A linearized Bregman method for compressive waveform inversion

Xintao Chai[1, 2]*, *Mengmeng Yang*[1], *Philip Witte*[1], *Rongrong Wang*[1], *Zhilong Fang*[1], *and Felix J. Herrmann*[1]
[1]*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia (UBC), Vancouver, British Columbia, Canada.*
[2]*China University of Petroleum (Beijing), State Key Laboratory of Petroleum Resources and Prospecting, CNPC Key Laboratory of Geophysical Exploration, Beijing, China.*

Summary

We present an implementation of a recently developed and relatively simple linearized Bregman method to solve the large-scale ℓ_1-norm sparsity-promoting Gauss-Newton (GN) full-waveform inversion (FWI) problem. Numerical experiments demonstrate that: the much simpler linearized Bregman method does a comparable and even superior job compared to a state-of-the-art ℓ_1-norm solver SPGℓ_1, which is previously used in the modified Gauss-Newton FWI; the linearized Bregman method is also more efficient (faster) than SPGℓ_1; the FWI result with the linearized Bregman method solving ℓ_1-norm sparsity-promoting problems to get the model updates is better than that obtained by solving ℓ_2-norm least-squares problems to get the model updates.

Introduction

Full-waveform inversion (FWI) aims at reaping the Earth's subsurface medium properties (e.g., velocity, density) from collected seismograms, which can be formulated as a least-squares (LS) problem (Tarantola, 1984; Pratt et al., 1998; Pratt, 1999; Virieux and Operto, 2009; Li et al., 2012, 2016)

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) := \frac{1}{2}\left\|\mathbf{D}-\mathbf{F}(\mathbf{m},\mathbf{S})\right\|_F^2 . \quad (1)$$

The medium parameter **m** is obtained by minimizing the LS misfit between the observed data **D** and the modeled data **F**(**m**, **S**). $\|.\|_F$ is the Frobenius norm. **F** is the wave-equation based nonlinear modeling operator, parameterized by the discrete and vectorized model **m** and the source signature matrix **S**. Without loss of generality, we assume the source wavelet to be known throughout this paper.

An important algorithm used to solve the non-linear least-squares FWI problem (cf. Equation 1) is the Gauss-Newton (GN) method (Pratt et al., 1998; Li et al, 2012), which utilizes the pseudoinverse of the reduced Hessian, given by the combined action of the Jacobian (linearized Born modeling) operator $\nabla\mathbf{F}(\mathbf{m}, \mathbf{S})$ and its adjoint (migration) $\nabla\mathbf{F}^*(\mathbf{m}, \mathbf{S})$. The superscript $(^*)$ denotes the adjoint. The GN method does not require explicit computation of the Hessian. However, in the frequency domain, each iteration for the GN subproblem requires $4K$ partial-differential-equation (PDE) solves, where $K = N_f \times N_s$, with N_f and N_s the number of frequencies and sources. To reduce the number of PDE solves, some researchers (e.g., Herrmann and Li, 2012) proposed to combine the sources **S** and observed data **D** into a much smaller volume by replacing equation 1 with

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) := \frac{1}{2}\left\|\underline{\mathbf{D}}-\mathbf{F}(\mathbf{m},\underline{\mathbf{S}})\right\|_F^2 , \quad (2)$$

where $\underline{\mathbf{S}}$ = **SE**, $\underline{\mathbf{D}}$ = **DE**. The underlined quantities refer to the simultaneous source experiments obtained by randomized source superimpositions via the action of a source-encoding matrix **E**, which has randomized Gaussian-distributed entries. $\mathbf{E} \in \mathbb{R}^{N_s \times N_s'}$, N_s' is the number of simultaneous sources, $N_s' \ll N_s$. Besides, we can use random subsets of shots as well. However, as discussed in Herrmann and Li (2012), an excessive subsampling will introduce source crosstalks into the inversion result.

The compressed sensing (CS) theory (Candès et al., 2006a, 2006b; Donoho, 2006) has shown that if the signals' energy is concentrated in a few large coefficients, these signals can be reconstructed from small amounts of data by solving sparsity-promoting problems. As we are subsampling over the source experiments (i.e., the input data are compressed) and the unknowns are sparse or compressible, this kind of FWI can be correspondingly called compressive waveform inversion. Herrmann et al. (2008) observed that curvelets, as a directional frame expansion, lead to sparsity of seismic images. With this observation and random subsampling related artifacts are not sparse in the curvelet domain, Li et al. (2012) proposed to remove source crosstalks by solving curvelet-domain sparsity-promoting problems.

The standard GN method exploits the convex-composite structure of equation 2 by linearizing the function (at an initial model $\mathbf{m}_0$) inside the convex ℓ_2-norm (Li et al., 2012)

$$\min_{\delta\mathbf{m}} \frac{1}{2}\left\|\underline{\delta\mathbf{d}}-\nabla\mathbf{F}(\mathbf{m}_0,\underline{\mathbf{S}})\delta\mathbf{m}\right\|_2^2 , \quad (3)$$

where $\underline{\delta\mathbf{d}}$ = vec($\underline{\delta\mathbf{D}}$), $\underline{\delta\mathbf{D}} = \underline{\mathbf{D}} - \mathbf{F}(\mathbf{m}_0, \underline{\mathbf{S}})$, vec(.) means to vectorize a matrix, and $\delta\mathbf{m}$ is the model update. Li et al. (2012) modify the GN subproblem by adding a curvelet sparsity-promotion operator, and recover the model updates by solving one-norm constrained (convex) optimization problems of the type

$$\min_{\mathbf{x}} \frac{1}{2}\left\|\underline{\delta\mathbf{d}}-\nabla\mathbf{F}(\mathbf{m}_0,\underline{\mathbf{S}})\mathbf{C}^*\mathbf{x}\right\|_2^2 \ \text{subject to} \ \|\mathbf{x}\|_1 \le \tau , \quad (4)$$

where $\mathbf{C}^*$ is the adjoint of the sparsifying transform **C**, determining $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$, with **x** a vector of transform coefficients. If no sparsity-promotion is employed, **C** = **I**, with **I** the identity matrix. $\|.\|_1$ and $\|.\|_2$ represent the vector ℓ_1 and ℓ_2 norms, respectively. Li et al. (2012) solve problem 4 with a spectral-gradient method, SPGℓ_1, which is an open-source Matlab solver for ℓ_1-regularized least-squares (van den Berg and Friedlander, 2008). However, SPGℓ_1 is

rather complicated that it is difficult to implement it on industry scale problems (Herrmann et al., 2015).

We adopt a recently developed and relatively simple linearized Bregman method (Cai et al., 2009; Goldstein and Osher, 2009; Yin, 2010; Lorenz et al., 2013, 2014) to solve the large-scale sparsity-promoting GNFWI subproblem (i.e., imposing sparsity on the model updates). The linearized Bregman method has received a lot of attention recently because of its efficiency and simplicity in solving ℓ_1-regularized problems. Lorenz et al. (2014) stated that the linearized Bregman method is especially useful for problems in which the linear measurements are slow and expensive to obtain. With the linearized Bregman method, Herrmann et al. (2015) presented a fast online least-squares migration (LSM), and concluded that the method provably converges to a mixed one-two-norm penalized solution while working on small subsets of data, making it particular suitable for large-scale and parallel industrial applications. When applied to GNFWI, we find that: the much simpler linearized Bregman method competes and even exceeds the performance of the sophisticated SPGℓ_1; the linearized Bregman method is also more efficient than SPGℓ_1.

The paper is organized as follows. We first introduce the theory of the linearized Bregman method used for FWI. Next, we compare the performance of the proposed method, followed by conclusions.

Theory

The linearized Bregman method aims at solving the following regularized version of the basis pursuit denoising (BPDN) problem (Yin, 2010; Lorenz et al., 2013, 2014)

$$\min_{\mathbf{x}} \lambda\|\mathbf{x}\|_1+\frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{subject to} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 \le \sigma^2, \qquad (5)$$

where $\mathbf{A} = \nabla\mathbf{F}(\mathbf{m}_0, \underline{\mathbf{S}})\mathbf{C}^*$, $\mathbf{b} = \text{vec}(\delta\underline{\mathbf{D}})$. The parameter λ is a trade-off factor determining the importance of the ℓ_1-norm compared to the ℓ_2-norm, and σ is a user-specified parameter that allows for noise in the data. Cai et al. (2009), Lorenz et al. (2013) proved the convergence of the linearized Bregman method. Goldstein and Osher (2009) discussed the benefits of the Bregman iteration for ℓ_1-regularized problems. It was shown in Yin (2010) that the solution to problem 5 is also a solution to the normal BPDN problem as long as λ is chosen large enough.

The algorithm 1 gives a pseudo-code of the linearized Bregman method used for GNFWI. Lines 7–12 explain the linearized Bregman iterations in detail. We can see that the main steps of the linearized Bregman method contain only two lines (i.e., lines 9-10). Therefore, it is much easier than SPGℓ_1. The linearized Bregman method is easy to program. A projection function $\Pi_\sigma(\mathbf{Ax}\text{-}\mathbf{b})$ (in line 9) is designed to handle the presence of noise, which involves orthogonal projection onto the ℓ_2-norm ball (Lorenz et al., 2013). The projection function $\Pi_\sigma(\mathbf{Ax}\text{-}\mathbf{b})$ is given by

$$\Pi_\sigma(\mathbf{Ax}\text{-}\mathbf{b}) = \max(0, \sigma/\|\mathbf{Ax}\text{-}\mathbf{b}\|_2)(\mathbf{Ax}\text{-}\mathbf{b}). \qquad (6)$$

For the stopping criteria (in line 3), it will stop when it has visited all the frequency batches. For the threshold λ (in line 10), we generally sort the transform coefficients in descending order, and set λ to be the value corresponding to a user-specified percentage (e.g., 5%) of the total number of the transform coefficients. We can also typically set λ as a fraction of the maximum value of the solution calculated at the first iteration, as Herrmann et al. (2015) did for LSM.

Algorithm 1. A linearized Bregman method for compressive waveform inversion

1: Input data: observed data **D**, source wavelet
   initial velocity model $\mathbf{v}_0$, $\mathbf{m}_0 = 1/(\mathbf{v}_0)^2$
2: Inversion parameters:
   number of simultaneous source experiments N_s'
   threshold λ, outer iteration limit $k_{\text{outerloop}}^{\max}$
   inner iteration limit $k_{\text{innerloop}}^{\max}$
3: **while** the stopping criteria not satisfied **do**
4: Initialize outer loop iteration index $k_o = 0$
5: **while** $k_o < k_{\text{outerloop}}^{\max}$ **do** // outer loop
6: Initialize inner loop iteration index $k_i = 0$
   $\mathbf{x}_{k_i} = \mathbf{g}_{k_i} = \mathbf{0}$, $\delta\mathbf{D} = \mathbf{D} - \mathbf{F}(\mathbf{m}_0, \mathbf{S})$
7: **while** $k_i < k_{\text{innerloop}}^{\max}$ **do** // inner loop
8: **new** random draw source encoding matrix **E**
   $\underline{\mathbf{S}} = \mathbf{SE}$ // subsampling of source experiments
   $\delta\underline{\mathbf{D}} = \delta\mathbf{D} \times \mathbf{E}$ // subsampling of data
   $\mathbf{A}_{k_i} = \nabla\mathbf{F}(\mathbf{m}_0, \underline{\mathbf{S}})\mathbf{C}^*$, $\mathbf{b}_{k_i} = \text{vec}(\delta\underline{\mathbf{D}})$
9: $\mathbf{g}_{k_i+1} = \mathbf{g}_{k_i} - t_{k_i}\mathbf{A}_{k_i}^*\Pi_\sigma\left(\mathbf{A}_{k_i}\mathbf{x}_{k_i} - \mathbf{b}_{k_i}\right)$
10: $\mathbf{x}_{k_i+1} = \text{shrink}\left(\mathbf{g}_{k_i+1}, \lambda\right)$ //soft thresholding
11: $k_i = k_i + 1$ //increment
12: **end while** // end inner loop
13: $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$, $\mathbf{m}_0 = \mathbf{m}_0 + \delta\mathbf{m}$ // model update
14: $k_o = k_o + 1$ //increment
15: **end while** // end outer loop
16: **end while**

The dynamic stepsize t (algorithm 1, line 9) is defined as

$$t_{k_i} = \left\|\mathbf{A}_{k_i}\mathbf{x}_{k_i} - \mathbf{b}_{k_i}\right\|_2^2 / \left\|\mathbf{A}_{k_i}^*\left(\mathbf{A}_{k_i}\mathbf{x}_{k_i} - \mathbf{b}_{k_i}\right)\right\|_2^2. \qquad (7)$$

The component-wise soft shrinkage function shrink(.) (in algorithm 1, line 10) is given by (Lorenz et al., 2013, 2014)

$$\text{shrink}(x, \lambda) = \max(|x| - \lambda, 0)\,\text{sign}(x). \qquad (8)$$

The implementation in this paper is 2D frequency domain FWI, however, the linearized Bregman method can be applied to 2D/3D time/frequency domain FWI/migration.

Numerical experiments

Marmousi II model

This experiment aims to demonstrate the superiority of sparsity-promoting by ℓ_1-norm. Keeping this purpose in mind, we use the LSQR algorithm (Paige and Saunders, 1986) to solve the following least-squares problem

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{subject to} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 \le \sigma^2, \tag{9}$$

to get the model updates. LSQR is an implementation of a conjugate-gradient type method for solving LS problems. Here, a strategy (e.g., signal-to-noise ratio, SNR) is adopted to evaluate the recovery quality. The SNR is calculated by

$$\text{SNR (dB)} = -20\log_{10}(\|\mathbf{v}_{\text{true}} - \mathbf{v}_{\text{inv}}\|_2 / \|\mathbf{v}_{\text{true}}\|_2) \tag{10}$$

where $\mathbf{v}_{\text{true}}$ and $\mathbf{v}_{\text{inv}}$ denote the actual and inverted velocity models, respectively. The size of the model (Figure 1a) is a 281×601 grid, with a grid spacing of $d_x = d_z$ = 12.5 m. We use a Ricker wavelet with a peak frequency of 10 Hz and a phase shift of 0.1 s. A total number of 26 frequencies in the frequency range [3 15.5] Hz are generated. All simulations are carried out with 301 shot and 601 receiver positions sampled at 25 m and 12.5 m intervals, respectively. We start with a smooth model (Figure 1b) and work with small batches of frequency data at a time, each using 6 frequencies and 10 randomly formulated simultaneous shots, moving from low to high frequencies in overlapping batches of 3. We use 10 GN outer iterations for each frequency batch. For each GN subproblem, we use 20 inner iterations of LSQR and the linearized Bregman method.

Comparing Figure 1c with 1d indicates that the FWI result with the linearized Bregman method solving ℓ_1-norm sparsity-promoting problems is better than the FWI result with LSQR solving ℓ_2-norm LS problems. This is due to the benefits of sparsity-promotion (Li et al., 2012, 2016), and also because using the ℓ_1-norm as a measurement of sparsity is better than using the ℓ_2-norm (Chen et al., 2001; Tu and Herrrmann, 2015a; Chai et al., 2014, 2016).

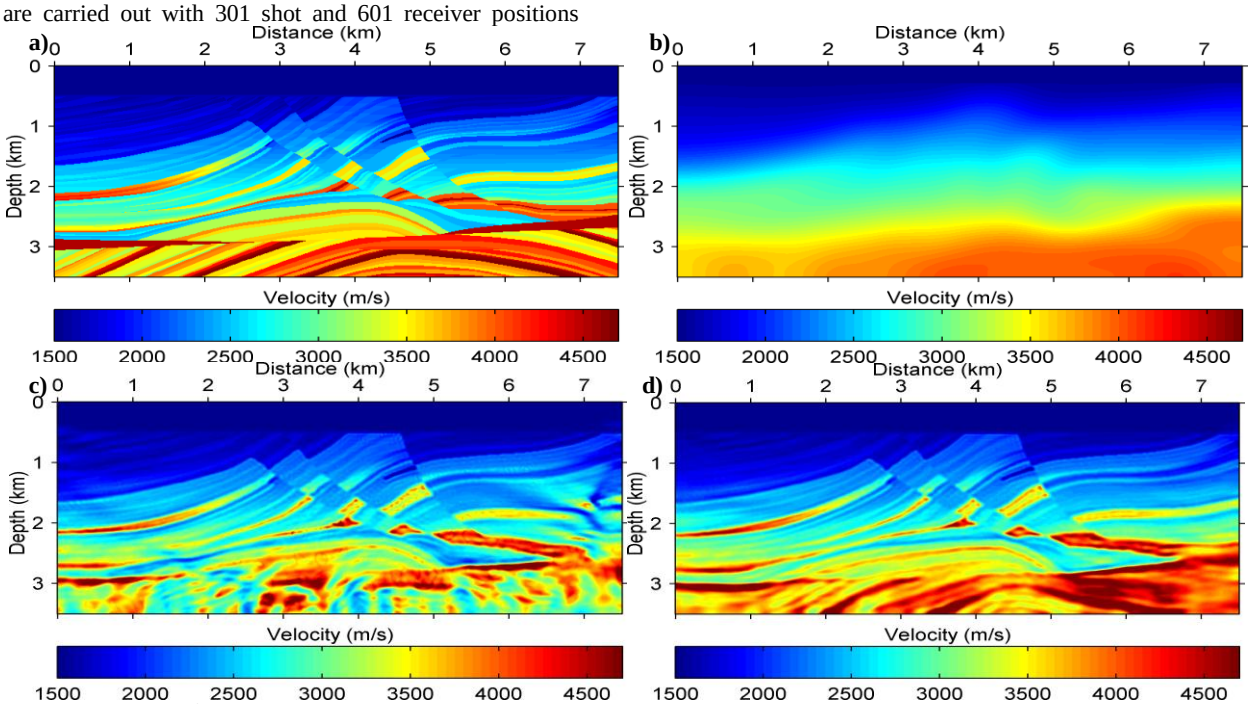

Figure 1. FWI results for a down-sampled Marmousi II model (Martin et al., 2006) dataset. (a) True velocity model. (b) Initial velocity model. (c) Inverted velocity model with LSQR, SNR = 16.2619 (dB). (d) Inverted velocity model with the linearized Bregman method, SNR = 18.3212 (dB).

BG Compass model

For this experiment, we use the same input data set and inversion parameter settings as Li et al. (2012). Specifically, the synthetic velocity model (Figure 2a) with a large degree of variability constrained by well data, is used to generate the observed data with a 12 Hz Ricker wavelet. The size of the model is a 205×701 grid, with a grid spacing of $d_x = d_z$ = 10 m. A total number of 58 frequencies in the frequency range [2.9 22.5] Hz are generated. All simulations are carried out with 351 shot and 701 receiver positions sampled at 20 m and 10 m intervals, respectively, yielding a maximum offset of 7 km. A smooth initial model without lateral information (Figure 2b) is used for FWI. The inversions are carried out sequentially in 10 overlapping frequency bands, each using 7 different randomly selected simultaneous shots and 10 selected frequencies. We use 10 GN iterations for each frequency batch. For each GN subproblem, we use 20 inner iterations of SPGℓ_1 and the linearized Bregman method. No changes are made to the codes and the results of Li et al. (2012).

The FWI results (in Figure 2c and 2d) are calculated using the same computing resources (1 node, 11 workers). The running time of FWI with the linearized Bregman method (17 hours) is much less than that with SPGℓ_1 (25.5 hours). Comparison of Figure 2c and 2d reveals that the structures and amplitudes produced by the linearized Bregman method are more clear and closer to the true model than that by SPGℓ_1. Therefore, the much simpler linearized Bregman method does an excellent job compared to SPGℓ_1.

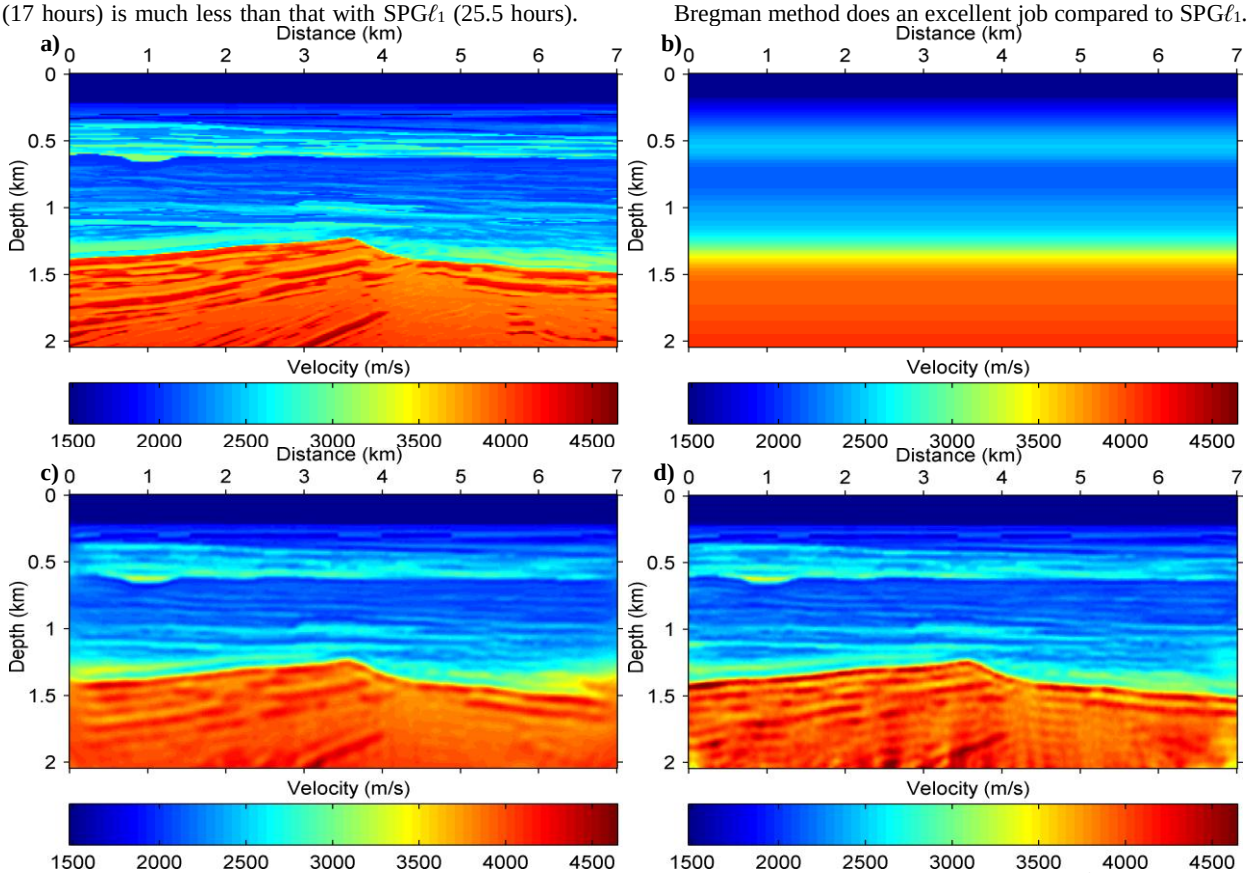

Figure 2. (a) True velocity model. (b) Initial model. (c) Inverted velocity model with the method developed in Li et al. (2012), for which SPGℓ_1 is used to solve the ℓ_1 sparsity-promotion problems to get the model updates. (d) Inverted velocity model with the linearized Bregman method.

Conclusions

We present an adaptation of a recently developed and relatively simple linearized Bregman method to solve the large-scale ℓ_1 sparsity-promoting GNFWI subproblem. Numerical experiments shows that: the much simpler linearized Bregman method does an excellent job compared to SPGℓ_1; the linearized Bregman method is more efficient than SPGℓ_1; the FWI result with the linearized Bregman method solving ℓ_1-norm sparsity-promoting problems is better than that obtained by solving ℓ_2-norm LS problems.

Acknowledgments

This work was financially supported in part by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of the SINBAD II project with the support of the member organizations of the SINBAD Consortium. We are grateful to Charles Jones from BG Group for providing us with the BG Compass velocity model. We thank Dr Xiang Li at PGS for helpful discussion about the modified GNFWI. We thank Dr Tristan van Leeuwen and Curt Da Silva for the frequency-domain modelling codes that we used in our inversion algorithm. We would like to thank the authors of SPOT (https://github.com/mpf/spot), SPGℓ_1 (https://github.com/mpf/spgl1), and CurveLab (http://www.curvelet.org). Xintao Chai sincerely thanks his home supervisor Dr Shangxu Wang for supporting his visiting to the University of British Columbia (UBC). Xintao Chai thanks all the people at UBC-SLIM for their beneficial discussion and generous help. Xintao Chai was financially supported in part by the National Key Basic Research Development Program (2013CB228600).

Cai, J. F., S. Osher, and Z. Shen, 2009, Linearized Bregman iterations for compressed sensing: Mathematics of Computation, **78**, 1515-1536.

Candès, E. J., J. Romberg, and T. Tao, 2006a, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207–1223, doi: 10.1002/cpa.20124.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006b, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, 861–899, doi: 10.1137/05064182X.

Chai, X., S. Wang, S. Yuan, J. Zhao, L. Sun, and X. Wei, 2014, Sparse reflectivity inversion for nonstationary seismic data: Geophysics, **79**, no. 3, V93-V105, doi: 10.1190/GEO2013-0313.1.

Chai, X., S. Wang, J. Wei, J. Li, and H. Yin, 2016, Reflectivity inversion for attenuated seismic data: Physical modeling and field data experiments: Geophysics, **81**, no. 1, doi: 10.1190/GEO2015-0250.1.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129–159, doi: 10.1137/S003614450037906X.

Donoho, D. L., 2006, Compressed sensing: IEEE Transactions on Information Theory, **52**, 1289–1306, doi: 10.1109/TIT.2006.871582.

Goldstein, T., and S. Osher, 2009, The split Bregman method for L1-regularized problems: Siam Journal on Imaging Sciences, **2**, 323-343, doi: 10.1137/080725891.

Herrmann, F. J., N. Tu, and E. Esser, 2015, Fast "online" migration with Compressive Sensing: EAGE Annual Conference Proceedings, doi: 10.3997/2214-4609.201412942.

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **60**, 696–712, doi: 10.1111/j.1365-2478.2011.01041.x.

Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008, Sparsity and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, 150–173, doi: 10.1016/j.acha.2007.06.007.

Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: Geophysics, **77**, no. 3, A13–A17, doi: 10.1190/GEO2011-0410.1.

Li, X., E. Esser, and F. J. Herrmann, 2016, Modified Gauss-Newton full-waveform inversion explained–-why sparsity-promoting updates do matter: accepted to be published in Geophysics.

Lorenz, D. A., F. Schöpfer, and S. Wenger, 2013, The linearized Bregman method via split feasibility problems: analysis and generalizations: Siam Journal on Imaging Sciences, **7**, 1237-1262.

Lorenz, D. A., S. Wenger, F. Schöpfer, and M. Magnor, 2014, A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing: Eprint Arxiv, 1347-1351.

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156-166, doi: 10.1190/1.2172306.

Paige, C. C., and M. A., Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43-71, doi: 10.1145/355984.355989.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888-901, doi: 10.1190/1.1444597.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341-362, doi: 10.1046/j.1365-246X.1998.00498.x.

Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015, doi: 10.1111/j.1365-2478.1984.tb00751.x.

Tu, N., and F. J. Herrmann, 2015a, Fast imaging with surface-related multiples by sparse inversion: Geophysical Journal International, **201**, 304-317, doi: 10.1093/gji/ggv020.

Tu, N., and F. J. Herrmann, 2015b, Fast least-squares imaging with surface-related multiples: Application to a North Sea data set: The Leading Edge, **34**, 788–794, doi: 10.1190/tle34070788.1.

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, doi: 10.1137/080714488.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

Yin, W., 2010, Analysis and generalizations of the linearized Bregman method: SIAM Journal on Imaging Sciences, **3**, 856-877, doi: 10.1137/090760350.