Randomized full-waveform inversion: a dimenstionality-reduction approach

Peyman P. Moghaddam and Felix J. Herrmann, University of British Columbia, Canada*

SUMMARY

Full-waveform inversion relies on the collection of large multi-experiment data volumes in combination with a sophisticated back-end to create high-fidelity inversion results. While improvements in acquisition and inversion have been extremely successful, the current trend of incessantly pushing for higher quality models in increasingly complicated regions of the Earth reveals fundamental shortcomings in our ability to handle increasing problem sizes numerically. Two main culprits can be identified. First, there is the so-called "curse of dimensionality" exemplified by Nyquist's sampling criterion, which puts disproportionate strain on current acquisition and processing systems as the size and desired resolution increases. Secondly, there is the recent "departure from Moore's law" that forces us to develop algorithms that are amenable to parallelization. In this paper, we discuss different strategies that address these issues via randomized dimensionality reduction.

INTRODUCTION

With the recent resurgence of full-waveform inversion (FWI), the costs of computing gradient and Newton updates have–aside from issues with non-uniqueness—become the major impediment withstanding successful application of this technology to industry-size data volumes. The cost of computing gradients depends on the size of the data volumes (number of sources and receivers) and on the discretization of the model. The challenge for Newton updates lies in the fact that the Hessian matrix is full and possibly indefinite (negative eigenvalues). Even though new preconditioning techniques have been developed for approximating the Hessian (Symes, 2008; Herrmann et al., 2009a), fast computations of the gradient remain illusive.

The computational cost for the gradient (read migration) is so high because it involves solutions of the wave equation on a large domain and for a large number of right-hand sides (sources). In 3D, this situation is exacerbated by the "curse of dimensionality", the large bandwidth of the Helmholtz discretization, and the large number of sources. Because of the increased bandwidth, it remains difficult to use direct solvers. This explains why time-domain finite differences are still the main "work horse" for FWI in industrial applications even though adjoint state methods rely on correlations between the source and residual wavefields, which require storage of the complete time history or other resource-demanding workarounds. Indirect Helmholtz solvers, on the other hand, address this issue and solve the wavefield for each frequency component independently. Until very recently these indirect solvers suffered from slow convergence and there are strong indications that this problem has been resolved (Erlangga and Nabben, 2008; Erlangga and Herrmann, 2008). However, like time-domain finite differences, implicit methods do not scale well with increasing numbers of sources. Therefore, FWI will remain very challenging for situations with very large numbers of sources.

In this paper, we will address this issue by means of dimensionality reduction, which comes in many different flavors including low-rank approximations (SVD) and transform-domain data compression with wavelets. However, these methods do not lend themselves particularly well for FWI.

For that reason, we chose a different randomized dimensionality reduction strategy, which exploits certain properties of random Gaussian matrices—i.e., matrices with *i.i.d.* Gaussian random entries and far fewer rows than columns. When applied to arbitrary signals, these matrices preserve most of the signal's energy and hence are capable of linearly encoding most of the signal's information in a much lower dimensional space. This profound idea has found a wide range of applications, solving problems ranging from measure concentration in pure mathematics (Johnson and Lindenstrauss, 1984), to fast matrix-vector multiplies in computer science (Drineas et al., 2007) to compressive sensing (CS Candès et al., 2006; Donoho, 2006) in signal processing, and to randomized Kaczmarz (Strohmer and Vershynin, 2009) in numerical linear algebra. In each of these approaches, the dimensionality is reduced by selecting only a subset of rows from a square Gaussian matrix. Inner products with these rows correspond to taking subsets of measurements.

For FWI, this approach corresponds to replacing many sequential sources by much fewer phase-encoded simultaneous sources (Romero et al., 2000; Herrmann et al., 2009b; Neelamani et al., 2008; Krebs et al., 2009a,a), possibly in combination with the removal of random subsets of angular frequencies (Sirgue and Pratt, 2004; Mulder and Plessix, 2004; Lin et al., 2008; Herrmann et al., 2009b). Judged by the recent resurgence of simultaneous sources in seismic data acquisition (Beasley, 2008; Berkhout, 2008; Krebs et al., 2009a; Herrmann et al., 2009b; Herrmann, 2009), this type of dimensionality reduction is not only practical for the reduction of the computational complexity of FWI (by limiting the number of sources and frequencies) but this method is also proving its applicability during the collection of field data (Krohn and Neelamani, 2008).

The advantage of dimensionality reduction comes, however, at the price of energy leakage in the form of source cross talk and aliases. At this point, the advantage of using Gaussian-like matrices, which include random phase encoding and randomly restricted Fourier transforms, comes into play because these random matrices turn these artifacts into harmless incoherent Gaussian noise. The larger the dimensionality reduction, the more leakage and the higher the noise level. Depending on the application, these noisy artifacts and amplitude effects can be mitigated and this can lead to significant speedups in wavefield simulations (Neelamani et al., 2008; Herrmann et al., 2009b) and in the computation of gradients (Romero et al., 2000; Krebs et al., 2009b; Herrmann and Li, 2010) and imaging (Herrmann and Li, 2010).

The main contribution of this paper is threefold. First, we show that FWI based on the Helmholtz equation has the advantage

that it allows us to form *super shots*. These *super shots* act as dimensionality reduced frequency-restricted right-hand-sides, which correspond to simultaneous sources made of superpositions of phase-encoded sources each with a different random set of frequencies. Second, we show that the artifacts related to randomized dimensionality reduction are incoherent noise like with a noise level that depends on the degree of dimensionality reduction. We show that this noise level decreases (as expected) for increasing numbers of *super shots* and increasing number of frequencies. Third, we show that dimensionality reduced FWI should be interpreted as an instance of stochastic optimization (Nemirovski et al., 2009). According to this theory, the averaged solution of the collection of all possible randomized FWI experiments with one (or more) *super shots* can be computed efficiently by drawing a new random experiment (e.g., a new single *super shot*) for each gradient update. We show that this observation can lead to remarkable speed up, confirming a recent observation reported by Krebs et al. (2009b). Because this method hinges on a delicate argument, we explore different scenarios where we combine the *sample average approximation*, which relies on different experiments for each iteration, with Monte-Carlo averaging amongst different instances of this approximation. This opens interesting perspectives towards massive parallelization of this algorithm while adding robustness. We illustrate our discussion by means of carefully designed examples.

THEORY

Full-waveform inversion (FWI) involves the solution of the following multi-experiment unconstrained optimization problem:

$$\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{P}-\mathscr{F}[\mathbf{m},\mathbf{Q}]\|_{2,2}^2, \tag{1}$$

where each column of $\mathbf{P}$ contains the observed data for one shot and all frequencies. The nonlinear operator $\mathscr{F}[\mathbf{m},\mathbf{Q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$ simulates data by solving the Helmholtz system $\mathbf{H}[\mathbf{m}]$ for all sources in the columns of $\mathbf{Q}$. To obtain observed data at the surface, we restrict the simulated data to the receiver positions with the detection operator $\mathbf{D}$. For simplicity, we assume that receivers for all sources are co-located. This underlying assumption is suitable specifically for land surveys. We also neglect surface-related multiples by using an absorbing boundary condition at the surface. Each iteration of FWI (see table 1) involves the computation of the gradient and the (approximate) inverse of the reduced Hessian ($\mathbf{H}_{\text{red}}^{\dagger}$) (Pratt et al., 1998; Sirgue and Pratt, 2004; Plessix, 2006). The model updates are weighted by α obtained by line search.

Each iteration of FWI is roughly equivalent to carrying out a full pre-stack least-squares migration. Given the size of current data volumes this proves prohibitive despite enormous increases in computational and memory storage capabilities. For this reason, FWI still remains illusive for realistically sized problems in particular in the light of continued push for higher resolution inversion for larger and larger models. We propose to address this problem created by this data deluge by leveraging the following four (recent) developments:

Result: Output estimate for the model $\mathbf{m}$
$\mathbf{m} \longleftarrow \mathbf{m}_0$; // initial model
while $\|\mathbf{P}-\mathscr{F}[\mathbf{m},\mathbf{Q}]\|_{2,2}^2 \geq \varepsilon$ **do**
  $\mathbf{g} \longleftarrow \mathbf{J}^*[\mathbf{m},\mathbf{Q}](\mathbf{P}-\mathscr{F}[\mathbf{m},\mathbf{Q}])$; // gradient
  $\delta\mathbf{m} \longleftarrow \mathbf{H}_{\text{red}}^{\dagger}\mathbf{g}$; // Gauss-Newton-Krylov update with CG
  $\mathbf{m} \longleftarrow \mathbf{m}+\alpha\delta\mathbf{m}$; // model update with line search
end

Algorithm 1: FWI by Gauss-Newton-Krylov

Preconditioners for Helmholtz: The success of FWI hinges on our ability to solve large scale 3D Helmholtz systems. Despite recent developments in explicit solvers (see e.g., Virieux et al., 2009, for an overview), the increased bandwidth in 3D poses a fundamental problem. Recent work by Erlangga and Nabben (2007); Erlangga and Herrmann (2008), has shown that preconditioners made of a combination of Laplacian shifts and deflation yield implicit solvers with a numerical complexity similar to that of time-domain finite differences. Unfortunately, this development only solves part of the puzzle because its computational costs increases linearly with the number of right-hand sides (sources). Therefore, we propose to use dimensionality reduction to address this important issue.

Randomized Kaczmarz: FWI involves the inversion of overdetermined systems of equations (the data space is typically much larger then the model space). In the numerical linear algebra community this is a well-known problem, which involves the solution of

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}, \tag{2}$$

with $\mathbf{A} \in \mathbb{R}^{m\times b}$ with $m \gg n$. Numerous solution methods have been developed for this problem ranging from direct methods to matrix-free indirect methods that only involve the action of $\mathbf{A}$ and its adjoint $\mathbf{A}^*$. All these methods, however, become prohibitively expensive when $\mathbf{A}$ is full. To address this issue, Kaczmarz (1937) proposes to calculate gradient updates that involve single rows of a properly normalized matrix $\mathbf{A}$ only—i.e, we iterate

$$\mathbf{x}^{k+1} = \mathbf{x}^k + (b_j - \mathbf{a}_i\mathbf{x}^k)\mathbf{a}_j^*, \tag{3}$$

with $\mathbf{a}_j$ the j^{th} row of the matrix $\mathbf{A}$. The algorithm proceeds by iterating sequentially over all rows of $\mathbf{A}$. After m iterations the algorithms restarts with the first row. This approach has been used successfully by Natterer (2001) to solve tomographical problems. In that application, the iterations run over different sources part of a multi-source experiment. Recently, Strohmer and Vershynin (2009) prove exponential convergence (read faster then Krylov methods such as conjugate gradients) for this algorithm for full random matrices. For our application, this suggests fast solvers for linear systems that consist of blocks for each randomly phase encoded simultaneous source experiment (Krebs et al., 2009b; Neelamani et al., 2008; Herrmann et al., 2009b). In that case, each iteration only involves one simultaneous source experiment at the time and this leads to significant reductions in memory use and in costs for matrix-vector multiplies.

Randomized dimensionality reduction: Gradient updates are extremely costly because each iteration requires the solution of the forward and time-reversed (adjoint) Helmholtz systems for each of the n_f frequencies and for each of the n_s sources.

Moreover, improvement in convergence of FWI require expensive inversions of the reduced Hessian (see e.g. Erlangga and Herrmann, 2009, and the references therein).

We address this problem by considering a number of randomized experiments (N) each of which consist of subsets of phase-encoded simultaneous sources with randomly selected frequencies. After random phase encoding, sequential sources are combined into one or more *super shots*. Depending on the computer architecture and the type of solver these randomized experiments may consist of a single *super shot*, or several *super shots* with the same frequencies, or different simultaneous shots with different subsets of frequencies each. Each randomized experiment is formed by applying

$$(\mathbf{RM})_i := (\mathbf{R}_i^{\Sigma}\mathbf{F}_{\Sigma}^{*}\,\mathrm{diag}(e^{\hat{\imath}\boldsymbol{\theta}_i})\mathbf{F}_{\Sigma}\otimes\mathbf{I}\otimes\mathbf{R}_i^{\Omega}\mathbf{F}_{\Omega}), \quad i=1\cdots N \quad (4)$$

to the source function $\mathbf{Q}$ and involves the solution of the Helmholtz system for a small subset of sources $n_s' \ll n_s$ and frequencies $n_f' \ll n_f$. For each experiment, these frequencies are randomly selected by the restriction matrix $\mathbf{R}_i^{\Omega}$. The simultaneous *super shots* themselves are obtained by phase encoding with random phases $\boldsymbol{\theta}_{1\cdots n_s'} = \mathrm{Uniform}([0,2\pi])$, followed by the selection of one or more simultaneous sources with the restriction matrix $\mathbf{R}_i^{\Sigma}$. Because the cost of evaluating the gradient and Newton updates depends on the number of sources (n_s) and frequencies (n_f), the computational costs are reduced as long as $N\times n_s' < n_s$ and $N\times n_f' < n_f$ (Herrmann et al., 2009b; Neelamani et al., 2008; Herrmann and Li, 2010; Krebs et al., 2009b). However, this speed up comes at a price, namely the smaller N, n_s', n_f' the more prominent the artifacts due to source crosstalk and aliasing. Due to the random construction of the dimensionality matrix in Equation 4, these artifacts are incoherent and noise like with a noise level that depends on the degree of dimensionality reduction. We used this property explicitly in Herrmann et al. (2009b); Herrmann and Li (2010) to reduce the costs of wavefield simulation and migration by using sparse inversion to remove the artifacts and to restore the amplitudes. In this paper, we study the behavior of FWI for different scenarios that involve these *super shots*.

Stochastic optimization: FWI with randomized dimensionality reduction via phase-encoded *super shots* should really be considered as a particular instance of a stochastic optimization problem where the problem in Equation 1 is replaced by

$$\min_{\mathbf{m}} \mathbb{E}\{\Phi(\mathbf{m},\xi)\} \quad (5)$$

with $\Phi(\mathbf{m},\xi) = \frac{1}{2}\|\mathbf{p}(\xi) - \mathscr{F}[\mathbf{m},\mathbf{q}(\xi)]\|_2^2$ and where $\mathbb{E}$ denotes stochastic expectation and where ξ identifies a single instance of a random simultaneous experiment defined by Equation 4—i.e., $\mathbf{p}(\xi) := (\mathbf{RM}\mathrm{vec}(\mathbf{P}))_{\xi}$ and $\mathbf{q}(\xi) := (\mathbf{RM}\mathrm{vec}(\mathbf{Q}))_{\xi}$. Aside from providing a general framework for (convex) optimization problems that goes well beyond the inversion of a overdetermined linear system (as in Kaczmarz), the main contribution of stochastic optimization (Robbins and Monro, 1951) is replacing solution strategies for Equation 5 based on exhaustive (Monte-Carlo) ensemble averaging by solution methods that for each update involve a different experiment. In our application, this corresponds to FWI where for each gradient update a new experiment is drawn, e.g., by selecting a new **RM** (see Algorithm 2). This second approach, known as the *sample average approximation* is distinctively different from selecting different experiments and solving them to completion, followed by ensemble averaging, which is known as the *stochastic approximation*. It is clear, that this latter method is significantly more expensive because it involves the solution of multiple FWI problems whereas the first approach involves iterations on a single experiment only. Recent work by Krebs et al. (2009b) confirms this observation and shows that good results can be obtained by iterating with a single simultaneous shot that is replaced after each iteration. Even though these results, as confirmed below, are very encouraging extreme care should be taken how to choose the step length in the updates and how to build in robustness with respect to noise. This suggests a possible combination of the *sample average* and *stochastic* approximation methods.

Result: Estimate for the model $\widetilde{\mathbf{m}}$
$\mathbf{m} \longleftarrow \mathbf{m}_0$; // initial model
$\{\Phi(\mathbf{m}), \nabla_{\mathbf{m}}\Phi(\mathbf{m})\} \longleftarrow \mathrm{Draw}(\mathbf{P},\mathbf{Q})$; // Setup experiment
while $\|\nabla_{\mathbf{m}}\Phi(\mathbf{m})\| \geq \varepsilon$ **do**
  $\mathbf{m} \longleftarrow \mathrm{Solve}(\Phi(\mathbf{m}, \nabla_{\mathbf{m}}\Phi(\mathbf{m}))$; // Warm start solve
  $\{\Phi(\mathbf{m}), \nabla_{\mathbf{m}}\Phi(\mathbf{m})\} \longleftarrow \mathrm{Draw}(\mathbf{P},\mathbf{Q})$; // New experiment
end

Algorithm 2: FWI with stochastic optimization

Randomized l-BFGS:
To test the above randomized solution strategies, let us consider vanilla gradient- and quasi-Newton based solvers that include a line search. At the k^{th} iteration, these solvers compute the following updates $\mathbf{m}^{k+1} := \mathbf{m}^k + \alpha_k\mathbf{g}^k$ and $\mathbf{m}^{k+1} := \mathbf{m}^k + \alpha_k\mathbf{H}^k\mathbf{g}^k$ with $\mathbf{H}^k$ and approximate inverse of the reduced Hessian. In this example, we use l-BFGS (Nocedal and Wright, 1999) to approximate this inverse.

Stylized experiments: We are interested in the following questions. First, how does inversion quality depend on the degree of dimensionality reduction? Second, how does FWI behave under the *sample average* and the *stochastic* approximation? Third, are the subsampling related artifacts Gaussian so that we can expect ensemble averaging to reduce the error?

Experimental setup: To answer these questions we run our tests on a part of Marmoussi model as shown in Figure 1(a). The model is 5000(m) long and 3000(m) deep with 10(m) grid spacing. We use the frequency-domain Helmholtz solver (Erlangga and Herrmann, 2008). We generate synthetic data with 113 shots with offset of $250-4749(m)$ and $40(m)$ shot spacing, 249 receivers with offset of $20-4980(m)$ and 20(m) receiver spacing. The source wavelet for all the shots is the Ricker wavelet with $10Hz$ central frequency. Total recording time is $3.6(s)$ with time sampling interval of $0.9(ms)$. As a reference for later comparison, we solve FWI with the initial model shown in Figure 1(b) for all shots with frequencies ranging from 4 to 33(Hz) with .5(Hz) sampling. The result after 18 iterations of the l-BFGS is plotted in Figure 2(a). The SNRs summarized in Table 1 are computed with respect to this reference result.

Observations: First, as expected the numbers in Table 1 confirm increasing recovery errors for increasing dimensionality reduction. For fixed sub-sampling ratios, however, we observe

improved results for decreasing frequency-to-shot ratios, which suggests that simultaneous shots contribute more to the solution. This observation is consistent with earlier observations (Herrmann et al., 2009b; Herrmann and Li, 2010). Second, the best result from this table gives a result that is difficult to distinguish from the reference image that took 88× longer to compute (cf. Figures 2(a) and 2(b)). This result has been obtained with randomized l-BFGS, which draws a new experiment for each iteration. Increasing the dimensionality reduction, leads to higher speedups but inferior inversion results. Third, we find that the artifacts are relatively incoherent , which opens the possibility for ensemble averaging of FWIs that are run independently. The SNRs for single-shot simultaneous source experiments are 27.1, 27.3 dB, respectively, yielding a SNR of their average of 27.5. This number is less then the SNR obtained by running the experiment with two simultaneous shots, which benefits more from the data yielding 27.8 dB.

Subsample ratio	0.0113	0.0028	0.0007
n'_f/n'_s	recovery error (dB)		
.25	**30.18**	**27.02**	**24.48**
1	**26.94**	**25.88**	**24.45**
4	**27.37**	**26.82**	**24.16**
Speed up (×)	88	352	1410

Table 1: Signal-to-noise ratios, $\text{SNR} = 20\log_{10}(\frac{\|\delta\mathbf{m}-\delta\widetilde{\mathbf{m}}\|_2}{\|\delta\mathbf{m}\|_2})$ for reconstructions with simultaneous-source experiment for different subsample and frequency-to-shot ratios.

DISCUSSION AND CONCLUSIONS

First of all, we do not claim in this paper that we solved the non-uniqueness problem that continues to plague full waveform inversion (FWI). Instead, we showed that significant improvements can be made towards the computational efficiency of FWI, which can be considered as much as an impediment to its success. We confirmed findings in the literature that report significant improvement of FWI for computations of the gradient that are based on different random realizations of simultaneous source experiments for each gradient update. This result is profound and has close connections with several different fields in mathematics. Even though, a rigorous analysis of the observed phenomenon is still largely lacking, the theory of Compressive Sensing offers interesting insights. First, it predicts that FWI improves for decreasing dimensionality reduction. Second, phase encoding and random frequency sampling lead to manageable incoherent artifacts. Third, using different randomized dimensionality reductions extract additional information, which improves convergence of the inversion. Fourth, dimensionality reduction allows us to work on much smaller subproblems.

ACKNOWLEDGMENTS

We wish to thank Eldad Haber, Yogi Erlangga, and Tim Lin for insightful discussions on this topic. This work was financed through the SINBAD II project with support from BG, BP, Chevron, Petrobras, and WesternGeco and was partly supported by a NSERC Discovery Grant (22R81254) and CRD Grant DNOISE (334810-05).

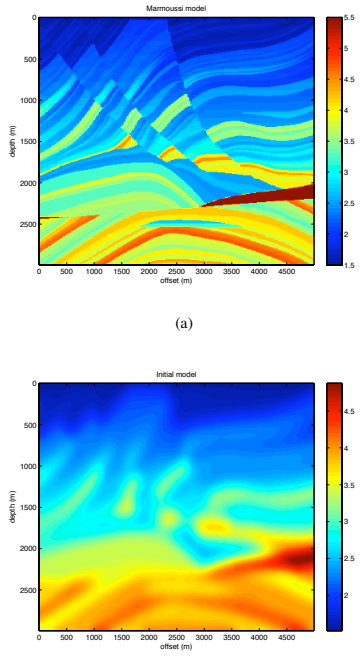

Figure 1: **(a)** Marmoussi Model. **(b)** Initial model.

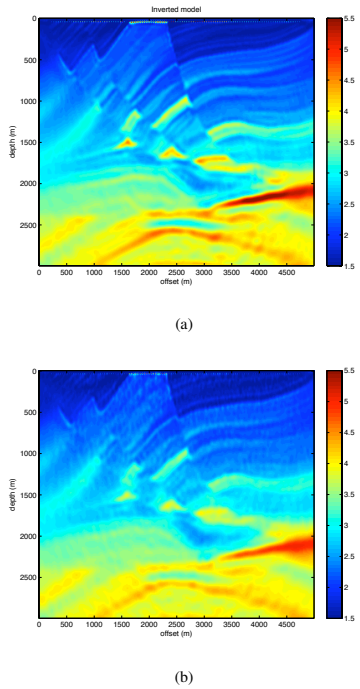

Figure 2: **(a)** Reference FWI result obtained with 18 iteration of l-BFGS, 113 sequential shots, 58 frequency components selected from 4 to 33 Hz. **(b)** Best recovery (largest SNR) from simulation results summarized in Table 1, corresponding to a speed up of 88×.

REFERENCES

Beasley, C. J., 2008, A new look at marine simultaneous sources: The Leading Edge, **27**, 914–917.

Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

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

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

Drineas, P., R. Kannan, and M. Mahoney, 2007, Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication: SIAM Journal on Computing, **36**, 132–157.

Erlangga, Y. A., and F. J. Herrmann, 2008, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion: SEG Technical Program Expanded Abstracts, SEG, 1957–1960.

———, 2009, Seismic waveform inversion with Gauss-Newton-Krylov method: SEG Technical Program Expanded Abstracts, SEG, 2357–2361.

Erlangga, Y. A., and R. Nabben, 2007, On multilevel projection krylov method for the preconditioned helmholtz system: Elec. Trans. Numer. Anal., submitted.

———, 2008, Multilevel projection-based nested Krylov iteration for boundary value problems: SIAM J. Sci. Comput., to appear.

Herrmann, F. J., 2009, Compressive imaging by wavefield inversion with group sparsity: SEG Technical Program Expanded Abstracts, SEG, SEG, 2337–2341.

Herrmann, F. J., C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, 2009a, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41.

Herrmann, F. J., Y. A. Erlangga, and T. Lin, 2009b, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35.

Herrmann, F. J., and X. Li, 2010, Randomized dimensionality reduction for full-waveform inversion: Presented at the EAGE 73th Conference & Exhibition, Barcelona,.

Johnson, W., and J. Lindenstrauss, 1984, Extensions of Lipschitz mappings into a Hilbert space: Contemporary mathematics, **26**, 1–1.

Kaczmarz, S., 1937, Angenäherte auflösung von systemen linearer gleichungen: Bull. Int. Acad. Polon. Sci. A, **355**, 357.

Krebs, J. R., J. E. Anderson, D. Hinkley, A. Baumstein, S. Lee, R. Neelamani, and M.-D. Lacasse, 2009a, Fast full wave seismic inversion using source encoding: , SEG, 2273–2277.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009b, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Krohn, C., and R. Neelamani, 2008, Simultaneous sourcing without compromise: Rome 2008, 70th EAGE Conference & Exhibition, B008.

Lin, T., E. Lebed, Y. A. Erlangga, and F. J. Herrmann, 2008, Interpolating solutions of the helmholtz equation with compressed sensing: SEG Technical Program Expanded Abstracts, SEG, 2122–2126.

Mulder, W., and R. Plessix, 2004, How to choose a subset of frequencies in frequency-domain finite-difference migration: Geoph. J. Int., **158**, 801–812.

Natterer, F., 2001, The mathematics of computerized tomography: Society for Industrial Mathematics.

Neelamani, N., C. Krohn, J. Krebs, M. Deffenbaugh, and J. Romberg, 2008, Efficient seismic forward modeling using simultaneous random sources and sparsity: SEG International Exposition and 78th Annual Meeting, 2107–2110.

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, 1999, Numerical optimization: Springer.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Pratt, R., C. Shin, and G. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space waveform inversion: Geoph. J. Int., **133**, 341–362.

Robbins, H., and S. Monro, 1951, A stochastic approximation method: The Annals of Mathematical Statistics, **22**, 400–407.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, no. 2, 426–436.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Strohmer, T., and R. Vershynin, 2009, A randomized Kaczmarz algorithm with exponential convergence: Journal of Fourier Analysis and Applications, **15**, 262–278.

Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, **73**, R23–R35. (10.1190/1.2836323).

Virieux, J., S. Operto, H. Ben-Hadj-Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar, 2009, Seismic wave modeling for seismic imaging: The Leading Edge, **28**, 538–544.