Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0).
Copyright (c) 2004 SLIM group @ The University of British Columbia.

Curvelet-domain preconditioned 'wave-equation' depth-migration with sparseness & illumination constraints

Felix Herrmann and Peyman Moghaddam, EOS, University of British Columbia

Abstract

A non-linear edge-preserving solution to the least-squares migration problem with sparseness & illumination constraints is proposed. The applied formalism explores Curvelets as basis functions. By virtue of their sparseness and locality, Curvelets not only reduce the dimensionality of the imaging problem but they also naturally lead to a dense preconditioning that almost diagonalizes the normal/Hessian operator. This almost diagonalization allows us to recast the imaging problem into a 'simple' denoising problem. As such, we are in the position to use non-linear estimators based on thresholding. These estimators exploit the sparseness and locality of Curvelets and allow us to compute a first estimate for the reflectivity, which approximates the least-squares solution of the seismic inverse scattering problem. Given this estimate, we impose sparseness and additional amplitude corrections by solving a constrained optimization problem. This optimization problem is initialized and constrained by the thresholded image and is designed to remove remaining imaging artifacts and imperfections in the estimation and reconstruction.

Introduction

Least-squares migration and migration deconvolution has been a topic that received a recent flare of interest [7, 8]. This interest is for a good reason because inverting for the normal operator (the demigration-migration operator) restores many of the amplitude artifacts related to acquisition and illumination imprints. However, the downside to this approach is that (unconstrained) least-squares tends to fit noise and smear the energy. In addition, artifacts may be created due to imperfections in the model and possible null space of the normal operator [11]. Regularization can alleviate some of these problems, but may go at the expense of resolution, certainly when quadratic penalty functions are used. By imposing sparseness constraints and illumination normalization on the imaged reflectivity [see for instance 13, 8, 11], significant progress has been made to improve the signal-to-noise ratio (SNR) and frequency content of seismic images.

In this paper, we combine above techniques with recent developments in Applied Computational Harmonic Analysis, non-linear estimation theory and global constrained optimization. Our approach is designed to (i) deal with substantial amounts of noise ($SNR \leq 0$); (ii) use the optimal (sparse & local) representation properties of Curvelets; (iii) exploit their almost diagonalization of the normal/Gramm/Hessian operator; (iv) use non-linear thresholding estimation techniques, supplemented by (v) constrained optimization on the estimated coefficients. This optimization not only imposes additional sparseness on the model but also restores the amplitudes within constraints given by the illumination [11].

The main motivation for this paper is to derive an imaging scheme that is robust in cases where there is substantial amounts of noise in the data space ($SNR \leq 0$). These noise-levels are intended to mimic situations where the high frequencies are severely deteriorated by the presence of noise. Without loss of generality, we will assume the additive noise to be white Gaussian. Refer for coherent noise removal (such as multiples) in the data space to other contributions by the first author to these Proceedings [see also 6].

As with adaptive subtraction [6], the key component of our algorithm is the use of basis functions that are local both in position and dip and that are well behaved under the operators. In that respect, we are in the business of finding the appropriate *full* preconditioners for the seismic imaging problem. If indeed, recently developed non-separable wavelets, such as Curvelets, can be used as preconditioners for the seismic imaging problem, then we can use non-linear estimators to significantly improve the SNR and resolution of noisy seismic images.

The paper is organized as follows. First, we briefly discuss the imaging problem and motivate why preconditioning is important, in particular when the preconditioner is defined by a generic transform (non-problem specific) that is close to an unconditional basis which is local and sparse on the model space. Next, we argue that diagonally weighted Curvelets are close to the ideal preconditioners for seismic imaging. We proceed by introducing a non-linear thresholding estimator that acts on the preconditioned image space. In the image space, the normal and covariance operators are close to being diagonalized by what one can call the 'Curvelet symbol' of the normal operator. Because thresholding on the coefficients of local and sparse basis functions approaches asymptotically minimax (minimize the maximum error given the worst possible Gaussian *prior* [see e.g. 9]), we obtain an edge-preserved least-squares image with good SNR by simply thresholding and weighting the migrated image. This image is subsequently used as a starting point for a global optimization scheme that is designed to impose sparseness (*via* a global sparseness constraint); remove imaging artifacts related to the basis-function decomposition, eliminate false thresholding and poor illumination. The method will be illustrated by a synthetic example using a 'wave-equation' depth-migration operator [11].

The seismic imaging problem preconditioned

In all generality, the seismic imaging problem can after linearization be cast into the following form for the forward model describing our data

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

where $\mathbf{K}$ is the wave-equation demigration operator based on the single-square-root equation [12]; $\mathbf{m}$ the model with the reflectivity and $\mathbf{n}$ white Gaussian noise. The pertaining inverse problem has the following general form [see e.g. 9, 14]

$$\hat{\mathbf{m}}: \quad \min_{\mathbf{m}} \frac{1}{2}\|\mathbf{d} - \mathbf{K}\mathbf{m}\|_2^2 + \mu_\lambda J(\mathbf{m}), \tag{2}$$

where $J(\mathbf{m})$ is an additional penalty functional that contains *prior* information on the model, such as particular sparseness constraints. The control-parameter μ_λ depends on the noise level λ.

Question now is how can we find the appropriate domain to solve this inverse problem, whose formal solution in the noise-free case can be found by taking the More-Penrose pseudo-inverse,

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

Several approaches exist to compute the solution of this least-squares problem. These solutions range from explicit analytical constructions for the pseudo-inverse of the normal operator [see e.g. 1, and the references therein] to iterative solutions with Conjugate-gradient type of

	FIO	ΨDO	$\mathbf{d}$, $\mathbf{m}$
Fourier	$\surd$	$\surd$	$\times$
Wavelets	$\times$	$\times$	$\times$
Curvelets	$\surd$	$\surd$	$\surd$

Table 1: Properties of Curvelet with regard to seismic imaging. As one can see Curvelets not only score representing the data and model but also score for the operators. Fourier is good for the operators (as the name FIO suggests) but fails representing the data/model sparsely.

algorithms [see e.g. 7, 8]. The first approaches derive from the observation that the leading-order behavior for the (de)-migration operators corresponds to that of certain Fourier Integral Operators (FIO's), while the composition of these two operators corresponds, under certain conditions, to an *invertible* elliptic pseudo-differential operator (ΨDO) [1]. The FIO's are responsible for moving the events from model to data space and *vice versa*, while the ΨDO can be interpreted as a generalized non-stationary filtering operator that does not (re)-move or create events (read singularities). While the first of these approaches, has the distinct advantage of being numerically fast, it requires an explicit construction of the operators and may suffer from the typical less than ideal data acquisition. The second approach is more flexible, since it only requires access to the migration and demigration operators and it may be less sensitive to the acquisition. However, noise, computational cost and inaccuracies in the operators may still be a limiting factor.

Therefore, we opt for a "combination" of afore mentioned methods by constructing explicit preconditioners that *almost* (as compared to FIO's and ΨDO's, which are purely diagonal in the Fourier domain) *diagonalize*, while also being *sparse* & *local* on the model space. In this way, we are not only well positioned for computing iterative solutions to the problem, but we are also able to use diagonal non-linear estimation techniques [9] in combination with global constrained optimization [10] in order to remove the noise; restore the amplitudes; impose sparseness constrains and incorporate the influence of illumination.

Our first priority is to find the appropriate preconditioners for the system in Eq. 3. To be more specific, we would need to construct preconditioners that use generic local basis functions that allow us to write

$$\mathbf{B}^T\mathbf{d} = \mathbf{B}^T\mathbf{B}\mathbf{x} + \mathbf{B}^T\mathbf{n}. \tag{4}$$

In this expression, $\mathbf{B} = \mathbf{K}\mathbf{P}$, $\mathbf{B}^T = \mathbf{P}^T\mathbf{K}^T$ are the preconditioned demigration and migration operators and the model is related to the new model according to the linear functional $\mathbf{x} = \mathbf{P}^T\mathbf{m}$. Successful preconditioning, yields an almost diagonalization for the normal equations, which corresponds to the following property $\mathbf{B}^T\mathbf{B} \approx \mathbf{I}$.

Question now is can we find a generic (i.e. not 'depending' on $\mathbf{K}$ and $\mathbf{K}^T$) close to orthonormal "miracle" basis-function transform, $\mathbf{C}$, that allows us to write the preconditioner in the following form: $\mathbf{P} = \mathbf{C}^T\mathbf{\Gamma}^{-1}$, where $\mathbf{\Gamma}^{-1}$ is a diagonal matrix, whose entries are defined by the regularized reciprocal of $\mathrm{diag}\{\Gamma\}$. Provided such a transform exists, we are not only in good shape to iteratively solve the least-squares problem, but, more importantly, we also will be able to explore that $\mathbf{C}$ and hence $\mathbf{P}$ are close to unconditional bases for the model space. Before going into detail how to construct non-linear estimators, we first introduce Curvelet frames that almost diagonalize the normal operator.

Curvelets and seismic imaging

Curvelets as proposed by [3], are a relatively new family of non-separable Wavelets that effectively represent seismic images with reflectors on piece-wise smooth curves ($\mathbf{m} \in C^2$). For these type of functions, Curvelets obtain near optimal nonlinear approximation rates. i.e. the error is $\mathcal{O}(m^{-2})$ as opposed to $\mathcal{O}(m^{-1/2})$ for Fourier, where m is the number of magnitude-sorted coefficients in the reconstruction.

Besides this grealy improved nonlinear approximation rate, Curvelets are local both in position and dip. This locality combined with the high approximation rate and the fact that Curvelets are close to an unconditional basis (for the above class of functions), allows the definition of diagonal estimators that asymptotically obtain minimax for denoising [see e.g. 9]. These non-linear estimators minimize the maximal error given the worst Bayesian *prior* [9] and are given by

$$\hat{\mathbf{x}} = \mathbf{C}^\dagger\Theta_\lambda\left(\mathbf{C}\mathbf{y}\right). \tag{5}$$

This basis function decomposition, followed by thresholding and subsequent reconstruction, solves for the model given noisy data, i.e. solves for $\mathbf{x}$ given $\mathbf{y} = \mathbf{x} + \mathbf{n}_\lambda$ with $\mathbf{n}_\lambda = N(0, \lambda^2)$, by applying a hard thresholding $\Theta_\lambda(\cdot)$, with a threshold level that depends on the standard deviation λ of the noise. Since Curvelets are local and sparse on the model, it is relatively easy to understand why this thresholding procedure works because the edges in the model end up in the large coefficients which survive the thresholding. The thresholding depends only on the magnitude of the coefficients and not on their sign or phase, which is a direct consequence of their (close to) unconditional-basis property. How do Curvelets obtain such a high non-linear approximation rate? Without being all inclusive [see for details 3], Curvelets are

- *multi-scale*, i.e. they live in different dyadic corona in the 2D Fourier-domain.
- *multi-directional*, i.e. they live on wedges within these corona.
- *anisotropic*, i.e. they obey the following scaling law width $\propto$ length^2.
- *directional selective* with # orientations $\propto \frac{1}{\sqrt{\text{scale}}}$.
- *local* both in (x, z) and (k_x, k_z).
- almost *orthogonal*, they are *tight* frames.

Besides their high non-linear approximation rate, Curvelet are well behaved under certain (homogeneous) operators [4, see the references therein]. This property has led to Quasi-SVD techniques, where the singular vectors of the operators are replaced by generic basis functions and their images under the operators. These methods work because the basis functions remain approximately invariant under the operators. As we can see from Fig. 1, this property also holds true for imaging operators, where not only Curvelets remain Curvelet-like, but where also a rapid decay for the Curvelet coefficients is observed. These observations are consistent with recent work by [2], where it is shown that FIO's are sparse in the Curvelet domain and, consequently, ΨDO's can be expected to be almost diagonal.

Non-linear diagonal estimation by thresholding

Obviously, above listed properties make Curvelets ideal for seismic imaging. Unfortunately, the background velocity-dependence and shear size of seismic data, makes it prohibitively expensive to use the Quasi-SVD methods because these techniques require a separate (de)-migration for each Curvelet to compute the quasi-singular values (the $\mathbf{\Gamma}$'s, see below). Instead, we settle for a formulation, where the normal and Covariance operators (for the noise) are approximately diagonalized. By using Curvelets in the definition for the preconditioners (cf. Eq. 4), where $\mathbf{C}\mathbf{K}^T\mathbf{K}\mathbf{C}^T \approx \mathrm{diag}\{\mathbf{C}\mathbf{K}^T\mathbf{K}\mathbf{C}^T\} = \mathrm{diag}\{\Gamma^2\}$, the preconditioned system can approximately be written in the form

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \tag{6}$$

Not only does the preconditioning take care off the operator, it also 'whitens' the noise, so we are set to use thresholding to obtain a minimax estimate for $\mathbf{x}$, using

$$\hat{\mathbf{x}}_0 = \Theta_\lambda\left(\mathbf{y}\right). \tag{7}$$

This expression is equivalent to the approximate solution for the model presented in [5]

$$\hat{\mathbf{m}}_0 = \mathbf{C}^\dagger\left(\mathbf{\Gamma}^2\right)^\dagger\Theta_{\lambda\Gamma}\left(\mathbf{C}\mathbf{K}^T\mathbf{d}\right) = \mathbf{P}\Theta_\lambda\left(\mathbf{P}^T\mathbf{K}^T\mathbf{d}\right) \tag{8}$$

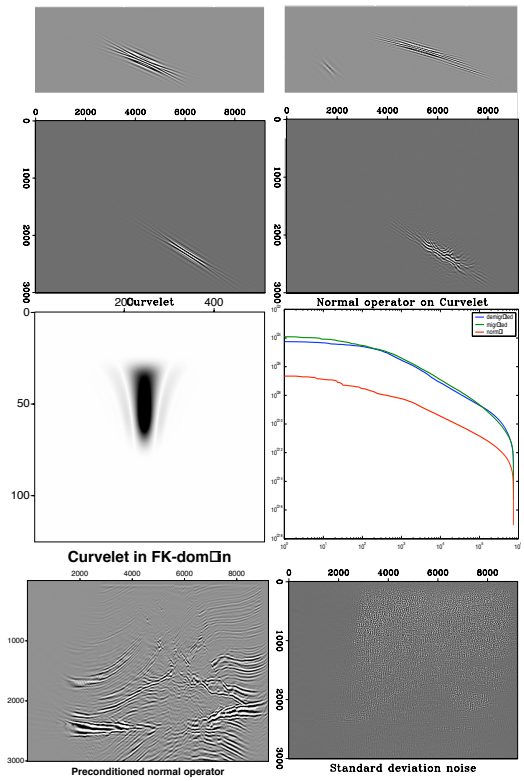

Fig. 1: Properties of the Curvelet Transform under imaging. **Top row:** Curvelet and its demigration for constant velocity Kirchoff. **Second row:** Same for 'wave-equation' migration Hessian in Marmousi model. For both cases, Curvelets remain Curvelet-like. **Third row:** FK-spectrum and plot of the rapid decay for the "off-diagonal" coefficients of the Curvelet-images. **Fourth row:** $\mathbf{A}\mathbf{x}_{\text{true}}$ which should be close to $\mathbf{I}\mathbf{m}_{true}$ and $\mathbf{C}^\dagger\mathbf{\Gamma}$, the inverse transform of the Curvelet 'symbol'.

in which the thresholded adjoint is corrected by the pseudo-inverse of Curvelet 'symbol' of the normal operator: $\left(\mathbf{\Gamma}^2\right)^\dagger$. The different steps that lead to the above diagonal estimate are summarized in Fig. 1 and 2. From the Fig. **??**, it is clear that Curvelets remain more or less curvelet-like under the operators while $\mathbf{A} \approx \mathbf{I}$.

Amplitude restoration by global optimization

So far, we have been able to make the argument that the bulk of the noise can be removed and amplitudes partly restored. However, the approach is approximate and one could expect improved results by computing the following pseudo-inverse

$$\hat{\mathbf{x}} = \mathbf{A}^\dagger \hat{\mathbf{x}}_0, \tag{9}$$

where $\mathbf{A} = \mathbf{B}^T\mathbf{B}$. Even tough this approach may seem very attractive, since we can argue that the preconditioning improves the condition number of the matrix (see other contributions by the authors to the Proceedings of this Conference), there remains the problem of left-over noise; high computational cost; shortcomings in the operator and the inclusion of sparseness constraints on the model: $\mathbf{x} = \mathbf{P}^T\mathbf{m}$. To accommodate for these issues, we formulate the following approximate (the operator $\mathbf{A}$ can be included) constrained-optimization problem (this approach was motivated by [3], who applies a similar technique for denoising)

$$\hat{\mathbf{m}}: \quad \min_{\mathbf{m}} J(\mathbf{m}) \quad \text{s.t.} \quad |\mathbf{x} - \hat{\mathbf{x}}_0|_\mu \leq \mathbf{e}_\mu \quad \forall \mu \tag{10}$$

with the tolerances set according to

$$\mathbf{e}_\mu = \begin{cases} \mathbf{I}_\mu & \text{if} \quad |\hat{\tilde{\mathbf{m}}}_0|_\mu \geq |\lambda \mathbf{I}|_\mu \\ \boldsymbol{\lambda}\mathbf{I}_\mu & \text{if} \quad |\hat{\tilde{\mathbf{m}}}_0|_\mu < |\lambda \mathbf{I}|_\mu. \end{cases} \tag{11}$$

In this formulation, the initial guess for the preconditioned model $\hat{\mathbf{x}}_0$ is updated such that the additional sparseness constraint is minimized. The optimization runs over all the index space μ and the tolerance differs by λ for those coefficients that have initially been thresholded to obtain the first estimate and those that survived the thresholding. The coefficients that were perceived to be noise are allowed to vary more as part of the optimization. This global optimization procedure is solved using an Augmented-Lagrangian technique [10] with the Lagrangian multipliers set by the thresholded result, $\hat{\mathbf{x}}_0$. The solution uses a Steepest Decent technique and involves in each iteration an inverse Curvelet transform combined with a weighting that depends on the diagonalized normal operator.

Above approach, depends heavily on how well the preconditioners are able to diagonalize the normal operator. Since, iterating with $\mathbf{A}$ is generally prohibitively expensive one may consider to use (i) the Lanczos tri-diagonalization method (see other contributions to these Proceedings) or (ii) the illumination-based normalization method [11], that uses the following diagonal weighting to approximate the normal operator

$$\mathbf{W}^{-2} = \text{diag}\{\hat{\mathbf{x}}_0\} \cdot [\text{diag}\{\mathbf{A}\hat{\mathbf{x}}_0\}]^{-1} \approx \mathbf{A}^\dagger \tag{12}$$

for a reference model that is close to the actual model. We use this diagonal preconditioner in the following constrained-optimization problem

$$\hat{\mathbf{m}}: \quad \min_{\mathbf{m}} J(\mathbf{m}) \quad \text{s.t.} \quad |\mathbf{W}^2\mathbf{x} - \hat{\mathbf{y}}_0|_\mu \leq \mathbf{e}_\mu \quad \forall \mu. \tag{13}$$

with $\hat{\mathbf{y}}_0 = \mathbf{A}\hat{\mathbf{x}}_0$. In this formulation, the initial guess for the reference model, which we naturally take to be the result obtained by thresholding, is updated such that the additional constraint is minimized. The elegance of this formulation lies in the fact that (i) explicit use is made of the almost diagonalization of the operators, i.e. the weighted Curvelets are non-diagonal preconditioners; (ii) the algorithm is constrained so that it can not wander off to far from the initial guess. Because of the minimax property of thresholding, we can be assured that the initial guess predominantly contains features with high SNR, rendering the above illumination normalization effective.

Example

To illustrate our proposed method, we included a synthetic example of a 'wave-equation' depth-migration for the Marmousi dataset. A Monte-Carlo sampling technique was used to compute the $\mathbf{\Gamma}^2$ [5, see for detail]. Fig. 2 includes the results of our imaging algorithm. We applied the algorithm to data with a SNR=0 and the noisy image is included in Fig. 2 (top). Next, we threshold the noisy image with Eq. 7, followed by applying the $\mathbf{\Gamma}^{-2}$-correction (cf. Eq. 8). The threshold parameter $\lambda = 2.5$. As we can see from Fig. 2 (second-third), the thresholding and subsequent correction have an influence. The thresholding removes the noise and preserves most of the edges. The correction restores part of the amplitudes and gives us a first approximation to the least-squares imaging problem. This approximation is used as a starting point for the constrained-optimization problem presented in Eq. 10. The results for this optimization are plotted in Fig. 2 (bottom). The L^1-constraint clearly removes more noise which goes at the expense of some details.

Conclusions and discussion

The methodology presented is an example of a divide-and-conquer approach. First, we obtain a reasonable estimate for the image by thresholding then we continue by restoring the amplitudes and removing the artifacts. Our approach builds on the premise that one stands a much better chance to solve an inverse problem into the appropriate

domain. The combination of migration with the Curvelet transform provides such a domain, where the energy is collapsed onto a limited number of coefficients that correspond to coherent features along which the Curvelets align. Since Curvelets are local in space and spatial frequency, thresholding can be used on the image. By using Curvelets to precondition the inverse scattering problem, we arrived at an elegant formulation that serves as a point of departure for our non-linear estimation procedure. Not only did we almost diagonalize the normal operator, turning the migration into an almost orthogonal transform, but we also included non-linear thresholding and global optimization in the formulation. Results, so far, for the Marmousi dataset are certainly promising but fall short somewhat of being conclusive. We hope to present compelling results at the Meeting that demonstrate the full potential of the presented method to deal with very noisy data ($SNR \leq 0$).

Acknowledgments

The authors would like to thank Emmanuel Candés and David Donoho for making their Digital Curvelet Transforms via Unequally-spaced Fourier Transforms available (presented at the ONR Meeting, University of Minnesota, May, 2003). We also would like to thank James Rickett for his migration code. This work was in part financially supported by a NSERC Discovery Grant.

References

[1] 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.

[2] E. J. Candès and L. Demanet. Curvelets and fourier integral operators. 2002. URL `http://www.acm.caltech.edu/~emmanuel/publications.html`. to appear Comptes-Rendus de l'Academie des Sciences, Paris, Serie I.

[3] 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. URL `http://www.acm.caltech.edu/~emmanuel/publications.html`.

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

[5] Felix J. Herrmann. Multifractional splines: application to seismic imaging. In Andrew F. Laine; Eds. Michael A. Unser, Akram Aldroubi, editor, *Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X*, volume 5207, pages 240–258. SPIE, 2003. URL `http://www.eos.ubc.ca/~felix/Preprint/SPIE03DEF.pdf`.

[6] Felix J. Herrmann and Eric Verschuur. Separation of primaries and multiples by non-linear estimation in the curvelet domain. In *Expanded Abstracts*. EAGE, 2004.

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

[8] H. Kuhl and M. D. Sacchi. Least-squares wave-equation migration for avp/ava inversion. *Geophysics*, 68(1):262–273, 2003.

[9] S. G. Mallat. *A wavelet tour of signal processing*. Academic Press, 1997.

[10] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer, 1999.

[11] James E. Rickett. Illumination-based normalization for wave-equation depth migration. *Geophysics*, 68, 2003. URL `http://link.aip.org/link/?GPY/68/1371/1`.

[12] James E. Rickett and Paul C. Sava. Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics*, 67, 2002. URL `http://link.aip.org/link/?GPY/67/883/1`.

[13] D. O. Trad. Interpolation and multiple attenuation with migration operators. *Geophysics*, 68(6):2043–2054, 2003.

[14] Curtis Vogel. *Computational Methods for Inverse Problems*. SIAM, 2002.

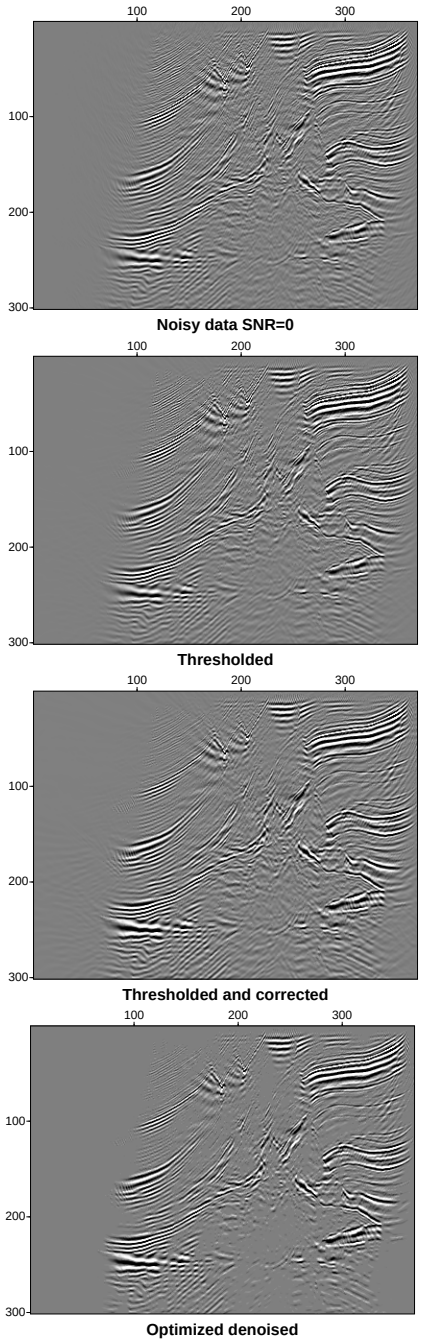

Fig. 2: Synthetic 'wave-equation depth-migration example for the Marmousi dataset. **Top:** Noisy image for $SNR = 0$ on the data space. **Second:** Image after thresholding (cf. 8). **Third:** Thresholded after correction for the 'symbol' of the normal operator (cf. 8). As we can see the amplitudes are partly restored. **Bottom:** Result after applying the constrained optimization (cf. 10) with $\lambda = 2.5$. The optimization clearly removes more noise while enhancing the sparseness. This procedure does, however, go at the expense of some details which is to be expected for this high noise level.