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

Mengmeng Yang\textsuperscript{1*},\textsuperscript{†}, Zhilong Fang\textsuperscript{2,†}, Philipp Witte\textsuperscript{3} and Felix J. Herrmann\textsuperscript{1,3}

\textsuperscript{1}School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30318, USA, \textsuperscript{2}Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, and \textsuperscript{3}School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA

Received February 2020, revision accepted July 2020

ABSTRACT
Least-squares reverse-time migration is well known for its capability to generate artefact-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 artefacts and restoring amplitudes. While powerful, 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 estimate 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.

Key words: Imaging, Inversion.

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 $\ell_2$-norm. Least-squares migration was initially introduced for ray-based methods (Schuster, 1983; Lambaré et al., 1992; Nemeth et al., 1999) to be followed by wave-equation based RTM (Valenciano, 2008; Dong et al., 2012; Zeng et al., 2014). Recent work has focussed on incorporating more accurate physics into least-squares RTM.
δF denote the model perturbation, so a single shot time samples over a time interval of T. The number of receivers is q.

Risk of overfitting the data when minimizing the ℓ2-norm of the data residual. This overfitting may introduce noise-related artefacts 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 (LB). Stochastic optimization allows us to work with small random subsets of shots, which limits the number of wave-equation solves – that is, passes through the data. Convergence is guaranteed (Herrmann et al., 2013; Bsemi Yang et al., 2016; Bsemi Witte et al., 2019c) by replacing the ℓ1-norm, by an elastic net consisting of a strongly convex combination of ℓ1- and ℓ2-norm objectives. Inclusion of the ℓ2-norm results in a greatly simplified algorithm involving LB 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 RTM (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; Bsemi 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 RTM and least-squares RTM. 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 (SPLS) migration with stochastic optimization, followed by our approach to on-the-fly source estimation during the iterations.

**From reverse time migration to least-squares reverse time migration**

Reverse-time migration (RTM) derives from a linearization (see, e.g. Mulder and Plessix, 2004) with respect to the background squared slowness. Following the notation used by Symes (2008), this linearization for the ith source reads

\[
\delta d_i = F_i(m_0 + \delta m, q) - F_i(m_0, q) \approx J_i(m_0, q) \delta m, 
\]

where the vectors \(\delta m, q\) and \(\delta d\) denote the model perturbation, the source-time function and the corresponding data perturbation that is known as the reflection data used in imaging, 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 $n_s \times n_t$. The non-linear forward modelling operator $F_i(m, q)$ for the $i$th source location involves the solution of the discretized acoustic wave equation

$$\left( m \odot \frac{\partial^2}{\partial t^2} - \Delta \right) u_i = P_i^\top q,$$

where $\odot$ represents the elementwise multiplication and $\Delta$ represents the discretized Laplacian. The linear operator $P_{i,i}$ restricts the wavefield for the $i$th source to the corresponding receiver locations, while the linear operator $P_{i,i}$ injects the source-time function at the location of the $i$th source in the computational grid. The Jacobian $J_i(m_0, q)$ is known as the Born modelling operator and is given by the derivative of $F_i(m, q)$ at the point of $m_0$. Applying the Jacobian $J_i(m, q)$ to the model perturbation $\delta m$ requires the solution of the following linearized equation:

$$\left( m_0 \odot \frac{\partial^2}{\partial t^2} - \Delta \right) \delta u_i = - \frac{\partial}{\partial t^2} (\delta m \odot u_i),$$

where the vector $\delta u_i$ corresponds to the wavefield perturbation for the $i$th 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 – that is, 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

$$\delta m_{\text{RTM}} = \sum_{i=1}^{n_s} J_i^\top \delta d_i,$$

with $n_s$ the number of shots, the adjoint of the Jacobian (denoted by the symbol $\top$) does not correspond to its inverse and $\delta m_{\text{RTM}}$ 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)), least-squares RTM (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 modelling operator, which can significantly mitigate these defects. Solutions of LS-RTM, $\hat{\delta m}$, involve minimization of the following unconstrained least-squares data-fitting problem:

$$\hat{\delta m} = \arg \min_{\delta m} \frac{1}{2} \sum_{i=1}^{n_s} \| J_i(m_0, q) \delta m - \delta d_i \|^2.$$

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 RTM (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 artefact-free images at the cost of two to three passes through the data. The concept of a data pass or epoch is adopted from the machine learning community (Bottou, 2010; Welling and Teh, 2011) where it refers to one complete presentation of the data set to a learning machine. In our case, it refers to touching all shot records once. For a single RTM over all $n_s$ shots this cost corresponds to $2n_s$ wave-equation solves.

**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 artefacts. 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

$$\min_x \| x \|_1, \quad \text{subject to } \sum_{i=1}^{n_s} \| J_i(m_0, q) C_j x - \delta d \|_2 \leq \sigma.$$  

In this formulation, known as the basis pursuit denoise (BPDN; Chen et al., 2001) problem, we include the sparsity-promoting $\ell_1$-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 \( (C^\top) \) and the above program seeks to find the sparsest curvelet coefficient vector that matches the data within the noise level \( \sigma \). While the above problem is known to produce high-fidelity results, its solution relies on iterations that involve a loop over all \( n_i \) 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 \( n_i \) 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 \( \ell_1 \)-norm objective by the strongly convex objective involving

\[
\min_{\mathbf{x}} \lambda_1 \| \mathbf{x} \|_1 + \frac{1}{2} \| \mathbf{x} \|_2^2
\]

subject to

\[
\sum_{i=1}^{n_i} \| J_i(m_0, \mathbf{q}) C^\top \mathbf{x} - \mathbf{d} \|_2 \leq \sigma,
\]

with the estimate for the image given by \( \hat{\mathbf{m}} = C^\top \hat{x} \) where \( \hat{x} \) 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 \( \mathcal{I}_k \subset \{1 \cdots n_i\} \) with cardinality \( |\mathcal{I}| = n'_i \ll n_i \). We choose these subsets without replacement.

For \( \lambda \to \infty \), which in practice means \( \lambda \) 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 \( \lambda \) on the dual variable \( (z_k) \) with a dynamic step length given by \( t_k = \| A_k \mathbf{x}_k - \mathbf{b}_k \|_2 \| A_k^\top (A_k \mathbf{x}_k - \mathbf{b}_k) \|_2 \) (Lorenz et al., 2014). During each iteration, as linearized Bregman (LB) iterations, the residual is projected on to an \( \ell_2 \)-norm ball with the radius \( \sigma \) through a projection operator \( \mathcal{P}_\sigma \). To avoid too many iterations, we set the threshold \( \lambda \) related to the trade-off between the \( \ell_1 \) and \( \ell_2 \)-norm objectives in equation (7), to a value that is not too large – that is, typically proportional to the maximum of \( |z_k| \) 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 next, the background velocity model also needs to be smooth so the tomographic artefacts are 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 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 \( w \) – that is, we have \( \mathbf{q} = w * \mathbf{q}_0 \) where the symbol \( * \) denotes the

---

**Algorithm 1. Linearized Bregman for SPLS-RTM**

1. Initialize \( \mathbf{x}_0 = 0, \mathbf{z}_0 = 0, \mathbf{q}, \lambda_1, \) batchsize \( n'_i \ll n_i \)
2. for \( k = 0, 1, \ldots \)
3. Randomly choose shot subsets \( \mathcal{I}_k \subset \{1 \cdots n_i\}, |\mathcal{I}| = n'_i \)
4. \( A_k = \{ J_i(m_0, \mathbf{q}) C^\top \} \subset \mathcal{I}_k \)
5. \( b_k = \{ \mathbf{d}_i \} \subset \mathcal{I}_k \)
6. \( t_k = \| A_k \mathbf{x}_k - \mathbf{b}_k \|_2 \| A_k^\top (A_k \mathbf{x}_k - \mathbf{b}_k) \|_2 \)
7. \( z_{k+1} = \mathbf{z}_k - t_k A_k^\top \mathcal{P}_\sigma (A_k \mathbf{x}_k - \mathbf{b}_k) \)
8. \( \mathbf{x}_{k+1} = S_{\lambda_1}(z_{k+1}) \)
9. end
10. Output: \( \hat{\mathbf{m}} = C^\top \mathbf{x}_{k+1} \)

**note:** \( S_{\lambda_1}(z_{k+1}) = \text{sign}(z_{k+1}) \max(0, |z_{k+1}| - \lambda_1) \)

© 2020 European Association of Geoscientists & Engineers, *Geophysical Prospecting*, 68, 2697–2711
temporal convolution. Because we assume one and the same source-time function for all shots, we can write

\[ J_i(m_0, w * q_0) = w * J(m_0, q_0), \tag{8} \]

for all sources \( i = 1 \ldots n_i \). In this expression, we use 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 – that is, all traces in the shot records.

Based on the above relationship, we propose to solve for \( w \) after each LB iteration (line 10 of Algorithm 2) via

\[ \min_w \sum_{i=1}^{n_i} \| w * J_i(m_0, q_0) C^T x - \delta d_i \|_2^2 + \| r \odot (w * q_0) \|_2^2, \tag{9} \]

with \( I_k \subset \{1 \ldots n_i \} \), \( |I| = n'_i \) a randomly chosen shot subset of shot records.

To prevent overfitting while fitting the generated data \( \hat{b}_k \) at the \( k \)th iteration to the observed data \( b_k \), we include a penalty term with a weighting vector \( \mathbf{r} \) whose elements are given by:

\[ r_l = r + \log(1 + e^{a(1 - b_l) \Delta t}) \quad \text{for} \quad l = 0 \ldots n - 1, \tag{10} \]

where \( \Delta t \) stands for the temporal sample interval and \( n \) for the number of time samples. In this expression, the scalar \( a \) determines the rate of growth after \( t_0 = l_0 \Delta t \). We choose \( l_0 \) such that oscillations related to overfitting are suppressed after this time \( t_0 \). This prevents overfitting and ensures the filters \( w_k \) to be short such that the estimated source-time function \( \mathbf{q} = w_k * q_0 \) remains short as well. The weight parameter \( r \) 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. As earlier, we solve the sparsity-promoting optimization problem via LB iterations, which now include a correlation (correlation denoted by the symbol \( * \) is the adjoint of convolution) with the current estimate for source-time function correction \( (w_k) \) in line 8. We initialize the source-time function correction with a discrete Delta distribution \( (w_0) = \delta \).

We refer to this method with on-the-fly source estimation as sparsity-promoting LS-RTM with source estimation (SPLS-RTM-SE).

### 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 (LB) 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 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 \( w \) and the curvelet coefficient vector \( x \). It is well known that this sort of bilinear dependence can give rise to ambiguities even though the vector \( x \) is sparse.

We exemplify this seismic bilinear relationship by defining \( Wx = b \), where the matrix \( A \in \mathbb{R}^{20,000 \times 10,000} \) is ill conditioned, with \( \text{rank}(A) = 500 \). The sparse vector \( x \in \mathbb{R}^{10,000 \times 1} \) has only 20 random non-zero elements. A block of the tall matrix, \( A_i \in \mathbb{R}^{500 \times 10,000} \), \( i \in [1 \ldots 40] \) serves as a proxy for the LB modelling operator \( J \), for the \( i \)th shot with only one single trace. We implement the trace-by-trace convolution via a Toeplitz matrix defined in terms of the filter \( w \in \mathbb{R}^{500 \times 1} \) acting on each \( A_i x \). The multiplication of the convolution matrix \( W \in \mathbb{R}^{20,000 \times 20,000} \) with \( A_i x \) compactly represents the repeated convolutions of the filter with all traces.
Figure 1 Comparison of solutions obtained with the linearized Bregman iterations (see Algorithm 2) for a fixed true source (denoted by the circles) 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.

This example, designed to jointly invert for \( x \) and \( w \), 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 \( A \) 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 \( w \), which in this case corresponds to a Ricker wavelet since we choose \( q_0 \) to be a delta Dirac. We also initialize the filter \( w \) 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’ (\( \hat{m} \)) and source functions (\( \hat{q} = w_{\text{final}} \ast q_0 \)) after normalization are included in Fig. 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 dot line in Fig. 1a and the dot line in Fig. 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 Fig. 1(d).
where 10% Gaussian noise is added. This result stresses the importance of including the penalty.

**Experiments on the modified Marmousi model**

To illustrate the performance and robustness with respect to noise of the proposed sparsity-promoting least-squares reverse-time migration source estimation (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 m × 5 m. To validate our source estimation, we first consider an idealized example where we have access linearized data—i.e., data generated with the linearized Born scattering operator. While sparsity-promoting imaging is relatively well behaved with respect to possible errors in the background velocity model (Tu and Herrmann, 2015a), we avoid complications by choosing a velocity model that is kinematically correct. We simulate 320 equally spaced sources positioned at a depth of 25 m. We use a minimum-phase source-time function with its significant spectrum ranging from 10 to 40 Hz as shown in Fig. 2. We use this type of source to generate linear data by applying the demigration operator \( J_i(m_0, q), i = 1 \cdots n_s \) to a bandwidth-limited medium perturbation \( \delta m \) 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-centred 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 \( q_0 \) for the source-time function (see Fig. 2a,b where the initial source-time function and its amplitude spectrum are depicted by dashed black lines). We need this initial source function to ensure that the finite-difference propagators remain stable. To ensure that 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 and 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 \( w \), we run Algorithm 2 for 40 iterations with a batch size of 8—that is, we use eight randomly selected sources per iteration without replacement. The total number of wave-equation solves is equivalent to touching each shot only once—that is, we make one pass through the data. To improve the convergence of the inversion, we employ preconditioners in both data and model domains (see Herrmann et al., 2009, for details). 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 \( \lambda \) 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.

The estimated source functions \( \hat{q} = w_{\text{final}} \ast q_0 \) and their amplitude spectra after applying an \( \ell_2 \)-norm normalization...
are included in Fig. 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 non-linear blind deconvolution, we do not have an explanation for this behaviour. While the noise dependence of the estimated source functions behaves somewhat aberrant, the recovered reflectivity behaves as expected (cf. Fig. 3a, b for images obtained with the true source and with the initial guess, and images in Figs. 4a–c 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 scheme is able to recover a high-resolution image (see Fig. 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 Fig. 3b). Energy is no longer focussed and the shape and locations of the imaged reflectors are off. However, the results included in Fig. 4 demonstrate that good results can be obtained when estimating the source function on the fly. The estimated reflectivity depicted in Figure 4(a) is close to the reflectivity obtained when we image with the true source function (cf. Figs. 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. Fig. 4a–c). 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 artefacts (Fig. 4c). These artefacts are due to the wrongly located curvelet coefficients that passed the threshold from the noise, and finally yield the large or small-scale curved artefacts (Herrmann and Hennenfent, 2008), which are as expected. Overall, these results confirm the robustness of our imaging algorithm in the situation where there is significant noise in the data.

To arrive at the estimated images in Fig. 4, we set the penalty parameters \( \nu = 1 \) and \( \alpha = 8 \) in Algorithm 2. After the

---

**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.
first source estimation in the second iteration, we reset the coefficients \( z \) and \( x \) 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, in Fig. 5, convergence plots for the relative error for the data residual (the relative \( \ell_2 \)-norm error between the observed data and the demigrated data for estimated reflectivity \( \hat{\mathbf{m}}_k \) convolved with the estimated filter \( \hat{\mathbf{w}}_k \)), and the relative model error (the \( \ell_2 \)-norm error between the true reflectivity and the recovered reflectivity, \( \| \mathbf{m}_k - \| \mathbf{m}_k - \hat{\mathbf{m}}_k \|_2 \|_2 \) 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 in Fig. 5a) is even better than the relative error in case the source function is known (solid line in Fig. 5a). Comparison between observed common-shot gathers and corresponding gathers generated by demigrating the inverted image with the true source and source estimation also confirms the validity of our approach (Fig. 6). While encouraging, these results are obtained for a relatively simple imaging experiment and for data that is obtained with linearized modelling 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 non-linear data.

Experiments on the Sigsbee model

Sparsity-promoting imaging algorithms such as SPLS-RTM (Algorithm 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 estimation in this scenario, we consider the challenging Sigsbee2A model of size 24.4 km \( \times \) 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 handle this challenging situation, we simulate full-waveform data for a marine acquisition without a free surface. Prior to imaging, we remove direct and turning waves by subtracting data modelled in the background velocity model. 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 15 Hz (see Fig. 7) and we record for 10 seconds.
Figure 6 Common-shot gathers from (a) the observed data, (b) synthetic data generated via linearized modelling of the inverted image with the true source, (c) the same but with the estimated source, (d) difference between (a) and (b), and (e) difference between (a) and (c).

As is customary during imaging under salt, we use a background velocity model that features salt with relatively strong and therefore reflecting boundaries. The true model perturbation we aim to invert is the difference between this background model and the true model. Due to the presence of salt in the background model, the incident wavefield contains reflections that give rise to unwanted tomographic low-frequency artefacts in the image. This problem is widely reported in the literature (e.g. Yoon and Marfurt, 2006; Guittion et al., 2007). To remove these imaging artefacts, we replace the conventional imaging condition for RTM by the inverse-scattering imaging condition (Op’t Root et al., 2012; Whitmore and Crawley, 2012; Witte et al., 2017). While this condition has been proven capable of removing tomographic artefacts during RTM (Whitmore and Crawley, 2012; Witte et al., 2017), least-squares RTM (Fang et al., 2017; Yang et al., 2018) and SPLS-RTM (Witte et al., 2017), it changes the linearized forward operator (the Jacobian $J$), 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.

Contrary to the situations where the conventional imaging condition is used, where the data residual only contains...
contributions from the wrong wavelet, linearized modelling with the inverse scattering condition yields data due to impedance perturbations and this 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 Figs. 7–9. 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 Fig. 7 again confirm the validity of our approach, yielding a reasonably accurate estimate for the source after only five iterations and subsequent normalization of the $\ell_2$-norm. Imaging results obtained after 20 iterations with 10% of the shots, which amounts to 2 data passes in total, are included in Figs. 8 and 9. Unlike a typical RTM image, images obtained by SPLS-RTM are well resolved and contain true amplitude. This is because we invert the linearized modelling 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 Fig. 8(a), obtained with Algorithm 1 with the true source function, and Fig. 8(a), which we compute with our hybrid method switching from Algorithm 2 to Algorithm 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 Fig. 9. Similar to the Marmousi experiments, we set the penalty parameters $\alpha = 8$ and $\nu = 1$, and the thresholding parameter $\lambda$ is set according to 10% of the maximum absolute amplitude level of $z_1$.

While the results of this challenging imaging problem shown in Fig. 8 are of high quality, some relatively minor low-wavenumber artefacts remain at the bottom of the least-squares RTM (LS-RTM) images in Fig. 8. These artefacts are due to poorer illumination of the deeper parts of the model. We expect these artefacts to disappear when increasing the number of iterations. More importantly, the image quality improves when estimating the source function on the fly (juxtapose Fig. 8a,b). This can be explained because the unknown source function acts as a model extension, which allows for an improved data fit at the early iterations.

**DISCUSSION**

Due to its computational costs, sparsity-promoting least-squares reverse-time migration (SPLS-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 artefacts. 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 stressed the importance of estimating the source
function, we believe that the sensitivity of least-squares RTM (LS-RTM) to errors in the background velocity model needs to be studied more carefully albeit early work on time-harmonic LS-RTM shows robustness with respect to these errors (Tu and Herrmann, 2015a). Moreover, recent work (Wang et al., 2017) has shown that additional robustness can be achieved by using an extended imaging condition that allows for time shifts. Finally, combining our approach with the method of on-the-fly Fourier transforms that avoids checkpointing 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 a serverless implementation of RTM (Witte et al., 2019a, 2019d) 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.

Figure 8 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.
SPLS-RTM with source estimation

Figure 9 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.

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

CONCLUSIONS

We proposed a scalable time-domain approach to sparsity-promoting least-squares reverse-time migration (SPLS-RTM) 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 artefacts, 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 artefact-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 RTMs involving all shots.

ACKNOWLEDGEMENTS

This research was funded by the Georgia Research Alliance and the Georgia Institute of Technology. And this work is a collaborative effort of all the co-authors. We confirm that there is no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

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

ORCID

Mengmeng Yang https://orcid.org/0000-0002-0034-8037

REFERENCES

Aoki, N. and Schuster, G.T. (2009) Fast least-squares migration with a deblurring filter. Geophysics, 74(6), WCA83–WCA93.
Aravkin, A.Y., van Leeuwen, T., Calandra, H. and Herrmann, F.J. (2012) Source estimation for frequency-domain FWI with robust penalties. In: 74th EAGE Conference and Exhibition Incorporating EUROPEC 2012. Houten, The Netherlands: European Association of Geoscientists & Engineers.
Aravkin, A.Y., van Leeuwen, T. and Tu, N. (2013) Sparse seismic imaging using variable projection. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE. pp. 2065–2069.
Baysal, E., Kosloff, D.D. and Sherwood, J.W. (1983) Reverse time migration. Geophysics, 48(11), 1514–1524.
Bednar, J.B., Bednar, C.J. and Shin, C. (2006) Two-way vs one-way: a case study style comparison. In: SEG Technical Program Expanded
Haber, E., Chung, M. and Herrmann, F. (2012) An effective method
for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM Journal on Optimization*, 22(3), 739–757.

Fang, Z., Wang, R. and Herrmann, F.J. (2018) Source estimation for wavefield-reconstruction inversion. *Geophysics*, 83(4), R345–R359.

Guitton, A., Kaelin, B. and Biondi, B. (2007) Least-squares attenuation of reverse-time-migration artifacts. *Geophysics*, 72(1), S19–S23.

Haber, E., Chung, M. and Herrmann, F. (2012) An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM Journal on Optimization*, 22(3), 739–757.

Herrmann, F.J., Brown, C.R., Erlangga, Y.A. and Moghaddam, P.P. (2009) Curvelet-based migration preconditioning and scaling. *Geophysics*, 74(4), A41–A46.

Herrmann, F.J. and Hennenfent, G. (2008) Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173(1), 233–248.

Herrmann, F.J. and Li, X. (2012) Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical Prospecting*, 60(4), 696–712.

Herrmann, F.J., Tu, N. and Esser, E. (2015) Fast “online” migration with compressive sensing. In: *Proceedings of the 77th EAGE Conference and Exhibition*. Houten, The Netherlands: European Association of Geoscientists & Engineers. pp. 1–5.

Hou, J. and Symes, W.W. (2016) Accelerating extended least-squares migration with weighted conjugate gradient iteration. *Geophysics*, 81(4), S165–S179.

Huang, Y., Nammour, R. and Symes, W. (2016) Flexibly preconditioned extended least-squares migration in shot-record domain. *Geophysics*, 81(5), S299–S315.

Lambaré, G., Virieux, J., Madariaga, R. and Jin, S. (1992) Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, 57(9), 1138–1154.

Lorenz, D.A., Schofper, F. and Wenger, S. (2014) The linearized Bregman method via split feasibility problems: analysis and generalizations. *SIAM Journal on Imaging Sciences*, 7(2), 1237–1262.

Louboutin, M., Lange, M., Luporini, F., Kukreja, N., Wine, P.A., Herrmann, F.J., et al. (2019) Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, 12(3), 1165–1187.

Luporini, F., Lange, M., Louboutin, M., Kukreja, N., Hückelheim, J., Yount, C. et al. (2018) Architecture and performance of devito, a system for automated stencil computation. *CoRR*, abs/1807.03032.

Mulder, W.A. and Plessix, R.-E. (2004) A comparison between one-way and two-way wave-equation migration. *Geophysics*, 69(6), 1491–1504.

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

Op’t Root, T., Stolk, C. and de Hoop, M. (2012) Linearized inverse scattering based on seismic reverse time migration. *Journal de Mathématiques Pures et Appliquées*, 98, 211–238.

Powell, W.B. (2014) Clearing the jungle of stochastic optimization. In: Newman, A. and Leung, J. (Eds.) *Bridging Data and Decisions*. Athens: Inform. pp. 109–137.

Pratt, R.G. (1999) Seismic waveform inversion in the frequency domain, Part I: theory and verification in a physical scale model. *Geophysics*, 64(3), 888–901.

Qu, Y., Huang, J., Li, Z., Guan, Z. and Li, J. (2017) Attenuation compensation in anisotropic least-squares reverse time migration. *Geophysics*, 82(6), S411–S423.

Rickett, J. (2013) The variable projection method for waveform inversion with an unknown source function. *Geophysical Prospecting*, 61(4), 874–881.

Schuster, G. (1983) Least-squares cross-well migration. In: *SEG Technical Program Expanded Abstracts 1993*. Tulsa. OK: Society of Exploration Geophysicists. pp. 110–113.
Sun, J., Fomel, S., Zhu, T. and Hu, J. (2016) Q-compensated least-squares reverse time migration using low-rank one-step wave extrapolation. Geophysics, 81(4), S271–S279.

Symes, W.W. (2008) Migration velocity analysis and waveform inversion. Geophysical Prospecting, 56(6), 765–790.

Tu, N., Aravkin, A.Y., van Leeuwen, T. and Herrmann, F.J. (2013) Fast least-squares migration with multiples and source estimation. In: Proceedings of the 75th EAGE Conference and Exhibition. Houten, The Netherlands: European Association of Geoscientists & Engineers.

Tu, N. and Herrmann, F.J. (2015a). Fast imaging with surface-related multiples by sparse inversion. Geophysical Journal International, 2011(1), 304–317.

Valenciano, A.A. (2008) Imaging by wave-equation inversion. PhD Thesis, Stanford University, Stanford, CA.

van Leeuwen, T., Aravkin, A.Y. and Herrmann, F.J. (2011) Seismic waveform inversion by stochastic optimization. International Journal of Geophysics, 2011. Available at: https://doi.org/10.1155/2011/689041.

Wang, M. and Xu, S. (2017) Least-squares reverse time migration with inaccurate velocity model. In: 2017 SEG International Exposition and Annual Meeting. Tulsa. OK: Society of Exploration Geophysicists.

Welling, M. and Teh, Y.W. (2011) Bayesian learning via stochastic gradient Langevin dynamics. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA. pp. 681–688.

Whitmore, N. (1983) Iterative depth migration by backward time propagation. In: SEG Technical Program Expanded Abstracts 1983. Tulsa. OK: Society of Exploration Geophysicists. pp. 382–385.

Whitmore, N. and Crawley, S. (2012) Applications of RTM inverse scattering imaging conditions. In: SEG Technical Program Expanded Abstracts 2012. Tulsa. OK: Society of Exploration Geophysicists. pp. 1–6.

Witte, P.A., Louboutin, M., Jones, C. and Herrmann, F.J. (2019a) Serverless seismic imaging in the cloud. In: Rice Oil and Gas High Performance Computing Conference 2020, Houston, TX.

Witte, P.A., Louboutin, M., Kukreja, N., Luporini, F., Lange, M., Gorman, G.J. et al. (2019b) A large-scale framework for symbolic implementations of seismic inversion algorithms in Julia. Geophysics, 84(3), F57–F71.

Witte, P.A., Louboutin, M., Luporini, F., Gorman, G.J. and Herrmann, F.J. (2019c) Compressive least-squares migration with on-the-fly fourier transforms. Geophysics, 84(5), 1–76.

Witte, P.A., Yang, M. and Herrmann, F.J. (2017) Sparsity-promoting least-squares migration with the linearized inverse scattering imaging condition. In: Proceedings of the 79th EAGE Conference and Exhibition. Houten, The Netherlands: European Association of Geoscientists & Engineers.

Yang, J., Hua, B., Williamson, P., Zhu, H., McMechan, G. and Huang, J. (2020) Elastic least-squares imaging in tilted transversely isotropic media for multicomponent land and pressure marine data. Surveys in Geophysics, 41: 805–833.

Yang, J. and Zhu, H. (2018) Least-squares reverse time migration using the impedance-sensitivity kernel. In 2018 SEG International Exposition and Annual Meeting. Tulsa. OK: Society of Exploration Geophysicists.

Yoon, K. and Marfurt, K.J. (2006) Reverse-time migration using the Poynting vector. Exploration Geophysics, 37(1), 102–107.

Zeng, C., Dong, S. and Wang, B. (2014) Least-squares reverse time migration: Inversion-based imaging toward true reflectivity. The Leading Edge, 33(9), 962–968.

Zhang, Y., Zhang, H. and Zhang, G. (2011) A stable TTI reverse time migration and its implementation. Geophysics, 76(3), WA3–WA11.