

## Automatic salt delineation — Wavefield Reconstruction Inversion with convex constraints

Ernie Esser, Lluís Guasch<sup>&</sup>, Tristan van Leeuwen<sup>†</sup>, Aleksandr Y. Aravkin<sup>‡</sup>, and Felix J. Herrmann<sup>\*</sup>

<sup>\*</sup>University of British Columbia Dept. of Earth, Ocean and Atmospheric Sciences; <sup>&</sup>Sub Salt Solutions Limited;

<sup>†</sup>Mathematical Institute at Utrecht University, <sup>‡</sup>IBM T.J. Watson Research Center

### SUMMARY

We extend full-waveform inversion by Wavefield Reconstruction Inversion by including convex constraints on the model. Contrary to the conventional adjoint-state formulations, Wavefield Reconstruction Inversion has the advantage that the Gauss-Newton Hessian is well approximated by a diagonal scaling, which allows us to add convex constraints, such as the box- and the edge-preserving total-variation constraint, on the square slowness without incurring significant increases in computational costs. As the examples demonstrate, including these constraints yields far superior results in complex geological areas that contain high-velocity high-contrast bodies (e.g. salt or basalt). Without these convex constraints, adjoint-state and Wavefield Reconstruction Inversion get trapped in local minima for poor starting models.

### INTRODUCTION

Time-harmonic acoustic full-waveform inversion (FWI, Tarantola (1984); Virieux and Operto (2009); Herrmann et al. (2013)) with the adjoint-state method derives from the following PDE-constrained optimization problem:

$$\min_{\mathbf{m}, \mathbf{u}} \sum_{sv} \frac{1}{2} \|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 \quad \text{such that } A_v(\mathbf{m})\mathbf{u}_{sv} = \mathbf{q}_{sv}, \quad (1)$$

where  $A_v(\mathbf{m})\mathbf{u}_{sv} = \mathbf{q}_{sv}$  denotes the discretized Helmholtz equation. Let  $s = 1, \dots, N_s$  index the sources and  $v = 1, \dots, N_v$  index frequencies. We consider the model,  $\mathbf{m}$ , which corresponds to the slowness squared, to be a real vector  $\mathbf{m} \in \mathbb{R}^N$ , where  $N$  is the number of points in the spatial discretization. For each source and frequency, the wavefields, sources and observed data are denoted by the complex-valued vectors  $\mathbf{u}_{sv} \in \mathbb{C}^N$ ,  $\mathbf{q}_{sv} \in \mathbb{C}^N$  and  $\mathbf{d}_{sv} \in \mathbb{C}^{N_r}$  respectively, where  $N_r$  is the number of receivers.  $P$  is the operator that restricts the wavefields onto the receiver locations. The discretized Helmholtz matrix has the form  $A_v(\mathbf{m}) = \omega_v^2 \text{diag}(\mathbf{m}) + L$ , where  $\omega_v$  is angular frequency and  $L$  is a discrete Laplacian.

The most common approach to solve Equation 1 is to eliminate the non-convex PDE constraints, which results in a parametric inversion problem over  $\mathbf{m}$ , requiring solves of the Helmholtz and adjoint Helmholtz systems for each iteration of the optimization. While elimination of these constraints undergirds the majority of current-day FWI approaches, it is computationally expensive and prone to local minima and therefore relying on unrealistically accurate starting models. To overcome these challenges, recent work proposes to reduce computational costs via randomized subsampling techniques [(Krebs et al., 2009; van Leeuwen et al., 2011; Li et al., 2012; Haber et al., 2012) and reliance on accurate starting models via so-called extensions. The goal of these extensions is to make sure that FWI does not paint itself in a corner for poor starting models that can not

fit the observed data. Instead, the extended formulations are designed to fit the data irrespective of the quality (within reason) of the starting model. Symes (2008) and later Biondi and Almomin (2013) accomplish this by extending the model space to include a subsurface offset coordinate while Mike Warner’s Adaptive Waveform Inversion (AWI, Warner and Guasch, 2014) introduces Wiener filters that live in the data-space and that are designed to fit the observed data irrespective of the quality of the starting model. The extended inversions, where the dimensionality of the search space is increased, are subsequently carried out by migration-velocity-analysis (MVA, Symes, 2008) type of focussing procedures that squeeze the expanded search space to something physical and thereby yielding model updates that are less prone to cycle skipping.

Wavefield Reconstruction Inversion (WRI, van Leeuwen and Herrmann, 2013a; van Leeuwen et al., 2014; van Leeuwen and Herrmann, 2013b) is also based on an extension. However, contrary to the afore mentioned extensions, WRI derives directly from Equation 1 by replacing the constraints by  $\ell_2$ -norm penalties yielding

$$\min_{\mathbf{m}, \mathbf{u}} \sum_{sv} \frac{1}{2} \|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2} \|A_v(\mathbf{m})\mathbf{u}_{sv} - \mathbf{q}_{sv}\|^2. \quad (2)$$

As opposed to conventional FWI, the optimization of Equation 2 is carried out over both the squared slownesses  $\mathbf{m}$  and the wavefields collected in  $\mathbf{u}$ . As in AWI, WRI follows an alternating optimization strategy where for a fixed model  $\mathbf{m}^n$  at the  $n^{\text{th}}$  iteration wavefields are calculated that fit both the observed data and the PDE. Given the solution of this data-augmented wave equation given by

$$\bar{\mathbf{u}}_{sv} = \arg \min_{\mathbf{u}_{sv}} \frac{1}{2} \|P\mathbf{u}_{sv} - \mathbf{d}_{sv}\|^2 + \frac{\lambda^2}{2} \|A_v(\mathbf{m}^n)\mathbf{u}_{sv} - \mathbf{q}_{sv}\|^2, \quad (3)$$

the model is updated by making a step that minimizes the PDE misfit,  $\frac{1}{2} \|A_v(\mathbf{m}^n)\bar{\mathbf{u}}_{sv} - \mathbf{q}_{sv}\|^2$ , given the solution of the data-augmented wave equation whose solutions  $\bar{\mathbf{u}}$  fit the observed data and the wave equation. The latter step in this alternating variable-projection approach (Aravkin and van Leeuwen, 2012), can be interpreted as a “focussing” by the wave equation itself yielding the following expressions for the gradient and Gauss-Newton Hessian (van Leeuwen and Herrmann, 2013a):

$$\mathbf{g}^n = \sum_{sv} \text{Re} \left\{ \lambda^2 \omega_v^2 \text{diag}(\bar{\mathbf{u}}_{sv})^* (A_v(\mathbf{m}^n)\bar{\mathbf{u}}_{sv} - \mathbf{q}_{sv}) \right\} \quad (4)$$

and

$$H_{sv}^n \approx \sum_{sv} \text{Re} \left\{ \lambda^2 \omega_v^4 \text{diag}(\bar{\mathbf{u}}_{sv}(\mathbf{m}^n))^* \text{diag}(\bar{\mathbf{u}}_{sv}(\mathbf{m}^n)) \right\}. \quad (5)$$

Compared to regular adjoint-state methods, the above Gauss-Newton Hessian is well approximated by the diagonal (because it is diagonally dominant) while the gradient requires only the

solution of the data-augmented wave equation and is made of correlating this solution with the PDE residual. Moreover, the gradient and Hessian of WRI can be accumulated in parallel and do not require storage of all wavefields as in full-space all-at-once methods (Haber and Ascher, 2001) used to solve Equation 1. The parameter  $\lambda$  in this formulation controls the PDE and data misfits and can be chosen to fit the observed data. Given the above expressions for the gradient and Gauss-Newton Hessian, we arrive at the following expression for scaled-gradient model updates (Bertsekas, 1999):

$$\begin{aligned} \Delta \mathbf{m} &= \arg \min_{\Delta \mathbf{m} \in \mathbb{R}^N} \Delta \mathbf{m}^T \mathbf{g}^n + \frac{1}{2} \Delta \mathbf{m}^T H^n \Delta \mathbf{m} + c_n \Delta \mathbf{m}^T \Delta \mathbf{m} \\ \mathbf{m}^{n+1} &= \mathbf{m}^n + \Delta \mathbf{m}, \end{aligned} \quad (6)$$

with  $c_n \geq 0$ . These updates correspond to ordinary gradient updates with  $H = 0$  and to the Newton method when  $c_n = 0$  and  $H$  the true Hessian. Since the Gauss-Newton Hessian approximation (cf. Equation 5) is diagonal, it can be easily incorporated into Equation 6 at essentially no additional computational cost. For this reason, the above quadratic approximation forms the basis for our proposed generalization of WRI that includes convex constraints.

### INCLUDING CONVEX CONSTRAINTS

While there are strong indications that WRI is, by virtue of its increased search space and ability to fit the data, less sensitive to the accuracy of starting models, its current formulation does not employ prior knowledge in the form of convex box- or edge-promoting total-variation constraints.

Inclusion of this type of constraints makes WRI better posed and can be accomplished by simply adding convex constraints  $\mathbf{m} \in C$ , where  $C$  is a convex set, to Equation 6. We now have

$$\begin{aligned} \Delta \mathbf{m} &= \arg \min_{\Delta \mathbf{m} \in \mathbb{R}^N} \Delta \mathbf{m}^T \mathbf{g}^n + \frac{1}{2} \Delta \mathbf{m}^T H^n \Delta \mathbf{m} + c_n \Delta \mathbf{m}^T \Delta \mathbf{m} \\ &\text{such that } \mathbf{m}^n + \Delta \mathbf{m} \in C. \end{aligned} \quad (7)$$

Incorporation of these constraints ensures that the next model iterate remains in the convex set ( $\mathbf{m}^{n+1} \in C$ ) but makes  $\Delta \mathbf{m}$  more difficult to compute. However, the problem remains tractable if  $C$  is easy to project onto or can be written as an intersection of convex constraints that are each easy to project onto. Because solving for  $\bar{\mathbf{u}}^n$  (Equation 3) is much more expensive, projections on these convex subproblems is unlikely to become a computational bottleneck. Quite the contrary, inclusion of these constraints could even speed up the overall method if it leads to fewer required model updates compared to solving the unconstrained problem.

According to Esser et al. (2013), the above approach works for fairly general Hessian approximations. For WRI specifically, where the Gauss-Newton Hessian is diagonal and made of the sum over all experiments (cf. Equation 5), we only need the diagonal entries of the Hessian to be bounded for  $\mathbf{m} \in C$ . In addition, line searches are not needed as long as  $c_n$  is small enough.

Since the approximate Hessian  $H^n$  is diagonal and positive, it is straightforward to add box constraints—i.e. the entries

of the vector  $\mathbf{m}$  are constrained to the interval  $[b, B]$  with  $b, B$  the minimum and maximum, or  $C_{\text{box}} = \{\mathbf{m} : m_i \in [b, B], i = 1 \dots N\}$  for short. While inclusion of these box constraints leads to cheap closed-form analytic expressions, imposing this constraint by itself is insufficient in situations where there is lots of source cross-talk, induced by the use of simultaneous sources to speed up the computations, or in the presence of high-contrast high-velocity unconformities such as salt or basalt. As we will show below, WRI yields better solutions when adding a combination of box- and total-variation norm constraints—i.e., we seek updates

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta \mathbf{m} \quad \text{s. t.} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}}, \quad (8)$$

where  $C_{\text{TV}} = \{\mathbf{m} : \|\mathbf{m}\|_{\text{TV}} \leq \tau\}$  with the discrete TV-norm (we reorganized  $\mathbf{m}$  into a  $N = N_1 \times N_2$  2D array) is defined as

$$\|\mathbf{m}\|_{\text{TV}} = \frac{1}{h} \sum_{ij} \sqrt{(m_{i+1,j} - m_{i,j})^2 + (m_{i,j+1} - m_{i,j})^2}. \quad (9)$$

As we will demonstrate below, jointly imposing bound and TV constraints, the latter by constraining the TV-norm to be less than some positive parameter  $\tau$ , drastically improves the performance of WRI. TV penalties are widely used in image processing to remove noise while preserving discontinuities (Rudin et al., 1992). It is also a useful regularizer in a wide variety of other inverse problems, especially when solving for piecewise constant or piecewise smooth unknowns. For example, TV regularization has been successfully used for electrical inverse tomography (Chung et al., 2005) and inverse wave propagation (Akcelik et al., 2002) and more recently in work by Guo et al. (2013). Although these approaches are somewhat similar, our formulation is different because we avoid expensive evaluations of dense Gauss-Newton Hessians. Instead, we impose the constraints in a cheap inner loop that does not involve additional wave-equation solves.

### SOLVING THE CONVEX SUBPROBLEMS

A key aspect of our work is that Equation 6, supplemented with convex constraints, is amenable to large-scale problems because it has an outer-inner loop structure. The outer loop concerns the evaluation of the expensive gradient and Hessian while the inner loop involves relatively cheap projections onto the constraints. An computationally efficient way to solve the inner convex subproblems is to modify Zhu and Chan (2008)'s primal-dual hybrid gradient method and find saddle points of the following Lagrangian (Esser et al., 2010; Chambolle and Pock, 2011; He and Yuan, 2012; Zhang et al., 2010):

$$\begin{aligned} \mathcal{L}(\Delta \mathbf{m}, \mathbf{p}) &= \Delta \mathbf{m}^T \mathbf{g} + \frac{1}{2} \Delta \mathbf{m}^T (H^n + 2c_n \mathbf{I}) \Delta \mathbf{m} \\ &\quad + \mathbf{p}^T D(\mathbf{m}^n + \Delta \mathbf{m}) - \tau \|\mathbf{p}\|_{\text{dual TV}} \end{aligned} \quad (10)$$

for  $\mathbf{m}^n + \Delta \mathbf{m} \in C_{\text{box}}$  and  $\mathbf{p}$  the Lagrange multiplier. In this expression,  $\|\mathbf{p}\|_{\text{dual TV}}$  is the dual norm of the TV-norm, which takes the maximum instead of the sum of the  $\ell_2$ -norm of the discrete gradients (denoted by  $D$ ) at each point of model  $\mathbf{m}$ . For more details on the specific implementations of these iterations including the projections on the TV-ball of size  $\tau$ , we refer to the technical report by Esser et al. (2014).

## PRACTICAL ASPECTS

The above scaled gradients with convex constraints provide the formal optimization framework with which we will carry out our experiments. Before demonstrating the performance of our approach, we discuss a number of practical considerations and strategies to expedite and terminate the iterative inversion and to make the inversion more robust to poor starting models and complex geology.

**Frequency continuation.** We divide the frequency spectrum into overlapping frequency bands of increasing frequency, in order to further robustify the WRI approach against cycle skipping.

**Stopping criterion.** As with all iterative schemes, some sort of stopping criterion is needed. For each frequency batch, we compute at most 25 outer iterations, each time solving the convex subproblem to convergence, stopping when  $\max(\frac{\|p^{k+1}-p^k\|}{\|p^{k+1}\|}, \frac{\|\Delta\mathbf{m}^{k+1}-\Delta\mathbf{m}^k\|}{\|\Delta\mathbf{m}^{k+1}\|}) \leq 1 \times 10^{-4}$ . We compute these outer iterations as long as the reduction of the overall objective is larger than some predefined fraction of its quadratic approximation  $\Delta\mathbf{m}^T \mathbf{g} + \frac{1}{2} \Delta\mathbf{m}^T (H^n + 2c_n I) \Delta\mathbf{m}$  at the current model iterate.

**Acceleration by randomized sampling.** To reduce the computational costs, we employ the widely-reported randomized sampling technique (Krebs et al., 2009; van Leeuwen et al., 2011; Haber et al., 2012) by considering  $N'_s \ll N_s$  random mixtures of the sources  $q_{sv}$  defined by

$$\mathbf{q}_{jv} = \sum_{s=1}^{N_s} w_{js} \mathbf{q}_{sv} \quad j = 1, \dots, N'_s, \quad (11)$$

where the weights  $w_{js} \in \mathcal{N}(0, 1)$  are drawn from a standard normal distribution. We modify the synthetic data according to  $\mathbf{d}_{jv}^s = P A_v^{-1}(\mathbf{m}) \mathbf{q}_{jv}$  and use the same strategy to solve the now much smaller optimization problem.

**Multiple passes through the data.** Despite the fact that WRI is expected to be less prone to local minima, restarting the optimizations with warm starts at the low frequencies again leads to superior results as reported by Peters et al. (2014). We employ this strategy.

**Relaxation of the TV constraint.** Having included constraints allows us to use a cooling technique during which the convex constraint, the TV-norm in this case, is gradually relaxed. While these type of continuation techniques have been used very successfully towards the solution of linear problems with convex constraints, devising these techniques for nonlinear problems is still a topic of open research. We found empirically that gradually relaxing the TV-constraint ( $\tau$ ) during the multiple passes through the data greatly improved our results.

## NUMERICAL EXPERIMENTS

As we have learned from the blind Chevron Gulf of Mexico dataset, FWI is severely challenged when applied to regions with high complexity for data that miss long offsets and low frequencies (Herrmann et al., 2013). To demonstrate the added value of WRI and of adding convex constraints to FWI or WRI,

we conduct a series of experiments on subsets of the 2004 BP velocity benchmark (Billette and Brandsberg-Dahl, 2005). These experiments are aimed at demonstrating the benefits of including TV-norm constraints for WRI in case of good and bad starting model, WRI compared to FWI, and multiple passes while relaxing the constraints.

**Good starting model.** Let us first consider, the results of WRI with and without TV-norm regularization for the upper left part of the BP model for a starting model obtained by smoothing. The results for the first and second pass are included in Figure 1 and were obtained by carrying out the inversions from 3Hz to 20Hz in overlapping batches of two. The bound constraints on the slowness squared are defined to correspond to minimum and maximum velocities of 1400 and 5000 m/s, respectively. The  $\tau$  parameter for the TV constraint is chosen to be .9 times the TV of the ground-truth model. To reduce computation time, we use only two simultaneous shots, which we redraw each time the model is updated. From these figures it is clear that including the edge-preserving TV-norm constraint greatly improves the results by removing the source crosstalk while clearly delineating the top and bottom of the salt. The figures also show that the inversion improves dramatically when making a second pass while relaxing the TV-norm constraint.

**Bad initial model.** As we can see from Figure 2, including TV-norm constraints also yields excellent results when starting with a model given by a simple gradient. The same applies when we approximate the Gauss-Newton Hessian of the adjoint-state method by Equation 5. This approximation, which is known as the pseudo Hessian, yields good but slightly inferior results compared to those of WRI. Both results were obtained via multiple passes and benefit tremendously from adding the TV-norm constraint and clearly delineate the top and bottom of the salt. (Results from adjoint-state FWI without the TV-norm constraint are very poor compared to the unconstrained WRI (not shown).)

## CONCLUSIONS AND FUTURE WORK

We presented a computationally feasible scaled-gradient projection algorithm for minimizing the quadratic penalty formulation of full-waveform inversion proposed by van Leeuwen and Herrmann (2013a) subject to additional convex constraints. We showed in particular how to solve convex subproblems that arise when adding total-variation and bound constraints on the model. The presented synthetic experiments clearly demonstrate that full-waveform inversion in complex geological areas can greatly improve by including convex constraints. This observation applies to conventional adjoint-state FWI, Wavefield Reconstruction Inversion (WRI), and Adaptive Waveform Inversion (AWI). Both the top and bottom of the salt are well recovered, which makes the presented method a candidate to replace labour intensive salt flooding techniques. There are also indications that the TV-norm constraint improves AWI from reflection data only as reported elsewhere in these proceedings.

In future work, we still need to find a practical way of selecting the regularization parameter  $\tau$ . The fact that significant discontinuities can be resolved using small  $\tau$  explains why our continuation strategy that gradually increases  $\tau$  yields im-

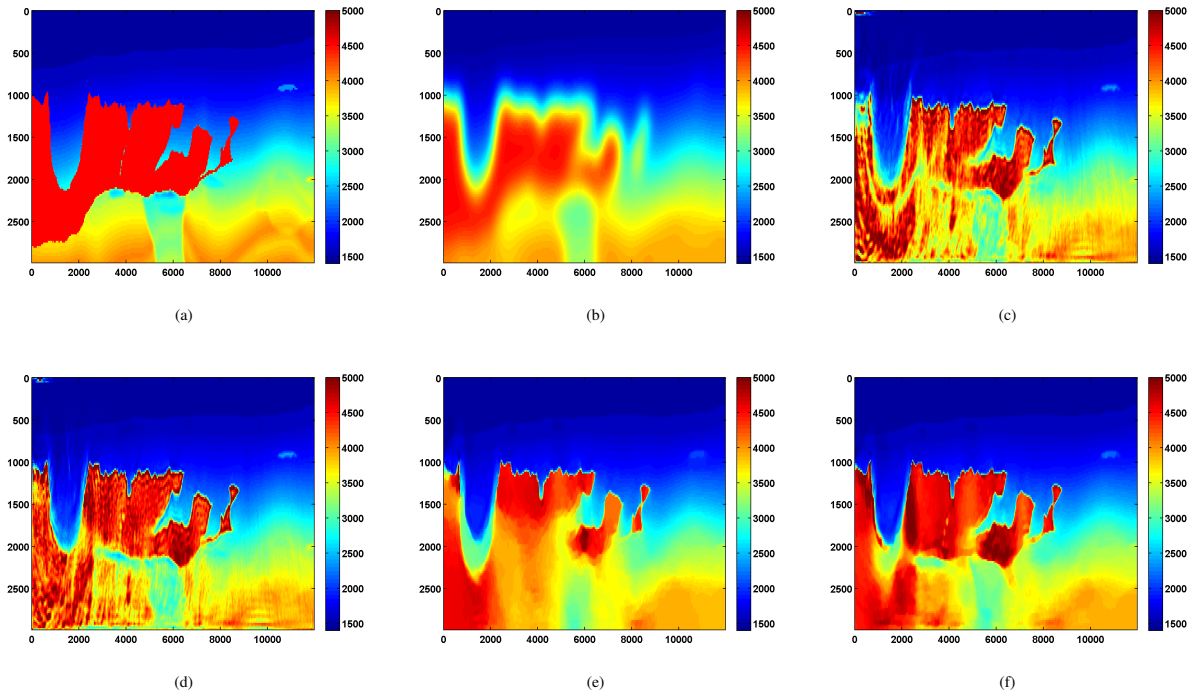


Figure 1: Top left portion of the BP 2004 velocity model (a), starting model (b), first pass w/o TV (c), second pass w/o TV (d), first pass w/ TV and  $\tau = 0.9\tau_{\text{true}}$  (e), second pass w/ TV  $\tau = 0.99\tau_{\text{true}}$  (f).

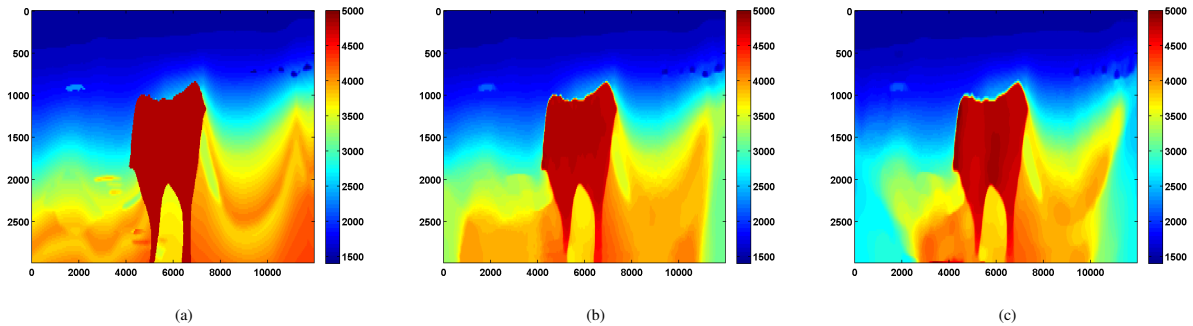


Figure 2: Middle portion of the BP 2004 BP velocity model (a), recovered velocity with WRI and TV-constraint, recovered model with adjoint-state FWI and TV-constraint.

proved results. These improvements stem from the fact that the TV-constraint introduces reflecting boundaries, which aid the inversion. A possible numerical framework for choosing the regularization parameter could be along the lines of the SPOR-SPG method in (van den Berg and Friedlander, 2011).

#### ACKNOWLEDGEMENTS

Ernie Esser (1980–2015), who passed away under tragic circumstances, was an extremely enthusiastic and talented researcher who was, above all, immensely generous with his time and ideas. This work is an attempt to share his work and is a reflection of what a true privilege it has been to have had the opportunity to work with Ernie. In earlier work, Ernie wished

to thank Bas Peters for insights about the formulation and implementation of WRI and Polina Zheglova for helpful discussions about the boundary conditions.

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 support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, Petrobras, PGS, Sub Salt Solutions Ltd, WesternGeco, and Woodside.

## REFERENCES

- Akcelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Supercomputing, *ACM/IEEE*, **00**, 1–15.
- Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: *Inverse Problems*, **28**, 115016.
- Bertsekas, D. P., 1999, *Nonlinear Programming*: Athena Scientific; 2nd edition.
- Billette, F., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th EAGE Conference & Exhibition, 13–16.
- Biondi, B., and A. Almomin, 2013, 200, *in* Tomographic full waveform inversion (TFWI) by extending the velocity model along the time-lag axis: 1031–1036.
- Chambolle, A., and T. Pock, 2011, A first-order primal-dual algorithm for convex problems with applications to imaging: *Journal of Mathematical Imaging and Vision*.
- Chung, E. T., T. F. Chan, and X.-C. Tai, 2005, Electrical impedance tomography using level set representation and total variational regularization: *Journal of Computational Physics*, **205**, 357–372.
- Esser, E., Y. Lou, and J. Xin, 2013, A Method for Finding Structured Sparse Solutions to Nonnegative Least Squares Problems with Applications: *SIAM Journal on Imaging Sciences*, **6**, 2010–2046.
- Esser, E., T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2014, A scaled gradient projection method for total variation regularized full waveform inversion: Technical report.
- Esser, E., X. Zhang, and T. F. Chan, 2010, A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science: *SIAM Journal on Imaging Sciences*, **3**, 1015–1046.
- Guo, Z., M. V. de Hoop, et al., 2013, Shape optimization and level set method in full waveform inversion with 3d body reconstruction: 2013 SEG Annual Meeting.
- Haber, E., and U. M. Ascher, 2001, Preconditioned all-at-once methods for large, sparse parameter estimation problems: *Inverse Problems*, **17**, 1847–1864.
- Haber, E., M. Chung, and F. Herrmann, 2012, An Effective Method for Parameter Estimation with PDE Constraints with Multiple Right-Hand Sides: *SIAM Journal on Optimization*, **22**, 739–757.
- He, B., and X. Yuan, 2012, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective: *SIAM Journal on Imaging Sciences*, 1–35.
- Herrmann, F. J., I. Hanlon, R. Kumar, T. van Leeuwen, X. Li, B. Smithyman, H. Wason, A. J. Calvert, M. Javanmehri, and E. T. Takougang, 2013, Frugal full-waveform inversion: From theory to a practical algorithm: *The Leading Edge*, **32**, 1082–1092.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, WCC177–WCC188.
- 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.
- Peters, B., F. J. Herrmann, and T. van Leeuwen, 2014, Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion: Presented at the EAGE.
- Rudin, L., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: *Physica D: Nonlinear Phenomena*, **60**, 259–268.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- van den Berg, E., and M. P. Friedlander, 2011, Sparse Optimization with Least-Squares Constraints: *SIAM Journal on Optimization*, **21**, 1201–1229.
- van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: *International Journal of Geophysics*, **2011**. (Article ID: 689041, 18pages).
- van Leeuwen, T., and F. J. Herrmann, 2013a, Mitigating local minima in full-waveform inversion by expanding the search space: *Geophysical Journal International*, **195**, 661–667.
- , 2013b, A penalty method for PDE-constrained optimization: Technical report.
- van Leeuwen, T., F. J. Herrmann, and B. Peters, 2014, A new take on FWI: wavefield reconstruction inversion: Presented at the EAGE.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC1–WCC26.
- Warner, M., and L. Guasch, 2014, 207, *in* Adaptive waveform inversion: Theory: 1089–1093.
- Zhang, X., M. Burger, and S. Osher, 2010, A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration: *Journal of Scientific Computing*, **46**, 20–46.
- Zhu, M., and T. Chan, 2008, An Efficient Primal-Dual Hybrid Gradient Algorithm For Total Variation Image Restoration: *UCLA CAM Report [08-34]*, 1–29.