Compressive least-squares migration with source estimation

Philipp A. Witte*, Mengmeng Yang* and Felix J. Herrmann

Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

Email: pwitte@eoas.ubc.ca; myang@eoas.ubc.ca; fherrmann@eoas.ubc.ca

* contributed equally

(July 25, 2017)

Running head: **Compressive LSM with source estimation**

ABSTRACT

Least-squares reverse-time migration is a powerful approach for true-amplitude seismic imaging of complex geological structures. The successful application of this method is hindered by its exceedingly large computational cost and required prior knowledge of the generally unknown source wavelet. We address these problems by introducing an algorithm for low-cost sparsity-promoting least-squares migration with source estimation. We adapt a recent algorithm from sparse optimization, which allows to work with randomized subsets of shots during each iteration of least-squares migration, while still converging to an artifact-free solution. We modify the algorithm to incorporate on-the-fly source estimation through variable projection, which lets us estimate the wavelet without additional PDE solves. The resulting algorithm is easy to implement and allows imaging at a fraction of the cost of conventional least-squares reverse-time migration, requiring only around two passes trough the data, making the method feasible for large-scale imaging problems with unknown source

wavelets.

INTRODUCTION

Reverse-time migration (RTM) is an increasingly popular wave-equation based seismic imaging algorithm that corresponds to applying the adjoint of the Born scattering operator to observed reflection data (Baysal et al., 1983; Whitmore, 1983). Without extensive preconditioning, applying the adjoint operator leads to an image with incorrect amplitudes, imprints of the source wavelet and therefore blurred reflectors. To overcome these issues, imaging can be formulated as a linear least-squares optimization problem, in which the mismatch between observed and modeled data is minimized (Schuster, 1993; Nemeth et al., 1999; Dong et al., 2012; Zeng et al., 2014). Conventional LS-RTM is computationally very expensive, because it requires migration and demigration of all shot records during each iteration. To save computational resources, shots can be subsampled or combined into simultaneous shots, which avoids having to treat every shot separately in each iteration (e.g. Romero et al., 2000; Tang and Biondi, 2009; van Leeuwen et al., 2011; Dai et al., 2011, 2012; Liu, 2013; Herrmann and Li, 2012; Lu et al., 2015).

An RTM image or LS-RTM gradient is calculated by forward modeling the source wavelet in the background model, backpropagating the observed data or data residual and applying an imaging condition. The source wavelet is generally unknown and therefore needs to be estimated prior to imaging or as part of the inversion process. Source estimation has been proposed in the context of frequency domain full-waveform inversion (FWI) (Pratt, 1999; Aravkin et al., 2012) and frequency domain LS-RTM (Tu et al., 2013), where a single complex number per frequency needs to be estimated. In the time domain, we are only aware of the work of Zhang et al. (2016), who formulate LS-RTM independently of the source wavelet, i.e. they eliminate the wavelet from the objective and minimize the misfit between

the predicted and true Green's functions. To our knowledge, there are no algorithms so far that combine time-domain LS-RTM and estimation of the full time-dependent source function.

We introduce an algorithm that addresses the two main challenges of time-domain LS-RTM: estimating the source wavelet and decreasing the overall computational cost. We extend our earlier work on sparsity-promoting imaging (SPLS-RTM) and FWI in the frequency domain (Li et al., 2012; Herrmann et al., 2015; Tu and Herrmann, 2015) and adapt an algorithm that has recently gained popularity in the context of compressive sensing. The linearized Bregman method (Yin, 2010) is an algorithm for recovering the sparse solution of an underdetermined linear system, which is closely related to the iterative soft-thresholding algorithm (ISTA). However, it has been mathematically shown that the algorithm can also be used to solve overdetermined linear systems and allows working with subsets of rows and data points in each iteration (Lorenz et al., 2014b). This means that during each LS-RTM iteration, we model and migrate only a small subset of randomly selected shots instead of the full data set and still converge to the true solution. The algorithm is simple to implement and in contrast to ISTA requires only a single and relatively easy to choose hyperparameter. Furthermore we incorporate source estimation via variable projection into this algorithm, which lets us estimate the source wavelet without additional PDE solves.

THE LINEARIZED BREGMAN METHOD FOR SPLS-RTM

In a series of previous publications (e.g. Herrmann and Li (2012), Tu and Herrmann (2015)), we have reformulated LS-RTM as a sparsity-promoting least-squares minimization problem

of the following form:

$$\underset{\delta\mathbf{m}}{\text{minimize}} \quad \|\mathbf{C}\,\delta\mathbf{m}\|_1$$

$$\text{subject to } \|\,\nabla\mathbf{F}(\mathbf{m}_0,\mathbf{q})\delta\mathbf{m} - \delta\mathbf{d}\|_2 \leq \sigma.$$

(1)

These optimization problems are called basis pursuit denoise (BPDN) problems and have recently gained large popularity in the context of compressive sensing (Donoho, 2006). The objective of this problem is to minimize the ℓ_1-norm of the curvelet coefficients of the seismic image $\delta\mathbf{m}$, obtained through the forward Curvelet transform $\mathbf{C}$. This is subject to the constraint that the synthetic data, given by the action of the Born modeling operator $\nabla\mathbf{F}$ on $\delta\mathbf{m}$, fits the observed reflections $\delta\mathbf{d}$ within some noise level σ. The Born modeling operator is a function of the background velocity model $\mathbf{m}_0$ and the source wavelet $\mathbf{q}$, which are for now assumed to be known. As shown in Li et al. (2012), enforcing sparsity on the seismic image leads to high quality results when working with randomization techniques, where the computational cost of LS-RTM can be significantly reduced by working with randomized subsets of sources during each iteration.

The main drawback of this approach is that optimization algorithms that solve the BPDN problem, such as the spectral projected gradient method with ℓ_1 (SPGℓ_1) (van den Berg and Friedlander, 2008) are difficult to implement and do not necessarily allow us to work with randomization techniques. Other solvers such as ISTA, work with modifications of the BPDN problem and solve an unconstrained formulation in which the constraint is added to the objective as a penalty. This introduces a penalty parameter that is difficult to choose and requires complex cooling techniques (i.e. the penalty parameter needs to be relaxed) in order to converge to the solution sufficiently quickly (Hennenfent et al., 2008).

An alternative algorithm, which has recently gained popularity in solving sparsity-promoting least-squares problems is the linearized Bregman method (Yin, 2010). The

algorithm solves a slightly modified version of the BPDN problem, in which an ℓ_2 penalty term is added to the objective::

$$\underset{\delta \mathbf{m}}{\text{minimize}} \quad \lambda \|\mathbf{C}\,\delta\mathbf{m}\|_1 + \frac{1}{2}\|\mathbf{C}\,\delta\mathbf{m}\|_2^2$$

$$\text{subject to } \|\,\nabla\mathbf{F}(\mathbf{m}_0,\mathbf{q})\delta\mathbf{m} - \delta\mathbf{d}\|_2 \leq \sigma. \tag{2}$$

The combination of ℓ_1- and ℓ_2-norm terms in the objective is referred to as an elastic net in machine learning and has the effect of making the objective function strongly convex (Lorenz et al., 2014a). This small change to the objective yields a relatively simple iterative scheme for minimizing the above optimization problem, in which the penalty parameter λ plays a completely different role than in ISTA. In ISTA, λ is a penalty parameter that controls the trade-off between the ℓ_1-norm of the (transform-domain) image and the data fidelity term and to solve the BPDN problem, we need $\lambda \to 0^+$ (i.e. λ goes to 0 from above). However, as discussed in Hennenfent et al. (2008), small values of λ lead to very slow convergence of ISTA and complex cooling techniques are required to improve its performance. A value of λ that is too large on the other hand, prevents any entries in the unknown to enter into the solution. In the linearized Bregman method on the other hand, λ controls the trade-off between the ℓ_1- and ℓ_2-norms of the unknown model vector. A value of λ that is too large only initially prevents new entries to enter into the solution, but as the iterations progress, larger and larger values start entering the solution, making the algorithm in some sense "self-tuning" (Herrmann et al., 2015). Choosing an appropriate value for λ in the linearized Bregman method is therefore easier and can be done automatically. For example, setting λ to the infinity norm of the gradient in the first iteration, leads to no value passing the first threshold, but the solution starts to build up from iteration two onwards.

Another important benefit of the linearized Bregman method is the possibility to work with subsets of blocks of rows of $\nabla\mathbf{F}$ (each corresponding to a source experiment) and $\delta\mathbf{d}$

during each iteration, while still converging to the correct solution (Lorenz et al., 2014b). For least-squares migration, this means that instead of treating all shot records during each iteration, we can use a randomized subset of fewer source positions, thus decreasing the number of PDE solves. This provides flexibility to the user to find a trade-off between the number of sources per iteration and the overall number of iterations. In the extreme case, it is possible to use only a single shot in each iteration, which then requires many iterations to reach an acceptable solution. We found that it is often sufficient to use between $\sim \frac{1}{10}$ and $\frac{1}{20}$ of the shots during each iteration and perform around ~ 20 overall iterations, i.e. on average every shot record is demigrated/migrated only once or twice, making our algorithm in terms of PDE solves several times cheaper than conventional LS-RTM.

ON-THE-FLY SOURCE ESTIMATION

The algorithm in the previous section assumes that the source wavelet $\mathbf{q}$ is known, which is generally not the case. Often some general information about the source function is known, i.e. the approximate frequency range, the general wavelet shape and the approximate length in time. We therefore base our source estimation algorithm on the assumption that an initial guess of the true wavelet $\mathbf{q}_0$ is available and that there exists a filter $\mathbf{w}$, such that convolving the initial guess with this filter yields the correct source wavelet: $\mathbf{q} = \mathbf{q}_0 * \mathbf{w}$, with the symbol $*$ denoting circular convolution over time. SPLS-RTM with source estimation can then be formulated as a multi-parameter optimization problem with the image $\delta \mathbf{m}$ and the filter $\mathbf{w}$ as unknowns. Optimization problems of this form require gradients of the objective with respect to both parameters, which not only requires a completely new optimization algorithm, but can also pollute the image during the early iterations, as the initial wavelet might be far off from the true one. We therefore follow the approach in Aravkin et al. (2012)

and Tu et al. (2013) and formulate the objective as a separable least-squares problem, which can be solved using the variable projection method (Golub and Pereyra, 2003). For this, we only replace the Jacobian in equation 2 by $\mathbf{w}(\delta\mathbf{m}) * [\nabla\mathbf{F}(\mathbf{m}_0, \mathbf{q}_0)\delta\mathbf{m}]$. This corresponds to modeling the data with the initial source wavelet $\mathbf{q}_0$, followed by a trace-by-trace convolution with the filter $\mathbf{w}(\delta\mathbf{m})$. The filter is a function of the current estimate of the image and is obtained by solving a small independent least-squares subproblem during each iteration:

$$\mathbf{w}_k = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{w} * \left[\nabla\mathbf{F}_k\big(\mathbf{m}_0, \mathbf{q}_0\big)\delta\mathbf{m}_k\right] - \delta\mathbf{d}_k\|^2. \tag{3}$$

As mentioned earlier, we want to work with randomized subsets of shots during each iteration, so $\nabla\mathbf{F}_k$ and $\delta\mathbf{d}_k$ denote the subset of the Jacobian and the data at the current iteration k and $\delta\mathbf{m}_k$ is the current estimate of the image. We want the estimated wavelet to be localized in time, since we know that the time signature of an impulsive source is typically short. Therefore the vector $\mathbf{w}$ for the filter is chosen to have the approximate length of the true wavelet.

The algorithm for time-domain SPLS-RTM with source estimation is outlined below. Each iteration involves choosing a randomized subset of shots and $\nabla\mathbf{F}_k$ and $\delta\mathbf{d}_k$ denote the current subset of the Jacobian and the data (lines $3-5$). Updating the gradient (line 8) requires a step size, which is chosen as $t_k = \|\mathbf{w}_k * \mathbf{d}_k - \widehat{\delta\mathbf{d}_k}\|/\|\nabla\mathbf{F}^\top\big(\mathbf{w}_k * \mathbf{d}_k - \widehat{\delta\mathbf{d}_k}\big)\|$ (Lorenz et al., 2014b). The operator $\mathcal{P}_\sigma$ in line 8 is a projection operator that projects the data residual onto the ℓ_2-ball of the size of the noise, i.e. it does not fit the data exactly. Calculating the gradient also involves the transpose action of the convolution operator (line 8), which is the correlation of $\mathbf{w}_k$ with the data (denoted by $\star$). The function S_λ (line 9) is the soft thresholding function.

Algorithm 1 Linearized Bregman for sparsity-promoting LS-RTM with source estimation.

1. Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{z}_0 = \mathbf{0}$, $\mathbf{q}_0$, λ, batchsize $n'_s \ll n_s$

2. **for** $k = 1, ..., n$

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

4. $\widehat{\nabla \mathbf{F}}_k = \{\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}_0)\}_{i \in \mathcal{I}}$

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

6. $\mathbf{d}_k = \widehat{\nabla \mathbf{F}}_k \mathbf{x}_k$

7. $\mathbf{w}_k = \text{argmin}_{\mathbf{w}} \|\mathbf{w} * \mathbf{d}_k - \widehat{\delta \mathbf{d}}_k\|^2$

8. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \widehat{\nabla \mathbf{F}}_k^\top \left[\mathbf{w}_k \star \mathcal{P}_\sigma (\mathbf{w}_k * \mathbf{d}_k - \widehat{\delta \mathbf{d}}_k) \right]$

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

10. **end**

$S_\lambda(\mathbf{z}) = \text{sign}(\mathbf{z}) \cdot \max(0, |\mathbf{z}| - \lambda)$

$\mathcal{P}_\sigma(\nabla \mathbf{F} \mathbf{x} - \delta \mathbf{d}) = \max\left(0, 1 - \frac{\sigma}{\|\nabla \mathbf{F} \mathbf{x} - \delta \mathbf{d}\|}\right) \cdot (\nabla \mathbf{F} \mathbf{x} - \delta \mathbf{d})$

NUMERICAL EXAMPLE

We test our algorithm for SPLS-RTM with source estimation on the Sigsbee 2A velocity model (Bashkardin et al., 2002) and show that our proposed algorithm is capable of accurately recovering the source wavelet and provides a high quality image with good illumination below the salt, while only requiring two passes through the data. We generate linearized observed data for 960 source locations with 25 m shot spacing and 10 seconds recording time. Each shot is recorded by 960 receivers, located at 25 m depth with a maximum offset of 10 km. We use a minimum phase source wavelet to generate the synthetic observed data with 15 Hz peak frequency (shown in figure 1).

For SPLS-RTM with source estimation, we have to define an initial wavelet from which

the true wavelet can be estimated reliably with our algorithm. We found this is the case for an initial wavelet with a wide frequency spectrum, meaning the initial wavelet contains a broad range of frequencies, while the shape of the spectrum, as well as the phase can be well recovered with our algorithm. We design the spectrum of the initial wavelet to have a plateau between 15-25 Hz, which is wider than the approximate peak frequency of the true wavelet. To show that we can recover the true source from an initial wavelet with a different phase and shape, we design the initial wavelet to be mixed-phase with various side lobes, while the true wavelet is minimum phase (figure 1).

For SPLS-RTM with and without source estimation we run 18 iterations of the linearized Bregman method with 100 sources per iteration, which corresponds to roughly two passes through the data. Our velocity model is a smoothed version of the true model, meaning it is kinematically correct. We use a few basic preconditioners to improve the convergence of our algorithm. In the model domain, we use a linear depth scaling preconditioner, as well as a topmute to remove the source/receiver imprints in the water column. In the data domain, we employ a data topmute to suppress the water bottom reflection, as well as a half integration to compensate for the order of the 2D normal operator (Herrmann et al., 2008). One challenge of the Sigsbee 2A model is the strong and shallow salt body, which causes significant backscattering of the forward modeled wavefield and leads to low-frequency artifacts in the gradients. We address this problem by using the linearized inverse scattering imaging condition (Op't Root et al., 2012; Whitmore and Crawley, 2012) instead of the crosscorrelation imaging condition. This modification is incorporated into both the imaging operator $\nabla \mathbf{F}^{\top}$ and the modeling operator $\nabla \mathbf{F}$, in order to preserve their adjointness (Witte et al., 2017). The only hyperparameter we need to choose is λ, which we set according to the maximum absolute value of the gradient in the first iteration, as mentioned in the theory
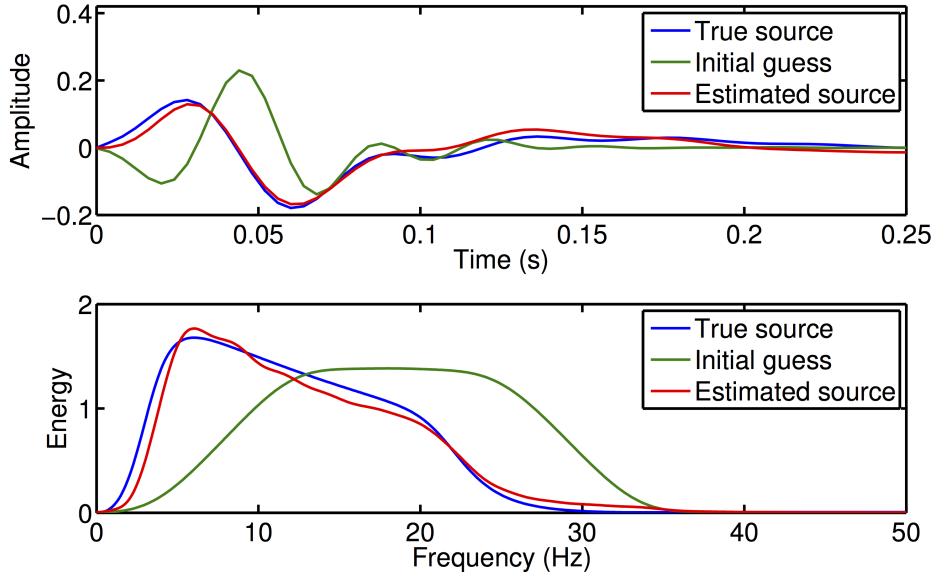
Figure 1: The top figure shows the true source wavelet, as well as the initial guess and the estimated wavelet. The bottom figure contains the frequency spectra of the respective wavelets.

section. We set λ to the $\frac{1}{10}$ of the maximum absolute value, which leads to values already entering the solution during the first iteration. We found that setting λ to be larger than this, requires too many iterations to build up an image that includes also the small amplitude reflectors.

DISCUSSION

Figures 2 and 3 show the conventional RTM image in comparison to the result after 18 iterations. The RTM image was obtained using the true source wavelet, whereas for the SP-LSRTM result, the source wavelet was estimated during inversion. We apply a vertical derivative to both final images to remove some residual low frequency artifacts. However, no derivatives or Laplacian filtering was used within the imaging algorithm itself. Whereas
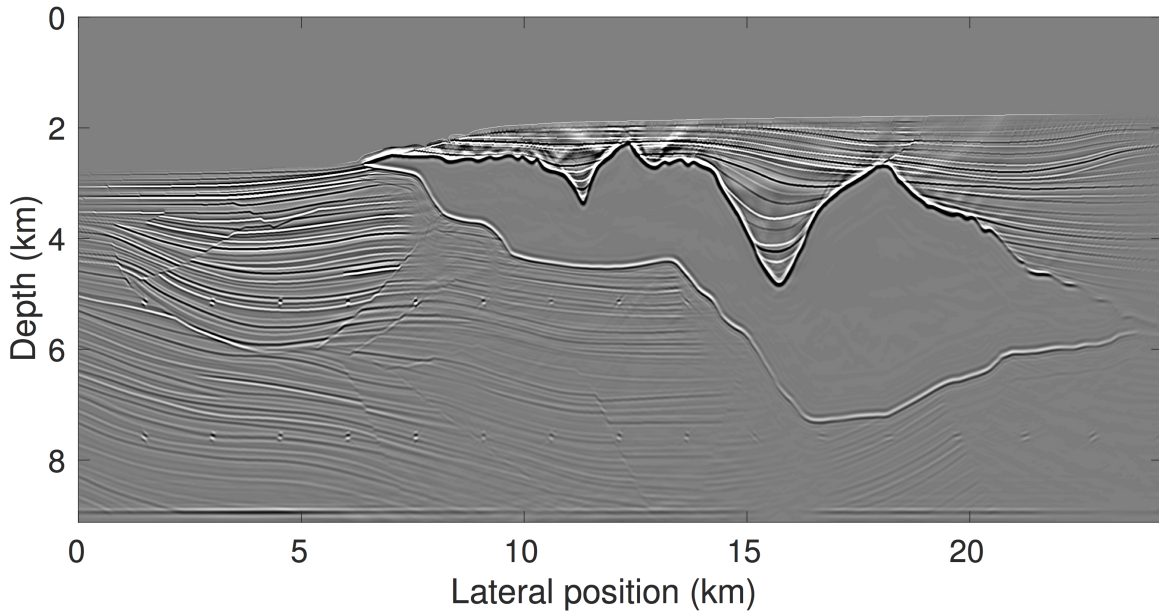
11

Figure 2: RTM image using the true source wavelet.

the RTM image suffers from typical migration artifacts like decaying amplitudes, blurred reflectors and poor illumination below the salt, the SPLS-RTM result has well balanced amplitudes and improved spatial resolution. The SPLS-RTM result was obtained with only two passes through the data, requiring only four times as many overall PDE solves as the RTM image.

Figure 1 shows the estimated wavelet, which is obtained by convolving the estimated filter with the initial guess of the wavelet. Both the time signature as well as the frequency spectrum are fairly close to the true wavelet. Because the spectrum of our initial wavelet is missing the lower frequencies of the true wavelet, a perfect recovery cannot be expected. The image and the source wavelet can both only be estimated up to some scaling factor, because it is possible to scale up the image and scale down the source without affecting the ℓ_2 norm of the data residual. Therefore the source functions in figure 1 are all normalized by their respective energies.
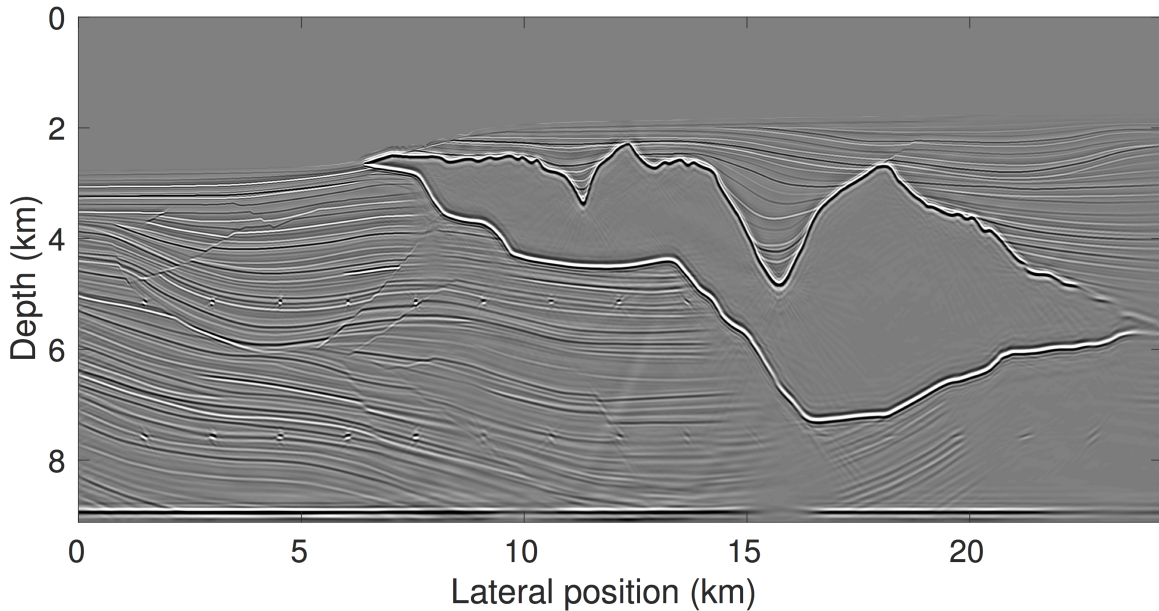
Figure 3: Sparsity-promoting LS-RTM with on-the-fly source estimation after 18 iterations.

CONCLUSIONS

Through leveraging recent advances in stochastic optimization and sparse inverse problems, we have introduced an algorithm for sparsity-promoting reverse-time migration with significantly reduced computational cost in comparison to conventional least-squares reverse-time migration. Our numerical example shows that our algorithm yields a high quality true amplitude image, even though we only use a small subset of shots during each iteration and limit the overall passes through the data to two. In contrast to earlier work, the algorithm relies on a solver with proven convergence, even when working with randomized subsets of data during each iteration. We also include source estimation into the imaging algorithm, which makes our approach feasible for 3D field data applications, where the source wavelet is generally unknown.

ACKNOWLEDGEMENTS

REFERENCES

Aravkin, A. Y., T. van Leeuwen, and N. Tu, 2012, Sparse seismic imaing using variable projection.

Bashkardin, V., S. Fomel, P. Karimi, A. Klokov, and X. Song, 2002, Sigsbee model: http://www.ahay.org/RSF/book/gallery/sigsbee/paper_html/paper.html. (accessed March 15 2017).

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

Dai, W., P. Fowler, and G. Schuster, 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **70**, no. 4.

Dai, W., W. Xin, and G. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: GEOPHYSICS, **76**.

Dong, S., J. Cai, M. Guo, S. Suh, Z. Zhang, B. Wang, and Z. Li, 2012, Least-squares reverse time migration: towards true amplitude imaging and improving the resolution: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5.

Donoho, D., 2006, Compressed sensing : Institute of Electrical and Electronics Engineers Transactions on Information Theory, **52**.

Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse problems, **19**, R1–R26.

Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Herrmann, 2008, New insights into one-norm solvers from the Pareto curve: GEOPHYSICS, **73**, A23–A26.

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., P. P. Moghaddam, and C. Stolk, 2008, Sparsity- and continuity- promoting

seismic image recovery with curvelet frames: Applied and Computational Harmonic Analysis, **24**, 150–173.

Herrmann, F. J., N. Tu, and E. Esser, 2015, Fast "online" migration with Compressive Sensing: 77th Conference and Exhibition, EAGE, Expanded Abstracts.

Li, X., A. Aravkin, T. van Leeuwen, and F. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: GEOPHYSICS, **77**, A13–A17.

Liu, Y., 2013, Multisource least-squares extended reverse time migration with preconditioning guided gradient method: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3709–3715.

Lorenz, D. A., F. Schöpfer, and S. Wenger, 2014a, The Linearized Bregman Method via Split Feasibility Problems: Analysis and Generalizations: Society for Industrial and Applied Mathematics Journal on Imaging Sciences, **7**, no. 2.

Lorenz, D. A., S. Wenger, F. Schöpfer, and M. Magnor, 2014b, A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing: Institute of Electrical and Electronics Engineers International Conference on Image Processing.

Lu, X., L. Han, J. Yu, and X. Chen, 2015, L1 norm constrained migration of blended data with the FISTA algorithm: Journal of Geophysics and Engineering, **12**, 620–628.

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

Op't Root, T. J., C. C. Stolk, and M. V. de Hoop, 2012, Linearized inverse scattering based on seismic reverse time migration: Journal de Mathematiques Pures et Appliquees, **98**, 211–238.

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

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: GEOPHYSICS, **65**, 426–436.

Schuster, G. T., 1993, Least-squares cross-well migration: 63th Annual International Meeting, SEG, Expanded Abstracts, 110–113.

Tang, Y., and B. Biondi, 2009, Least-squares migration/inversion of blended data: 79th Annual International Meeting, SEG, Expanded Abstracts.

Tu, N., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2013, Fast least-squares migration with multiples and source estimation: 75th Conference and Exhibition, EAGE, Expanded Abstracts.

Tu, N., and F. Herrmann, 2015, Fast imaging with surface-related multiples by sparse inversion: Geophysical Journal International, **201**, 304–417.

van den Berg, E., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.

van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic Waveform Inversion by Stochastic Optimization: International Journal of Geophysics.

Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 1983 SEG Annual Meeting, Expanded Abstracts.

Whitmore, N. D., and S. Crawley, 2012, Applications of RTM inverse scattering imaging conditions: 82nd Annual International Meeting, SEG, Expanded Abstracts.

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

Yin, W., 2010, Analysis and Generalizations of the Linearized Bregman Method: SIAM Journal on Imaging Sciences, **3**, 856–877.

Zeng, C., S. Dong, and B. Wang, 2014, Least-squares reverse time migration: Inversion-based imaging toward true reflectivity: The Leading Edge, **33**, no. 9, 962–964,966,968.

Zhang, Q., H. Zhou, H. Chen, and J. Wang, 2016, Least-squares reverse time migration with and without source wavelet estimation: Journal of Applied Geophysics, **143**, 1–10.