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

Felix J. Herrmann[†], Peyman P. Moghaddam[†] and Chris Stolk[‡]

[†] *Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC, Canada*
[‡] *Department of Applied Mathematics, University of Twente, the Netherlands*

E-mail: fherrmann@eos.ubc.ca

A nonlinear singularity-preserving solution to seismic image recovery with sparseness and continuity constraints is proposed. The method explicitly explores the curvelet transform as a directional frame expansion that, by virtue of its sparsity on seismic images and its invariance under the Hessian of the linearized imaging problem, allows for a stable recovery of the migration amplitudes from noisy data. The method corresponds to a preconditioning that corrects the amplitudes during a post-processing step. 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 'wave-equation' migration code on a synthetic dataset.

Key Words: Seismic imaging, image recovery, curvelet transform, overcomplete signal representations, inversion, preconditioning, nonlinear optimization, ℓ_1 minimization, anisotropic diffusion.

1. INTRODUCTION

1.1. The seismic imaging problem

The primary focus of seismic imaging is to locate singularities in the earth's elastic properties from seismic data recorded at the surface [1; 15; 44; 41; 2; 3]. A seismic survey consists of multiple seismic experiments in which both the location of the sources, s, and receivers, r, are varied along the surface. After linearization and by ignoring the source and receiver signatures, the discretized forward model that generates seismic data can be written as

$$\mathbf{d} = \mathbf{K}\mathbf{m} + \mathbf{n}. \tag{1}$$

In this single-scattering expression, $\mathbf{m}(\mathrm{x})$ represents the (singular) fluctuations in the earth's acoustic properties with respect to an appropriately chosen smoothly varying background velocity model (the density of mass is assumed constant) [15; 2]. These fluctuations are referred to as the model or reflectivity and seismic imaging aims to recover both the locations and the relative amplitudes of the velocity fluctuations from seismic data. The model is a sampled function of d-spatial dimensions $\mathrm{x} = (x_1, \cdots, x_d)$ and the discretized scattering matrix $\mathbf{K}$ generates the data vector, which corresponds to a discretization of a $(2(d-1)+1)$-dimensional data volume, sampled along the $(d-1)$-dimensional source and receiver coordinates, s and r and along the positive time axis t. The additive pseudo-random noise vector, $\mathbf{n}$ accounts for possible measurement errors and is assumed to be given by Gaussian white noise, $\mathbf{n} \in N(0, \sigma)$, with zero mean and σ standard deviation. The assumption of white noise can be relaxed to colored noise. Without loss in generality, we limit the discussion to imaging in two dimensions ($d = 2$), where the data vec-

tor is a sampling of a three-dimensional data volume and where the model vector represents a discretized two-dimensional seismic 'image'.

The main purpose of this paper is to recover the relative amplitudes for the image from seismic data that is contaminated with noise. The forward model given in Eq. 1 is derived from a linearization of a data misfit functional. The Jacobian of this functional [see e.g. 34] corresponds, in the noise-free case, to applying the adjoint of the linearized scattering operator to the data vector (cf. Eq. 1), i.e.,

$$\mathbf{y} = \mathbf{K}^T\mathbf{d} \tag{2}$$

with the symbol ^{T} denoting the matrix transpose. Because $\mathbf{K}$ is near unitary, i.e., $\mathbf{K}^T\mathbf{K} \approx \mathbf{I}$, the migrated image, $\mathbf{y}$, is generally a good approximation to the reflectivity.

With the increased demand for high-quality images, the above approximation of $\mathbf{K}^T\mathbf{K} \approx \mathbf{I}$ is no longer justifiable because it may lead to amplitude errors. An extensive literature has emerged to restoring the migration amplitudes by inverting the Gram or normal matrix $\mathbf{K}^T\mathbf{K}$. These approaches are known as 'least-squares migration' [32; 10; 16; 27; 28; 34] and involve computation of the Moore-Penrose pseudo inverse (denoted by the symbol †)

$$\mathbf{m} = \overbrace{\left(\mathbf{K}^T\mathbf{K}\right)^{-1}}^{\Psi\text{DO}} \underbrace{\mathbf{K}^T}_{\text{FIO}}\mathbf{d} = \mathbf{K}^\dagger\mathbf{d}. \tag{3}$$

of the scattering or demigration matrix $\mathbf{K}$. In this expression, the amplitudes are restored by inverting the Hessian also known as the Gram matrix or normal operator.

Unfortunately, seismic data volumes are large (for $d = 3$ typically Tera to Peta Bytes) making it difficult to calculate the inverse of the Hessian. In addition, the Hessian it too big to be constructed explicitly and is too expensive to be evaluated as part of an iterative Krylov-subspace solver to invert the Hessian [see e.g. 32].

There exists a large body of literature on approximating the inverse Hessian and the approaches differ in the assumptions that are being made. There are high-frequency approximations that lead to explicit formulae for the gradient and inverse Hessian in terms of Fourier integral (FIOs) and pseudo-differential operators (ΨDOs) [44; 16; 42]. Even though these high-frequency approaches proof to be very useful, i.e., they invert for the Hessian by a diagonal weighting in the Fourier domain, they do not account for finite-frequency effects present in bandwidth-limited seismic data. Gradient-based 'wave-equation' methods have become feasible with the advent of larger and faster computers, yielding migration operators that include finite-frequency effects [43; 35; 34]. Even though 'wave-equation' migration is rapidly being adapted as the imaging tool of choice, the computation of the Hessian remains computationally illusive, precluding Krylov subspace-based approximations to the inverse of the Hessian.

In recent work, [34] proposes a preconditioning where the inverse of the Hessian can, under certain conditions, successfully be approximated by a diagonal matrix with migration weights. When applied to migrated data this diagonal matrix restores the amplitudes in the image. We follow a similar approach, where the Hessian is approximated by a diagonal weighting. Our approach differs, however, with respect to the domain in which and how the diagonal weighting is applied.

We argue that a stable amplitude recovery scheme can be obtained when an atomic decomposition is used that not only sparsely represents the unknown image, through a fast decay for the magnitude-sorted coefficient vector, but whose atoms are also invariant under the action of the Hessian. We show that in that case the amplitude weighting can be performed as part of a nonlinear optimization procedure which exploits these properties and hence adds stability. This stability includes insensitivity to noise and the ability to recover the amplitudes from incomplete migrated data.

Our diagonal approximation derives from sparsity and invariance properties of the curvelet transform [see e.g., 7] under the Hessian. This eigenfunction-like behavior, together with the sparsity property, allows for a nonlinear solution of the amplitude-recovery problem. This combination of sparsity and invariance forms the key element of the wavelet-vaguelette method (WVD), proposed by [29] and later [22; 30] to invert scale-invariant homogeneous operators such as the Hilbert and Radon transforms. We propose to apply these results to the seismic imaging problem, using an overcomplete signal representation by curvelets.

This work is an extension to earlier work [26] in which the recovery of seismic images is discussed in less detail. Even though, our method derives from recent work on approximating imaging operators [see e.g. 6; 21], our main focus is to recover the migration amplitudes from noisy data and not the compression of the scattering and migration operators.

1.2. Stable seismic image recovery with sparseness and continuity constraints

With the increasing demand for hydrocarbons, there is a constant push for higher resolution seismic images with improved signal-to-noise ratios and accurately recovered amplitudes. The presence of noise in seismic data recordings, possible limitations in the acquisition, such as finite aperture (longest distance source-receiver) and missing source-/receiver-positions, call for a stable formulation for the seismic image recovery problem. Motivated by recent results on stable signal recovery for natural images from incomplete and noisy data [see e.g. 19; 9; 14], the seismic image recovery problem is formulated as a nonlinear optimization problem. During the optimization, sparsity in the transformed domain as well as continuity along imaged reflectors, are jointly promoted. Both penalties are part of the following

optimization problem

$$\mathbf{P}: \quad \begin{cases} \min_{\mathbf{x}} J(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \hat{\mathbf{m}} = \left(\mathbf{A}^T\right)^\dagger \hat{\mathbf{x}}, \end{cases} \tag{4}$$

where the sparsity vector $\mathbf{x}$ is optimized with respect to the penalty functional $J(\mathbf{x})$ and the data misfit. The penalty functional $J(\mathbf{x})$ contains both sparseness- and continuity-enhancing penalty terms. The to be recovered sparsity vector contains the unknown coefficients of a transformed domain that is sparse on the model. The sparsity vector is related to the migrated input data $\mathbf{y} = \mathbf{K}^T\mathbf{d}$ through the sparsity synthesis matrix $\mathbf{A}$, which is designed to permit a sparse representation for the migrated data, i.e.,

$$\begin{aligned} \mathbf{y} &= \mathbf{K}^T\mathbf{K}\mathbf{m} + \mathbf{e} \\ &\simeq \mathbf{A}\mathbf{A}^T\mathbf{m} + \mathbf{e} \\ &= \mathbf{A}\mathbf{x}_0 + \mathbf{e} \end{aligned} \tag{5}$$

with $\mathbf{x}_0$ sparse and $\mathbf{e} := \mathbf{K}^T\mathbf{n}$. We will show that for the appropriate choice for $\mathbf{A}$ in terms of a transform that is sparse on the model and invariant under the Hessian, the sparsity vector $\mathbf{x}_0$ can be recovered by solving the above optimization problem. This solution fits the migrated data to within a noise-level dependent tolerance, ε. By applying the pseudo inverse to the estimated vector, $\hat{\mathbf{x}}$ the amplitude-corrected image is obtained by nonlinear optimization. We reserved the 'hat' symbol ($\hat{}$) to denote estimates that are found through optimization.

Because our method is based on curvelets that localize both in space and spatial frequency, our approach holds the middle between high-frequency asymptotic methods, where the inverse of the Hessian is computed explicitly in the Fourier domain (by inverting a ΨDO [see e.g. 44]), and the recently proposed preconditioners that approximately invert the Hessian by a diagonal matrix in the physical domain [34]. Besides sufficient smoothness on the background medium, our method does not

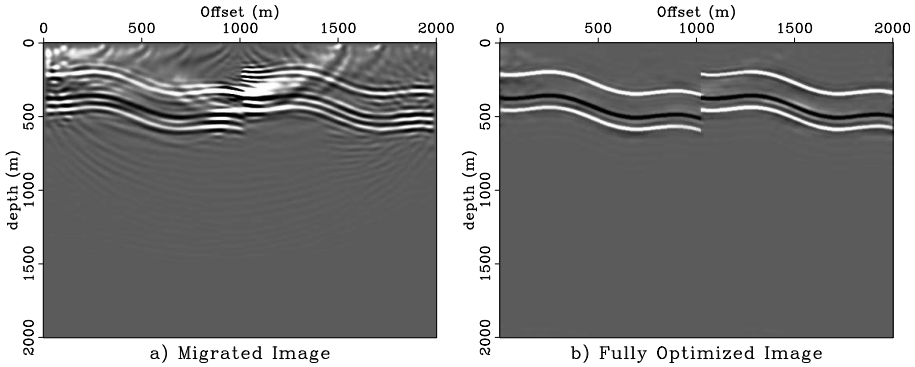

FIG. 1. Comparison between conventional imaging and the proposed image-amplitude recovery method. **(a)** image obtained by conventional 'wave-equation' migration $\mathbf{K}^T\mathbf{d}$ on data.; **(b)** Result of sparseness- and continuity-promoted image recovery.

require infinite-frequencies nor does it require infinite aperture or stringent conditions on the acquisition. Based on heuristic arguments, we will demonstrate that the above optimization procedure leads to a stable recovery of the migration amplitudes with a reduction of the imaging artifacts accomplished by the continuity-enhancing penalty term. As the images in Fig. 1 demonstrate, the proposed method arguably leads to an improvement in the image quality.

1.3. Outline

In section 2, we briefly sketch the main ingredients of gradient-based seismic imaging, followed by a discussion on approximate sparsity of seismic images in the curvelet domain, the high-frequency behavior of the Hessian and its near diagonalization in this domain. We then proceed in section 3 by discussing the canonical denoising problem for orthornormal and overcomplete sparsity representations. We show that the constrained optimization problem can be solved by a series of unconstrained optimization problems based on the iterative thresholded Landweber descend method. During this optimization, sparsity in the curvelet domain is exploited by minimizing a ℓ_1-penalty term. Through preconditioning, it is shown in section 4 that the seismic image recovery problem can be solved by thresholding on

the coefficients of a diagonally weighted orthonormal transform. By analogy with denoising for overcomplete representations, this result with preconditioning is extended towards the case where the Hessian is approximately inverted by a weighted overcomplete image representation that is approximately sparse on the model and approximately invariant under the Hessian. Corrections for possible migration artifacts, due to bad illumination and side-band effects [8], are made by complementing the sparsity-enhancing penalty functional with a continuity-promoting penalty functional. This approach leads in section 6 to our sparseness- and continuity-enhancing imaging scheme, which is subjected to a numerical performance evaluation in section 7.

2. SEISMIC IMAGE RECOVERY AND SPARSITY FRAMES

A variation on the wavelet-vaguelette (WVD) method adapted to seismic imaging problem is presented. Because of the large dimensions of the imaging operators, the method is derived for the normal equation, i.e. the method uses migrated data as input which leads to a reduction of the problem size. The crux of WVD methods lies in transforming the unknown image with respect to a transform that is not only sparse on the model but which is also invariant under the operator. [29], followed by many others, used this invariance, to formulate the solution of inverse problems by thresholding. A similar argument can be made for seismic image recovery based on the invariance of curvelets under the Hessian. Before discussing the sparsity and diagonal approximation of the Hessian by the curvelet transform, a more detailed introduction is given on gradient-based imaging.

Gradient-based imaging. Irrespective of its implementation, the scalar wave equation (constant density) corresponds, when ignoring source-/receiver characteristics, to a nonlinear functional

$$\mathbf{F}[s] : \; s(\mathrm{x}) \rightarrow d(\mathrm{s}, \mathrm{r}, t) \tag{6}$$

relating the slowness $s(\mathrm{x}) = c^{-1}(\mathrm{x})$ with $c(\mathrm{x})$ the compressional wavespeed [ms^{-1}], to the data $d(\mathrm{s},\mathrm{r},t)$. The functional $\mathbf{F}[s]$ involves the solution of the acoustic wave equation, followed by a restriction to measurements recorded at the surface [see for details 34]. Formally, we are interested in the inverse of this functional $\mathbf{F}^{-1}[d] : d(\mathrm{s},\mathrm{r},t) \rightarrow s(\mathrm{x})$. Following [43; 35; 34], this functional is linearized with respect to a smooth background medium $\bar{\mathbf{s}}$ and a small (singular) perturbation. This perturbation corresponds to the discretized image, $\mathbf{m} = \mathbf{s}^2 - \bar{\mathbf{s}}^2$, and is proportional to the reflectivity. The linearization coincides with the well-known single-scattering (Born) approximation. After linearization, the migration operator and the Hessian follow from the data misfit functional

$$J(s) = \frac{1}{2}\|\mathbf{d} - \mathbf{F}[\mathbf{s}]\|_2^2. \tag{7}$$

The negation of the linearized Jacobian of this functional defines migration,

$$\mathbf{y} = -\nabla_{\mathbf{m}} J(\mathbf{m}) = \mathbf{K}[\bar{\mathbf{s}}]^T \mathbf{d}. \tag{8}$$

The matrix $\mathbf{K}[\bar{\mathbf{s}}]^T$ is the adjoint of the scattering matrix $\mathbf{K}[\bar{\mathbf{s}}] \in \mathbb{R}^{L \times M}$ with L and M the length of the data and image vectors. $\nabla_{\mathbf{m}}$ is the discretized gradient. Explicit reference is made to the background slowness model, $\bar{\mathbf{s}}$, to illustrate the nonlinear dependence of the scattering matrix and its adjoint on this model. Since this background slowness is assumed to be known, this dependence is dropped for notational convenience.

The process of migration is overdetermined because of the reduction in dimensionality between the data and the image, e.g. for $d = 2$ the three-dimensional data volume is mapped to a two-dimensional image. Consequently, the data vector exceeds the size of the model, i.e., $L \gg M$. Despite this overdeterminism, seismic images suffer from a degradation in the amplitudes, which can be corrected by including the inverse of the Hessian,

$$\mathbf{m} = -\mathbf{\Psi}^{\dagger} \nabla_{\mathbf{m}} J(\mathbf{m}) \tag{9}$$

with $\mathbf{\Psi} := \mathbf{K}^T\mathbf{K} \in \mathbb{R}^{M\times M}$. For an invertible Hessian and noise-free data, this expression is equivalent to the least-squares migration result presented in Eq. (3).

Instead of making (asymptotic) assumptions on the acquisition and the frequency that yield approximations for the inverse Hessian, a formulation will be derived that exploits an overcomplete signal representation to recover the amplitudes from noisy migrated data

$$\mathbf{y} = \mathbf{\Psi}\mathbf{m} + \mathbf{e} \tag{10}$$

that contains contributions from the Hessian both on the 'image' as well as on the noise. The expectation for the covariance of this noise term equals, $\mathbf{Cov}[\mathbf{e}, \mathbf{e}] := \mathbf{E}\{\mathbf{e}\mathbf{e}^T\} = \sigma^2\mathbf{\Psi}$ with $\mathbf{E}$ the expectation operators with respect to the probability density function of $\mathbf{n} \in N(0, \sigma)$. This covariance is proportional to the Hessian. The aim of this paper is to approximately invert the Hessian by diagonal preconditioning in the curvelet domain.

Zero-order imaging operators. In the high-frequency limit, the scattering matrix and the Hessian can under certain conditions on the medium and ray-geometry be considered as discretized version of FIO's [44]. Both $\mathbf{K}$ and $\mathbf{K}^T$ are for $d = 2$ FIO's of order $\frac{1}{4}$, while the leading behavior for their composition, the Hessian $\mathbf{\Psi}$, corresponds to that of an order-one invertible elliptic ΨDO. The action of the ΨDO on a nice function f is given by

$$\mathbf{T}f(\mathrm{x}) = \int e^{i\mathrm{k}\cdot\mathrm{x}} a(\mathrm{x},\mathrm{k})\hat{f}(\mathrm{k})\mathrm{dk}. \tag{11}$$

This operator is diagonal in the Fourier domain and involves a multiplication by the space-spatial frequency dependent symbol $a(\mathrm{x},\mathrm{k})$. Notice that this multiplication corresponds to a nonstationary convolution.

To make this ΨDO amiable to an approximation by curvelets, the following substitutions are made for the scattering operator and the model: $\mathbf{K} \mapsto \mathbf{K}\,(-\Delta)^{-1/2}$

and $\mathbf{m} \mapsto (-\Delta)^{1/2}\,\mathbf{m}$ with $((-\Delta)^{\alpha} f)\hat{}\,(\mathrm{k}) = |\mathrm{k}|^{2\alpha} \cdot \hat{f}(\mathrm{k})$. After these substitutions[1], the Hessian $\mathbf{\Psi}$ becomes zero-order. Before detailing the approximate diagonalization of the Hessian, we first discuss the sparsity of curvelets on seismic images.

Curvelet sparsity frames for seismic images. Seismic images (see e.g. Fig. 2) contain singularities along piecewise smooth curves. This sort of images are difficult to compress with respect to known orthonormal expansions such as Fourier and discrete wavelet transforms [7; 17]. The prototype waveforms that make up these transforms are not rich enough to sparsely represent seismic images, they either lack a multiscale structure or directionality.

Recently developed curvelet transforms compose 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 transforms obtains near optimal sparsity on the bandwith-limited[2] imaged reflectivity. The curvelet transform is overcomplete because the number of rows with waveforms exceeds the number of samples in the image. By using the fast discrete curvelet transform [FDCT by warping 5], the image is perfectly reconstructed after decomposition by applying the adjoint of the curvelet transform, i.e., we have $\mathbf{r} = \mathbf{C}^T\mathbf{C}\mathbf{r}$ for arbitrary $\mathbf{r}$. The curvelet transform matrix and its adjoint are given by $\mathbf{C}$ and $\mathbf{C}^T$ and these transforms define the sparsity synthesis and analysis matrices according $\mathbf{A} := \mathbf{C}^T \in \mathbb{R}^{M\times N}$ and $\mathbf{A}^T := \mathbf{C}$. For this choice of curvelet transform, the pseudo inverse equals the adjoint, i.e., $\mathbf{C}^T = \mathbf{C}^{\dagger}$, which means that the FDCT by warping is a numerical isometry, i.e., the collection of curvelets in the overcomplete

[1] Alternatively, the operators can be made zero-order by composing the data side with a 1/2-order fractional integration along the time coordinate, i.e., $\mathbf{K} \mapsto \partial_t^{-1/2}\mathbf{K}$ [see e.g. 2].

[2] Because of bandlimitation of the source, seismic data and hence seismic images are always limited in bandwidth, yielding an imaged reflectivity that is relatively smooth in the direction along the reflectors and oscillatory across.

signal representation $\mathbf{A}$ forms a tight frame with moderate redundancy (a factor of roughly 8 for $d = 2$). Even though the energy is preserved $\|\mathbf{r}\| = \|\mathbf{Cr}\|$, the curvelet representation is overcomplete ($N >> M$) and hence $\mathbf{CC}^T$ is a projection, which makes it difficult to recover the sparsity vector $\mathbf{x}_0$ from $\mathbf{r} = \mathbf{C}^T\mathbf{x}_0$.

Recent work on overcomplete signal representations [see e.g. 39; 24] has shown that ℓ_1-norm minimization leads to a stable recovery of a vector $\mathbf{x}_0$ that is approximately sparse. An image admits an approximately sparse composition $\mathbf{r} = \mathbf{Ax}_0$ when the magnitude sorted entries of the sparsity vector $\mathbf{x}_0$ display a power-law decay [see e.g. 24; 9]. The faster this decay, the faster the decay for the nonlinear approximation error. The nonlinear approximation error is given by the energy difference between the original function and its reconstruction using the largest entries in the sorted sparsity vector. The faster this error decays as a function of the number of largest entries, the higher the nonlinear approximation rate.

For two-dimensional functions ($f \in \mathbb{R}^2$) that contain singularities along piecewise twice differentiable curves, e.g., faults and pinchouts are allowed, curvelets obtain a near optimal nonlinear approximation rate [see e.g. 7]. Besides log-terms this rate is $\mathcal{O}(P^{-2})$ for curvelets and $\mathcal{O}(P^{-1/2})$ for Fourier with P the number of largest entries. This rate for the curvelets is the same as the optimal rate. Seismic images arguably belong to this class of functions as one can see from the reflectivity for the EAGE Marmoussi model plotted in Fig. 2. The near optimality for the above infinite dimensional case, translates in practice to a fast decay for the approximation error of the empirical curvelet coefficients. Fig. 2 shows that the decay for the approximation error of the EAGE Marmoussi model is clearly faster for the FDCT than for approximations based on the Fourier and the discrete wavelet transforms.

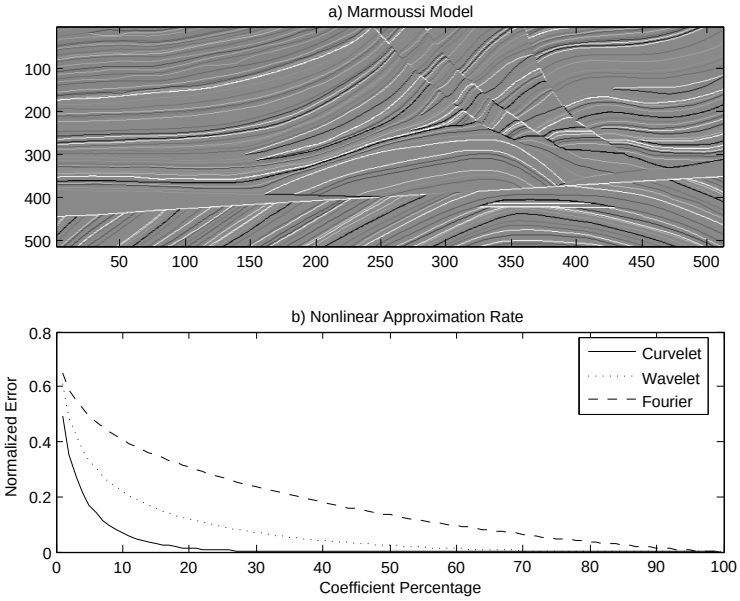

FIG. 2. Nonlinear approximation rate for the curvelet transform compared to the Fourier and discrete wavelet transforms. **(a)** reflectivity of the EAGE Marmousi model; **(b)** nonlinear approximation rate for curvelets (solid), Fourier (dashed) and discrete wavelets (dotted). The decay for the curvelet transform clearly compares favorably to these other two transforms.

Diagonal approximation of the Hessian. The approach we are taking to recover the migration amplitudes revolves around the following approximation

$$\mathbf{\Psi r} \simeq \mathbf{C}^T \mathbf{\Gamma}[\mathbf{r}_0]\mathbf{\Gamma}[\mathbf{r}_0]\mathbf{C r} := \mathbf{A}\mathbf{A}^T\mathbf{r} \tag{12}$$

for the Hessian on a proper[3] test vector $\mathbf{r}$. Suppose that this approximate relation (denoted by the symbol $\simeq$) holds for a particular test vector $\mathbf{r}$ to within the noise level, then the above definitions for the sparsity synthesis and analysis matrices ($\mathbf{A} := \mathbf{C}^T\mathbf{\Gamma}$ and $\mathbf{A}^T := \mathbf{\Gamma C}$) bring us in the position to formulate a stable recovery scheme for the amplitudes of the seismic image. Not only, can the sparsity of the curvelet transform be exploited but the above approximation also allows for a stable approximate inverse for the Hessian. The recovery requires knowledge of the

[3]Not 'too far' from the reference vector $\mathbf{r}_0$, used to compute the diagonal weighting matrix $\mathbf{\Gamma}[\mathbf{r}_0]$.

diagonal weighting matrix, $\mathbf{\Gamma}[\mathbf{r}_0]$, which for now is assumed to exist and calculable. The following three arguments support this claim.

First, there is the argument that diagonal approximations of the Hessian, provide, under certain conditions on the acquisition and velocity model, an adequate approximation for the Hessian [see e.g. 34, and the references therein]. This assumption underlies migration preconditioning, used to compute the pseudo inverse of the scattering operator [28].

Second, there is the method of illumination-based normalization [36], dating back to ideas by Symes and reported by [13]. In that approach, the inverse of the Hessian is approximated by a diagonal weighting matrix which plays the role of a preconditioning. This weighting matrix is calculated for a particular reference vector $\mathbf{r}_0$ and involves an additional de-migration and migration step. The diagonal is estimated by 'dividing' the demigrated-migrated image by this reference vector. The closer this reference vector $\mathbf{r}_0$ is to the unknown image, the better the approximation. Provided $\mathbf{\Psi}$ is reasonably well behaved, the diagonal approximation for the inverse is accurate when the reference vector is sufficiently close to the unknown image. In practice, the migrated image $\mathbf{y}$ serves as a good guess for the unknown reference vector. In that sense, this approach is somewhat reminiscent of Krylov-subspace methods.

The third and most important argument is formed by the invariance of curvelets with respect to high-frequency wave propagation, a statement that dates back to the work of [40] and [38]. Curvelets, by virtue of their multiscale and multidirectional construction, are well suited to capture the evolution of wavefronts. Their second dyadic partitioning, the partitioning of the dyadic coronae into angular wedges [40; 38; 5], and their parabolic scaling relation, guarantee that curvelets remain invariant under the action of FIO's. These FIO's represent high-frequency solutions of the wave equation and the invariance translates into a dispersion-free propagation of curvelets. This property was used by [26; 21] and holds for smoothly

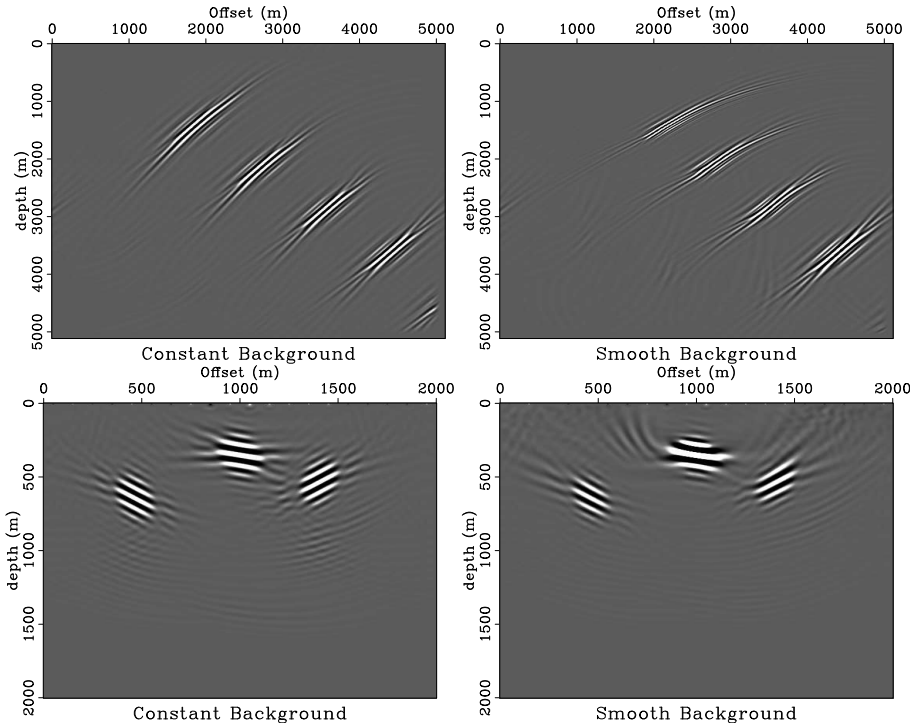

FIG. 3. Invariance of curvelets under imaging. **(a-b)** Snapshots for different time evolutions for the solution of the acoustic wave equation with an initial condition given by a spatial source distribution given by a single curvelet fired at $t = 0$. **(a)** background medium with a constant velocity; **(b)** smoothly varying velocity. (**c-d**) Hessian on a relflectivity given by a few curvelets. (**c**) constant background velocity; **(d)** smoothly background velocity.

varying media, where curvelets remain curvelet-like for small propagation distances compared to their essential support. Indeed, recent work by [6] shows that the time-evolution operator for wave propagation is diagonal dominant in the curvelet domain. This result [Theorem 1.1 of 6, to be precise] serves as our main argumentation for the above approximation of the Hessian. This argument is powerful because it makes no assumptions on the reference vector. Moreover, the Hessian is a reasonably well-behaved zero-order ΨDO, which corresponds to a FIO with a trivial phase function. The action of this ΨDO corresponds to that of a nonstationary filter (cf. Eq. (11)) which does not move the imaged reflectors, only changes their amplitudes. This filter is approximated in the curvelet domain by a diagonal

weighting with the square of the $\mathbf{\Gamma}$. The entries on the diagonal of the $\mathbf{\Gamma}$ correspond to the quasi-singular values, which play a predominant role in [29]'s WVD method. Since the ΨDO is zero-order, the 'wavelets' and 'vaguelettes' each correspond to curvelets. The quasi-singular values are absorbed in the weighted synthesis and analysis matrices $\mathbf{A}$ and $\mathbf{A}^T$.

The actual computation of the diagonal weighting matrix $\mathbf{\Gamma}$ is more involved than a simple 'division', and includes the solution of a regularized least-squares problem. The cost of the algorithm is one demigration-migration for each reference vector. The results in this paper are based on a diagonal estimated for one reference vector only. Details of this estimation are deferred to a separate paper [31]. By itself, the diagonal approximation in Eq. (12) is not new and was first introduced in [26]. In that work, the diagonal was calculated with an expensive Monte-Carlo sampling technique that involved multiple migrations with white noise realizations. Throughout this paper, the $\mathbf{\Gamma}$ is assumed to be given.

Examples. A number of numerical experiments are conducted to further substantiate the above claim. In Fig. 3, snapshots are shown for the time evolution of the solution of the acoustic wave equation for a constant and a smoothly varying velocity model. The initial condition is given by a spatial source distribution given by a single curvelet that is fired at $t = 0$. Fig. 3 (**a-b**) show that the curvelets indeed remain curvelet-like for moderate propagation distances. The invariance of the curvelets under the Hessian is verified by demigating and subsequently migrating the reflectivity consisting of three curvelets with a 'wave-equation' migration code. Fig. 3 (**c-d**) shows that the curvelets remain curvelet like under the action of the Hessian for a constant and smoothly varying velocity model.

To test the validity of the diagonal approximation in Eq. (12), we conduct an experiment based on the velocity model and reflectivity depicted in Fig. 4 (**a-b**). This figure includes a velocity model consisting of a single low-velocity lens, representa-

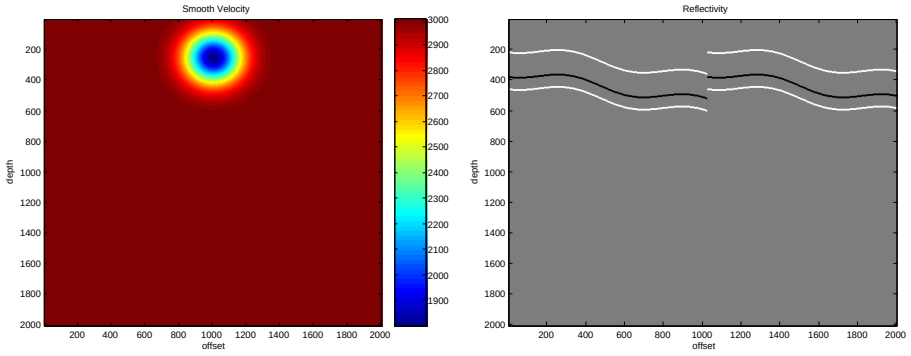

FIG. 4. The smooth velocity model and broadband reflectivity. **(a)** The smooth background model with the low-velocity lens; **(b)** The broadband reflectivity.

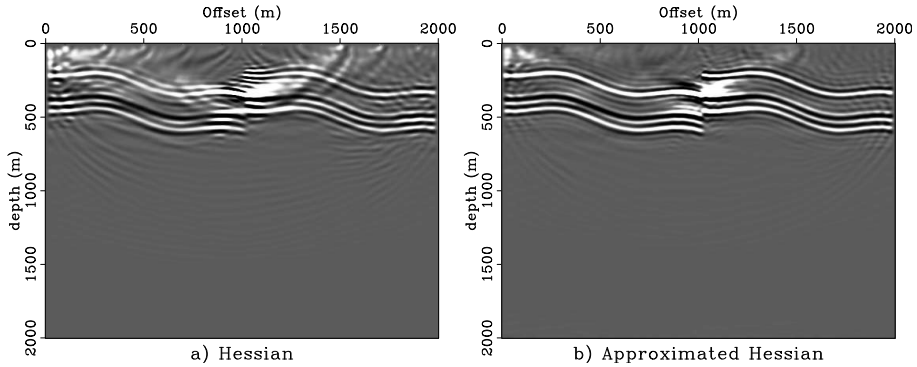

FIG. 5. Approximation of the Hessian. **(a)** Action of the Hessian on the reflectivity; **(b)** The same but approximated according Eq. (12). Notice that the main characteristics of the Hessian are captured. The relative ℓ_2-error between the actual and the approximate Hessian is approximately 20 %.

tive of a gas lens in the overburden, and a broadband reflectivity consisting of three curved reflectors with a fault. From this reflectivity data is generated and subsequently imaged by migration. This image is used for the reference vector which is used to calculate the $\mathbf{\Gamma}$. In Fig. 5 a comparison is made between, the action of the Hessian on this reference vector $\mathbf{r}_0$ and our diagonal approximation (cf. Eq. (12)). Fig. 5 shows that our approximation captures most of the Hessian. The relative ℓ_2-error between the actual and the approximate Hessian is approximately 20 %.

The observed invariance of curvelets under the Hessian and hence the ability to approximate the Hessian with a diagonal weighting in the curvelet domain is key

to our nonlinear formulation of the seismic amplitude recovery problem. As with the wavelet-vaguelette decomposition (WVD) or quasi-SVD methods proposed by [29; 22; 30], the fact that curvelets are near eigenfunctions of the Hessian is used to invert the Hessian nonlinearly, using not only the invariance property but also the near sparsity of curvelets on the model. Before deriving the recovery scheme, we first briefly present the canonical denoising problem as a recovery problem.

3. STABLE SIGNAL RECOVERY

The denoising problem is first discussed for orthonormal basis and then for overcomplete sparsity representations.

3.1. The denoising problem

Separating a deterministic signal component from incoherent, typically Gaussian noise, is the canonical inverse problem where one aims to recover the unknown image $\mathbf{m}$ from noisy data given by

$$\mathbf{y} = \mathbf{m} + \mathbf{n} \tag{13}$$

with the noise term as defined before.

Following recent developments in theoretical image processing, the image $\mathbf{m}$ can be recovered from noisy and incomplete data (e.g. by considering only a subset [see e.g. 9] of the measurement vector $\mathbf{y}$) when the image permits a sparse representation, i.e.,

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \tag{14}$$

for a sparsity vector $\mathbf{x}_0$ with its sorted entries decaying rapidly. This recovery involves the solution of the following convex optimization program

$$\mathbf{P}_1 : \quad \begin{cases} \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \hat{\mathbf{m}} = \mathbf{A}\mathbf{x}, \end{cases} \tag{15}$$

which is remarkable robust under noise and missing data. As long as $\mathbf{A}$ adheres to certain conditions, the solution of this unconstrained optimization problem lies to within the noise level [see e.g. 9; 24]. The optimization problem $\mathbf{P}_1$ is known as the constrained variation of the basis-pursuit denoising algorithm (BPDN) [11].

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 $\mathbf{n}$. Since $n_{1\cdots M} \in N(0, \sigma^2)$, 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 $n \cdot \sigma^2$ and variance $\sqrt{2M} \cdot \sigma^2$. By choosing $\epsilon^2 = \sigma^2(M + \nu\sqrt{2M})$ with $\nu = 2$, we remain within the mean plus or minus two standard deviations.

3.2. Orthonormal sparsity representation

In case $\mathbf{m}$ permits a sparse representation in an orthonormal basis, i.e., $\mathbf{A} := \mathbf{S}^T$, $\mathbf{P}_1$ permits an explicit solution. For instance, this is the case when $\mathbf{S} = \mathbf{W}$ with $\mathbf{W}$ an orthonormal discrete wavelet transform. To establish this observation, we first associate with this optimization problem the following unconstrained optimization problem

$$\hat{\mathbf{m}} = \arg\min_{\mathbf{m}} \|\mathbf{y} - \mathbf{m}\|_2^2 + \lambda\|\mathbf{S}\mathbf{m}\|_1 \tag{16}$$

with $\lambda = \sqrt{\epsilon^2/(M + \nu\sqrt{2M})} \cdot \sqrt{\log 2M} = \sigma \cdot \sqrt{\log 2M}$ [18; 11]. Because of the orthonormality of $\mathbf{S}$, Eq. (16) can also be written in the wavelet domain

$$\begin{cases} \min_{\mathbf{v}} \|\mathbf{u} - \mathbf{v}\|_2^2 + \lambda\|\mathbf{v}\|_1 \\ \hat{\mathbf{m}} = \mathbf{S}^T\hat{\mathbf{v}} \end{cases} \tag{17}$$

with $\mathbf{u} := \mathbf{S}\mathbf{y}$ and $\mathbf{v} := \mathbf{S}\mathbf{m}$. This optimization problem can be solved by a single soft-thresholding operation [see e.g. 30; 14]

$$\hat{\mathbf{m}} = \mathbf{S}^T S_\lambda^s(\mathbf{S}\mathbf{y}), \tag{18}$$

with the soft thresholding operator given by

$$S^s_\lambda(x) := \begin{cases} x - \text{sign}(x)\lambda & |x| \geq \lambda \\ 0 & |x| < \lambda. \end{cases} \tag{19}$$

3.3. Overcomplete sparsity transforms

Denoising based on orthonormal transforms often does not give the most pleasing results. Compared to decimated orthonormal wavelets, non-decimated wavelets are known to give superior denoising results for functions with point-singularities. For non-decimated wavelets, the synthesis matrix contains more columns (wavelets) than rows (data points), i.e., $\mathbf{A} := \mathbf{S}^\dagger = \mathbf{W}^T \in \mathbb{R}^{M \times N}$ with $N = M \cdot \log M \gg M$ is overcomplete. The recovery of the sparsity vector becomes underdetermined and Eq.'s (16) and (17) are no longer equivalent, an observation also made by [23].

Since the constrained optimization problem $\mathbf{P}_1$ extends to overcomplete representations, this formulation is used to recover the sparsity vector $\mathbf{x}_0$. For seismic images, the curvelet transform is arguably a good choice. By virtue of the tightness of the FDCT, we have $\mathbf{A} := \mathbf{S}^\dagger$ with $\mathbf{S} := \mathbf{C}$ or $\mathbf{A} = \mathbf{C}^T$.

Following [24], the constrained optimization problem ($\mathbf{P}_1$), is replaced by a series of simpler unconstrained optimization problems

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

These optimization problems depend on the Lagrange multiplier λ, which is not known. 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. The optimal $\hat{\mathbf{x}}$ is found for the largest λ for which $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2 \leq \epsilon$. 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 14]. We refer to [20; 45] for the recovery conditions for Eq.'s (15) and (20).

3.4. Solution by the cooling method based on iterative thresholding

Following [14; 24] and ideas dating back to [25], Eq. (20) is solved by an iterative thresholding technique that derives from the Landweber descend method. After m iterations of the outer cooling loop, the estimated coefficient vector is computed for fixed λ by the following inner loop

$$\mathbf{x}^{m+1} = S_{\lambda}^{s}\left(\mathbf{x}^{m} + \mathbf{A}^{T}(\mathbf{y} - \mathbf{A}\mathbf{x}^{m}\right) \tag{21}$$

with $\lambda = \lambda_m$. As shown by [14], 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 synthesis and subsequent analysis. The details of the cooling algorithm are presented in Table. 1.

Initialize:

$m = 0$; $\mathbf{x}^0 = \mathbf{A}^T\mathbf{y}$;

Choose: L, $\lambda_1 > \lambda_2 > \cdots$

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

  $m = m + 1$;

  $\mathbf{x}^m = \mathbf{x}^{m-1}$;

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

    $\mathbf{x}^m = S_{\lambda_m}^{s}\left(\mathbf{x}^m + \mathbf{A}^T\left(\mathbf{y} - \mathbf{A}\mathbf{x}^m\right)\right)$ {Iterative thresholding}

  end for

end while

$\hat{\mathbf{m}} = \mathbf{A}\mathbf{x}^m$.

TABLE 1
The cooling method with $\lambda_1 > \lambda_2 > \cdots$ the series of decreasing Lagrange multipliers. The inner loop is repeated L times.

4. SEISMIC AMPLITUDE RECOVERY BY NONLINEAR OPTIMIZATION

Let us now return to the seismic imaging problem by considering the normal equation

$$\mathbf{y} = \mathbf{\Psi m} + \mathbf{e}, \tag{22}$$

defining the input to our amplitude recovery algorithm. Compared to the above denoising problem for orthonormal sparsity representations, there are two complications, namely, there is an operator involved and second the noise term $\mathbf{e} = \mathbf{K}^T\mathbf{n}$ is colored. It is shown that by preconditioning with an orthonormal basis the recovery of $\mathbf{m}$ can again be formulated as a thresholding problem. This formulation forms the basis for our extension, valid for overcomplete curvelets.

4.1. Preconditioned formulation for orthonormal transform

To start the argument, suppose that Eq. (12) holds exactly for a sparsity representation based on an orthonormal basis, say wavelets. In that case, we have on a test vector $\mathbf{r}$

$$\mathbf{\Psi r} = \mathbf{A}\mathbf{A}^T\mathbf{r} \tag{23}$$

with $\mathbf{A} := \mathbf{W}^T\mathbf{\Gamma}$. Since $\mathbf{A} := \mathbf{S}^\dagger$, we find that

$$\mathbf{\Psi}^{-1}\mathbf{r} = \mathbf{S}^T\mathbf{S}\mathbf{r} \tag{24}$$

for the diagonal elements in $\mathbf{\Gamma}$ sufficiently bounded away from zero and $\mathbf{S} = \mathbf{A}^\dagger = \mathbf{\Gamma}^{-1}\mathbf{W}$. This latter expression follows from the orthogonality of the wavelet transform, i.e.,$\mathbf{W}^\dagger = \mathbf{W}^T$, and the assumed invertability of the $\mathbf{\Gamma}$. By making the following substitutions for the modeling operator, $\mathbf{K} \mapsto \mathbf{F} = \mathbf{K}\mathbf{S}^T$, and the model, $\mathbf{m} \mapsto \mathbf{x} = \mathbf{A}^T\mathbf{m}$, a preconditioned normal equation is obtained

$$\begin{aligned} \mathbf{F}^T\mathbf{d} &= \mathbf{F}^T\mathbf{F}\mathbf{x} + \mathbf{F}^T\mathbf{n} \\ \mathbf{u} &= \mathbf{I}\mathbf{v} + \tilde{\mathbf{n}}. \end{aligned} \tag{25}$$

By virtue of the isometry $\mathbf{F}^T\mathbf{F} = \mathbf{I}$, which follows from Eq. (23) and the orthonormality of $\mathbf{W}$, this expression regains the form of the canonical denoising problem discussed earlier. The orthonormality of $\mathbf{W}$ implies $\mathbf{W}\mathbf{K}^T\mathbf{K}\mathbf{W}^T = \mathbf{\Gamma}^2$ and from that it follows that the noise term $\tilde{\mathbf{n}} = \mathbf{F}^T\mathbf{n}$ is again white, i.e., $\mathbf{Cov}[\tilde{\mathbf{n}}, \tilde{\mathbf{n}}] = \sigma^2\mathbf{I}$.

By setting λ as before, the sparsity vector and the amplitude-corrected image are recovered by solving

$$\begin{cases} \min_{\mathbf{v}} \|\mathbf{u} - \mathbf{v}\|_2^2 + \lambda\|\mathbf{v}\|_1 \\ \hat{\mathbf{m}} = (\mathbf{A}^T)^\dagger \hat{\mathbf{v}}. \end{cases} \tag{26}$$

Comparing these expressions with those in Eq.(17) shows that these expressions are the same except for the inverse transform now defined through the pseudo inverse. The solution of Eq. (26) can again be written in terms of a soft-thresholding operation

$$\hat{\mathbf{m}} = \left(\mathbf{A}^T\right)^\dagger S_\lambda^s\left(\mathbf{u}\right) = \mathbf{S}^H S_\lambda^s\left(\mathbf{S}\mathbf{y}\right), \tag{27}$$

where the data is migrated with the preconditioned migration operator, $\mathbf{u} = \mathbf{F}^T\mathbf{d}$, followed by a recovery through soft thresholding and a reconstruction by the pseudo inverse. By virtue of Eq.(24), this sequence of operations is equivalent to decomposing the migrated data, $\mathbf{y}$, with respect to the weighted sparsity transform $\mathbf{S}$, followed by the same soft-thresholding operation, and an inverse transform with $\mathbf{S}^T$. According to Eq. (24), applying the weighted forward and inverse transforms without the thresholding, corresponds to inverting the Hessian in two steps, namely, by the weighted forward and inverse transforms. The inserted thresholding removes the noise (cf. Eq. (18)). Even though the noise is initially colored, the thresholding is still applicable because the weighted transform whitens the spectrum of $\mathbf{e}$, i.e., $\mathbf{Cov}[\mathbf{Se}, \mathbf{Se}] = \sigma^2\mathbf{I}$. This property holds by virtue of Eq.'s (23), (24) and the above isometry.

Clearly, the above estimation is reminiscent of the WVD-based thresholding. Because the Hessian is zero-order the vaguelettes correspond to wavelets and Eq.(27)

corresponds to the expression for the WVD estimator. The above thresholding result was first presented in [26], as an approximate solution to the recovery of the amplitudes in seismic images. These results were presented for non-orthonormal curvelets and for the case where Eq. (23) only approximately holds. Even though, thresholding techniques can still be applied to transforms other than orthonormal [as shown by 23], diagonally-weighted overcomplete sparsity transforms are no longer tight frames with normalized columns, requiring a different approach. In the next section, a formulation is presented that replaces the thresholding by nonlinear optimization, which accounts for the overcompleteness of the weighted curvelet frame.

4.2. Preconditioned formulation for overcomplete curvelet frames

So far, it was established that the FDCT provides a transform well adapted to seismic imaging. Not only are seismic images sparse in the curvelet domain, the curvelets also remain curvelet like under the Hessian. Question now is how to incorporate these results in a formulation of the seismic amplitude recovery problem given the inexact nature of Eq. (23) and the overcompleteness of the curvelet transform. For reasons similar to the denoising problem, the coefficient vector can no longer be recovered by thresholding. Instead, the recovery requires the solution of a constrained optimization problem with the weighted synthesis matrix defined as $\mathbf{A} := \mathbf{C}^T\mathbf{\Gamma}$, with a diagonal weighting $\mathbf{\Gamma}$ that may contain small entries on its diagonal. Both the overcompleteness and small entries result in a sparsity transform matrix $\mathbf{S}$, involving an pseudo inverse, that can no longer be stably calculated and Eq. (24) no longer applies.

As demonstrated with the denoising problem, the recovery of the sparsity vector for an orthonormal $\mathbf{A}$ corresponded to taking the forward transform, followed by applying a soft-thresholding operation and the inverse transform (cf. Eq. (18)). These results carried over to thresholding the migrated data, by virtue of the precondition-

ing and orthonormality of the sparsity transform (cf. Eq. (27)). By analogy with the denoising problem, where the explicit solution by thresholding is replaced by the unconstrained optimization problem for overcomplete transforms, we propose to recover the amplitude-corrected image by solving the following unconstrained optimization problem for decreasing λ's

$$\mathbf{P}'_\lambda : \quad \begin{cases} \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ \hat{\mathbf{m}} = \left(\mathbf{A}^T\right)^\dagger \hat{\mathbf{x}}. \end{cases} \tag{28}$$

The first line of this program corresponds to the ℓ_1-norm penalized inverse of $\mathbf{A}$, which can be associated with the weighted forward transform and thresholding in Eq. (27). In both cases, the sparsity vector is recovered by minimizing the ℓ_1-norm subject to that the data is matched in the ℓ_2-sense to within the tolerance ε. The subsequent inverse transform by $\mathbf{S}^T$ for the thresholding is analogous to the second line in Eq. (28), where the inverse transform is computed via $\left(\mathbf{A}^T\right)^\dagger$. The claim is that $\mathbf{P}'_\lambda$ is analogous to the preconditioned system defined for the orthonormal case. We use this correspondence as the primary motivation for our approximate formulation of the amplitude recovery problem by nonlinear optimization of $\mathbf{P}'_\lambda$.

The main assumption is that the curvelets remain approximately sparse for the imaged data, i.e., the imaged data again permits a sparse representation $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}$ with the sorted entries in $\mathbf{x}_0$ rapidly decaying. This sparsity vector only exist if curvelets remain sufficiently invariant under the Hessian. Each of the two steps – finding the sparsest set of coefficients and the subsequent reconstruction – corrects for the amplitudes by approximately inverting the 'square-root' of the Hessian. The minimization of $\mathbf{P}'_\lambda$ only enhances the sparsity aspect of $\mathbf{P}$. In the next section, an additional continuity-promoting penalty functional is included that removes possible artifacts.

5. CONTINUITY-ENHANCED SEISMIC IMAGE RECOVERY

The above formulation provides a stable framework for the approximate recovery of the amplitudes for seismic images $\mathbf{m}$ without the necessity of multiple evaluations of the Hessian. Of course, this recovery is only accurate when Eq. (12) provides a sufficiently accurate approximation to the Hessian. The accuracy of this approximation depends on the complexity of the background model, the acquisition geometry and the closeness of the reference $\mathbf{r}_0$ to the actual unknown model $\mathbf{m}$.

Since we do not have control over either of these factors, the issue remains how to limit spurious curvelet artifacts. These artifacts are either related to so-called "pseudo Gibb's" phenomena (or better side-band effects [8], inherent to the curvelet or other harmonic transforms), or to instabilities caused by small entries in the diagonal of $\mathbf{\Gamma}$. These small entries are typically due to lack of insonification, i.e. source-/receiver combinations in the data that do no lead to wave paths that reach every point in the subsurface.

To reduce these spurious artifacts, the sparsity enhancing penalty functional in Eq. (28) is complemented with a penalty functional that enhances the continuity of the image along the imaged reflectors. For this purpose, a 'wavefront-set' enhancing penalty functional is introduced. This functional exploits the predominant relative (piece-wise) smoothness of reflectors in the tangential direction. To promote this feature and to preserve the oscillatory behavior in the normal direction, an anisotropic smoothing technique is proposed that differs from commonly used edge-preserving penalty functionals such as total variation (TV) [see e.g. 12; 37]. Instead, an anisotropic smoothing technique is used that penalizes fluctuations in the tangential direction only. This anistropic smoothing is performed by defining the following anisotropic penalty term as part of the optimization,

$$J_c(\mathbf{m}) = \|\mathbf{\Lambda}^{1/2}\nabla_d \mathbf{m}\|_p \tag{29}$$

with ∇_d the discretized gradient matrix. The matrix $\mathbf{\Lambda}$ is location dependent (see Fig. 6 for an example) and rotates the gradient towards the approximate tangent

to the reflecting surface. This rotation matrix depends either on the smooth background velocity model, ($\bar{\mathbf{b}} := \bar{\mathbf{c}} = \bar{\mathbf{s}}^{-1}$), or on the smoothed reference vector $\bar{\mathbf{b}} := \mathbf{r}_0$ and is given by

$$\mathbf{\Lambda}[\bar{\mathbf{b}}] = \frac{1}{\|\nabla_d \bar{\mathbf{b}}\|_2^2} \left\{ \begin{pmatrix} \mathbf{D}_2\bar{\mathbf{b}} \\ -\mathbf{D}_1\bar{\mathbf{b}} \end{pmatrix} \begin{pmatrix} \mathbf{D}_2\bar{\mathbf{b}} & -\mathbf{D}_1\bar{\mathbf{b}} \end{pmatrix} \right\}. \tag{30}$$

with $\mathbf{D}_i$ the discretized derivative in the i^{th} coordinate direction. For $p = 2$, the penalty term in Eq. 29 corresponds to anisotropic diffusion [4], penalizing the ℓ_2-norm of the rotated discrete gradient on the image. For $p = 1$, the penalty functional corresponds to an anisotropic TV, which penalizes the ℓ^1-norm of the gradient in the direction along the reflectors. Both penalty terms are active only in the tangential direction and differ from conventional norms that penalize fluctuations in the normal direction. In our case, those penalties could potentially remove the imprint of the fine-structure of the Earth's reflectors by eliminating the oscillations. The two approaches differ in the sense that anistropic diffusion tends to smooth jumps in the amplitude along the reflections while anistropic TV preserves these discontinuities. In this paper, we will limit ourselves in the examples to $p = 2$.

6. SPARSITY- AND CONTINUITY-ENHANCED SEISMIC IMAGE RECOVERY

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

$$\mathbf{P}: \quad \begin{cases} \min_{\mathbf{x}} J(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \hat{\mathbf{m}} = \left(\mathbf{A}^T\right)^\dagger \mathbf{x} \end{cases} \tag{31}$$

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

$$J(\mathbf{x}) = \alpha J_s(\mathbf{x}) + \beta J_c(\mathbf{x}), \quad \alpha + \beta = 1 \tag{32}$$

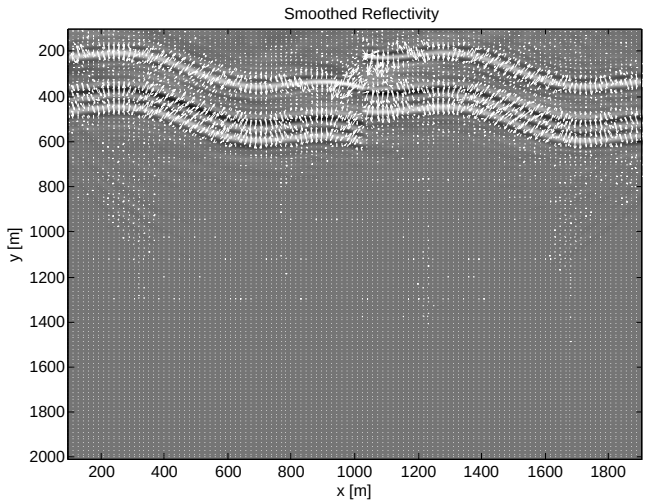

FIG. 6. Gradient for the smoothed migrated image (white ↑'s). The direction perpendicular to the gradient defines the tangent direction allong which the additional anisotropic smoothing is applied.

with α, $\beta \geq 0$. The $J_s(\mathbf{x}) = \|\mathbf{x}\|_{1,\mathbf{W}} := \sum_{j=1}^{N} |w_j x_j|$ is the weighted ℓ_1-norm with $w_j \geq 0$, $j = 1 \cdots N$, which is used to account for limitations in the acquisition, limiting the spatial frequency content of the image. Instabilities related to this limitation are removed by heavily penalizing the entries in $\mathbf{x}$ that correspond to the locations where the diagonal entries of $\mathbf{\Gamma}$ are small. The weights of $\mathbf{w}$ are set according to $w_j = \infty$ for $j \in \mathcal{M}_0$ and $w_j = 1$ otherwise with $\mathcal{M}_0 = \{j : \Gamma_j \leq \delta\}$ with δ a small threshold value.

The second term in the penalty term is given by $J_c(\mathbf{x}) = \|\mathbf{\Lambda}^{1/2} \left(\mathbf{A}^T\right)^\dagger \mathbf{x}\|_p$ with $p = 1$ or $p = 2$. This term corresponds to Eq. (29) and includes a synthesis with the pseudo inverse of $\mathbf{A}^T$, because the optimization is carried out over $\mathbf{x}$ and not over the model vector $\mathbf{m}$. This pseudo inverse is calculated with the LSQR-algorithm [33] and typically requires only a few iterations. The scaling term μ is determined by a line search and depends on the emphasis of the sparsity ℓ_1-norm penalty term with respect to the continuity penalty term.

The above nonlinear optimization problem (**P**) is again solved with a cooling method as discussed in section 3.1. This method consists of a series of thresh-

olded Landweber iterations that solve a series of unconstrained subproblems for decreasing λ. Since this method only requires knowledge on 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 $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}$, 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 $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$:

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{x}) ; \tag{33}$$

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

$$\mathbf{x} \leftarrow S^s_{\lambda \mathbf{W}}(\mathbf{x}); \tag{34}$$

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

$$\mathbf{x} \leftarrow \mathbf{x} - \mu \nabla_{\mathbf{X}} J_c(\mathbf{x}) \tag{35}$$

with

$$\nabla_{\mathbf{X}} J_c(\mathbf{x}) = 2\mathbf{A}^\dagger \nabla \cdot \left(\mathbf{\Lambda} \nabla \left(\left(\mathbf{A}^T \right)^\dagger \mathbf{x} \right) \right). \tag{36}$$

Remember that steps 1 & 2 for $\|\mathbf{A}\| \leq 1$ converge to the solution of Eq. (28) for a fixed λ. Because a solution based on these steps only, may contain additional entries compared to $\mathbf{x}_0$, the continuity-enhancing penalty is minimized as well. This additional anisotropic penalty term, removes artifacts related to regions in the curvelet space that are badly illuminated, i.e., those regions that correspond to small entries in the diagonal $\mathbf{\Gamma}$, or to side-band effects. During step 3, the coefficients are updated according to the gradient of the anisotropic diffusion norm. The μ is found by conducting a line search

$$\min_{\mu} \frac{1}{2}\|\mathbf{y} - \mathbf{A}(\mathbf{x} - \mu \nabla_{\mathbf{X}} J_c(\mathbf{x}))\|_2^2 + \lambda\alpha \|\mathbf{x} - \mu \nabla_{\mathbf{X}} J_c(\mathbf{x})\|_1 + \lambda\beta J_c(\mathbf{x} - \mu \nabla_{\mathbf{X}} J_c(\mathbf{x})). \tag{37}$$

In Table 2, the different steps for our algorithm are given. Typically, a linear decay is chosen for the λ's.

Initialize:

$m = 0$;

$\mathbf{x}^0 = \mathbf{A}^T\mathbf{y}$;

$\mathbf{y} = \mathbf{K}^T\mathbf{d}$;

Choose:

M and L

$\lambda_1 > \lambda_2 > \cdots > \lambda_M$

while $\|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}\|_2 > \epsilon$ and $m < M$ **do**

  $m = m + 1$;

  $\mathbf{x}^m = \mathbf{x}^{m-1}$;

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

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

  end for

  Line search according to Eq. 37;

  $\mathbf{x}^m = \mathbf{x}^m - \mu\nabla_{\mathbf{x}^m} J_c(\mathbf{x}^m)$;

end while

$\hat{\mathbf{m}} = \left(\mathbf{A}^T\right)^\dagger \hat{\mathbf{x}}$.

TABLE 2

Sparsity-and continuity-enhancing recovery of seismic amplitudes.

7. NUMERICAL EXPERIMENTS

The proposed algorithm is tested on the data set generated by the three reflectors included in Fig. 4 (**b**). Noise was added to yield a signal-to-noise ratio of 0 dB. The background veloxity model depicted in Fig. 4 (**a**) contains a lens and has a grid spacing of 4 m in each direction. The size of the model is 2×2 km. The synthetic

data set is made out of 20 equally-spaced shots with 100 m spacing and 501 equally-spaced receivers on each grid point. Each receiver records 512 samples with 4 ms sample rate. The source function is given by a trapezoidal filter with a central frequency of 50 Hz. The results of the migrated image are depicted in Fig. 7 (**a**) and contains imaging artifacts as well as amplitude degradation. Amplitudes recovered with the sparseness promoting weighted penalty functional ($\beta, \mu = 0$) are presented in Fig. 7 (**b**). The amplitudes are recovered with our algorithm and the image contains less image artifacts. The result for the joined sparsity- and continuity-enhanced image recovery are presented in Fig. 7 (**c**). Remaining image artifacts are removed by the anisotropic diffusion and the amplitudes display, as expected, less variations along the imaged reflectors. The resolution is also improved as part of our procedure. As can be observed from Fig. 8, the quality of the data, generated by demigrating the image found by solving the nonlinear optimization program **P** is significantly improved. This visual improvement is reflected in the signal-to-noise ratio which improves from 0 dB to 10 dB. The results where obtained for $\alpha = 0.6$ and $\beta = 0.4$.

8. CONCLUSIONS

The method presented in this paper combines the sparsity of curvelets and their invariance under the Hessian to formulate a stable recovery for the seismic amplitudes. During this recovery the Hessian is approximately inverted. Compared to other approaches for 'true-amplitude' migration, our method not only imposes fewer conditions on the acquisition and the background velocity model but it also brings the amplitude correction problem within the context of stable signal recovery. As long as the background velocity model is sufficiently smooth and there is a reference vector available close enough to the actual reflectivity, the Hessian can be approximated by a diagonal weighting in the curvelet domain. The diagonal approximation can be used to formulate the amplitude recovery in terms of a

constrained optimization problem. The amplitude-corrected image is obtained by solving this optimization problem designed to promote both the sparsity of curvelets on the image and the smoothness along the imaged reflectors. The invariance of the curvelets under the Hessian preserves the sparsity. The cost of computing the diagonal approximation is one de-migration-migration per reference vector which is much less compared to the costs involved with Krylov-subspace based least-squares migration. The results show an overall improvement of the image quality. The joined sparsity- and continuity-enhanced image has little artifacts, improved resolution and amplitude recovery.

The fact that the recovery method is stable with respect to missing data opens the exiting perspective of reducing the number of number of rows of the migration operator. In a future paper, we will report on this additional feature of our proposed method.

ACKNOWLEDGMENT

The authors would like to thank the authors of the Fast Discrete Curvelet Transform for making this code available at `www.curvelets.org`. We also would like to thank Dr. William Symes for making his migration code available to us. This work was in part financially supported by NSERC Discovery Grant 22R81254 of Felix J.Herrmann and was carried out as part of the SINBAD project with support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and Shell. The authors would also like to thank the Institute of Pure and Applied Mathematics at UCLA supported by the NSF under grant DMS-9810282.

References

1. G. Beylkin. The inverse problem and applications of the generalized Radon transform. *Comm. Pure Appl. Math.*, 37:579–599, 1984.

2. N. Bleistein, J. Cohen, and J. Stockwell. *Mathematics of Multidimensional Seismic Imaging, Migration and Inversion*. Springer, 2001.

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

4. G. C. Fehmers C. and Hocker. Fast structural interpretation with structure-oriented filtering. *Geophysics*, 68(4), 2003.

5. E. Candes, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. 2005.

6. E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Comm. Pure Appl. Math*, 58(11):1472–1528, 2005.

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

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

9. E.J. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. 2005. to appear in Comm. Pure Appl. Math.

10. G. Chavent and R. Plessix. An optimal true-amplitude least-squares prestack depth-migration operator. *Geophysics*, 64(2):508–515, 1999.

11. S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 43(1):129–159, 2001.

12. J. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(05):826–844, 1973.

13. J. Claerbout and D. Nichols. Spectral preconditioning. Technical report, Stanford Exploration Project, 1994.

14. I. Daubechies, M. Defrise, and C. de Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constrains. *CPAM*, pages 1413–1457, 2005.

15. M. de Hoop and N. Bleistein. Generalized radon transform inversions for reflectivity in anisotropic elastic media. *Inverse Problems*, 13(3):669–690, 1997.

16. M. de Hoop and S. Brandsberg-Dahl. Maslov asymptotic extension of generalized radon transform inversion in anisotropic elastic media: a least-squares approach. *Inverse problems*, 16(3):519–562, 2000.

17. M. Do and M. Vetterli. *Beyond wavelets*, chapter Contourlets. Academic Press, 2002.

18. D. L. Donoho. De-noising by soft thresholding. *IEEE Trans. Inform. Theory*, 41:613–627, 1995.

19. D. L. Donoho. Extensions of compressed sensing. 2004.

20. D. L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 2006.

21. H. Douma and M. de Hoop. Wave character preserving pre-stack map migration using curvelets. In *Expanded Abstracts*, Tulsa, 2004. Soc. Expl. Geophys.

22. J. Candès E and D. L. Donoho. Recovering Edges in Ill-posed Problems: Optimality of Curvelet Frames. *Ann. Statist.*, 30:784–842, 2000.

23. M. Elad. Why simple shrinkage is still relevant for redundant representations. *IEEE Trans. Inform. Theory*, 2005.

24. M. Elad, J. Starck, P. Querre, and D. L. Donoho. Simulataneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA). *Appl. Comput. Harmon. Anal.*, 2005.

25. M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Processing*, 2003.

26. F. J. Herrmann. Multifractional splines: application to seismic imaging. In Michael Unser, Akram Aldroubi, Andrew, and Laine, editors, *Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X*, volume 5207, pages 240–258. SPIE, 2003.

27. J. Hu, G. T. Schuster, and P.A. Valasek. Poststack migration deconvolution. *Geophysics*, 66(3):939–952, 2001.

28. H. Kuhl and M. D.Sacchi. Least-squares wave-equation migration for AVP/AVA inversion. *Geophysics*, 68(1):262–273, 2003.

29. D. L.Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *App. and Comp. Harmonic Analysis*, 2, 1995.

30. N. Y. Lee and B. J. Lucier. Wavelet methods for inverting the Radon transform with noisy data. *IEEE T*, 10:79–94, 2001.

31. P. Moghaddam, C. C. Stolk, and F. J. Herrmann. Curvelet-domain migration preconditioning. to appear.

32. T. Nemeth., C. Wu, and G. T. Schuster. Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221, 1999.

33. C. C. Paige and M. A. Saunders. Lsqr: An algorithm for sparse linear equations and sparse least squares. *ACM TOMS*, 8(1):43–71, 1982.

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

35. R. Pratt, C. Shin, and G. Hicks. Gauss-Newton and full Newton methods in frequency-space waveform inversion. *Geoph. J. Int.*, 133:341–362, 1998.

36. J. E. Rickett. Illumination-based normalization for wave-equation depth migration. *Geophysics*, 68, 2003.

37. O. Schertzer. Scale-Space Methods and Regularization for Denosing and Inverse Problems. *Advances in Imaging and Electron Physics*, 128:445–530, 2003.

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

39. J. L. Starck, M. Elad, and D. L. Donoho. Redundant multiscale transforms and their application to morphological component separation. *Advances in Imaging and Electron Physics*, 132, 2004.

40. E. Stein. *Harmonic Analysis: Real-Variable Methods, Orthogonalityand Oscillatory Integrals*. Princeton University Press, 1993.

41. C. C. Stolk. Microlocal analysis of a seismic linearized inverse problem. *Wave Motion*, pages 267–290, 2000.

42. C. C. Stolk and W. W. Symes. Smooth objective functionals for seismic velocity inversion. *Inverse Problems*, 19(1):73–89, 2003.

43. A. Tarantola. *Inverse Problem Ttheory: Methods for Data Fitting and Model Parameters Estimation*. Elsevier, 1987.

44. A.P.E. ten Kroode, D.-J Smit, and A.R. Verdel. A microlocal analysis of migration. *Wave Motion*, 28, 1998.

45. T. Tropp. Just relax: convex programming methods for subset selection and sparse approximation. *IEEE Trans. Inform. Theory*, 2006.

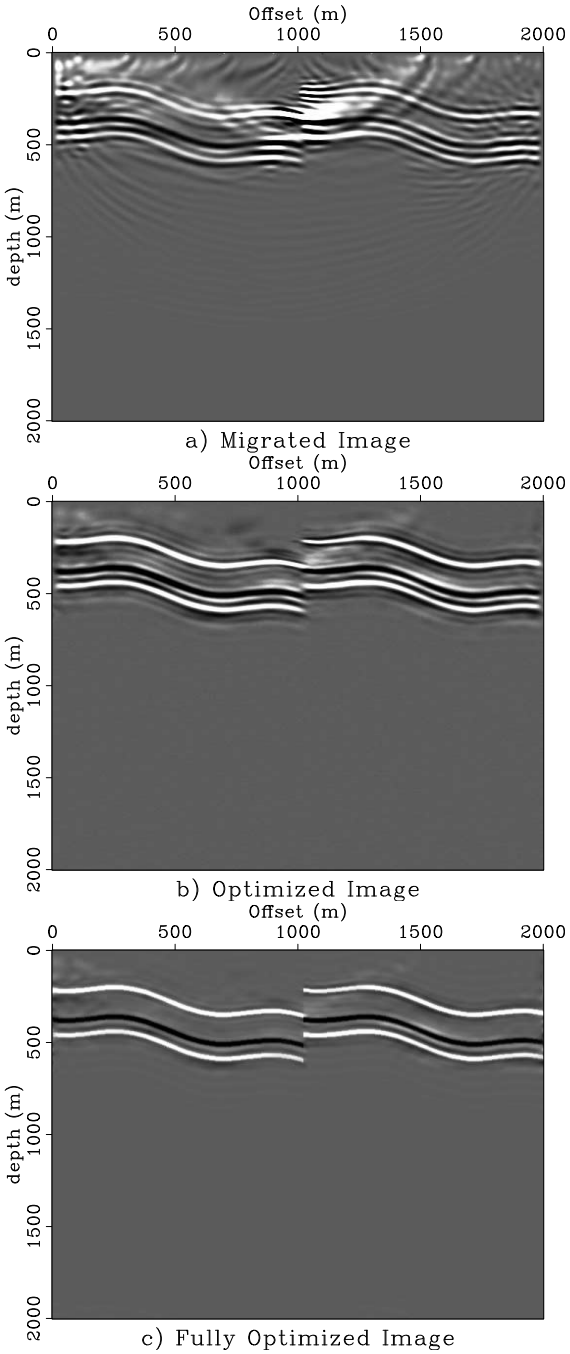

FIG. 7. Comparision between the migrated image and the image recovered by sparsity and continuity promoting imaging for data with a signal-to-noise ratio of 0 dB. **(a)** the migrated image; **(b)** recovered image using the ℓ_1 penalty term; **c** the recovered image for sparsity and continuity penalty term; Notice the improvement in recovered amplitudes and resolution obtained by solving **P** for $\alpha = 0.6$ and $\beta = 0.4$.

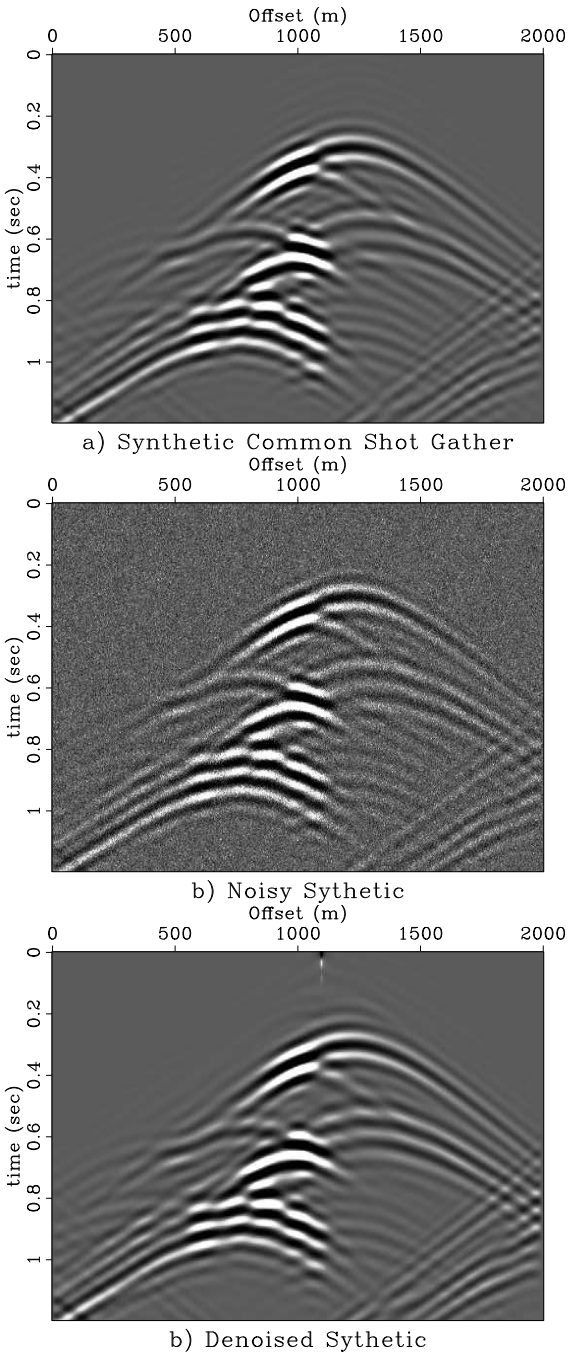

FIG. 8. Comparision between demigrated shot records for the three-reflector model. **(a)** the demigrated data; **(b)** the demigrated data with noise added to yield a signal-to-noise ratio of 0 dB; **c** shot record pertaining to recovered image depicted in Fig. 7 (**c**). Notice that the data is denoised reflected in the improvement of the signal-to-noise ratio from 0 dB to 10 dB.