# Fast randomized full-waveform inversion with compressive sensing

*Xiang Li, Aleksandr Y. Aravkin, Tristan van Leeuwen, and Felix J. Herrmann*

## ABSTRACT

Wave-equation based seismic inversion can be formulated as a nonlinear inverse problem where the medium parameters are obtained via minimization of a least-squares misfit functional. The demand for higher resolution models in more geologically complex areas drives the need to develop techniques that explore the special structure of full-waveform inversion to reduce the computational burden and to regularize the inverse problem. We meet these goals by using ideas from compressive sensing and stochastic optimization to design a novel Gauss-Newton method, where the updates are computed from random subsets of the data via curvelet-domain sparsity promotion. Application of this idea to a realistic synthetic shows improved results compared to quasi-Newton methods, which require passes through all data. Two different subset sampling strategies are considered: randomized source encoding, and drawing sequential shots firing at random source locations from marine data with missing near and far offsets. In both cases, we obtain excellent inversion results compared to conventional methods at reduced computational costs.

## MOTIVATION

Full-waveform inversion (FWI) can be formulated as the separable parameter-estimation problem

$$\underset{\mathbf{m}}{\text{minimize}}\ \Phi(\mathbf{m}) := \left\{ \frac{1}{2K} \sum_{i=1}^{K} \|\mathbf{d}_i - \mathcal{F}_i[\mathbf{m}, \mathbf{q}_i]\|_2^2 = \frac{1}{2} \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}]\|_F^2 \right\}, \qquad (1)$$

with $\mathbf{d}_i$ monochromatic shot records of the Earth response to monochromatic sources $\mathbf{q}_i$, $\mathcal{F}_i[\mathbf{m}, \mathbf{q}_i]$, $i = 1 \cdots K$ monochromatic nonlinear forward operators, and $K = N_f \cdot N_s$, with $N_f, N_s$ the number of frequencies and sources, respectively. In the acoustic constant-density case, this operator is parameterized by the unknown velocity model $\mathbf{m}$ and involves the inversion of a large system of linear equations.

Solving for the velocity model is challenging for several reasons. First, the solution is non-unique due to "cycle skipping". This phenomenon gives rise to local minima in the objective function. Second, the data are incomplete because low frequencies and

certain offsets are missing. Third, iterative methods for equation 1 are prohibitively expensive, since they require many PDE solves.

To address these issues, we use the following properties of FWI (cf. equation 1):

- linearity with respect to the sources that gives equation 1 its separable structure
- transform-domain sparsity on the updates, which allows us to fill in the null space of the Hessian, to remove the source crosstalk and to restore the amplitudes
- convex composite structure of $\Phi(\mathbf{m})$—it is a composition of the convex $\ell_2$-norm with the smooth $\boldsymbol{\mathcal{F}}$, and thus admits the standard Gauss-Newton (GN) algorithm as well as sparsity-promoting variants.

Our main contribution is to combine these properties with existing multiscale continuation methods (Bunks et al., 1995) that are widely employed to solve equation 1. The outcome is a formulation that allows us to reduce the number of PDE solves, to mitigate the effects of source crosstalk, to "fill in" the nullspace of the wave equation Hessian, and to speed up progress of the algorithm by using ideas from stochastic optimization (Haber et al., 2010; van Leeuwen et al., 2011a).

Our paper is organized as follows. First, we discuss dimensionality reduction techniques that allow algorithms to work with only a portion of the data at each iteration. We then exploit the convex-composite structure of equation 1 to design a reduced GN subproblem. Next, we modify this reduced subproblem using ideas from Compressive Sensing (CS)(Candès et al., 2006b; Donoho, 2006; Mallat, 2009), i.e. we remove source crosstalk by curvelet-domain sparsity promotion. Finally, we improve the results by drawing new encodings after solving each modified GN subproblem. We evaluate the performance of this algorithm on a realistic 2-D synthetic.

# METHODOLOGY

An important class of algorithms used to solve the FWI problem are the GN methods that involve the pseudo-inverse of the reduced Hessian given by the combined action of the Jacobian operator $\nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k;\mathbf{Q}]$ and its adjoint. Even though this method does not require explicit computation of the Hessian, each iteration for the GN subproblem requires the solution of $4K$ PDE's.

## Dimensionality reduction by randomized source superposition

To reduce the number of required PDE solves, we combine the sources and data into a smaller volume with $K' \ll K$ simultaneous experiments by replacing equation 1

with

$$\underset{\mathbf{m}}{\text{minimize}}\ \underline{\Phi}(\mathbf{m}) := \left\{ \frac{1}{2K'} \sum_{i=1}^{K'} \|\mathbf{D}\mathbf{w}_i - \mathcal{F}_i[\mathbf{m}, \mathbf{Q}\mathbf{w}_i]\|_2^2 = \frac{1}{2}\|\underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m}, \underline{\mathbf{Q}}]\|_F^2 \right\}, \quad (2)$$

where $\{\underline{\mathbf{D}}, \underline{\mathbf{Q}}\} \overset{\text{def}}{=} \{\mathbf{DW}, \mathbf{QW}\}$ (Moghaddam and Herrmann, 2010; Haber et al., 2010; van Leeuwen et al., 2011a). If we choose the random weights $\mathbf{w} = [w_1, \cdots, w_{N_s}]^T$ in $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_{K'}]$ such that the expectation $\mathrm{E}\{\mathbf{w}\mathbf{w}^T\}$ equals the identity matrix, we have $\mathrm{E}\{\underline{\Phi}(\mathbf{m})\} = \Phi(\mathbf{m})$. Because equation 2 can be interpreted as a sample-average, it represents an approximation of this expectation with an error that depends on $K'$. This error, which results in energy leaking towards source crosstalk, decreases for larger $K'$ and this property underlies the method of the sample-average approximation (SAA, Bertsekas and Tsitsiklis, 1996; Nemirovski et al., 2009).

Unfortunately the randomization in equation 2 defeats the purpose of making FWI faster because the error of SAA is known to decay slowly as a function of increasing $K'$. To overcome this issue, Krebs et al. (2009) proposed an approach reminiscent of the stochastic approximation where different weights are drawn for each gradient update. This method of stochastic-gradient descent is relatively well understood, and available convergence theories rely on specialized step lengths and averaging over previous model iterates (Bertsekas and Tsitsiklis, 1996; Haber et al., 2010; van Leeuwen et al., 2011a). Unfortunately, this averaging significantly slows down the progress of the algorithm. Therefore, we propose a solution method that relies on transform-domain sparsity promotion instead of averaging to reduce the error induced by the source crosstalk.

## Exploiting the convex-composite structure

The standard GN method exploits the convex-composite structure of equation 2 by linearizing the function inside the convex $\ell_2$-norm. We modify the standard GN subproblem by adding transform-domain sparsity promotion, and instead recover the updates by solving the following constrained (convex) optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}}\, \frac{1}{2}\|\underline{\delta\mathbf{D}} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}; \underline{\mathbf{Q}}]\mathbf{S}^H\mathbf{x}\|_F^2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau, \quad (3)$$

where $\underline{\delta\mathbf{D}} = \underline{\mathbf{D}} - \boldsymbol{\mathcal{F}}[\mathbf{m}]\underline{\mathbf{Q}}$, $\mathbf{S}^H$ is the inverse of the sparsifying transform determining the model update $\delta\mathbf{m} = \mathbf{S}^H\mathbf{x}$, and $\mathbf{x}$ is a vector of transform coefficients. The constraint enforces the $\ell_1$-norm of $\mathbf{x}$ to be smaller then some constant $\tau$.

## Dimensionality-reduction and compressive sensing

As long as signals exhibit structure—e.g., transform-domain compressibility where the signal's energy is concentrated in a few large coefficients—CS theory tells us that we

can recover these signals from severe undersampling by solving a sparsity-promoting program.

Unfortunately, this recovery does not hold for arbitrary subsamplings. Instead, CS prescribes sampling matrices that roughly behave like matrices with Gaussian entries. In that case, subsampling artifacts are shaped into white Gaussian "noise" and the sparsity-promoting recovery separates the signal from these "noisy" interferences.

The dimensionality reduction outlined in the previous section follows these guidelines because we typically draw the random weights from a Gaussian distribution. Because our definition of the "CS sampling matrix" includes the wave-equation Jacobian there is an important difference between CS and our setting. We need to invert a tall system of equations to obtain the GN updates, so for CS-type of arguments to hold, we need $\nabla\boldsymbol{\mathcal{F}}^H[\mathbf{Q}]\nabla\boldsymbol{\mathcal{F}}[\mathbf{Q}]$ to be near unitary with incoherent off-diagonals. Since this property is already true for $\mathbf{W}\mathbf{W}^H$, it would then hold for $\nabla\boldsymbol{\mathcal{F}}^H[\mathbf{Q}\mathbf{W}]\nabla\boldsymbol{\mathcal{F}}[\mathbf{Q}\mathbf{W}]$.

For high frequencies, this argument holds because the wave-equation Hessian acts locally as a directional filter as long as we are in the range of the Jacobian (Herrmann et al., 2008). This means that the dimensionality reduction keeps the problem in the regime where the insights from CS are applicable (Herrmann and Li, 2011, 2012). Consequently, it is useful to modify dimensionality-reduced GN subproblems to incorporate sparsity-promotion.

## MODIFIED GAUSS-NEWTON

To arrive at a practical and fast GN formulation, we need to address the following issues: (i) convergence guarantees, (ii) selection of the proper sparsifying transform, (iii) selection of the one-norm solver and sparsity levels, and (iv) data-volume reduction to decrease the number of PDE solves. We discuss each of these issues by considering the pseudo code of Algorithm 1 in detail.

**Proof of convergence [lines 2–9, Algorithm 1]:** The convergence algorithm for the standard GN method is well known (cf. Burke, 1990), and can be shown to apply to our modified algorithm as long as the sparsity levels $\tau_k$ remain bounded and the weights $\mathbf{W}$ remain the same for each iteration—i.e., $\mathbf{W}^k = \mathbf{W}$ for all $k$ (Herrmann et al., 2011).

**Sparsifying transform [lines 6–7, Algorithm 1]:** Selection of the appropriate sparsifying transform for the updates is important for two reasons. First, transform-domain sparsity leads to a concentration of the update's energy into a few large transform-domain coefficients. Second, the wave-equation reduced Hessian—also known as the demigration-migration operator—has a null space and requires regularization to stabilize its inversion. Here, transform-domain sparsity-promotion serves

as a prior that fills in the nullspace (see e.g., Daubechies et al. (2005) or chapter 11, Mallat (2009)).

To guarantee a fast decay for the magnitude-sorted transform coefficients on the updates, we require the transform to detect wavefronts, possibly with conflicting dips, and to be nearly *invariant* under wave propagation. This allows us to sparsely represent both the medium perturbations and the updates even in situations where the current model iterate is far from the true model. Finally, to avoid non-physical artifacts at the boundaries of the model, we use a mirror extended discrete curvelet transform (Candès et al., 2006a; Demanet and Ying, 2007), which decomposes the updates with respect to a collection of multiscale and multidirectional *localized plane waves*. We denote the 2D mirror extended curvelet transform by the symbol $\mathbf{C}_2$.

**One-norm solver and relaxation [lines 5-6, Algorithm 1]** An essential component of our algorithm requires solution of sparsity-inducing GN subproblems (cf. equation 3). We solve these subproblems with a spectral-gradient method (SPG, see e.g. van den Berg and Friedlander, 2008).

Because solutions of the GN subproblems depend on the sparsity level $\tau$—i.e., small $\tau$'s lead to sparse solutions with a large residue—we need to carefully select the $\tau^k$ for each GN subproblem. For this purpose, we follow van den Berg and Friedlander (2008); Hennenfent et al. (2008) and introduce the function $v(\tau)$, which parametrizes the $\ell_2$-norm of the optimal residual in equation 3 as a function of the sparsity level $\tau$. For each GN subproblem, this curve is decreasing, convex, and smooth (van den Berg and Friedlander, 2008), which allows us to obtain reasonable values of $\tau^k$ for each subproblem. The energy of the residual depends on the current iterate $\mathbf{m}^k$ and on the solution of the GN subproblem, which is unknown (van den Doel and Ascher, 2012). However, because $\delta\mathbf{m}^k$ is a descent direction (Burke, 1990), this residual will go down with each GN solve. We calculate the $\tau^k$ using the first Newton iteration for solving $v_k(\tau) = 0$, i.e. $\tau_k = -v_k(0)/v_k'(0)$, with $v'$ given by van den Berg and Friedlander (2008), Thm. 2.1. This yields

$$\tau^k = \frac{\|\underline{\delta\mathbf{D}}^k\|_2}{\|\mathbf{C}_2\nabla\boldsymbol{\mathcal{F}}^H[\mathbf{m}^k;\underline{\mathbf{Q}}]\underline{\delta\mathbf{D}}^k\|_\infty}. \tag{4}$$

**Algorithm speed-up using stochastic optimization [lines 3 and 7, Algorithm 1]** While the proposed dimensionality reduction technique allows us to solve the GN subproblems with a reasonable computational effort, the overall costs of the inversion remains prohibitively expensive. Therefore, we rely on insights from stochastic optimization (Haber et al., 2010; van Leeuwen et al., 2011a) by drawing independent $\mathbf{W}$'s for each GN subproblem. This extends our work on the linear case of least-squares migration, where these renewals led to a significant improvement in the convergence (Herrmann and Li, 2011, 2012). It is not clear how to select the steplength $\gamma^k$ when redrawing $\mathbf{W}^k$ at every iteration. Motivated by stochastic-

gradient descent (Bertsekas and Tsitsiklis, 1996), we use a fixed steplength $\gamma_k = 1$, which works well in practice.

## THE BG COMPASS MODEL

To test our inversion algorithm in a realistic setting, we consider a synthetic velocity model with a large degree of variability constrained by well data. We use this model to generate data with a 12 Hz Ricker wavelet. We use a smooth starting model without lateral information (Fig. 1(b)) and we start the inversion at 2.9 Hz.

All simulations are carried out with 350 shot and 701 receiver positions sampled at 20m and 10m intervals, respectively, yielding a maximum offset of 7km. We simulate wavefields by solving the 9-point discretized Helmholtz system (Jo et al., 1996) with a direct factorization method. To avoid local minima and to improve convergence, the inversions are carried out sequentially in 10 overlapping frequency bands on the interval $2.9 - 22.5$Hz (Bunks et al., 1995), each using 7 different randomly selected simultaneous shots and 10 selected frequencies. We use ten GN iterations for each frequency band. For each GN subproblem, we use roughly 20 iterations of SPG. Unfortunately, the computational costs of GN with all data are prohibitive. Instead, we compare our method to a limited-memory quasi-Newton method (l-BFGS, Nocedal and Wright, 2006). This benchmark uses approximately twice the number of PDE solves per iteration compared to the new algorithm. The results for l-BFGS and our algorithm with and without independent redraws for the $\mathbf{W}$'s are included in Figure 2. We make the following observations from these results. First, the results from the modified GN have higher resolution and recover the different layers more accurately. This is remarkable, and demonstrates the ability of curvelet sparsity promotion on the updates to fill in the null space of the reduced Hessian. Second, the renewals remove visual artifacts in the estimated velocities. For an example, see the area under the main unconformity between lateral distances of $1000 - 3000$ m in Figure 2(b) and

**Result**: Output estimate for the model $\mathbf{m}$

$k \longleftarrow 0$; $\mathbf{m}^k \longleftarrow \mathbf{m}_0$ ; // initial model

**while** not converged **do**

  $\{\underline{\mathbf{D}}^k, \underline{\mathbf{Q}}^k\} \longleftarrow \{\mathbf{D}\mathbf{W}^k, \mathbf{Q}\mathbf{W}^k\}$ with $\mathbf{W}^k \in N(0,1)$ ; // indep. draw.

  $\underline{\delta\mathbf{D}}^k \longleftarrow \underline{\mathbf{D}}^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]$; // residual

  $\tau^k \longleftarrow \|\underline{\delta\mathbf{D}}^k\|_F \,/\, \|\mathbf{C}_2 \nabla \boldsymbol{\mathcal{F}}^*[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\underline{\delta\mathbf{D}}^k\|_\infty$; // update $\tau$

  $\delta\mathbf{x} \longleftarrow \begin{cases} \arg\min_{\delta\mathbf{x}} \frac{1}{2}\|\underline{\delta\mathbf{D}}^k - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\mathbf{C}_2^H \delta\mathbf{x}\|_F^2 \\ \text{subject to} \quad \|\delta\mathbf{x}\|_1 \le \tau^k \end{cases}$

  $\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \mathbf{C}_2^H \delta\mathbf{x}$ ; // update with line search

  $k \longleftarrow k + 1$;

**end**

**Algorithm 1**: Dimensionality-reduced Gauss Newton with sparsity promotion

2(c).

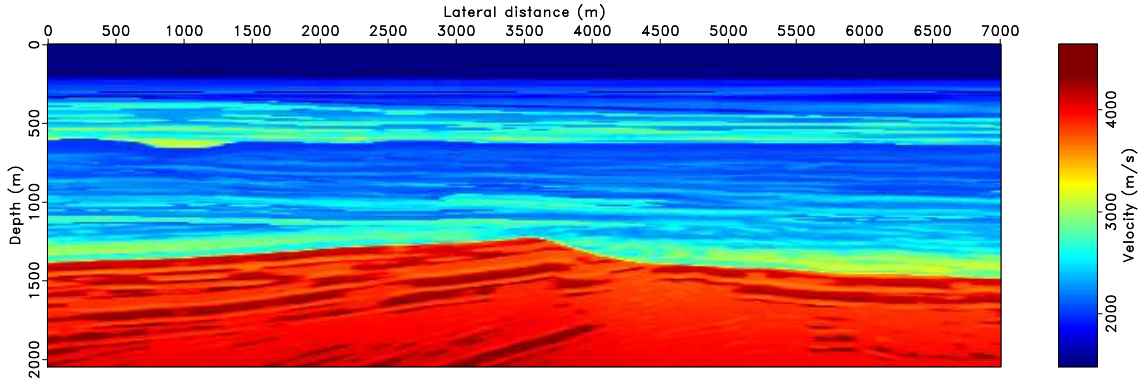
Even though the above example shows that excellent inversion results are attainable working with randomized superpositions, the dimensionality reduction relies on fixed-spread acquisition where each source sees the same receivers. Unfortunately, our reliance on this type of reduction technique limits the applicability of our approach to marine data where the sources and receivers both move and where near and far offsets are missing. This dependence can be avoided if we replace the Gaussian i.i.d. columns of the $\mathbf{W}$'s by randomly selected columns of the Dirac basis. This choice, which corresponds to selecting random subsets of sequential shots, allows dimensionailty reduction for marine data (van Leeuwen et al., 2011b; Aravkin et al., 2012), without relying on correlation/phase-based misfit functionals (Routh et al., 2011). To test the performance of this type of dimensionality reduction for a marine acquisition with offsets from $100 - 3000\,\mathrm{m}$, we rerun the above GN examples with and without renewals. The results are included in Figure 3 and clearly illustrate the importance of changing the random subsets of shots for the different GN updates. The results suggest that simultaneous source sampling is more efficient than using random sequential shots — note the slight loss in resolution and quality of the recovered discontinuities comparing Figures 2(b) and 3(a).

## DISCUSSION

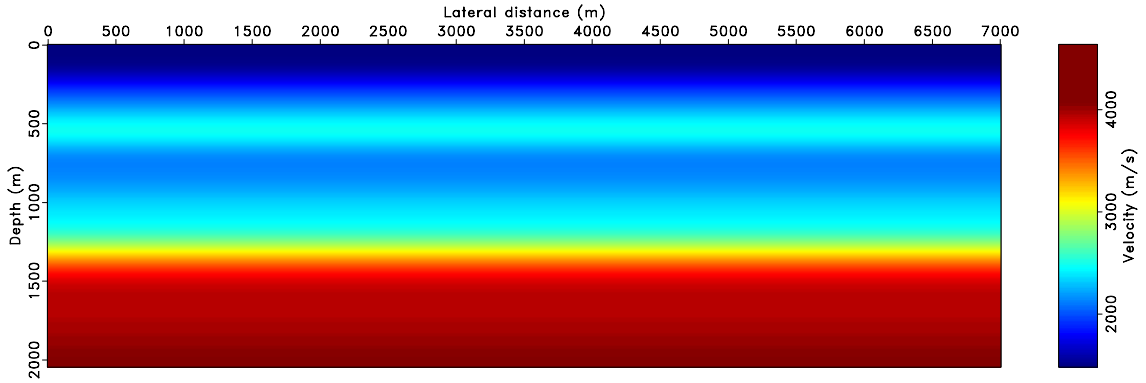
Efforts to control source artifacts related to fast inversions that act on subsets of the data rely mostly on averaging. In contrast, we employ sparsity-promotion on the model updates to remove the interference related to the dimensionality reduction. In that sense, our method is somewhat reminiscent of "gradient preconditioning" (see e.g. Fichtner, 2011) via smoothing but it differs because we leverage curvelet-domain sparsity promotion, which preserves wavefront-like features, removes source crosstalk, and does not need model-specific information such as local dips (Guitton et al., 2010). An alternative approach is to impose sparsity on the model itself, rather than the updates. This requires a different formulation and algorithm, and we leave this for future work.

## CONCLUSIONS

We introduced an efficient algorithm to solve the full-waveform inversion (FWI) problem by incorporating insights from convex optimization, stochastic optimization, and compressive sensing. By exploiting the convex-composite, multiscale, and separable structure of FWI, we modified the Gauss-Newton (GN) method to produce a new algorithm that permits us to consider GN updates as compressive-sensing experiments with dramatically reduced numbers of sources. The speedup depends on the type of wave simulator (time/frequency domain). Because we work with small subsets of data, our method also reduces the memory imprint and opens the possibility to store

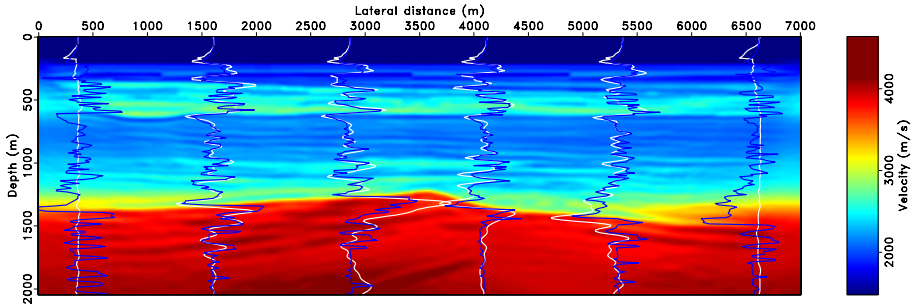



(a)

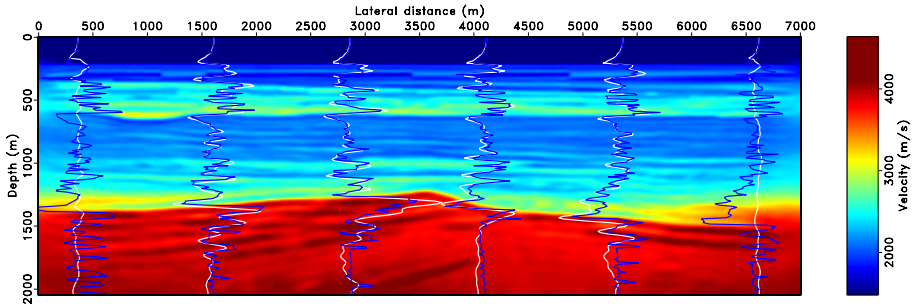



(b)

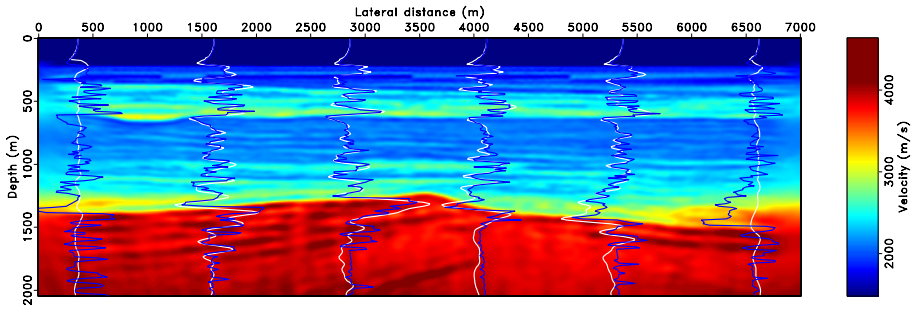
Figure 1: BG Compass model. **(a)** original model ($\mathbf{m}$), **(b)** initial model ($\mathbf{m}_0$) used to start FWI.




(a)

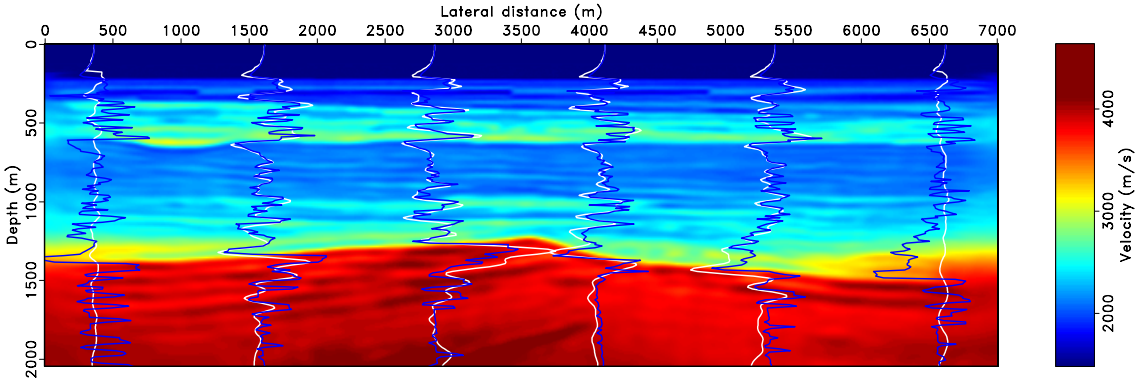



(b)

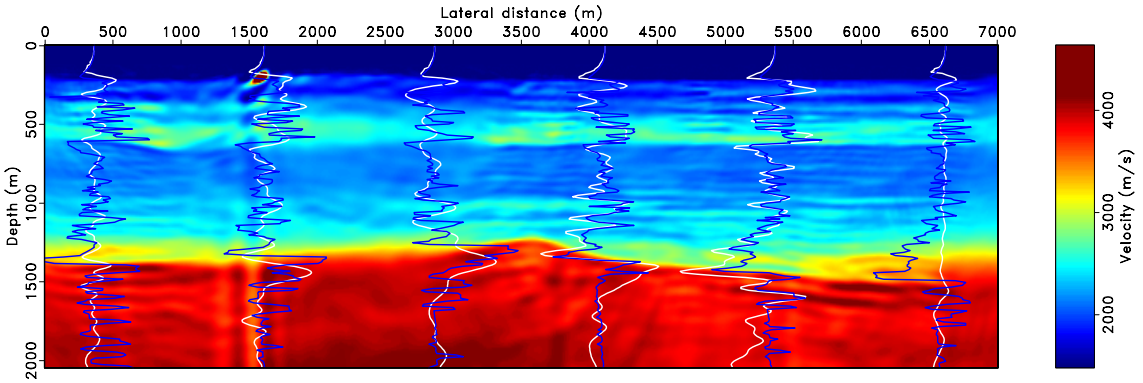



(c)

Figure 2: Full-waveform inversion results starting from 2.9Hz over 10 frequency bands. **(a)** Inverted result for all data using l-BFGS. **(b)** Inverted result with the modified GN method using 7 simultaneous shots and 10 frequencies. **(c)** the same as **(b)** but without renewals. The vertical traces show the differences between the true (blue)/reconstructed (white) model and the initial model.




(a)




(b)

Figure 3: Full-waveform inversion results starting from 2.9Hz over 10 frequency bands. **(a)** Inverted result with the modified GN method using 7 randomly selected sequential shots and 10 frequencies. **(b)** the same as **(a)** but without renewals.

wavefields instead of computing them on the fly. For fixed randomized subsets of data, we were able to establish convergence of our method. We also demonstrated a significant improvement of the inversion results obtained by selecting independent randomized subsets of the data for each GN update. While we argue that these renewals remove possible correlations between the subsampling and model iterates, a formal convergence proof of the optimization algorithm still needs to be established.

Application of our algorithm to a complex synthetic data set leads us to the following conclusions. First, dimensionality reduced GN with curvelet-domain sparsity promotion yields higher-quality inversion results than do quasi-Newton methods. Second, sparse recovery in combination with randomized dimensionality reduction allows us to speed FWI significantly by iterating on small subsets of the data only. Third, we were able to obtain inversion results from reduced experiments based on either randomized simultaneous sources or on randomized subsets of sequential sources. The latter has the advantage of being suitable for marine acquisition at the cost of a moderately inferior inversion result relative to fixed-spread acquisition. Finally, we find that renewals make a significant difference, in particular for dimensionality reduction based random subsets of sequential shots.

## ACKNOWLEDGMENTS

We are grateful to Charles Jones from BG for providing us with the BG Compass velocity model. We would also like to thank the authors of SPOT, SPGL1, and CurveLab. This work was supported by Grant Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (375142-08). The industrial sponsors of the Seismic Laboratory for Imaging and Modelling (SLIM): BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA., and WesternGeco are also gratefully acknowledged.

## REFERENCES

Aravkin, A. Y., X. Li, and F. J. Herrmann, 2012, Fast seismic imaging for marine data: IEEE International Conference on Acoustics, Speech, and Signal Processing.

Bertsekas, D., and J. Tsitsiklis, 1996, Neuro-Dynamic programming: Athena Scientific.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Burke, J., 1990, Numerical optimization: University of Washington course notes.: http://www.math.washington.edu/∼burke/crs/516/index.html.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, no. 3, 861–899.

Candès, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Comm. Pure Appl. Math., **59**, 1207–1223.

Daubechies, I., M. Defrise, and C. de Mol, 2005, An iterative thresholding algorithm for linear inverse problems with a sparsity constrains: Comm. Pure Appl. Math., 1413–1457.

Demanet, L., and L. Ying, 2007, Curvelets and wave atoms for mirror-extended images: SPIE Conference Series, **6701**.

Donoho, D. L., 2006, Compressed sensing: IEEE Trans. Inform. Theory, **52**, 1289–1306.

Fichtner, A., 2011, Full Seismic Waveform Moldelling and Inversion: Springer.

Guitton, A., G. Ayeni, and G. Gonzales, 2010, A preconditioning scheme for full waveform inversion: SEG, Expanded Abstracts, **29**, 1008–1012.

Haber, E., M. Chung, and F. J. Herrmann, 2010, An effective method for parameter estimation with PDE constraints with multiple right hand sides: Technical Report TR-2010-04, UBC-Earth and Ocean Sciences Department.

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**, no. 4.

Herrmann, F. J., and X. Li, 2011, Efficient least-squares migration with sparsity promotion: EAGE Technical Program Expanded Abstracts.

———, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing.: Geophysical Prospecting, to appear.

Herrmann, F. J., X. Li, A. Y. Aravkin, and T. van Leeuwen, 2011, A modified, sparsity promoting, Gauss-Newton algorithm for seismic waveform inversion: Proceedings of the SPIE, **81380V**.

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.

Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator: Geophysics, **61**, 529–537.

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.

Mallat, S. G., 2009, A Wavelet Tour of Signal Processing: the Sparse Way: Academic Press.

Moghaddam, P. P., and F. J. Herrmann, 2010, Randomized full-waveform inversion: a dimenstionality-reduction approach: SEG, Expanded Abstracts, **29**, 977–982.

Nemirovski, A., A. Juditsky, G. Lan, and A. Shapiro, 2009, Robust stochastic approximation approach to stochastic programming: SIAM Journal on Optimization, **19**, 1574–1609.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer Verlag.

Routh, P., J. Krebs, S. Lazaratos, A. Baumstein, S. Lee, Y. H. Cha, I. Chikichev, N. Downey, D. Hinkley, and J. Anderson, 2011, Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data: SEG, Expanded Abstracts, **30**, 2433–2438.

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 den Doel, K., and U. Ascher, 2012, Adaptive and stochastic algorithms for EIT

and DC resistivity problems with piecewise constant solutions and many measurements: SIAM J. Sc. Comput. (to appear).

van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011a, Seismic waveform inversion by stochastic optimization: International Journal of Geophysics, article ID: 689041.

van Leeuwen, T., M. Schmidt, M. Friedlander, and F. Herrmann., 2011b, A hybrid stochastic-deterministic optimization method for waveform inversion: EAGE Technical Program Expanded Abstracts.