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

Mengmeng Yang[*], Philipp Witte, Zhilong Fang and Felix J. Herrmann
Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

SUMMARY:

Traditional reverse-time migration (RTM) gives images with wrong amplitudes and low resolution. Least-squares RTM (LS-RTM) on the other hand, is capable of obtaining true-amplitude images as solutions of ℓ_2-norm minimization problems by fitting the synthetic and observed reflection data. The shortcoming of this approach is that solutions of these ℓ_2 problems are typically smoothed, tend to be overfitted, and computationally too expensive because it requires compared to standard RTM too many iterations. By working with randomized subsets of data only, the computational costs of LS-RTM can be brought down to an acceptable level while producing artifact-free high-resolution images without overfitting the data. While initial results of these "compressive imaging" methods were encouraging various open issues remain including guaranteed convergence, algorithmic complexity of the solver, and lack of on-the-fly source estimation for LS-RTMs with wave-equation solvers based on time-stepping. By including on-the-fly source-time function estimation into the method of Linearized Bregman (LB), on which we reported before, we tackle all these issues resulting in a easy-to-implement algorithm that offers flexibility in the trade-off between the number of iterations and the number of wave-equation solves per iteration for a fixed total number of wave-equation solves. Application of our algorithm on a 2D synthetic shows that we are able to obtain high-resolution images, including accurate estimates of the wavelet, for a single pass through the data. The produced image, which is by virtue of the inversion deconvolved with respect to the wavelet, is roughly of the same quality as the image obtained given the correct source function.

INTRODUCTION

Reverse-time migration (RTM) approximates the inverse of the linearized Born modelling operator by its adjoint. As a consequence, RTM without extensive preconditioning—generally in the form of image domain scalings and source deconvolution prior to imaging—cannot produce high-resolution true-amplitude images. By fitting the observed reflection data in the least-squares sense, least-squares migration, and in particular least-squares reverse-time migration (LS-RTM), overcomes these issues except for two main drawbacks withstanding their succesful application. Firstly, the computational cost of conducting demigrations and migrations iteratively is very large. Secondly, minimizing the ℓ_2 norm of the data residual can lead to model perturbations with artifacts caused by overfitting. These are due to the fact that LS-RTM involves the inversion of a highly overdetermined system of equations where it is easy to overfit noise in the data.

One approach to avoid overfitting is to apply regularization to the original formulation of LS-RTM, and search for the sparsest possible solution by minimizing the ℓ_1-norm on some sparsifying representation of the image. Motivated by the theory of Compressive Sensing (Donoho, 2006), where sparse signals are recovered from severe undersamplings, and considering the huge cost of LS-RTM we turn the overdetermined imaging problem into a underdetermined one by working with small subsets of sources at each iteration of LS-RTM, reducing the computational costs.

Herrmann et al. (2008) found that, as a directional frame expansion, curvelets lead to sparsity of seismic images in imaging problems. This property led to the initial formulation of curvelet-based "Compressive Imaging" (Herrmann and Li, 2012), which formed the bases of later work by Tu et al. (2013) who included surface-related multiples and on-the-fly source estimation via variable projection. While this approach represents major progress the proposed method, which involves drawing new independent subsets of shots after solving each one-norm constrained subproblem, fails to continue to bring down the residual. This results in remaining subsampling related artifacts. In addition, the proposed method relies on a difficult to implement one-norm solver. To overcome these challenges, Herrmann et al. (2015) motivated by Lorenz et al. (2014) relaxed the one-norm objective of Basis Pursuit Denoise (Chen et al., 2001) by an objective given by the sum of the λ-weighted ℓ_1- and ℓ_2-norms. This seemingly innocent change resulted in a greatly simplified implementation that no longer relies on relaxing the one-norm constraint and more importantly provably continues to make progress towards the solution irrespective of the number of randomly selected sources participating in each iteration.

Inspired by this work, we propose to extend our earlier work in frequency domain to the time domain, which is more appropriate for large scale 3D problems and to include on-the-fly source estimation via variable projection. Both generalizations are completely new and challenging because the estimation of the time-signature of the source wavelet no longer involves estimating a single complex number for each temporal frequency (Tu et al., 2013).

Our paper is organized as follows. First, we formulate sparsity-promoting LS-RTM as a Basis Pursuit Denoising problem formulated in the curvelet domain. Next, we relax the ℓ_1-norm and describe the relative simple iterative algorithm that solves this optimization with the mixed ℓ_1- ℓ_2-norm objective. We conclude this theory section by including estimation of the source-time signature that calls for additional regularization prevent overfitting of the wavelet. We evaluate the performance of the proposed method on the synthetic SIGSBEE2A model (Paffenholz et al., 2002) where we compare our inversion results to LS-RTM with the true source function and standard RTM.

THEORY

Imaging with linearized Bregman

The original sparsity-promoting LS-RTM has the same form, albeit it is overdetermined rather than underdetermined, as Basis Pursuit Denoise (BPDN, Chen et al. (2001)), which is expressed as below

$$\min_{\mathbf{x}} \ \|\mathbf{x}\|_1$$
$$\text{subject to} \quad \sum_i \|\nabla \mathbf{F}_i(\mathbf{m}_0, q_i) \mathbf{C}^T \mathbf{x} - \delta \mathbf{d}_i\| \le \sigma, \quad (1)$$

where the vector $\mathbf{x}$ contains the sparse curvelet coefficients of the unknown model perturbation. The symbols $\|\cdot\|_1$ and $\|\cdot\|$ denote ℓ_1 and ℓ_2 norms, respectively. The matrix $\mathbf{C}^T$ is the transpose of curvelet transform $\mathbf{C}$. The vectors $\mathbf{m}_0$ stands for the background velocity model and $q_i = q_i(t)$ is the time dependent source function for the ith shot. The matrix $\nabla \mathbf{F}_i$ and vector $\delta \mathbf{d}_i$ represent the linearized Born modelling operator and observed reflection data along the receivers for the ith shot. The $\sum_i$ runs over all shots. The parameter σ is the noise level of the data.

Despite the success of BPDN in Compressive Sensing, where matrix-vector multiplies are generally cheap, solving Equation 1 in the seismic setting where the evaluation of the $\nabla \mathbf{F}_i$'s involve wave-equation solves is a major challenge because it is computationally infeasible. To overcome this computational challenge, Herrmann and Li (2012) proposed, motivated by ideas from stochastic gradients, to select subsets of shots by replacing the sum over $i = 1 \cdots n_s$, with n_s the number of shots, to a sum over randomly selected index sets $\mathscr{I} \in [1 \cdots n_s]$ of size $n_s' \ll n_s$. By redrawing these subsets at certain points of the ℓ_1-norm minimization, Li et al. (2012) and Tu et al. (2013) were able to greatly reduce the computational costs but at the expense of giving up convergence when the residual becomes small.

For underdetermined problems, Cai et al. (2009) presented a theoretical analysis that proves convergence of a slightly modified problem for iterations that involve arbitrary subsets of rows (= subsets $\mathscr{I}$). Herrmann et al. (2015) adapted this idea to the seismic problem and early results suggest that the work of Cai et al. (2009) can be extended to the overdetermined (seismic) case. With this assumption, we propose to replace the optimization problem in Equation 1 by

$$\min_{\mathbf{x}} \ \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x}\|^2$$
$$\text{subject to} \sum_i \|\nabla \mathbf{F}_i(\mathbf{m}_0, q_i) \mathbf{C}^T \mathbf{x} - \delta \mathbf{d}_i\| \le \sigma. \quad (2)$$

Mixed objectives of the above type are known as elastic nets in machine learning and for $\lambda \to \infty$—which in practice means λ large enough—solutions of these modified problems converge to the solution of Equation 1. While adding the ℓ_2-term may seem an innocent change, it completely changes the iterations of sparsity-promoting algorithms including the role of thresholding (See Herrmann et al. (2015) for more discussion).

Pseudo code illustrating the iterations to solve Equation 2 are summarized in Algorithm 1.

Algorithm 1 Linearized Bregman for LS-RTM

1. Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{z}_0 = \mathbf{0}$, q, λ, batchsize $n_s' \ll n_s$
2. **for** $k = 0, 1, \cdots$
3.   Randomly choose shot subsets $\mathscr{I} \in [1 \cdots n_s]$, $|\mathscr{I}| = n_s'$
4. $\mathbf{A}_k = \{\nabla \mathbf{F}_i(\mathbf{m}_0, q_i) \mathbf{C}^T\}_{i \in \mathscr{I}}$
5. $\mathbf{b}_k = \{\delta \mathbf{d}_i\}_{i \in \mathscr{I}}$
6. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^T P_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
7. $\mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
8. **end**
note: $S_\lambda(\mathbf{z}_{k+1}) = \text{sign}(\mathbf{z}_{k+1}) \max\{0, |\mathbf{z}_{k+1}| - \lambda\}$
    $\mathscr{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k) = \max\{0, 1 - \frac{\sigma}{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|}\} \cdot (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$

In Algorithm 1, $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|^2 / \|\mathbf{A}_k^T(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|^2$ is the dynamic stepsize and $S_\lambda(\mathbf{z})$ is the soft thresholding operator (Lorenz et al., 2014), and $\mathscr{P}_\sigma$ projects the residual onto a ℓ_2-norm ball given by the size of the noise σ. To avoid too many iterations, the threshold λ, which has nothing to do with the noise level but with the relative importance of the ℓ_1 and ℓ_2-norm objectives, should be small enough, usually set proportional to the level of the maximum of $\mathbf{z}_k$ to let entries of $\mathbf{z}_k$ enter into the solution.

On-the-fly source estimation

Algorithm 1 requires knowledge of the source-time signatures q_i to be successfully employed. Unfortunately, this information is typically not available. Following our earlier work on source estimation in time-harmonic imaging and full-waveform inversion (Tu and Herrmann, 2015; van Leeuwen et al., 2011), we propose an approach where after each model update, we estimate the source-time function by solving a least-squares problem that matches predicted and observed data.

For the source estimation, we assume that we have some initial guess $q_0 = q_0(t)$ for the source, which convolved with a filter $w = w(t)$ gives the true source function q.

$$\min_{\mathbf{x}} \ \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x}\|^2$$
$$\text{subject to} \sum_i \|\nabla \mathbf{F}_i(\mathbf{m}_0, w * q_0) \mathbf{C}^T \mathbf{x} - \delta \mathbf{d}_i\| \le \sigma, \quad (3)$$

where $*$ stands for convolution along time. The filter w is unknown, and needs to be estimated in every Bregman step by solving an additional least-squares subproblem. Generally, estimating w requires numerous recomputations of the linearized data. This can be avoided by assuming that the linearized data for the current estimated of w is given by modelling the data with the initial source q_0 and convolving with the filter w afterwards: $\nabla \mathbf{F}_i(\mathbf{m}_0, w * q_0) \approx w * \nabla \mathbf{F}_i(\mathbf{m}_0, q_0)$, where the right hand side stands for trace by trace convolution of w with the shot gather traces.

We expect the estimated source signature to decay smoothly to zero with few oscillations within a short duration of time. Therefore in our subproblem, we add a penalty term $\|r(w * q_0)\|^2$, where r is a weight function that penalizes non-zeros entries at later times. In our example we choose r to be a logistic loss function (Rosasco et al., 2004) as below. Furthermore we add a second penalty term $\lambda_2 \|w * q_0\|^2$ to control the overall

energy of the estimated wavelet. Our subproblem for source estimation is then given by

$$\min_w \sum_i \|w * \nabla \mathbf{F}_i(\mathbf{m}_0, q_0) - \delta \mathbf{d}_i\|^2$$
$$+ \|r(w * q_0)\|^2 + \lambda_2 \|w * q_0\|^2. \tag{4}$$

and the weight function r is defined as

$$r = \begin{cases} \log(1 + e^{\alpha(t - t_0)}) & \text{if } t \le t_0 \\ t - \log(1 + e^{\alpha(t_0 - t)}) & \text{if } t > t_0. \end{cases} \tag{5}$$

The algorithm to solve the sparsity-promoting LS-RTM with source estimation using linearized Bregman is summarized in Algorithm 2.

Algorithm 2 LB for LS-RTM with source estimation

1. Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{z}_0 = \mathbf{0}$, q_0, λ, λ_2, batchsize $n_s' \ll n_s$, weights r
2. **for** $k = 0, 1, \cdots$
3.   Randomly choose shot subsets $\mathscr{I} \in [1 \cdots n_s], |\mathscr{I}| = n_s'$
4. $\mathbf{A}_k = \{\nabla \mathbf{F}_i(\mathbf{m}_0, q_0) \mathbf{C}^T\}_{i \in \mathscr{I}}$
5. $\mathbf{b}_k = \{\delta \mathbf{d}_i\}_{i \in \mathscr{I}}$
6. $\tilde{\mathbf{d}}_k = \mathbf{A}_k \mathbf{x}_k$
7. $\mathbf{w}_k = \arg\min_{\mathbf{w}} \sum_{\mathscr{I}} \|\mathbf{w} * \tilde{\mathbf{d}}_k - \mathbf{b}_k\|^2 + \|r(\mathbf{w} * \mathbf{q}_0)\|^2 + \lambda_2 \|\mathbf{w} * \mathbf{q}_0\|^2$
8. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^* \left(\mathbf{w}_k \star P_\sigma (\mathbf{w}_k * \tilde{\mathbf{d}}_k - \mathbf{b}_k) \right)$
9. $\mathbf{x}_{k+1} = S_\lambda (\mathbf{z}_{k+1})$
10. **end**

In Algorithm 2, $\star$ stands for correlation, which is the adjoint operation of convolution. Since the initial guess of $\mathbf{x}$ is zero, we omit the filter estimation in the first iteration and update $\mathbf{x}_1$ with the initial guess of the source. In the second iteration, after the filter is estimated for the first time (line 7), $\mathbf{z}_1$ is reset to zero, and $\mathbf{z}_2$ and $\mathbf{x}_2$ are updated according to lines $8 - 9$. This prevents that the imprint of the initial (wrong) source wavelet, pollutes the image updates of later iterations. The subproblem in line 7 can be solved by formulating the optimality conditions and solving for w_k directly.

NUMERICAL EXPERIMENTS

To test our algorithm, we conduct two types of experiments. First, we test the linearized Bregman method for sparsity promoting LS-RTM for the case with a known wavelet q and a wrong wavelet q_0. We then compare these results with an example in which the source function is unknown and estimated in the described way. For the experiments, we use the top left part of the SIGSBEE2A model, which is 2918m by 4496m and discretized with a grid size of 7.62m by 7.62m. The model perturbation, which is the difference between the true model and a smoothed background model, is shown in Figure 1a. We generated linearized, single scattered data using the born modelling operator. The experiment consists of 295 shot positions with 4 seconds recording time. Each shot is recorded by 295 evenly spaced receivers at a depth of 7.62m and a receiver spacing of

15.24m, yielding a maximum offset of 4km. For later comparisons with the LS-RTM images, we also generate a traditional RTM image (Figure 1b), which illustrates the wrong amplitudes and blurred shapes at interfaces and diffraction points that are typically associated with RTM images. Using a wrong source wavelet in RTM enhances artifacts even more and destroys the energy's focusing at reflectors and diffractors, as shown in Figure 1c. The wrong wavelet is the initial guess plotted with red line in Figure 3.

In our LS-RTM experiments, we perform 40 iterations and in each iteration draw 8 randomly selected shots. Therefore, after 40 iterations, approximately all shot gathers have been used one time (~one pass through the data). To accelerate the convergence, we apply a depth preconditioner to the model updates, which compensates for the spherical amplitude decay (Herrmann et al., 2008).

LS-RTM with linearized Bregman

In our first test, we perform LS-RTM via the linearized Bregman method using both a correct source wavelet q (dashed yellow line, Figure 3) and a wrong wavelet q_0 (solid red line, Figure 3). The true source wavelet q is formed by convolving a 15 Hz Ricker wavelet with the wavelet q_0, which has a spectrum with a flat shape ranging from 12 to 28 Hz and exponential tapering between 0~12HZ and 28~40HZ (solid red line, Figure 3 bottom).

The only parameter that needs to be provided by the user for the algorithm, is the weighting parameter λ. In our experiments, we determine λ in the first iteration, by setting $\lambda = 0.1 * \max(|\mathbf{z}|)$, which allows roughly 90 percent of the entries in $\mathbf{z}$ to pass the soft thresholding in the first iteration. Thresholding too many entries of $\mathbf{z}$ leads to slower convergence and more iterations, whereas a value of λ that is too small allows noise and subsampling artifacts to pass the thresholding.

The inversion result of the first experiment with the correct source wavelet is shown in Figure 2a. It clearly shows that the faults and reflectors are all accurately imaged. The interfaces are sharp because the influence of the source wavelet has been removed by the inversion. The data residual decays as a function of the iteration number (Figure 4a), although some jumps occur due to redrawing the shot records in iteration. The model error (Figure 4b) shows the same behaviour. The LS-RTM image with the wrong wavelet on the other hand has strong artifacts, blurred layers and the wrongly located diffractors, as shown in Figure 2c. This test indicates the importance of the correct source in LS-RTM.

Linearized Bregman with source estimation

To solve the problem with the unknown source wavelet, we use the linearized Bregman method with source estimation as described in algorithm 2. The initial guess for the source wavelet q_0 is shown in Figure 3. For the source estimation, the user also needs to supply the parameter α in the damping function (equation 5). From our experience, once α is big enough, the recovered source is not sensitive to the change of α. In our test, we set $\alpha = 8$ and $\lambda_2 = 1$. The recovered source (blue dash line) closely matches the correct source time

(a) Model perturbation

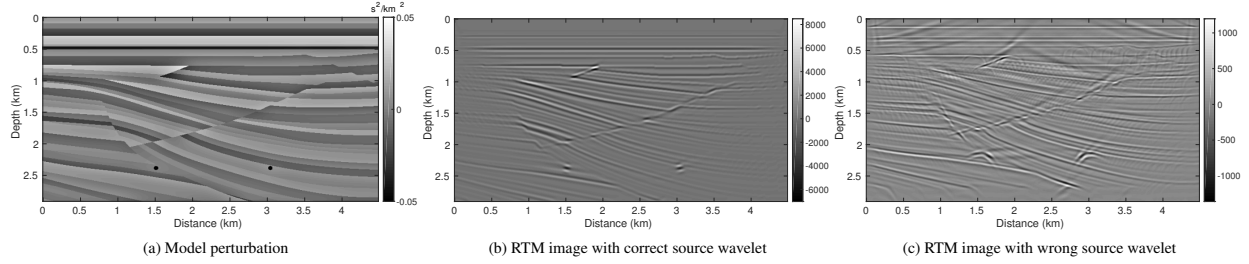(b) RTM image with correct source wavelet

(c) RTM image with wrong source wavelet

Figure 1: Model modified from Sigsbee2A and the corresponding RTM images with correct and wrong source wavelet.

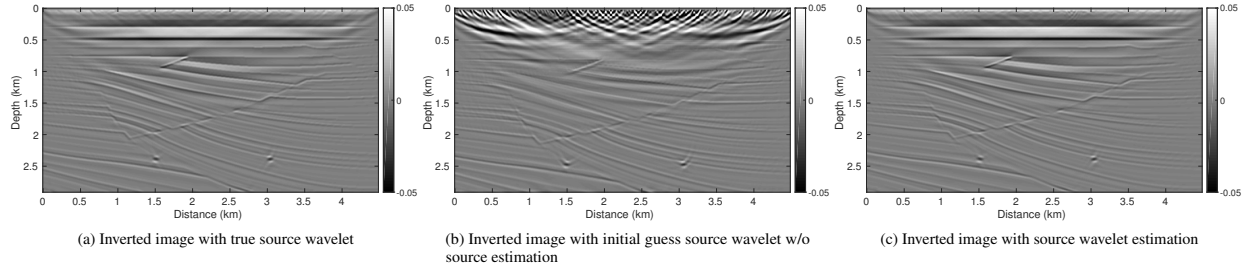(a) Inverted image with true source wavelet

(b) Inverted image with initial guess source wavelet w/o source estimation

(c) Inverted image with source wavelet estimation

Figure 2: LS-RTM image results with roughly one data pass

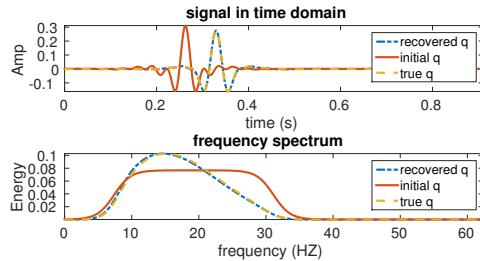function, but some small differences remain in the frequency spectrums.

Figure 3: Recovered source result from the combining algorithm with linearized Bregman

With the same strategy for choosing λ as in experiment 1, the inversion result shown in Figure 2c looks as good as the result from the first experiment in which the correct source was used. The residual as a function of the iterations (Figure 4a) is similar to the residual function with the correct source, with the beginning being slightly larger due to the wrong initial guess wavelet. The same is true for the model error in Figure 4b. However, after 40 iterations, both the final data residual and model error are in the range as the residual and error from the experiment with the correct source wavelet.

CONCLUSIONS

In this work, we perform sparsity promoting LS-RTM in the time-domain using the linearized Bregman method. This algorithm is easy to implement and allows to work with random subsets of data, which allows to significantly reduce the cost of LS-RTM. We provide an example in which we show that we are

(a) Residuals along iterations
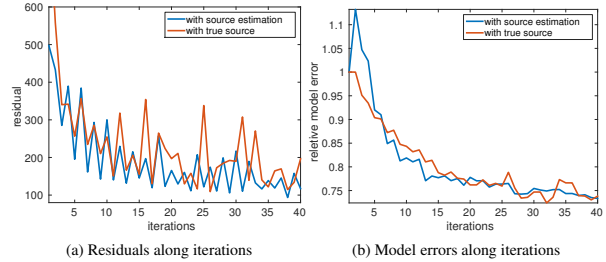
(b) Model errors along iterations

Figure 4: Residuals and model errors along iterations

able to obtain a sharp image without the imprint of the wavelet and true amplitudes at moderate computational cost, which is in the range of one conventional RTM image. Furthermore, this method allows to incorporate source estimation into the inversion, by solving an small additional least-squares subproblem in each Bregman iteration. We propose a modified version of the LB algorithm including source estimation and show that the image we obtain is as good as the LS-RTM result in which we use the correct source wavelet.

ACKNOWLEDGEMENTS

REFERENCES

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.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM review, **43**, 129–159.

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

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: Presented at the EAGE Annual Conference Proceedings, UBC, UBC. ((EAGE, Madrid)).

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

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.

Paffenholz, J., J. Stefani, B. McLain, and K. Bishop, 2002, Sigsbee_2a synthetic subsalt dataset-image quality as function of migration algorithm and velocity model error: Presented at the 64th EAGE Conference & Exhibition.

Rosasco, L., E. De Vito, A. Caponnetto, M. Piana, and A. Verri, 2004, Are loss functions all the same?: Neural Computation, **16**, 1063–1076.

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 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013.

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

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