T]^T$. The block-diagonal matrix $\boldsymbol{\Lambda}$ is location dependent (see Fig. 10, which plots the gradients) and rotates the gradient towards the tangents of the reflecting surfaces. This rotation matrix is given by

$$\boldsymbol{\Lambda}[\mathbf{r}] = \frac{1}{\|\boldsymbol{\nabla}\mathbf{r}\|_2^2 + 2\upsilon} \left\{ \begin{pmatrix} +\mathbf{D}_2\mathbf{r} \\ -\mathbf{D}_1\mathbf{r} \end{pmatrix} (+\mathbf{D}_2\mathbf{r} \quad -\mathbf{D}_1\mathbf{r}) + \upsilon \, \mathbf{Id} \right\}, \tag{29}$$

with $\mathbf{D}_i$ the discretized derivative in the ith coordinate direction and υ a parameter that controls the fluctuations for regions where the gradient is small. Following [3], this control parameter is set proportional to the median of $|\boldsymbol{\nabla}\mathbf{r}|$ with $|\ |$ the length of each gradient vector (white arrows in Fig. 10). Similar to the diagonal approximation, a reference vector derived from the migrated image (cf. Eq. (24)) is used to calculate the tangential directions of the reflecting surfaces.

By combining the two different penalty terms that promote sparsity and continuity, we finally arrive at our formulation for the seismic-amplitude recovery problem

$$\mathbf{P}_\epsilon: \quad \begin{cases} \min_\mathbf{x} J(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leqslant \epsilon, \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \mathbf{x}, \end{cases} \tag{30}$$

in which the composite penalty term $J(\mathbf{x})$ is given by

$$J(\mathbf{x}) = \alpha J_s(\mathbf{x}) + \beta J_c(\mathbf{x}), \tag{31}$$

with $\alpha, \beta \geqslant 0$ and $\alpha + \beta = 1$. The $J_s(\mathbf{x}) = \|\mathbf{x}\|_1$ is the ℓ_1-norm. The second term in the penalty term is given by $J_c(\mathbf{x}) = \|\boldsymbol{\Lambda}^{1/2} \boldsymbol{\nabla} (\mathbf{A}^T)^\dagger \mathbf{x}\|_2^2$. Because the optimization is carried out over $\mathbf{x}$ and not over the model vector $\mathbf{m}$, this expression includes a pseudo-inverse that is calculated with a few iterations of the LSQR algorithm [35].

4.1.1. The cooling method

The above nonlinear optimization problem ($\mathbf{P}_\epsilon$) is solved with a cooling method (see, e.g., [39]). This method consists of a series of thresholded Landweber iterations that solve a series of unconstrained subproblems for decreasing λ. Since this method only requires knowledge of the Jacobians at each iteration, it is relatively straightforward to include the Jacobian of the additional continuity-enhancing penalty functional $J_c(\mathbf{x})$. For data given by Eq. (26), the iterations of the cooling method for a particular cooling parameter λ consist of the following three main steps:

Step 1: update of the Jacobian of $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$:

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}); \tag{32}$$

Step 2: project onto the ℓ_1 ball $S = \{\|\mathbf{x}\|_1 \leqslant \|\mathbf{x}_0\|_1\}$ by soft thresholding

$$\mathbf{x} \leftarrow T_\lambda(\mathbf{x}); \tag{33}$$

Step 3: project onto the anisotropic diffusion ball $C = \{\mathbf{x}: J(\mathbf{x}) \leqslant J(\mathbf{x}_0)\}$ by

$$\mathbf{x} \leftarrow \mathbf{x} - \beta \boldsymbol{\nabla}_\mathbf{x} J_c(\mathbf{x}) \tag{34}$$

with

$$\boldsymbol{\nabla}_\mathbf{x} J_c(\mathbf{x}) = 2\mathbf{A}^\dagger \boldsymbol{\nabla} \cdot \left(\boldsymbol{\Lambda} \boldsymbol{\nabla} \left((\mathbf{A}^T)^\dagger \mathbf{x} \right) \right). \tag{35}$$

Table 2
Sparsity- and continuity-enhancing recovery of seismic amplitudes

Initialize:
$m = 0$;
$\mathbf{x}^0 = \mathbf{0}$;
$\mathbf{y} = \mathbf{K}^T \mathbf{d}$;
Choose:
M and L
$\|\mathbf{A}^T \mathbf{y}\|_\infty > \lambda_1 > \lambda_2 > \cdots$
while $\|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2 > \epsilon$ *do*
$\quad m = m + 1$;
$\quad \mathbf{x}^m = \mathbf{x}^{m-1}$;
$\quad$ *for* $l = 1$ to L *do*
$\quad\quad \mathbf{x}^m = T_{\lambda_m}(\mathbf{x}^m + \mathbf{A}^T(\mathbf{y} - \mathbf{x}^m))$ {iterative thresholding}
$\quad$ *end for*
$\quad$ *Anisotropic descent update*:
$\quad\quad \mathbf{x}^m = \mathbf{x}^m - \beta \nabla_{\mathbf{x}^m} J_c(\mathbf{x}^m)$;
end while
$\tilde{\mathbf{x}} = \mathbf{x}_m$; $\tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \tilde{\mathbf{x}}$

Remember that steps 1 and 2 with $\|\mathbf{A}\| \leqslant 1$ converge to the solution of Eq. (27) for a fixed λ. During step 3, the coefficients are updated according to the gradient of the anisotropic diffusion norm designed to reduce spurious artifacts. The different steps of our algorithm are summarized in Table 2. To ensure sparsity, the algorithm is started by setting $\mathbf{x}^0 = \mathbf{0}$.

5. Practical considerations

The presented method depends on the availability of a reference vector with an image that is sufficiently close to the unknown model. As in most Krylov-subspace solvers, we derive the reference vector from the migrated image (matched filter) as an initial guess. Since waves are subject to spherical spreading, a depth-dependent correction is applied to the migrated image (cf. Eq. (24)). This corrected image serves as input for the diagonal estimation procedure and the calculation of the anisotropic diffusion norm. The image recovery procedure consists of the following steps:

(1) Obtain the reference vector, $\mathbf{r}$, by imaging the data with Eq. (24), followed by applying the depth correction;
(2) Use the reference vector, $\mathbf{r}$, to estimate the diagonal matrix, $\text{diag}(\tilde{\mathbf{u}})$, with the procedure outlined in Table 1;
(3) Define the synthesis and analysis matrices $\mathbf{A}$ and $\mathbf{A}^T$;
(4) Calculate the rotation matrix, $\mathbf{\Lambda}$ (cf. Eq. (29)), for the continuity promoting penalty functional;
(5) Solve the optimization problem $\mathbf{P}_\epsilon$ (Eq. (30)) with the algorithm outlined in Table 2. This yields the estimate for the amplitude-corrected image, $\tilde{\mathbf{m}}$;
(6) Optionally goto 1 with $\mathbf{r} := \tilde{\mathbf{m}}$.

6. Numerical experiments

The recovery method is applied to a synthetic imaging example for data of the SEG/EAGE AA$'$ salt model [1,34]. The original velocity model and its smoothed version that defines the background velocity model, used by the migration operators, are included in Fig. 7. To avoid multiple reflections, the background velocity model is averaged over a window of size 720×720 m. The reflectivity (velocity perturbation) is defined as the band-pass filtered difference between the smoothed velocity model and a slightly less smoothed velocity model. The density is assumed to be constant. Figure 8a shows the reflectivity model that is 3.6 km by 15.3 km with a grid spacing of 24 m in the vertical
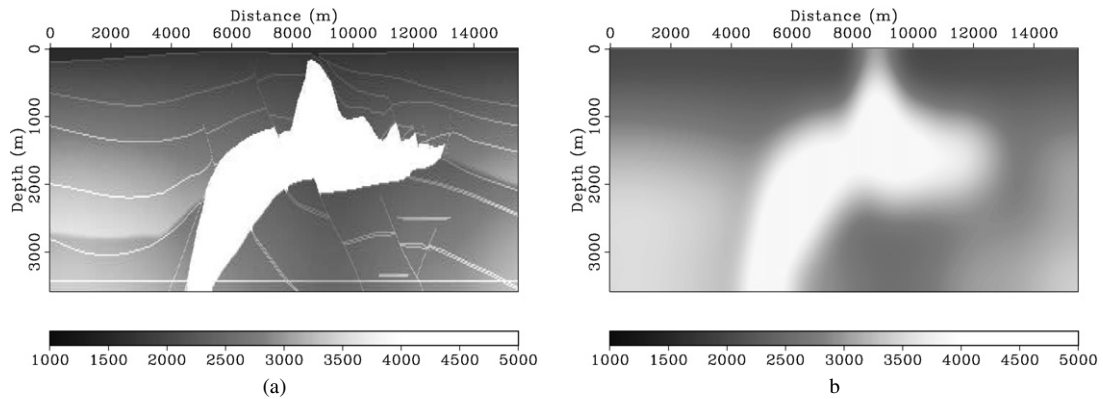
Fig. 7. The SEG/EAGE AA′ salt model [1]. (a) The original velocity model. (b) The smoothed velocity model with a window size of 720×720 m. Remark: Imaging this model is a challenge because of the presence of the high-velocity (bright) salt.

and horizontal coordinate directions. This reflectivity was used to generate data with the linearized Born modeling operator.

6.1. The migrated image

A two-way reverse-time migration code with checkpointing was used to conduct the experiments. This two-dimensional ($d = 2$) time-domain code, developed by [43], has absorbing boundary conditions and is based on the adjoint-state method. The migration operator is derived from the gradient of a descent algorithm for least-squares inversion (see [43] and the references therein). The data is generated with the adjoint of the migration operator, which corresponds to the linearized Born approximation. This Born modeling operator generates data without non-linear events such as multiple reflections. The migration operator and its adjoint are made zero-order by applying a half-order fractional integration on the source wavelet. We choose a source wavelet with a relatively broad temporal frequency band [5–60] Hz. The data consists of 324 shots with 176 receivers each with a shot at every 48 m, yielding a maximum offset of 4.224 km. Each geophone records 626 time samples with a sample interval of 8 ms, which amounts to 5 s of data.

The 8000 time-step linearized Born modeling takes about 64 min with 68 CPU's on a MPI cluster, while the migration takes about 294 min. These computation times explain why calculating the pseudoinverse of the normal operator becomes computationally prohibitive in real applications where the images are three dimensional ($d = 3$).

Figure 8 shows the original reflectivity and the migrated image obtained with Eq. (24). The image suffers from amplitude deterioration due to spherical spreading and bad illumination. Despite the overall amplitude deterioration for steep reflectors and events at increasing depths, the migration resolves most of the singularities in the image (cf. Figs. 8a and 8b).

6.1.1. Diagonal estimation

After applying corrections for the spherical spreading, the migrated data serves as the reference vector. This reference vector is used to compute the diagonal approximation for the normal operator and for the calculation of the gradients. Because the image deterioration between the reference vector (Fig. 9a) and the demigrated-migrated reference vector (Fig. 9b) is similar to the amplitude deterioration observed when comparing the 'unknown' true and imaged reflectivities (cf. Fig. 8), the images in Figs. 9a and 9b should be adequate as input for the diagonal estimation procedure outlined in Table 1. To avoid instabilities related to small entries in the estimated diagonal diag $\tilde{\mathbf{u}}$, we set $\mathbf{\Gamma} = \text{diag}\,(\tilde{\mathbf{u}} + \delta)/\delta$ with $\delta = 0.2$. The result of applying the estimated diagonal to the bandwidth-limited reflectivity of Fig. 9a is included in Fig. 9c and shows that most of the imprint of the normal operator is captured by the diagonal approximation.
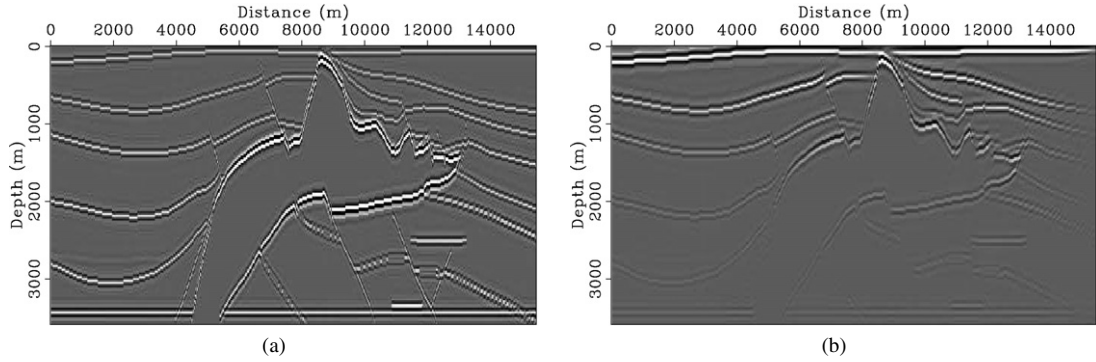
Fig. 8. Imaging of the SEG/EAGE AA′ salt model [1]. (a) Reflectivity defined in terms of the band-pass filtered difference between the smoothed velocity model and a slightly less smoothed velocity model. (b) The imaged reflectivity according to Eq. (24). The main reflection events are present but suffer from deteriorated amplitudes, especially under the high-velocity salt and for steep reflectors and faults.
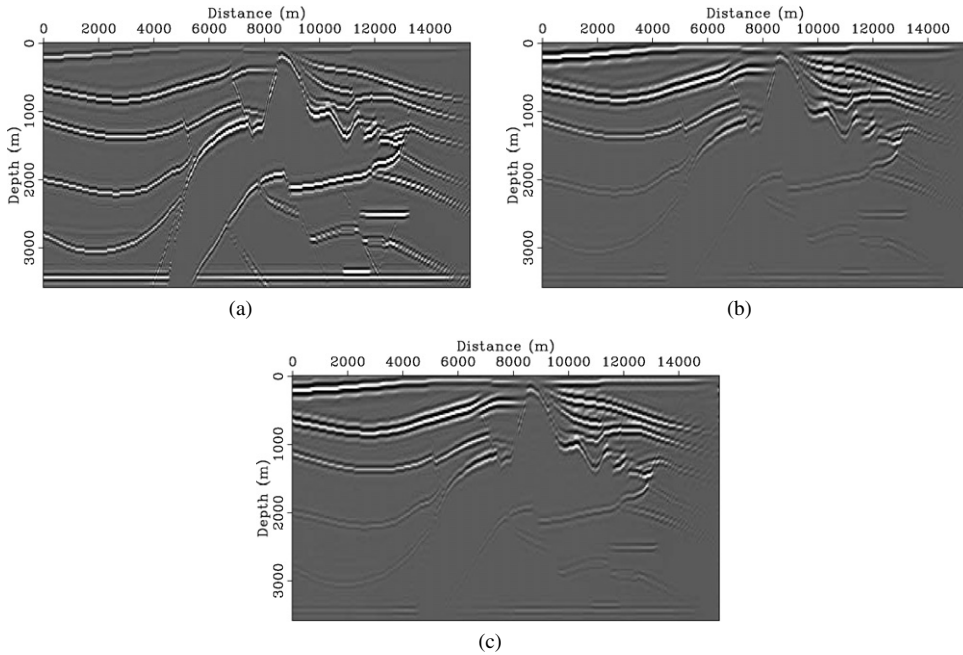

Fig. 9. Images that are input to the diagonal approximation scheme outlined in Table 1. (a) The reference vector, $\mathbf{r}$, derived from Fig. 8b, after applying corrections for the spherical-divergence and receiver-array taper. (b) The demigrated-migrated reference vector, $\mathbf{b} = \boldsymbol{\Psi}\mathbf{r}$, that serves as 'data' for the diagonal estimation. (c) Diagonal approximation on the original bandwidth-limited reflectivity plotted in (a). This diagonal approximation by $\mathbf{AA}^T\mathbf{m}$ captures the normal operator, $\boldsymbol{\Psi}\mathbf{m}$, quite well.

6.2. Seismic amplitude recovery for the SEG AA′ model

With the diagonal approximation in place, the amplitude corrections are applied with the procedure presented in Table 2. The anisotropic diffusion penalty term is calculated from the gradients plotted in Fig. 10. The results of the amplitude recovery are summarized in Fig. 11. Figure 11a depicts the result obtained by only applying a ℓ_1-norm penalty ($\beta = 0$). The image is calculated by lowering the Lagrange multiplier λ in 20 steps from a value that corresponds to a threshold that removes all but 5% of the coefficients to a Lagrange multiplier that only removes 1% of the curvelet coefficients. For each intermediate λ, the inner loop is repeated 5 times ($L = 5$). The result of this norm-one optimization are plotted in Fig. 11a. The image shows a clear improvement over the image plotted in Fig. 9a. Not only is the spherical divergence corrected in a stable manner, but the amplitude of the steep reflections events are recovered as well. There are, however, some small remaining artifacts related to side-band effects [12]. By including
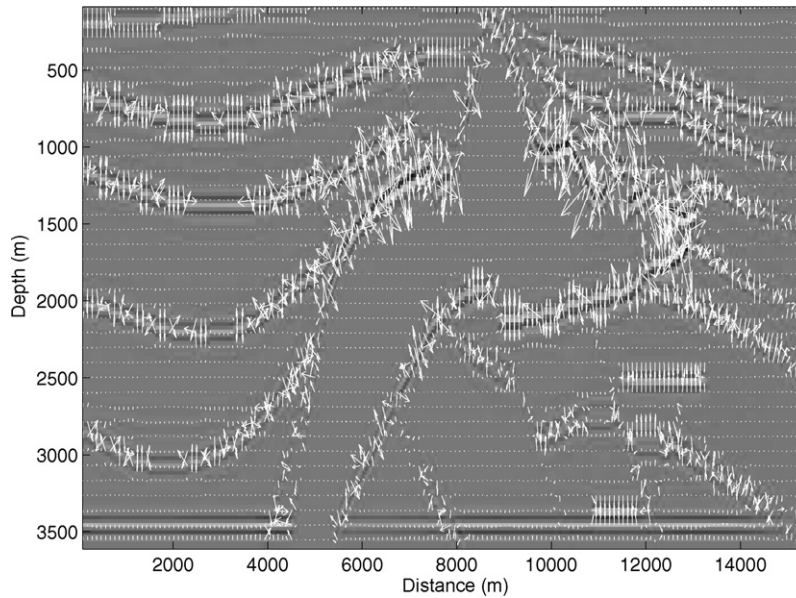
Fig. 10. Gradient vectors for the reference vector **r** plotted in Fig. 9a. The gradients (white ↑'s) are used to calculate the tangential directions along which the additional anisotropic smoothing is applied.
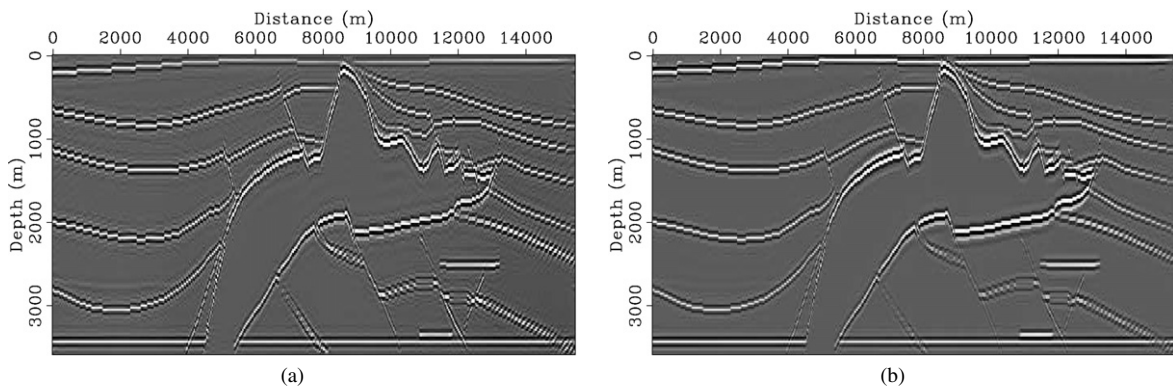

Fig. 11. Images after nonlinear recovery ($\mathbf{P}_\epsilon$). (a) Result with the ℓ_1-norm recovery only ($\beta = 0$). (b) Recovery with the combined ℓ_1 and continuity recovery for $\alpha = \beta = 1/2$. The amplitudes are recovered. The anisotropic diffusion successfully removes the artifacts.

the continuity-enhancing penalty functional most of the aforementioned artifacts can be removed as can be seen from Fig. 11b. In this example, the sparsity- and continuity-enhancing norms are weighted equally, i.e., $\alpha = \beta = 1/2$ in Eq. (31). To further illustrate the migration amplitude recovery, detailed plots for different traces of the recovered the amplitudes are included in Fig. 12. Figure 12a compares vertical traces of the original, migrated and amplitude recovered images. The traces are normalized with respect to the first reflector and the plots show that the amplitudes are nicely recovered. Similarly, the amplitudes along the horizontal reflector at $z = 3432$ m, are nicely recovered. The improvement in the recovered amplitudes is also apparent from Fig. 12c that contains the average vertical-wavenumber amplitude spectra of the original, the imaged and the recovered reflectivity.

6.3. Seismic amplitude recovery from noisy data

So far, the examples were noise free. In practice, seismic data volumes contain several sources of clutter, ranging from coherent nonlinear events, such as multiple reflections, to incoherent measurement errors. In this paper, only the recovery from incoherent contamination is considered. Sources of coherent clutter can be removed by using curvelet-based signal separation techniques reported elsewhere [29]. An image computed from data with a signal-to-noise
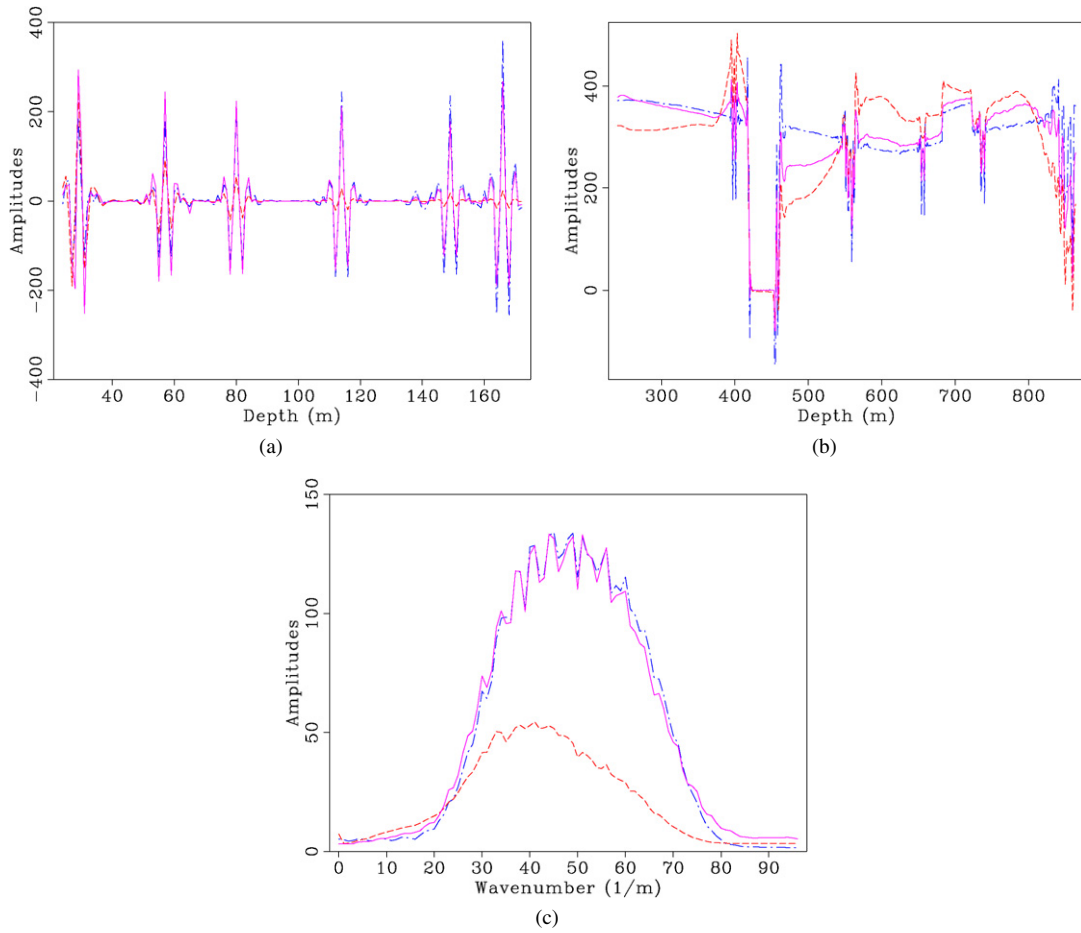
Fig. 12. Recovery of the amplitudes after normalization with respect to the first event. (a) Seismic traces of the original reflectivity (dot-dashed line), the migrated reflectivity (dashed line) and the recovered reflectivity (solid line). (b) Amplitudes along the bottom most horizontal reflector. (c) Average amplitude spectra of the depth-dependent reflectivity. The original and recovered spectra are normalized to match. Remarks: The nonlinear recovery corrects for the amplitudes and restores the spectrum.

ratio (SNR) of 3 dB is included in Fig. 13a. As shown in Eq. (24), the image now contains an additional contribution due to the noise. This noise contribution is clearly visible throughout the image as a nonstationary clutter and leads to a deterioration of the image quality. The clutter in the image space, however, is significantly smaller than in the data space, as can be observed from the noisy shot record plotted in Fig. 14b. The difference in noise levels between the data and image spaces can be explained by the 154-fold redundancy of the data space compared to the image space ($L = 154 \times M$).

The result for the amplitude recovery of the noisy data are included in Fig. 13b and this figure shows that the migration amplitudes can stably be recovered, leading to a significant improvement in the SNR for the image (9.2 dB for the image in Fig. 13b as opposed to 1.5 dB for the noisy image which includes the error due to the imprint of the normal operator). This improved image was obtained using the algorithm of Table 2, with the same settings as in the noise-free example except for the lowest value of the Lagrange multiplier, which is now set to a larger value calculated from the noise level. From Fig. 13b, one can see that the noise contamination has largely been removed. Moreover, the amplitudes have been restored in particular for the steep events at depth. Data generated from the estimated image, $\tilde{\mathbf{d}} = \mathbf{K}\tilde{\mathbf{m}}$ shows a significant removal of the noise (cf. Figs. 14b–14c), with reflection events that match the noise-free data plotted in Fig. 14a. This visual improvements leads to an improvement of SNR for the data of 19.2 dB.
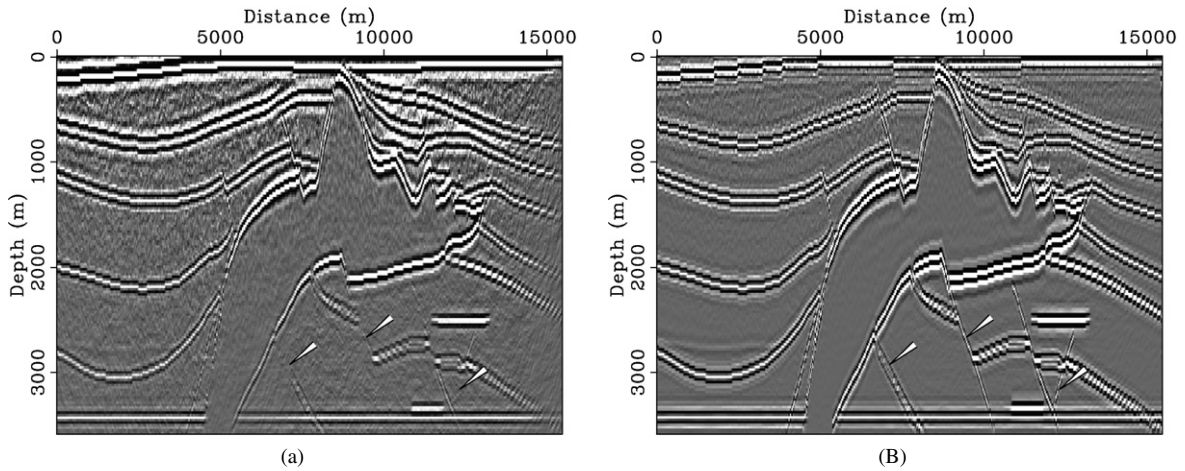
Fig. 13. Image amplitude recovery for noisy data (SNR 3 dB). (a) Noise image according to Eq. (24). (b) Image after nonlinear recovery from noisy data ($\mathbf{P}_\epsilon$). The clearly visible nonstationary noise in (a) is removed during the recovery while the amplitudes are also restored. Steeply dipping reflectors denoted by the arrows are also well recovered.
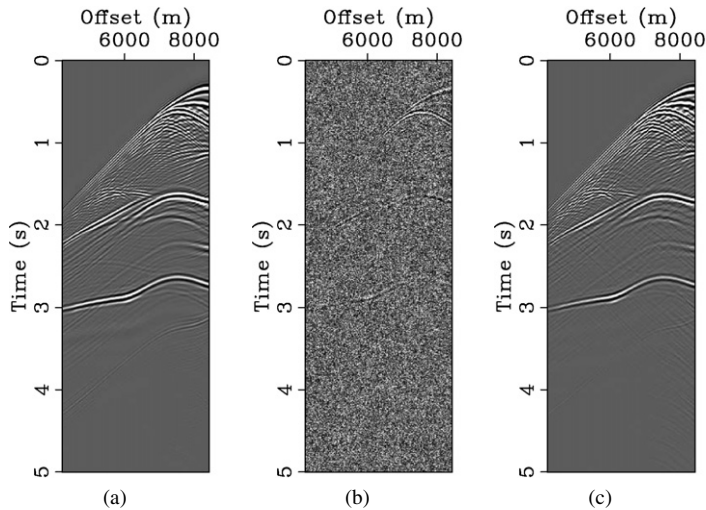

Fig. 14. 'Denoising' of a shot record. (a) The noise-free data received by the receiver array. (b) Noisy data with a SNR of 3 dB. (c) Forward modeled data after amplitude recovery. Observe the significant improvement in the data quality, reflected in an increase for the SNR of 19.2 dB.

7. Conclusions

7.1. Initial findings

The method presented in this paper combines the compression of images by curvelets with the approximate invariance of this transform under the normal operator. This combination allows for a formulation of a stable recovery method for seismic amplitudes. During the recovery, the normal operator is approximately inverted. Compared to other approaches for migration scaling, the presented method (i) includes a theoretical bound on the L^2-error for the diagonal approximation in the curvelet domain; (ii) prescribes a procedure for the estimation of the diagonal from numerical implementations of the imaging operators; (iii) formulates the amplitude recovery problem as a nonlinear optimization problem, where the inversion of the diagonalized normal operator is regularized by imposing sparsity in the curvelet domain and continuity along the imaged reflectors.

As long as the background velocity model is sufficiently smooth and there is a reference vector available close enough to the actual reflectivity, the normal operator can be approximated by a diagonal weighting in the curvelet

domain. The theoretical approximation error is scale-dependent and decreases for finer scales, an observation consistent with the microlocal properties of the normal operator and curvelets in the high-frequency limit. Estimation of the diagonal exploits smoothness of the symbol by penalizing neighboring entries in the diagonal that fluctuate. These fluctuations are inconsistent with the smoothness of the symbol. For a sufficiently large Lagrange multiplier, the regularized least-squares estimation for the diagonal leads to positive entries. Numerical experiments with the diagonal approximation showed a moderate-sized constant for the bound of the theoretical error.

The diagonal approximation is used to formulate the seismic amplitude recovery in terms of a constrained optimization problem. The amplitude-corrected image is obtained by solving this sparsity- and continuity-promoting optimization problem. The approximate invariance of curvelets under the normal operator preserves the sparsity. The cost of computing the diagonal approximation is one demigration–migration per reference vector, which is much less compared to the cost of Krylov-based least-squares inversion. The recovery results show an overall improvement of the image quality. The joined sparsity- and continuity-enhanced image has diminished artifacts, improved resolution and recovered amplitudes.

7.2. Extension to prestack imaging

The discussion has been on 'post-stack' images obtained after applying the zero-offset imaging condition. When the velocity model is correct, images are smooth along the redundant coordinate in prestack angle gathers. This smoothness property has been used in velocity analyses methods such as differential semblance [44] and is suitable for the framework laid down in this paper. Besides the ℓ_1 and continuity norms, the amplitude recovery can also exploit the smoothness along the redundant angle coordinate. The fact that smoothness translates to sparsity is especially exciting because it allows to further exploit the results from the field of compressed sensing.

7.3. Recent related work

During the revision, we were informed of recent work by William Symes on optimal scaling of reverse-time migration [42]. This work and recent work by [26] attempt to invert the normal operator by estimating a diagonal approximation from a reference vector, typically given by the migrated image, and the demigrated–migrated reference vector. Smoothness of the diagonal approximation is also imposed by both authors by parameterizing the diagonal as a smooth function in the spatial domain. Our approach goes several steps further by imposing smoothness in phase space and by making use of the curvelet transform that allows for the inversion of the diagonal using the methods of stable image recovery.

7.4. Choice of the reference vector

Having access to a proper reference vector is a prerequisite for the success of the method presented in this paper. Selection of the appropriate reference is somewhat reminiscent of finding a good initial guess for Krylov subspace methods. In this paper, we took a data-driven approach by using the migrated image (matched-filter result) as a first guess. Our formulation allows for multiple reference vectors and we hope to report on appropriate selection strategies for reference vectors in the future. This investigation will include an analysis of the sensitivity of our method to the accuracy of the velocity model.

Acknowledgments

Group, BP, Chevron, ExxonMobil and Shell. The first two authors would also like to thank the Institute of Pure and Applied Mathematics at UCLA, supported by the NSF under grant DMS-9810282, for their hospitality. The research of the third author was supported by the Netherlands Organization for Scientific Research, through a VIDI grant No. 639.032.509.

Appendix A. Proofs of Lemma 1 and Theorem 1

We first prove Theorem 1 from Lemma 1, we then prove the lemma.

Proof of Theorem 1. C^T acting on a vector $\mathbf{c} = \{c_\mu\}_{\mu \in \mathcal{M}}$ in ℓ^2 is simply given by $C^T \mathbf{c} = \sum_\nu c_\nu \varphi_\nu$. Hence,

$$C^T \mathbf{D}_\Psi C \varphi_\mu = \sum_\nu (C \varphi_\mu)_\nu a(x_\nu, \xi_\nu) \varphi_\nu.$$

The difference $(\Psi(x, D) - C^T \mathbf{D}_\Psi C) \varphi_\mu$ can be written as, using that $C^T C = \mathrm{Id}$,

$$\sum_\nu (C \varphi_\mu)_\nu \big(\Psi(x, D) - a(x_\nu, \xi_\nu) \big) \varphi_\nu.$$

The vector $(C\varphi_\mu)_\nu$ is bounded in $\ell^2(\mathcal{M})$ of vectors with curvelet indices, and its entries can only be nonzero if $|\mu| - 2 \leqslant |\nu| \leqslant |\mu| + 2$, since the support of two curvelets in the Fourier domain is disjoint when curvelets are two or more scales apart. It follows that

$$\big\| \big(\Psi(x, D) - C^T \mathbf{D}_\Psi C \big) \varphi_\mu \big\|_{L^2(\mathbb{R}^n)}^2 \leqslant C_6 \sum_\nu \big\| (C \varphi_\mu)_\nu \big(A(x, D) - a(x_\nu, \xi_\nu) \big) \varphi_\nu \big\|_{L^2(\mathbb{R}^d)}^2$$

$$\leqslant C_7 \sum_\nu 2^{-|\nu|} \big| (C \varphi_\mu)_\nu \big|^2 \leqslant C_8 2^{-|\mu|/2}. \qquad \square$$

Proof of Lemma 1. Possibly after a coordinate transformation, which does not affect the pseudodifferential nature of Ψ, we may assume that $\xi_\mu = (0, \dots, 0, \lambda)$, we will write $\xi' = (\xi_1, \dots, \xi_{d-1})$.

We may take $|\nu|$ greater than some minimum value, so that the support in the wavenumber domain is away from $\xi = 0$. We can then assume that the principal symbol is homogeneous of order 0 in the support of the curvelet. On the support of the curvelet in the Fourier domain the symbol can be written as $a(x, \frac{\xi'}{\xi_n})$. We apply a preparation theorem to this term. There are C^∞ functions $b_n(x, \xi'/\xi_d)$, $n = 1, \dots, 2d - 1$, such that

$$a\left(x, \frac{\xi'}{\xi_d}\right) - a\left(x_\nu, \frac{\xi'_\nu}{\xi_{\nu,d}}\right) = \sum_{n=1}^{d} (x - x_\nu)_n b_n\left(x, \frac{\xi'}{\xi_d}\right) + \sum_{n=1}^{d-1} \left(\frac{\xi_n}{\xi_d} - \frac{\xi_{\nu,n}}{\xi_{\nu,d}}\right) b_{d+n}\left(x, \frac{\xi'}{\xi_d}\right).$$

Therefore, we can write

$$\left(\Psi\left(x, \frac{D'}{D_d}\right) - a\left(x_\nu, \frac{\xi'_\nu}{\xi_{\nu,d}}\right) \right) \varphi_\nu = \left[\sum_{n=1}^{d} (x - x_\nu)_n b_n\left(x, \frac{D'}{D_d}\right) + \sum_{n=1}^{d-1} b_{d+n}\left(x, \frac{D'}{D_d}\right) \frac{D_n}{D_d} \right] \varphi_\nu. \tag{A.1}$$

Consider first the contribution for one of the b_n with $1 \leqslant n \leqslant d$. The operator acting on the curvelet reads

$$(x - x_\nu)_n b_n\left(x, \frac{D'}{D_d}\right) \varphi_\nu = b_n\left(x, \frac{D'}{D_d}\right)(x - x_\nu)_n \varphi_\nu + \left[(x - x_\nu)_n, b_n\left(x, \frac{D'}{D_d}\right) \right] \varphi_\nu.$$

Because of the support and decay properties of curvelets we have $\|(x - x_\nu)_n \varphi_\nu\|_{L^2} \leqslant C_9^{-|\nu|/2}$, therefore

$$\left\| b_n\left(x, \frac{D'}{D_d}\right)(x - x_\nu)_n \varphi_\nu \right\|_{L^2(\mathbb{R}^d)} \leqslant C_{10} 2^{-|\nu|/2}.$$

By the calculus of pseudodifferential operators, the operator $[(x - x_\nu)_n, b_n(x, D'/D_d)]$ is a pseudodifferential operator of order -1, which is continuous $H^{-1}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ therefore

$$\left\| \left[(x - x_\nu)_n, b_n\left(x, \frac{D'}{D_d}\right) \right] \varphi_\nu \right\|_{L^2(\mathbb{R}^d)} \leqslant C_{10} \|\varphi_\nu\|_{H^{-1}(\mathbb{R}^d)} \leqslant C_{11} 2^{-|\nu|}.$$

Adding the previous two estimates, we find that

$$\left\| (x - x_v)_n b_n \left(x, \frac{D'}{D_d} \right) \varphi_v \right\|_{L^2(\mathbb{R}^d)} \leqslant C_{12} 2^{-|v|/2}. \tag{A.2}$$

We now consider the b_n, $d + 1 \leqslant n \leqslant 2d - 1$. We must estimate $b_{d+n}(x, D'/D_d) \frac{D_n}{D_d} \varphi_v$. By (11) we have $\| \frac{D_n}{D_d} \varphi_v \|_{L^2(\mathbb{R}^n)} \leqslant C_{13} 2^{-|v|/2}$. Therefore,

$$\left\| b_{d+n} \left(x, \frac{D'}{D_d} \right) \frac{D'}{D_d} \varphi_v \right\|_{L^2(\mathbb{R}^d)} \leqslant C_{14} 2^{-|v|/2}. \tag{A.3}$$

From (A.1) and the estimates (A.2) and (A.3) for the summands, we find the result. $\square$

References

[1] F. Aminzadeh, J. Brac, T. Kunz, 3-D Salt and Overthrust Model, SEG/EAGE 3-D Modeling Series 1, SEG, Tulsa, 1997.
[2] G. Beylkin, The inversion problem and applications of the generalized Radon transform, Comm. Pure Appl. Math. 37 (1984) 579–599.
[3] M.J. Black, G. Sapiro, D. Marimont, Robust anisotropic diffusion, IEEE Trans. Signal Process. 7 (3) (1998) 421–432.
[4] N. Bleistein, J. Cohen, J. Stockwell, Mathematics of Multidimensional Seismic Imaging. Migration and Inversion, Springer, 2001.
[5] S. Brandsberg-Dahl, M.V. de Hoop, Focusing in DIP and AVA compensation on scattering-angle/azimuth common image gathers, Geophysics 68 (1) (2003) 232–254.
[6] E.J. Candès, D.L. Donoho, Continuous curvelet transform I: Resolution of the wavefront set, Appl. Comput. Harmon. Anal. 19 (2005) 162–197.
[7] E.J. Candès, L. Demanet, D.L. Donoho, L. Ying, Fast discrete curvelet transforms, SIAM Multiscale Model. Simul. 5 (3) (2006) 861–899.
[8] E.J. Candès, D.L. Donoho, Continuous curvelet transform II: Discretization and frames, Appl. Comput. Harmon. Anal. 19 (2005) 198–222.
[9] E.J. Candès, D.L. Donoho, Curvelets—A surprisingly effective nonadaptive representation for objects with edges, in: L.L. Schumaker, et al. (Eds.), Curves and Surfaces, Vanderbilt Univ. Press, 2000.
[10] E.J. Candès, D.L. Donoho, Recovering edges in ill-posed problems: Optimality of curvelet frames, Ann. Statist. 30 (2000) 784.
[11] E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure Appl. Math. 57 (2) (2004) 219–266.
[12] E.J. Candès, F. Guo, New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction, Signal Process. (2002) 1519–1543.
[13] E.J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[14] G. Chavent, R. Plessix, An optimal true-amplitude least-squares prestack depth-migration operator, Geophysics 64 (2) (1999) 508–515.
[15] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 43 (1) (2001) 129–159.
[16] J. Claerbout, F. Muir, Robust modeling with erratic data, Geophysics 38 (05) (1973) 826–844.
[17] J. Claerbout, D. Nichols, Spectral preconditioning, technical report, Stanford Exploration Project, 1994.
[18] I. Daubechies, M. Defrise, C. de Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints, Comm. Pure Appl. Math. 57 (11) (2005) 1413–1457.
[19] M.V. de Hoop, N. Bleistein, Generalized radon transform inversions for reflectivity in anisotropic elastic media, Inverse Problems 13 (3) (1997) 669–690.
[20] M.V. de Hoop, S. Brandsberg-Dahl, Maslov asymptotic extension of generalized radon transform inversion in anisotropic elastic media: A least-squares approach, Inverse Problems 16 (3) (2000) 519–562.
[21] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition, Appl. Comput. Harmon. Anal. 2 (1995) 101–126.
[22] D.L. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory 52 (2006) 6–18.
[23] M. Elad, J.L. Starck, P. Querre, D.L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), Appl. Comput. Harmon. Anal. 19 (2005) 340–358.
[24] G.C. Fehmers, C.F. Höcker, Fast structural interpretation with structure-oriented filtering, Geophysics 68 (4) (2003) 1286–1293.
[25] M. Figueiredo, R. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Trans. Image Process. (2003).
[26] A. Guitton, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators, Geophysics 69 (4) (2004) 1017–1024.
[27] G. Hennenfent, F.J. Herrmann, Seismic denoising with non-uniformly sampled curvelets, Comp. Sci. Eng. 8 (3) (2006) 16–25.
[28] F.J. Herrmann, Multifractional splines: Application to seismic imaging, in: M. Unser, A. Aldroubi (Eds.), Wavelets: Applications in Signal and Image Processing X, in: Proc. SPIE, vol. 5207, SPIE, 2003, pp. 240–258.
[29] F.J. Herrmann, U. Boeniger, D.J. Verschuur, Nonlinear primary-multiple separation with directional curvelet frames, Geophys. J. Int. 170 (2007) 781–799.
[30] J. Hu, G.T. Schuster, P.A. Valasek, Poststack migration deconvolution, Geophysics 66 (3) (2001) 939–952.
[31] H. Kuhl, M. Sacchi, Least-squares wave-equation migration for AVP/AVA inversion, Geophysics 68 (1) (2003) 262.
[32] N.Y. Lee, B.J. Lucier, Wavelet methods for inverting the Radon transform with noisy data, IEEE Trans. Image Process. 10 (2001) 79–94.

[33] T. Nemeth, C. Wu, G.T. Schuster, Least-squares migration of incomplete reflection data, Geophysics 64 (1) (1999) 208–221.
[34] M. O'Brien, S. Gray, Can we image beneath salt? Leading Edge 15 (1) (1997) 17–22.
[35] C.C. Paige, M.A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, ACM TOMS 8 (1) (1982) 43–71.
[36] R. Plessix, W. Mulder, Frequency-domain finite difference amplitude-preserving migration, Geophys. J. Int. 157 (2004) 975–987.
[37] J.E. Rickett, Illumination-based normalization for wave-equation depth migration, Geophysics 68 (4) (2003) 1371–1379.
[38] O. Schertzer, Scale-space methods and regularization for denosing and inverse problems, Adv. Imag. Electron Phys. 128 (2003) 445–530.
[39] J.L. Starck, M. Elad, D.L. Donoho, Redundant multiscale transforms and their application to morphological component separation, Adv. Imag. Electron Phys. 132 (2004).
[40] C.C. Stolk, Microlocal analysis of a seismic linearized inverse problem, Wave Motion 32 (2000) 267–290.
[41] C.C. Stolk, W.W. Symes, Smooth objective functionals for seismic velocity inversion, Inverse Problems 19 (1) (2003) 73–89.
[42] W.W. Symes, Optimal scaling for reverse time migration, Geophysics, in press.
[43] W.W. Symes, Reverse time migration with optimal checkpointing, Geophysics 72 (5) (2007) SM213–SM221.
[44] W.W. Symes, J. Carazzone, Velocity inversion by differential semblance optimization, Geophysics 56 (1991) 2061–2073.
[45] A.P.E. ten Kroode, D.-J. Smit, A.R. Verdel, A microlocal analysis of migration, Wave Motion 28 (2) (1998) 149–172.
[46] R. Tibshirani, Regression shrinkage and selection via the LASSO, J. Roy. Statist. Soc. B 58 (1) (1997) 267–288.
[47] J.A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, IEEE Trans. Inform. Theory 52 (3) (2006) 1030–1051.
[48] Y. Tsaig, D.L. Donoho, Extensions of compressed sensing, Signal Process. 86 (3) (2006) 549–571.
[49] L. Ying, L. Demanet, E.J. Candés, 3D Discrete Curvelet Transform, in: Proc. SPIE, vol. 5914, SPIE, 2005, p. 591413.