Time-domain sparsity promoting least-squares reverse time migration with source estimation

Mengmeng Yang[1][*][#], Zhilong Fang[2][*], Philipp Witte[3] and Felix J. Herrmann[1,3]

[1] School of Earth and Atmospheric Sciences, Georgia Institute of Technology

[2] Department of Mathematics, Massachusetts Institute of Technology

[3] School of Computational Science and Engineering, Georgia Institute of Technology

[*] Equally contributed

[#] To whom correspondence should be addressed: myang64@gatech.edu

(February 18, 2020)

Running head: **SPLS-RTM with source estimation**

ABSTRACT

Least-squares reverse time migration is well-known for its capability to generate artifact-free true-amplitude subsurface images through fitting observed data in the least-squares sense. However, when applied to realistic imaging problems, this approach is faced with issues related to overfitting and excessive computational costs induced by many wave-equation solves. The fact that the source function is unknown complicates this situation even further. Motivated by recent results in stochastic optimization and transform-domain sparsity-promotion, we demonstrate that the computational costs of inversion can be reduced significantly while avoiding imaging artifacts and restoring amplitudes. While powerfull, these new approaches do require accurate information on the source-time function, which is often lacking. Without

this information, the imaging quality deteriorates rapidly. We address this issue by presenting an approach where the source-time function is estimated on the fly through a technique known as variable projection. Aside from introducing negligible computational overhead, the proposed method is shown to perform well on imaging problems with noisy data and problems that involve complex settings such as salt. In either case, the presented method produces high resolution high-amplitude fidelity images including an estimates for the source-time function. In addition, due to its use of stochastic optimization, we arrive at these images at roughly one to two times the cost of conventional reverse time migration involving all data.

INTRODUCTION

Reverse-time migration (RTM) is a popular wave equation-based seismic imaging methodology where the inverse of the linearized Born scattering operator is approximated by applying its adjoint directly to the observed reflection data (Baysal et al., 1983; Whitmore, 1983). Because the adjoint does not equal the pseudo inverse, conventional RTM produces images with incorrect amplitudes. Among the factors that contribute to low-fidelity amplitudes, the imprint of the temporal bandwidth limitation of the typically unknown source wavelet features prominently and so does the fact that the Born scattering operator is not inverted. To overcome these issues, we formulate our imaging problem as a linear least-squares inversion problem where the difference between observed and predicted data is minimized in the ℓ_2-norm (Schuster, 1983; Nemeth et al., 1999; Dong et al., 2012; Zeng et al., 2014). While least-squares migration is a powerful technique, its succesful application to industry-scale problems is hampered by three key issues. First, iterative demigrations (i.e. Born modeling) and migrations become computationally prohibitively expensive when carried out over all shots. Second, we run the risk of overfitting the data when minimizing the ℓ_2-norm of the data residual. This overfitting may introduce noise-related artifacts in inverted images. Third, while the source location is generally well known, the temporal source function is often not known accurately. Because imaging relies on knowing the source function, this may have a detrimental effect on the image and makes it necessary to come up with source estimation methodology. Since we carry out our imaging iteratively, we propose to estimate the wavelet on the fly as we build up the image.

We address the issue of computational feasibility by combining techniques from stochastic optimization (van Leeuwen et al., 2011; Haber et al., 2012; Powell, 2014), curvelet-domain

sparsity-promotion (Herrmann and Li, 2012), and online convex optimization (Lorenz et al., 2014) with linearized Bregman. Stochastic optimization allows us to work with small random subsets of shots, which limits the number of wave equation solves—i.e., passes through the data. Convergence is guaranteed (Herrmann et al., 2015; Yang et al., 2016, ; Witte et al., 2019c) by replacing the ℓ_1-norm, by an elastic net consisting of a strongly convex combination of ℓ_1- and ℓ_2-norm objectives. Inclusion of the ℓ_2-norm results in a greatly simplified algorithm involving linearized Bregman iterations, which corresponds to gradient descent on the dual variable supplemented by a simple soft thresholding operation (Cai et al., 2009; Yin, 2010) with a threshold that is fixed. We refer to this method as sparsity-promoting least-squares reverse-time migration (SPLS-RTM).

In addition to the high computational cost, the lack of accurate knowledge on the unknown temporal source signature may also adversely affect the performance of the inversion. Errors in the source signature lead to erroneous residuals, which in turn result in inaccurately imaged reflectors, which now may be positioned wrongly or may have the wrong amplitude or phase. To mitigate these errors, we need an embedded procedure where the source signature is updated along with the image during the inversion (Pratt, 1999; Aravkin et al., 2012, 2013; Fang et al., 2018) using a technique known as variable projection (Rickett, 2013; van Leeuwen et al., 2014). For time-harmonic imaging, variable projection involves the estimation of the source function by solving a least-squares problem for each frequency separately. Since the unknown for each frequency in that case is a single complex-valued variable, this process is simple and has resulted in an accurate estimation and compensation for the source-time function (Tu et al., 2013, ; Fang et al., 2018). Unfortunately, the situation is more complicated during imaging in the time domain, where we have to estimate the complete source signature during each iteration. For this purpose, we integrate early work by Yang et al. (2016) and

inverse-scattering method (Witte et al., 2019c) and achieve an approach that is suitable for realistic imaging scenarios that may include salt.

Our work is outlined as follows. First, we introduce the basic equations for time-domain reverse time migration and least-squares reverse time migration. To overcome the computational cost associated with the latter, we introduce a stochastic optimization method with sparsity promotion. This method is designed to provide an image at a fraction of the cost. Next, we extend this approach so it includes on-the-fly source estimation. This allows us to remove the requirement of the source function. We conclude by presenting a number of synthetic case studies designed to demonstrate robustness with respect to noisy data and to complex imaging scenarios that include salt.

METHODOLOGY

Since our approach hinges on cost-effective least-squares imaging, we first introduce our formulation of sparsity-promoting least-squares migration with stochastic optimization, followed by our approach to on-the-fly source estimation during the iterations.

From RTM to LS-RTM

Reverse time migration derives from a linearization (see e.g. Mulder and Plessix (2004)) with respect to the background squared slowness. For the ith source this linearization reads

$$\delta\mathbf{d}_i = \mathbf{F}_i(\mathbf{m}_0 + \delta\mathbf{m}, \mathbf{q}) - \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \approx \mathbf{J}_i(\mathbf{m}_0, \mathbf{q})\delta\mathbf{m}, \tag{1}$$

where the vectors $\delta\mathbf{m}$, $\mathbf{q}$, and $\delta\mathbf{d}$ denote the model perturbation, the source-time function, and the corresponding data perturbation, respectively. We model the data for n_t time samples over a time interval of T s. The number of receivers is n_r so a single shot record is of

size $n_t \times n_r$. The nonlinear forward modeling operator $\mathbf{F}_i(\mathbf{m}, \mathbf{q})$ for the ith source location involves the solution of the discretized acoustic wave equation

$$\left(\mathbf{m} \odot \frac{\partial^2}{\partial t^2} - \Delta\right) \mathbf{u}_i = \mathbf{P}_{\text{s},i}^\top \mathbf{q},$$

$$\mathbf{P}_{\text{r},i} \mathbf{u}_i = \mathbf{d}_i,$$

(2)

parameterized by the squared slowness collected in the vector $\mathbf{m}$ (for simplicity, we kept the density constant and we used the symbol $\odot$ to denote elementwise multiplication.) The symbol Δ represents the discretized Laplacian and the linear operator $\mathbf{P}_{\text{r},i}$ restricts the wavefield for the ith source to the corresponding receiver locations, while the linear operator $\mathbf{P}_{\text{s},i}^\top$ injects the source time function at the location of the ith source in the computational grid. The Jacobian $\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})$ is known as the Born modeling operator and is given by the derivative of $\mathbf{F}_i(\mathbf{m}, \mathbf{q})$ at the point of $\mathbf{m}_0$. Applying the Jacobian $\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})$ to the model perturbation $\delta\mathbf{m}$ requires the solution of the following linearized equation:

$$\left(\mathbf{m}_0 \odot \frac{\partial^2}{\partial t^2} - \Delta\right) \delta\mathbf{u}_i = -\frac{\partial^2}{\partial t^2}(\delta\mathbf{m} \odot \mathbf{u}_i),$$

$$\mathbf{P}_{\text{r},i} \delta\mathbf{u}_i = \delta\mathbf{d}_i,$$

(3)

where the vector $\delta\mathbf{u}_i$ corresponds to the wavefield perturbation for the ith source.

The goal of seismic imaging is to estimate the model perturbations from observed data. We can expect this reconstruction process to be successful in situations where the above linear approximation is accurate—i.e., the background velocity model needs to be sufficiently accurate, which we assume it is. We also need accurate knowledge on the source function, an important aspect we will address below.

While the above linearization allows us to create an image via

$$\delta\mathbf{m}_{\text{RTM}} = \sum_{i=1}^{n_s} \mathbf{J}_i^\top \delta\mathbf{d}_i \tag{4}$$

with n_s the number of shots, the adjoint of the Jacobian (denoted by the symbol $^\top$) does not correspond to its inverse and $\delta\mathbf{m}_{\mathrm{RTM}}$ will suffer from wavelet side lobes and inaccurate and unbalanced amplitudes (Mulder and Plessix, 2004; Bednar et al., 2006; Hou and Symes, 2016). Unlike RTM (Equation 4), LS-RTM (Aoki and Schuster, 2009; Herrmann and Li, 2012; Tu and Herrmann, 2015b) reconstructs the model perturbation by computing the pseudo-inverse of the Born modeling operator, which can significantly mitigate these defects. LS-RTM iteratively solves the following least squares data-fitting problem:

$$\underset{\delta\mathbf{m}}{\text{minimize}} \; \frac{1}{2} \sum_{i=1}^{n_{\mathrm{s}}} \|\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})\delta\mathbf{m} - \delta\mathbf{d}_i\|^2. \tag{5}$$

Compared to Equation 4, the above minimization requires multiple evaluations of the Jacobian and its adjoint, which becomes rapidly computationally prohibitive for large 2D and 3D imaging problems as the number of sources n_s grows. This in part explains the relatively slow adaptation of least-squares reverse time migration (cf. Equation 5) by industry. As we show below, we overcome this problem by combining ideas from stochastic optimization and sparsity promotion (Herrmann et al., 2015; Yang et al., 2016; Witte et al., 2019c), which allow us to obtain artifact-free images at the cost of two to three passes through the data.

Stochastic optimization with sparsity promotion

As we mentioned above, the minimization of Equation 5 over all n_s shots is computationally prohibitively expensive. In addition, the minimization is unconstrained and misses regularization to battle the adverse effects of noise and the null space (missing frequencies and finite aperture) associated with solving the least-squares imaging problems of the type listed in Equation 5. To address these two problems, we combine ideas from stochastic optimization, during which we only work on randomized subsets of shots during each iteration, and ideas

from sparsity-promoting optimization designed to remove the imprint of the null space and source subsampling related artifacts. As we have learned from the field of compressive sensing (Candès et al., 2006; Donoho, 2006; Candès and Wakin, 2008), transform-domain sparsity promotion is a viable technique to remove subsample related noise in imaging via

$$
\begin{aligned}
&\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{x}\|_1, \\
&\text{subject to} \quad \sum_{i=1}^{n_s} \|\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})\mathbf{C}^\top \mathbf{x} - \delta \mathbf{d}_i\|_2 \leq \sigma.
\end{aligned}
\tag{6}
$$

In this formulation, known as the Basis Pursuit Denoise (BPDN, Chen et al. (2001)) problem, we include the sparsity-promoting ℓ_1-norm as the objective on the curvelet coefficients $\mathbf{x}$ of the image. These coefficients are related to the linearized data via the adjoint of the curvelet transform ($\mathbf{C}^\top$) and the above program seeks to find the sparsest curvelet coefficient vector that matches the data within the noise level σ. While the above problem is known to produce high-fidelity results, its solution relies on iterations that involve a loop over all n_s shots.

Stochastic gradient descent (e.g. Haber et al., 2012, in the context of seismic inversion) is a widely used tool to make unconstrained optimization problems of the type included in Equation 5 computationally feasible by computing the gradient of Equation 5 for randomized subsets of shots, using a given a batch size that corresponds to the number of shots used per iteration. This approach allows to minimize Equation 5 in very few epochs, using only few passes through data consisting of n_s shot records, as long as the step lengths adhere to certain conditions to guarantee convergence. Unfortunately, this complicates the solution of BPDN. To avoid this complication, we reformulate, following Cai et al. (2009), Equation 6 by replacing its convex ℓ_1-norm objective by the strongly convex objective involving

$$
\begin{aligned}
&\underset{\mathbf{x}}{\text{minimize}} \quad \lambda_1 \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|_2^2 \\
&\text{subject to} \quad \sum_{i=1}^{n_s} \|\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})\mathbf{C}^\top \mathbf{x} - \delta \mathbf{d}_i\|_2 \leq \sigma
\end{aligned}
\tag{7}
$$

with the estimate for the image given by $\delta \hat{\mathbf{m}} = \mathbf{C}^\top \hat{\mathbf{x}}$ where $\hat{\mathbf{x}}$ is the minimizer of the above optimization problem. The mixed objective in this problem is known as an elastic net in machine learning, which offers convergence guarantees (see Lorenz et al. (2014)) in situations where we work during each iteration with different randomized subsets of shots indexed by $\mathcal{I}_k \subset [1 \cdots n_s]$ with cardinality $|\mathcal{I}| = n_s' \ll n_s$. We choose these subsets without replacement.

For $\lambda \to \infty$, which in practice means λ large enough, iterative solutions of Equation 7 as summarized in Algorithm 1 converge to the solution of Equation 6, even in situations where we work with randomized subsets of shots. Compared to iterative solutions of Equation 6, the iterations (lines 7–8 in Algorithm 1) correspond to iterative thresholding with a fixed threshold λ on the dual variable ($\mathbf{z_k}$) with a dynamic step length given by $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$ (Lorenz et al., 2014). During each iteration, known as linerarized Bregman iterations, the residual is projected onto an ℓ_2-norm ball with the radius σ through a projection operator $\mathcal{P}_\sigma$. To avoid too many iterations, we set the threshold λ, related to the the tradeoff between the ℓ_1 and ℓ_2-norm objectives in Equation 7, to a value that is not too large—i.e., typically proportional to the maximum of $|\mathbf{z}_k|$ at the first iteration ($k = 1$). As reported by Yang et al. (2016) and Witte et al. (2019c), high quality images can be obtained running Algorithm 1 for a few epochs as long as the source time function $\mathbf{q}$ and background velocity model are sufficiently accurate. As we will show below, the background velocity model also needs to be smooth so the tomography-related imaging is avoided.

On-the-fly source estimation

In practice, we unfortunately do not have access to the source time function $\mathbf{q}$ required by Algorithm 1. Following our earlier work on source estimation in time-harmonic imaging and

---
Algorithm 1 Linearized Bregman for SPLS-RTM
---
1. Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{z}_0 = \mathbf{0}$, $\mathbf{q}$, λ_1, batchsize $n'_s \ll n_s$

2. **for** $k = 0, 1, \cdots$

3.   Randomly choose shot subsets $\mathcal{I}_k \subset [1 \cdots n_s]$, $|\mathcal{I}| = n'_s$

4. $\mathbf{A}_k = \{\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})\mathbf{C}^\top\}_{i \in \mathcal{I}_k}$

5. $\mathbf{b}_k = \{\delta \mathbf{d}_i\}_{i \in \mathcal{I}_k}$

6. $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$

7. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^\top \mathcal{P}_\sigma (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$

8. $\mathbf{x}_{k+1} = S_{\lambda_1}(\mathbf{z}_{k+1})$

9. **end**

10. **Output:** $\delta \hat{\mathbf{m}} = \mathbf{C}^\top \mathbf{x}_{k+1}$

note: $S_{\lambda_1}(\mathbf{z}_{k+1}) = \text{sign}(\mathbf{z}_{k+1}) \max\{0, |\mathbf{z}_{k+1}| - \lambda_1\}$

$$\mathcal{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k) = \max\{0, 1 - \tfrac{\sigma}{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|}\} \cdot (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$$
---

full-waveform inversion (van Leeuwen et al., 2011; Tu and Herrmann, 2015a), we propose an approach during which we estimate the source-time signature after each model update by solving a least-squares problem that matches predicted and observed data via a time-domain filter.

To keep our time-domain wave equation solvers with finite differences numerically stable (In our implementation, we used Devito (https://www.devitoproject.org) for our time-domain finite difference simulations and gradient computations (Luporini et al., 2018; Louboutin et al., 2019) and JUDI (https://github.com/slimgroup/JUDI.jl) as an abstract linear algebra interface to our Algorithms (Witte et al., 2019b)), we introduce an initial guess for the source time function $\mathbf{q}_0$ with a bandwidth limited spectrum that is flat over the frequency range of

interest. Under some assumptions on the source time function, we can write the true source

time function as the convolution between the initial guess and the unknown filter $\mathbf{w}$—i.e., we

have $\mathbf{q} = \mathbf{w} * \mathbf{q}_0$ where the symbol $*$ denotes the temporal convolution. Because we assume

one and the same source time function for all shots, we can write

$$\mathbf{J}_i(\mathbf{m}_0, \mathbf{w} * \mathbf{q}_0) = \mathbf{w} * \mathbf{J}_i(\mathbf{m}_0, \mathbf{q}_0) \tag{8}$$

for all sources $i = 1 \cdots n_s$. In this expression, we make use of linearity of the wave equation

with respect to its source. To simplify the notation, we also overload the temporal convolution

(denoted by the symbol $*$) to apply to all data—i.e. all traces in the shot records.

Based on the above relationship, we propose to solve for $\mathbf{w}$ after each linearized Bregman

iteration (line 10 of Algorithm 2) via

$$\min_{\mathbf{w}} \sum_{i \in \mathcal{I}_k} \|\mathbf{w} * \mathbf{J}_i(\mathbf{m}_0, \mathbf{q}_0)\mathbf{C}^T\mathbf{x} - \delta\mathbf{d}_i\|_2^2 + \|\mathbf{r} \odot (\mathbf{w} * \mathbf{q}_0)\|_2^2 \tag{9}$$

with $\mathcal{I}_k \subset [1 \cdots n_s], |\mathcal{I}| = n_s'$ a randomly chosen shot subset of shot records.

To prevent overfitting while fitting the generated data $\tilde{\mathbf{b}}_k$ at the kth iteration to the

observed data $\mathbf{b}_k$, we include a penalty term $\mathbf{r}$ consisting of an exponential weighting vector

as given by:

$$\mathbf{r}(t) = \nu + \log(1 + e^{\alpha(t-t_0)}). \tag{10}$$

In this expression, the scalar α determines the rate of growth after $t = t_0$. We choose t_0

such that oscillations related to overfitting are suppressed after this time. This prevents

overfitting and ensures the filters $\mathbf{w}_k$ to be short such that the estimated source time function

$\mathbf{q} = \mathbf{w}_k * \mathbf{q}_0$ remains short as well. The weight parameter ν penalizes the energy of the

estimated source $\mathbf{q}$, which also helps to alleviate the ill-conditioning of this sub-problem.

We summarize the different steps of our approach in Algorithm 2 below. As earlier, we

11

solve the sparsity-promoting optimization problem via linearized Bregman iterations, which now include a correlation (correlation denoted by the symbol $\star$ is the adjoint of convolution) with the current estimate for source time function correction ($\mathbf{w}_k$) in line 8. We initialize the source time function correction with a discrete Delta distribution ($\mathbf{w}_0 = \delta$). We refer to this method with on-the-fly source estimation as sparsity-promoting LS-RTM with source estimation (SPLS-RTM-SE).

Algorithm 2 LB for LS-RTM with source estimation

---

1. Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{w}_0 = \delta$, $\mathbf{z}_0 = \mathbf{0}$, $\mathbf{q}_0$, $\mathbf{r}(t)$, batchsize $n_s' \ll n_s$, $\mathbf{r}$

2. **for** $k = 0, 1, \cdots$

3.   Randomly choose shot subsets $\mathcal{I}_k \subset [1 \cdots n_s], |\mathcal{I}| = n_s'$

4. $\mathbf{A}_k = \{\mathbf{J}_i(\mathbf{m}_0, \mathbf{q}_0)\mathbf{C}^\top\}_{i \in \mathcal{I}_k}$

5. $\mathbf{b}_k = \{\delta\mathbf{d}_i\}_{i \in \mathcal{I}_k}$

6. $\tilde{\mathbf{b}}_k = \mathbf{A}_k\mathbf{x}_k$

7. $t_k = \|\tilde{\mathbf{b}}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top(\tilde{\mathbf{b}}_k - \mathbf{b}_k)\|_2^2$

8. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k\mathbf{A}_k^\top\left(\mathbf{w}_k\star\mathcal{P}_\sigma(\mathbf{w}_k * \tilde{\mathbf{b}}_k - \mathbf{b}_k)\right)$

9. $\mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$

10. $\mathbf{w}_{k+1} = \arg\min_{\mathbf{w}} \|\mathbf{w} * \tilde{\mathbf{b}}_k - \mathbf{b}_k\|_2^2 + \|\mathbf{r} \odot (\mathbf{w} * \mathbf{q}_0)\|_2^2$

11. **end**

12. **Output:** $\hat{\mathbf{q}} = \mathbf{w}_{k+1} * \mathbf{q}_0$ and $\hat{\delta\mathbf{m}} = \mathbf{C}^\top\mathbf{x}_{k+1}$

NUMERICAL EXPERIMENTS

In this experiment section, we demonstrate the viability of our approach by means of carefully designed synthetic examples. We start by showing that linearized Bregman iterations with on-the-fly source estimation are indeed able to jointly estimate the source and the sparse vector

of image curvelet coefficients. Next, we consider an imaging experiment on the Marmousi model emphasizing the importance of including the source function and the influence of noise. We conclude by introducing a practical workflow that is capable of handling salt-related imaging problems.

Stylized example

While imaging with linearized Bregman (LB) iterations has resulted in high-fidelity true-amplitude images in complex models (Witte et al., 2019c), the viability of this alternative sparsity-promoting approach has not yet been verified in combination with on-the-fly source estimation. For this purpose, we examine the performance of LB with source estimation on a simplified stylized example. As we can see, Equation 8 implies a bilinear dependence of the reflected data on both the unknown filter $\mathbf{w}$ and the curvelet coefficient vector $\mathbf{x}$. It is well known that this sort of bilinear dependence can give rise to ambiguities even though the vector $\mathbf{x}$ is sparse.

We exemplify this seismic bilinear relationship be defining $\mathbf{WAx} = \mathbf{b}$, where the matrix $\mathbf{A} \in \mathcal{R}^{20000 \times 10000}$ is ill-conditioned, with rank($\mathbf{A}$) = 500. The sparse vector $\mathbf{x} \in \mathcal{R}^{10000 \times 1}$ has only 20 random non-zero elements. A block of the tall matrix, $\mathbf{A}_i \in \mathcal{R}^{500 \times 10000}, i \in [1 \ldots 40]$ serves as a proxy for the LB modeling operator $\mathbf{J}_i$ for the ith shot with only one single trace. We implement the trace-by-trace convolution via a Toeplitz matrix defined in terms of the filter $\mathbf{w} \in \mathcal{R}^{500 \times 1}$ acting on each $\mathbf{A}_i \mathbf{x}$. The multiplication of the convolution matrix $\mathbf{W} \in \mathcal{R}^{20000 \times 20000}$ with $\mathbf{Ax}$ compactly represents the repeated convolutions of the filter with all traces.

This example, designed to jointly invert for $\mathbf{x}$ and $\mathbf{w}$, aims to exhibit the capability of our

Algorithm 2 to carry out seismic imaging and on-the-fly source estimation. To demonstrate the effect of the penalty term in line 10 of Algorithm 2, we compare sparsity-promoting solutions for the fixed true wavelet to solutions with on-the-fly source estimation with and without the additional penalty. During each iteration, we randomly choose 10% of the blocks of the tall matrix $\mathbf{A}$ and we run five passes through the data in total. After some parameter testing, we choose the following values for the penalty parameters: $\lambda = 1, \nu = 1$ and $\alpha = 8$. We find that different choices for these penalty parameters have little effect on our inversion results. Finally, the time parameter t_0 is set according to the approximate duration of the filter $\mathbf{w}$, which in this case corresponds to a Ricker wavelet since we choose $\mathbf{q}_0$ to be a delta Dirac. We also initialize the filter $\mathbf{w}$ with a normalized Dirac. Because of the well-known amplitude ambiguity inherent to blind deconvolution problems, we normalize the ℓ_2-norms of the estimated source and reflectivity.

Pairs of estimated sparse "reflectivities" ($\hat{\delta\mathbf{m}}$) and source functions ($\hat{\mathbf{q}} = \mathbf{w}_{final} * \mathbf{q}_0$) after normalization are included in Figure 1 . We can draw the following conclusions from these results. First, for the noise-free data, the LB iterations are able to recover the sparse "reflectivity" and source function well up to a constant single amplitude factor, which we correct by normalizing its ℓ_2-norm. Second, the estimated source function and reflectivity become noisy (cf. the dotted line in Figure 1a and the dot line in Figure 1b) when we do not include a penalty enforcing the estimated filter to be short in time. Finally, the method is robust with respect to noise as we can see from Figures 1c and 1d where 10% Gaussian noise is added. This result stresses the importance of including the penalty.

[Figure 1 about here.]

Experiments on the modified Marmousi model

To illustrate the performance and robustness with respect to noise of the proposed SPLS-RTM-SE method, we consider a model with complex layered stratigraphy. We derive this imaging example from the well-known synthetic Marmousi model (Brougois et al., 1990), which is 3.2 km deep and 8.0 km wide, with a grid size of 5×5 m. To avoid imaging artifacts, we use a background velocity that is kinematically correct. We simulate 320 equally spaced sources positioned at a depth of 25m. We use a minimum phase source time function with its significant spectrum ranging from 10 to 40 Hz as shown in Figure 2. We use this type of source to generate linear data by applying the demigration operator ($\mathbf{J}_i(\mathbf{m}_0, \mathbf{q})$, $i = 1 \cdots n_s$) to a bandwidth limited medium perturbation $\delta\mathbf{m}$ given by the difference between two smoothings of the true medium (Huang et al., 2016). We record data at 320 equally spaced co-located receivers. To assess the sensitivity to noise, we create two additional data sets by adding zero-centered Gaussian noise whose energies are 50% and 200% of the simulated linear data respectively.

Contrary to source estimation in the frequency domain, we need an initial source function $\mathbf{q}_0$ for the source time function (see Figure 2a and 2b where the initial source time function and its amplitude spectrum are depicted by dashed black lines). We need this initial source function to make sure that the finite-difference propagators remain stable. To make sure we do not exceed the valid frequency range of our simulations, we choose the frequency band of the initial source time function broader than the true one. To circumvent bias, we initialize the time function with a flat amplitude spectrum between $20 - 50$ Hz. To allow for a realistic scenario, we apply a phase shift to this initial guess making it mixed phase and non symmetric .

To carry out the alternating inversion for the reflectivity and unknown filter $\mathbf{w}$, we run Algorithm 2 for 40 iterations with a batch size of 8—i.e., we use 8 randomly selected sources per iteration without replacement. The total number of wave equation solves is equivalent to touching each shot only once—i.e., we make one pass through the data. To improve the convergence of the inversion, we employ preconditioners in both the data and model domains (see Herrmann et al., 2009, for detail). To remove the imprint of the sources/receivers on the image, we also include a top mute to our operators. Also, we apply a mute to the data to suppress the dominating water bottom reflection and long offsets. Finally, we choose the thresholding parameter λ to be 10% of the maximum value of the first gradient to avoid unnecessary extra iterations resulting from a threshold value that is too large or small.

[Figure 2 about here.]

The estimated source functions $\hat{\mathbf{q}} = \mathbf{w}_{final} * \mathbf{q}_0$ and their amplitude spectra after applying an ℓ_2-norm normalization are included in Figure 2. Overall we can see that the source functions are well recovered despite the presence of noise. For a small amount of noise, the estimated spectrum is the same as the one obtained from the noise-free data while the source function obtained from data with a high noise level is less smooth, but closer to the true source function. Other than the fact that we are dealing with nonlinear blind deconvolution, we do not have an explanation for this behavior. While the noise dependence of the estimated source functions behaves somewhat aberrant, the recovered reflectivity behaves as expected (cf. Figures 3a and 3b for images obtained with the true source and with the initial guess and images 4a – 4c obtained with on-the-fly source estimation for noise-free and noisy data.) We can make the following observations from these experiments: first, it is important to image with the correct source even when the data is noise free. While our sparsity-promoting

16

scheme is able to recover a high-resolution image (see Figure 3a) when the source function corresponds to the true source, the image quality deteriorates rapidly if the amplitude and phase spectra of the wavelet are wrong (see Figure 3b). Energy is no longer focussed and the shape and locations of the imaged reflectors are off. However, the results included in Figure 4 demonstrate that good results can be obtained when estimating the source function on the fly. The estimated reflectivity depicted in Figure 4a is close to the reflectivity obtained when we image with the true source function (cf. Figures 3a and 4). Moreover, the estimated images are, as expected, relatively insensitive to noise in the data albeit the imaged reflectivity for the high noise case somewhat deteriorated (cf. Figures 4a – 4c). Contrary to the imaging result for the wrong initial source function, the reflectors are positioned correctly and have the correct phase, shape, and amplitude, even in situations of substantial noise although at the expense of some remaining low- and high-frequency artifacts. The latter are related to the use of the curvelet transform and are to be expected. Overall, these results confirm the robustness of our imaging algorithm in the situation where there is significant noise in the data.

[Figure 3 about here.]

[Figure 4 about here.]

To arrive at the estimated images in Figure 4, we set the penalty parameters $\nu = 1$ and $\alpha = 8$ in Algorithm 2. After the first source estimation in the second iteration, we reset the coefficients $\mathbf{z}$ and $\mathbf{x}$ to zero to avoid spending too many iterations on correcting wrongly located reflectors from the first iteration in which the initial guess of the source wavelet is used. In addition to the visual quality of the estimated images, convergence

plots for the relative error for the data residual (the relative ℓ_2-norm error between the observed data and the demigrated data for estimated reflectivity $\delta\hat{\mathbf{m}}$ convolved with the estimated filter, $\frac{\|\mathbf{w}_k*\tilde{\mathbf{b}}_k-\mathbf{b}_k\|_2}{\|\mathbf{b}_k\|_2}$) and the relative model error (the ℓ_2-norm error between the true reflectivity and the recovered reflectivity, $\frac{\|\delta\hat{\mathbf{m}}_k-\delta\mathbf{m}\|_2}{\|\delta\mathbf{m}\|_2}$) confirm our observation that Algorithm 2 is capable of providing high quality images in the absence of precise knowledge on the source function and in the presence of substantial noise. Our approach arrives at these least-squares images at the cost of a single data pass. Understandably, the algorithm starts off with a large relative residual and model error due to the wrong initial guess for the source function. As Algorithm 2 progresses, these relative errors continue to decay and are comparable to the convergence plots for the true source function. Because on-the-fly source estimation improves our ability to adapt to the data, the relative data residual for the noise-free case (dashed line) is even better then the relative error in case the source function is known (solid line). While encouraging, these results are obtained for a relatively simple imaging experiment and for data that is obtained with linearized modeling via demigration. In other words, we commit an inversion crime. In the next section, we will show that the proposed method also performs well in more complicated settings with nonlinear data.

[Figure 5 about here.]

Experiments on the Sigsbee model

Sparsity-promoting imaging algorithms such as SPLS-RTM (Algoritm 2) are designed to handle complex imaging scenarios with strong velocity contrasts and strong lateral velocity variations. Examples of such scenarios are salt plays where reflections underneath the salt are of interest. To demonstrate the viability of our imaging approach with on-the-fly source

18

estimation in this scenario, we consider the challenging Sigsbee2A model of size 24.4×9.2 km. This model contains a large salt body and a number of faults and point diffractors. To demonstrate the capability of our approach to handel this challenging situation, we simulate nonlinear data for a marine acquisition without a free surface. We model 960 sources in total, with each shot record being recorded by an array of 320 receivers with 25 m receiver spacing, a maximum offset of 8 km and a towing depth of 15 m. We use a source wavelet with a peak frequency at 15Hz (see Figure 6) and we record for 10 s.

As is customary during imaging under salt, we use a background velocity model that features salt with relatively strong and therefore reflecting boundaries. We approximate linear data by using this background velocity model to generate data, which we subtract from the simulated data in the true Sigsbee2A model (i.e. from the observed data). Due to the presence of salt in the background model, the incident wavefield contains reflections that give rise to unwanted tomographic low-frequency artifacts in the image. This problem is widely reported in the literature (e.g. Yoon and Marfurt (2006); Guitton et al. (2007)). To remove these imaging artifacts, we replace the conventional imaging condition for RTM by the inverse-scattering imaging condition (Stolk et al., 2012; Whitmore and Crawley, 2012; Witte et al., 2017). While this condition has been proven capable of removing tomographic artifacts during RTM (Whitmore and Crawley, 2012; Witte et al., 2017) and sparsity-promoting least-squares RTM (Witte et al., 2017), it changes the linearized forward operator (the Jacobian $\mathbf{J}_i$), resulting in an inconsistent system. Contrary to RTM with the conventional imaging condition, imaging with the inverse scattering imaging condition corresponds to estimating perturbations in the impedance, rather than in the velocity.

Unfortunately, the difference in which quantity is being imaged, is problematic for our proposed on-the-fly source estimation, which tries to correct for inconsistencies between

19

observed and predicted data. Contrary to the situations where we use the conventional imaging condition, the data residual now contains contributions from the wrong wavelet and the linearized imaging condition, which leads to wrong estimates for the unknown source function. We overcome this problem via a hybrid iterative algorithm where we switch imaging conditions during the iterations as outlined in Algorithm 2. To estimate the source function, we first iterate with the conventional imaging condition. Since the convergence to the source function is fast, we switch after five iterations to the scattering imaging, but keep the estimated source function fixed. Basically, we jump from Algorithm 2 to Algorithm 1.

Results of this hybrid approach are summarized in Figures 6 – 8. As before, we compare our results with on-the-fly source estimation to SPLS-RTM for the true source function. The initial guess and estimated wavelets in Figure 6 again confirm the validity of our approach, yielding a reasonably accurate estimate for the source after only five iterations and subsequent normalization of the ℓ_2-norm. Imaging results obtained after twenty iterations with 10% of the shots, which amounts to two data passes in total, are included in Figures 7 and 8. Unlike a typical RTM image, images obtained by SPLS-RTM are well resolved and contain true amplitude. This is because we invert the linearized modeling operator, which compensates for the source, finite aperture, and propagation effects. As before, we include preconditioners and mutes to improve the conditioning number of the linear system. Comparisons of Figure 7a, obtained with Algoritm 1 with the true source function, and Figure 7b, which we compute with our hybrid method switching from Algoritm 2 to Algoritm 1 after five iterations, show near identical results, thus confirming the validity of the proposed approach. These observations are confirmed by trace-by-trace comparisons in Figure 8. Similar to the Marmousi experiments, we set the penalty parameters $\alpha = 8$, and $\nu = 1$, and the thresholding parameter λ is set according to 10% of the maximum absolute amplitude level of $\mathbf{z}_1$.

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

DISCUSSION

Due to its computational costs, sparsity-promoting LS-RTM has been an expensive proposition and is for this reason not yet widely adapted while this method is capable of achieving images with high-amplitude fidelity and fewer artifacts. With the proposed work, we are able to come up with an alternative low-cost approach combining techniques from stochastic optimization and sparsity-promotion with on-the-fly source estimation using the technique of variable projection. Compared to earlier work, addressing the memory demands of (LS-)RTM via on-fly-Fourier transforms (Witte et al., 2019c), we tackle the problem of on-the-fly source estimation in the time domain. Because we use industry-strength time-domain finite-difference propagators provided by Devito (Luporini et al., 2018; Louboutin et al., 2019) and exposed in the Julia programming language by JUDI (Witte et al., 2019b), our approach scales in principle to large 3D industrial problems. While we address the importance of estimating source function, we believe that the sensitivity of LS-RTM to errors in the background velocity model needs to be studied as well albeit early work on time-harmonic LS-RTM showing some robustness with respect to these errors (Tu and Herrmann, 2015a). Combining our approach with the method of on-the-fly Fourier transforms is also a topic that needs further study.

While carrying out full scale 3D (LS-)RTM experiments is generally out of reach for academia where access to large high-performance compute is often limited, recent work on

21

a serverless implementation of RTM (Witte et al., 2019a,d) has shown that industry-scale workloads can be run in the cloud leveraging the power of Devito. In the not too distant future, we plan to demonstrate the presented method on an industry-scale imaging problem using tilted transversely-isotropic propagators. This would truly exemplify the power of modern code bases and linear algebra abstractions as utilized by JUDI (Witte et al., 2019b). This framework gives us flexibility for instance to switch to more involved 3D propagators or to estimating source-time functions that are allowed to vary along the survey.

Finally, since sparsity-promoting LS-RTM carries out inversions, we expect to be able to obtain images from sparsely sampled data, e.g. data collected with sparse ocean bottom nodes and (multi-)source vessel simultaneous recordings. We plan to report on these aspects in the not too distant future.

CONCLUSION

We proposed a scalable time-domain approach to sparsity-promoting least-squared reverse time migration with on-the-fly source estimation in principle suitable for industrial 3D imaging problems. The presented approach leverages recently developed techniques from convex optimization and variable projection that greatly reduce costs and the necessity to provide an estimate for the source function. As a result, our approach is capable of generating high-fidelity true-amplitude images including source estimates at the cost of roughly one to two migrations involving all data.

By means of carefully designed experiments in 2D, we were able to demonstrate that our method is capable of handling noisy data and complex imaging settings such as salt. We were able to image under salt, which is often plagued by low-frequency tomographic

artifacts, by switching between applying the conventional imaging condition initially, followed by iterations that apply the inverse-scattering condition. In this way, we estimated the source function first while creating an artifact-free image with later iterations during which the imaging condition was switched while keeping the source function fixed.

Because the presented method relies on time-domain propagators, we anticipate it will be able to scale to large 3D industrial imaging problems. Because 3D imaging with full-azimuthal sparse data typically provided good illumination of the reservoir, we expect the proposed methodology to produce high fidelity results at a cost of roughly one to two reverse time migrations involving all shots.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Aoki, N., and G. T. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, WCA83–WCA93.

Aravkin, A. Y., T. van Leeuwen, H. Calandra, and F. J. Herrmann, 2012, Source estimation for frequency-domain fwi with robust penalties: Presented at the 74th EAGE Conference and Exhibition incorporating EUROPEC 2012.

Aravkin, A. Y., T. van Leeuwen, and N. Tu, 2013, Sparse seismic imaging using variable projection: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2065–2069.

Baysal, E., D. D. Kosloff, and J. W. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

Bednar, J. B., C. J. Bednar, and C. Shin, 2006, Two-way vs one-way: a case study style comparison, *in* SEG Technical Program Expanded Abstracts 2006: Society of Exploration Geophysicists, 2343–2347.

Brougois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1990, Marmousi, model and data: EAGE workshop-practical aspects of seismic data inversion, European Association of Geoscientists & Engineers, cp–108.

Cai, J.-F., S. Osher, and Z. Shen, 2009, Convergence of the linearized Bregman iteration for L1-norm minimization: Mathematics of Computation, **78**, 2127–2136.

Candès, E. J., et al., 2006, Compressive sampling: Proceedings of the international congress of mathematicians, Madrid, Spain, 1433–1452.

Candès, E. J., and M. B. Wakin, 2008, An introduction to compressive sampling: IEEE signal processing magazine, **25**, 21–30.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit:

SIAM review, **43**, 129–159.

Dong, S., J. Cai, M. Guo, S. Suh, Z. Zhang, B. Wang, and e. Z. Li, 2012, Least-squares reverse time migration: Towards true amplitude imaging and improving the resolution, *in* SEG Technical Program Expanded Abstracts 2012: Society of Exploration Geophysicists, 1–5.

Donoho, D. L., 2006, Compressed sensing: IEEE Transactions on information theory, **52**, 1289–1306.

Fang, Z., R. Wang, and F. J. Herrmann, 2018, Source estimation for wavefield-reconstruction inversion: Geophysics, **83**, R345–R359.

Guitton, A., B. Kaelin, and B. Biondi, 2007, Least-squares attenuation of reverse-time-migration artifacts: Geophysics, **72**, S19–S23.

Haber, E., M. Chung, and F. Herrmann, 2012, An effective method for parameter estimation with PDE constraints with multiple right-hand sides: SIAM Journal on Optimization, **22**, 739–757.

Herrmann, F. J., C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41–A46.

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical prospecting, **60**, 696–712.

Herrmann, F. J., N. Tu, and E. Esser, 2015, Fast "online" migration with compressive sensing: Presented at the Proceedings of the 77th EAGE Conference and Exhibition.

Hou, J., and W. W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, S165–S179.

Huang, Y., R. Nammour, and W. Symes, 2016, Flexibly preconditioned extended least-squares migration in shot-record domain: Geophysics, **81**, S299–S315.

Lorenz, D. A., F. Schopfer, and S. Wenger, 2014, The linearized bregman method via split feasibility problems: Analysis and generalizations: SIAM Journal on Imaging Sciences, **7**, 1237–1262.

Louboutin, M., M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman, 2019, Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration: Geoscientific Model Development, **12**, 1165–1187.

Luporini, F., M. Lange, M. Louboutin, N. Kukreja, J. Hückelheim, C. Yount, P. Witte, P. H. J. Kelly, F. J. Herrmann, and G. J. Gorman, 2018, Architecture and performance of devito, a system for automated stencil computation: CoRR, **abs/1807.03032**.

Mulder, W. A., and R.-E. Plessix, 2004, A comparison between one-way and two-way wave-equation migration: Geophysics, **69**, 1491–1504.

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

Powell, W. B., 2014, Clearing the jungle of stochastic optimization, *in* Bridging data and decisions: Informs, 109–137.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, 874–881.

Schuster, G., 1983, Least-squares cross-well migration, *in* SEG Technical Program Expanded Abstracts 1993: Society of Exploration Geophysicists, 110–113.

Stolk, C., M. De Hoop, et al., 2012, Linearized inverse scattering based on seismic reverse time migration: Journal de Mathématiques Pures et Appliquées, **98**.

Tu, N., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2013, Fast least-squares

migration with multiples and source estimation: Presented at the Proceedings of the 75th EAGE Conference and Exhibition.

Tu, N., and F. J. Herrmann, 2015a, Fast imaging with surface-related multiples by sparse inversion: Geophysical Journal International, **201**, 304–317.

———, 2015b, Fast least-squares imaging with surface-related multiples: Application to a north sea data set: The Leading Edge, **34**, 788–794.

van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: International Journal of Geophysics, **2011**.

van Leeuwen, T., A. Y. Aravkin, F. J. Herrmann, M. Li, J. Rickett, and A. Abubakar, 2014, Reply to the discussion: Geophysics, **79**, X11–X15.

Whitmore, N., 1983, Iterative depth migration by backward time propagation, *in* SEG Technical Program Expanded Abstracts 1983: Society of Exploration Geophysicists, 382–385.

Whitmore, N., and S. Crawley, 2012, Applications of rtm inverse scattering imaging conditions, *in* SEG technical program expanded abstracts 2012: Society of Exploration Geophysicists, 1–6.

Witte, P. A., M. Louboutin, C. Jones, and F. J. Herrmann, 2019a, Serverless seismic imaging in the cloud. Submitted to Rice Oil and Gas High Performance Computing Conference 2020 on November 27, 2019.

Witte, P. A., M. Louboutin, N. Kukreja, F. Luporini, M. Lange, G. J. Gorman, and F. J. Herrmann, 2019b, A large-scale framework for symbolic implementations of seismic inversion algorithms in julia: Geophysics, **84**, F57–F71.

Witte, P. A., M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann, 2019c, Compressive least-squares migration with on-the-fly fourier transforms: Geophysics, **84**,

1–76.

Witte, P. A., M. Louboutin, H. Modzelewski, C. Jones, J. Selvage, and F. J. Herrmann, 2019d, An event-driven approach to serverless seismic imaging in the cloud. Submitted to IEEE Transactions on Parallel and Distributed Systems on August 20, 2019.

Witte, P. A., M. Yang, and F. J. Herrmann, 2017, Sparsity-promoting least-squares migration with the linearized inverse scattering imaging condition: Presented at the Proceedings of the 79th EAGE Conference and Exhibition.

Yang, M., P. Witte, Z. Fang, and F. Herrmann, 2016, Time-domain sparsity-promoting least-squares migration with source estimation, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 4225–4229.

Yin, W., 2010, Analysis and generalizations of the linearized bregman method: SIAM Journal on Imaging Sciences, **3**, 856–877.

Yoon, K., and K. J. Marfurt, 2006, Reverse-time migration using the Poynting vector: Exploration Geophysics, **37**, 102–107.

Zeng, C., S. Dong, and B. Wang, 2014, Least-squares reverse time migration: Inversion-based imaging toward true reflectivity: The Leading Edge, **33**, 962–968.

LIST OF FIGURES

Figure 1: Comparison of solutions obtained with the LB iterations (see Algorithm 2) for a fixed true source (denoted by the dash-dot line) and for on-the-fly source estimations with and without penalties. We obtained results with five passes through the data. Our method is capable of estimating the "reflectivity" (a) and "source function" (b) after normalizing the ℓ_2-norm. The proposed method is also robust with respect to additive noise as we can see in (c) and (d). We add 10% Gaussian noise.
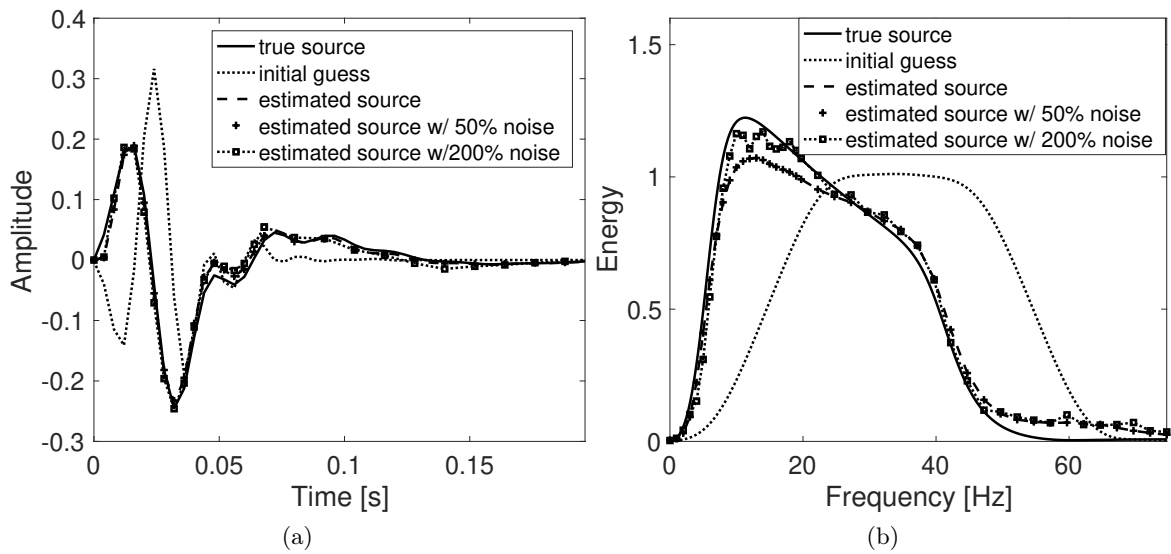
Figure 2: Comparison between true, initial, and estimated source time functions ($\mathbf{q}_0$, $\hat{\mathbf{q}} = \mathbf{w}_{final} * \mathbf{q}_0$) and their associated amplitude spectra. (a) The time signatures and (b) the frequency spectra. The estimated source time functions and spectra are obtained from noise-free data and from data to which zero-centered Gaussian noise is added with energy ranging from 50% to 200% of the simulated linear data.
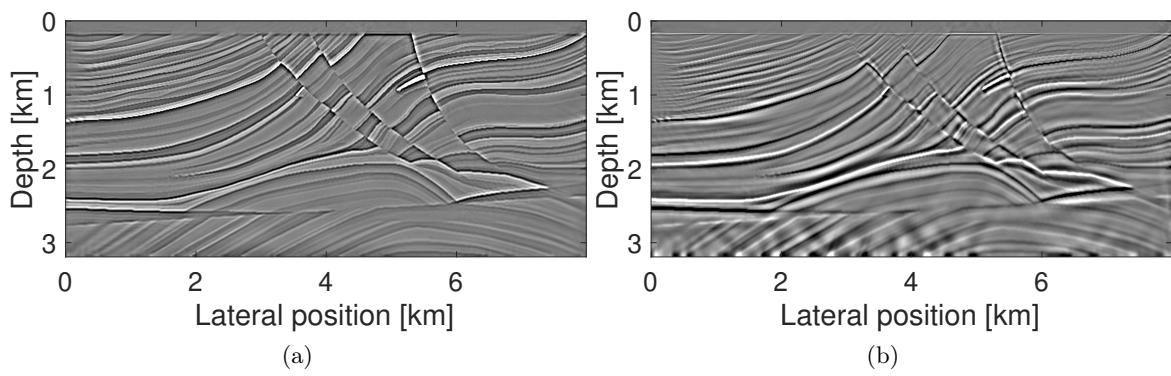
Figure 3: Inverted images with (a) the true source and (b) initial source, generated by the Marmousi model.
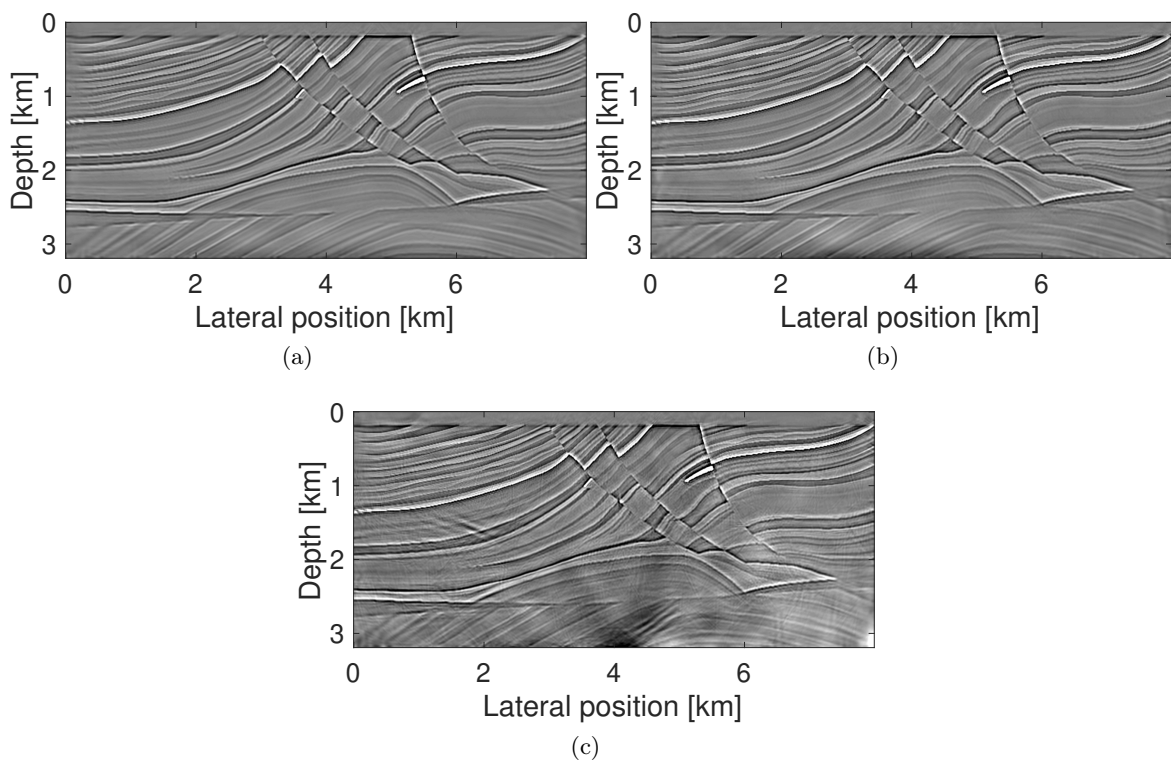
Figure 4: Inverted images with source estimation (Algorithm 2) for (a) noise-free data, (b) data with 50% noise and (c) data with 200% noise, generated from the Marmousi model.
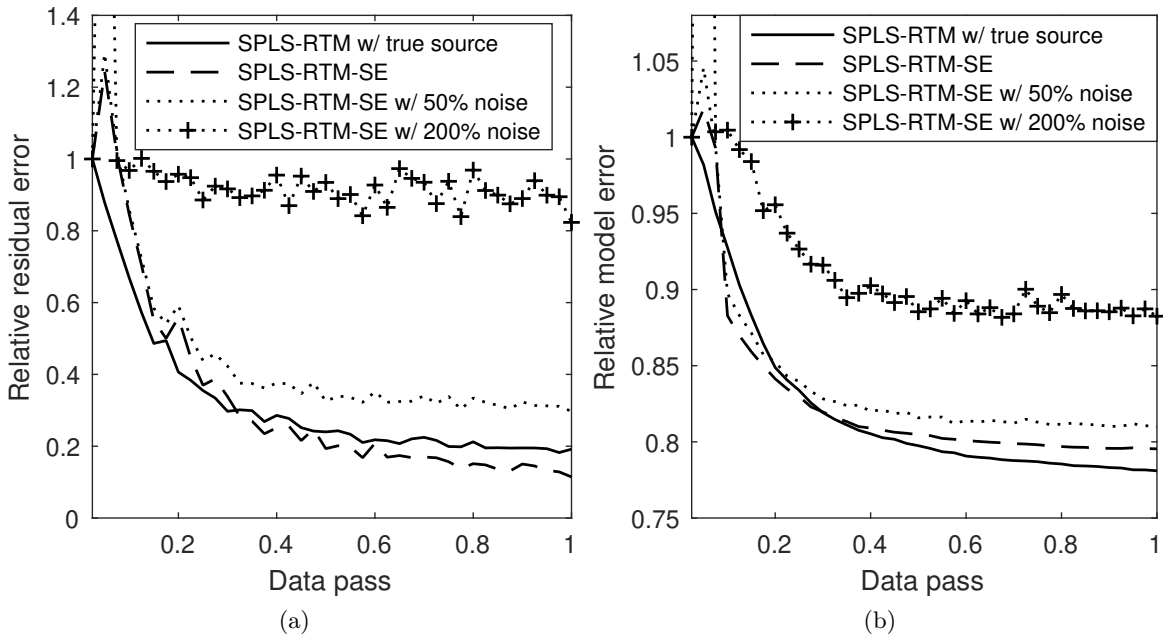
Figure 5: Convergence plots for the relative residual error (a)—i.e., the relative ℓ_2-norm error between the observed data and the demigrated data for estimated reflectivity $\delta\hat{\mathbf{m}}_k$ convolved with the estimated filter $\mathbf{w}_k$, and the relative model error (b)—i.e., the relative ℓ_2-norm error between the true reflectivity and the recovered reflectivity.
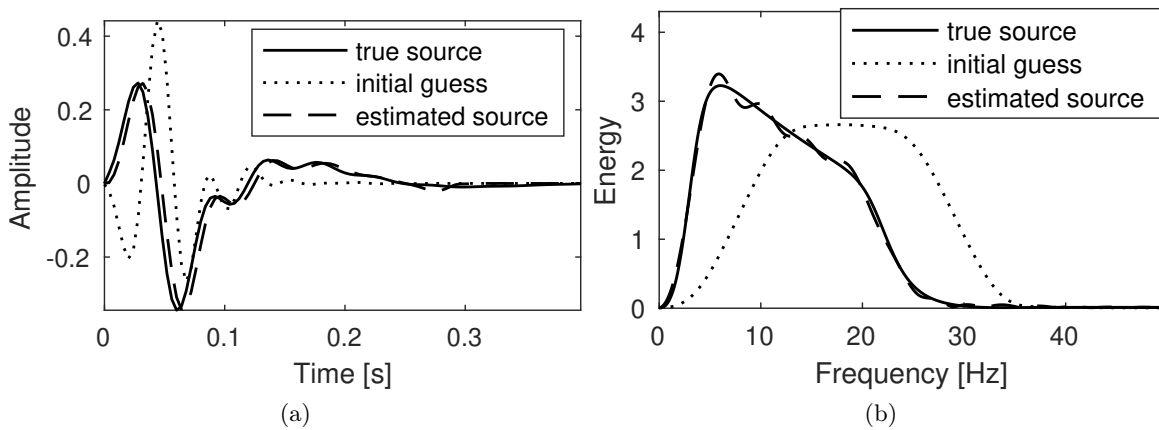
Figure 6: Comparison of true, initial, and estimated source time functions $\left(\mathbf{q}_0, \hat{\mathbf{q}} = \mathbf{w}_{final} * \mathbf{q}_0\right)$ and their associated amplitude spectra. (a) The time signatures and (b) the frequency spectra. The estimated source time functions and spectra are obtained during the first five iterations with the conventional imaging condition.
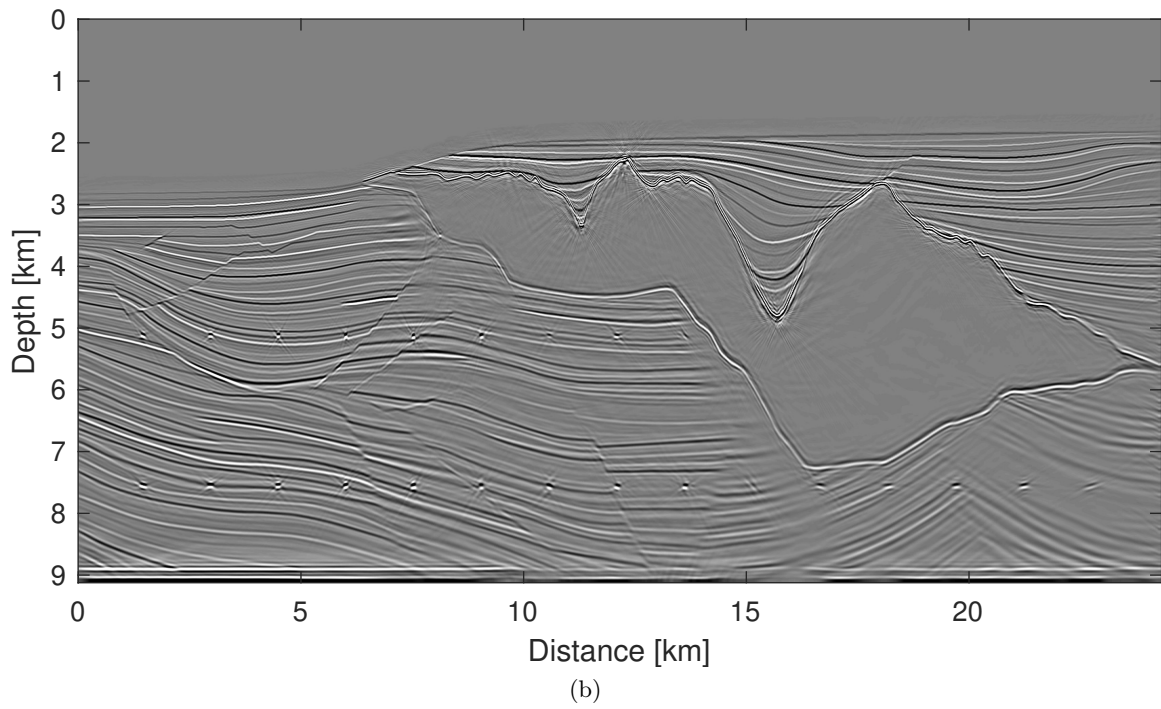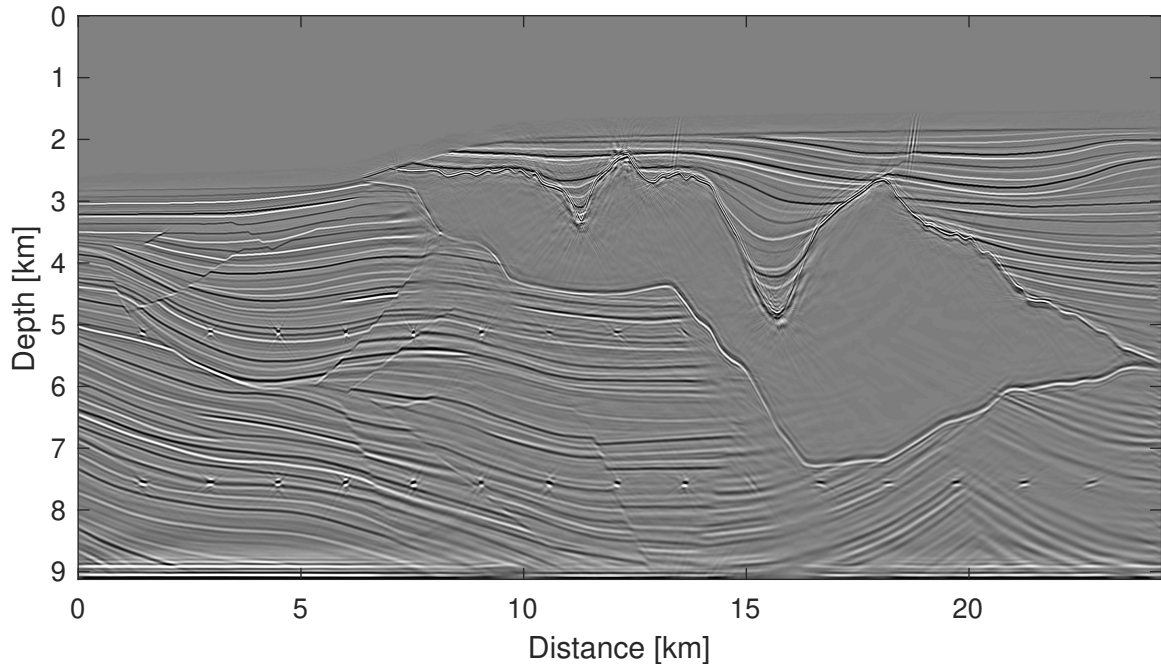
Figure 7: Comparison between (a) SPLS-RTM with the true source wavelet and (b) SPLS-RTM with on-the-fly source estimation. In both cases, we conduct total 20 iterations amounting to two data passes.
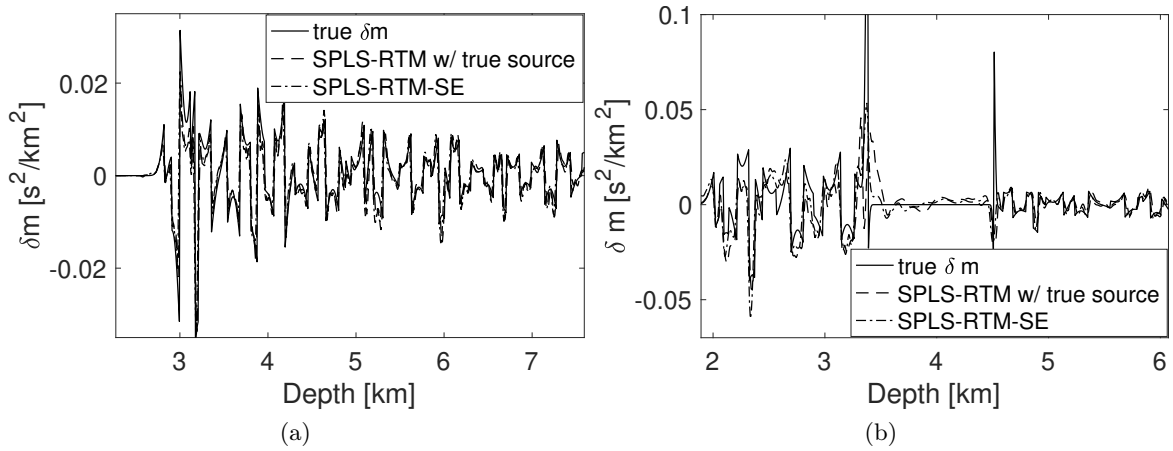
Figure 8: Trace by trace comparisons between the true model perturbation, the images from SPLS-RTM with the true source and with on-the-fly source estimation. The traces in (a) and (b) are extracted from lateral positions $x = 4.5$ and $x = 11.3$ km, respectively.