

Full-waveform inversion from compressively recovered model updates

Xiang LI* and Felix J.Herrmann* EOS-UBC

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 size 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 lower our expectations to compute ourselves out of this. In this paper, we address this situation by randomized dimensionality reduction, which we adapt from the field of compressive sensing. In this approach, we combine deliberate randomized subsampling with structure-exploiting transform-domain sparsity promotion. Our approach is successful because it reduces the size of seismic data volumes without loss of information. With this reduction, we compute Newton-like updates at the cost of roughly one gradient update for the fully-sampled wavefield.

INTRODUCTION

With the recent resurgence of full-waveform inversion, the cost of computing gradient and Newton updates has become—aside from issues with non-uniqueness—one of the major impediments withstanding successful application of this technology to industry-size data volumes. The cost of computing the gradient accurately depends on the size of the data the discretization of the model while the challenge for the Newton updates lies in the fact that the Hessian matrix is full and possibly indefinite (negative eigenvalues). Finally, FWI calls for some sort of regularization that imposes prior information. Because the elastic properties in the earth contain singularities (zero-, first, and fractional-order discontinuities (Herrmann, 2003; Herrmann et al., 2001; Herrmann, 2001)) that trace curved interfaces, this prior information will be in the form of one-norms on wavelet or curvelet coefficients (Herrmann et al., 2008).

To overcome these impediments, we follow an approach that tackles both these issues replacing Newton updates with sparsity-promoting programs that for each outer ‘Newton’ update invert the adjoint of the Jabian while minimizing the one-norm on transform-domain coefficients of the updates. The proposed methodology follows in the footsteps of attempts towards cost reductions for the computation of gradients, which are central to imaging and inversion, through phase-encoded simultaneous sources (Romero et al., 2000; Herrmann et al., 2009b), possibly in combination with the removal of subsets of angular frequencies (Sirgue and Pratt, 2004; Mulder and Plessix, 2004; Lin et al., 2008; Herrmann et al., 2009b). The advantage of this approach are fourfold.

First, sparsity-promoting programs for the inversion of linear systems of equations allow us to tap into recent results from compressive sensing (CS in short throughout the paper, Candès et al., 2006; Donoho, 2006)—where the argument is made, and rigorously proven—that compressible signals can be recovered from severely sub-Nyquist samplings by solving a sparsity promoting program. These developments in conjunction with the recent resurgence of simultaneous sources (Beasley, 2008; Berkhout, 2008; Krebs et al., 2009a; Herrmann et al., 2009b; Herrmann, 2009) allow us to replace conventional impulsive sources by a limited number of simultaneous phase-encoded sources to reduce the computational complexity of computing the gradients (migration). These randomly phase-encoded sources render source crosstalk harmless by turning these artifacts into incoherent noise. This leads to a significant speedup in the computation of the model updates as reported by the authors (Herrmann and Li, 2010). Second, the recovery of the gradient from simultaneous data by sparsity-promoting recovery entails inversion of the adjoint of the Jacobian, which replaces the computation of Gauss-Newton updates with the reduced Hessian without forming this Hessian. Normally, the costs of these updates are prohibitive because they are roughly equivalent to least-squares migration for each Gauss-Newton update. However, we overcome this problem by using simultaneous sources and by solving the sparsity-promoting program with controlled misfit. This makes the problem easier to solve and this limits the number of iterations of the one-norm solver (SPGL1 - Berg and Friedlander, 2008) for the first iterations. This reduces the risk of overfitting the data and this is a common approach in inverse problem where the reduced Hessian is inverted with a limited number of iterations of the conjugate gradient solver (CG - Akcelik et al., 2002; Erlangga and Herrmann, 2009; Burstedde and Ghattas, 2009). Third, promoting curvelet-domain sparsity on the coefficients of the updates rather than on the model—e.g. by adding transform-domain sparsity as a prior to Equation 1—simplifies the optimization because the linearized problem is convex and has a global minimum. Fourth, our approach opens the possibility of using *restarts* by drawing of a new random CS experiment for each outer loop. This means that we apply a new dimensionality reduction for each model update. This approach is reminiscent of the Kaczmarz method (Kaczmarz, 1937) method that has been successfully applied by Natterer (2001) in tomography and that has recently been extended to the inversion of random matrices by Strohmer and Vershynin (2009). Our approach is also related to the *stochastic approximation* method in stochastic optimization (Robbins and Monro, 1951) and was recently proposed as part of FWI by Krebs et al. (2009a).

THEORY

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

$$\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]\|_{2,2}^2, \quad (1)$$

compressive FWI

where each column of \mathbf{P} contains the observed data for one shot and all frequencies. The nonlinear operator $\mathcal{F}[\mathbf{m}, \mathbf{Q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$ simulates data by solving the Helmholtz system \mathbf{H} for all sources in the columns of \mathbf{Q} . We restrict the simulated data to the receiver positions with the detection operator \mathbf{D} to obtain observed data. For simplicity, we assume that the sources are known and co-located with the receivers. 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 pseudo 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 a line search.

Result: Estimate for the model $\tilde{\mathbf{m}}$

```

m ← m0; // initial model
while  $\|\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]\|_{2,2} \geq \epsilon$  do
   $\tilde{\mathbf{g}} \leftarrow \mathbf{J}^*[\mathbf{m}, \mathbf{Q}](\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}])$ ; // gradient
   $\delta\mathbf{m} \leftarrow \mathbf{H}_{\text{red}}^{\dagger}\tilde{\mathbf{g}}$ ; // Gauss-Newton-Krylov update with CG
  m ← m +  $\alpha\delta\mathbf{m}$ ; // model update with line search
end

```

Algorithm 1: FWI by Gauss-Newton

Dimensionality reduction: Unfortunately, the solution of the above minimization problem with Gauss-Newton is extremely costly because each iteration requires the solution of the forward and time-reversed (adjoint) Helmholtz system for each of the n_f frequencies and for each of the n_s sources. Moreover, improvements in convergence of these schemes require expensive inversions of the reduced Hessian (see e.g. Erlangga and Herrmann, 2009, and the references therein).

Here, we address these problems by combining dimensionality-reduction strategies with recovery based on sparsity promotion. We reduce the data volume and hence the size of the Helmholtz system 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

$$(\mathbf{RM})_i := (\mathbf{R}_i^{\Sigma}\mathbf{F}_{\Sigma}^* \text{diag}(e^{i\theta_i})\mathbf{F}_{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_i^{\Omega}\mathbf{F}_{\Omega}), \quad i = 1 \dots N \quad (2)$$

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^{Ω} . The simultaneous *super shots* themselves are obtained by phase encoding with random phases $\boldsymbol{\theta}_{1 \dots n'_s} = \text{Uniform}([0, 2\pi])$, followed by the selection of one or more simultaneous sources with the restriction matrix \mathbf{R}_i^{Σ} . 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). After

applying this sampling operator, the nonlinear least-squares problem reduces to $\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]\|_{2,2}^2$, where the underbarred quantities refer to compressively sampled wavefields—i.e., $\delta\mathbf{P} := \mathbf{RM}(\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]) = \mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]$ (see Herrmann et al., 2009b, for details).

After this subsampling, FWI involves solutions of the dimensionality reduced Helmholtz system over $N \times n'_s \ll n_s$ sources with $N \times n'_f \ll n_f$ frequencies. As shown in Herrmann and Li (2010) this leads to cost reductions for wavefield simulation and migration as reported in the literature (Herrmann et al., 2009b; Herrmann and Li, 2010).

Dimensionality-reduced gradient updates: FWI involves gradients (cf. Table 1),

$$\mathbf{g} = \Re \left(\sum_{\omega} \omega^2 \sum_s (\tilde{\mathbf{u}} \odot \mathbf{v})_{s,\omega} \right) = \mathbf{J}^*[\mathbf{m}, \mathbf{Q}]\delta\mathbf{P}, \quad (3)$$

where \mathbf{u} , \mathbf{v} are the forward modeled source and reverse-time modeled residual wavefields for each frequency. Grosse modo, these gradient updates correspond to migrations given by the adjoint of the linearized multi-experiment Born scattering operator $\mathbf{J}[\mathbf{m}, \mathbf{Q}]$. To illustrate how simultaneous sources affect the gradient (\mathbf{g}), we simulate for all frequencies linearized data for a single impulsive shot and for a single simultaneous shot. This is an instance of one block of \mathbf{RM} . From Figure 1, we can clearly see that the migrated image from the simultaneous-source experiment is well resolved. However, there is clearly leakage from the image towards incoherent noise-like artifacts.

Subsample ratio	0.015	0.006	0.002
n'_f/n'_s	recovery error (dB)		
5	17.44 (1.32)	11.66 (0.78)	6.83 (-0.14)
1	17.53 (1.59)	11.89 (1.05)	7.19 (0.15)
0.2	18.22 (1.68)	12.11 (1.32)	7.46 (0.27)
Speed up (\times)	66	166	500

Table 1: Signal-to-noise ratios, $\text{SNR} = 20 \log_{10}(\frac{\|\delta\mathbf{m} - \delta\tilde{\mathbf{m}}\|_2}{\|\delta\mathbf{m}\|_2})$ for reconstructions with the wavelet sparsity transform for different subsample and frequency-to-shot ratios. SNRs for ℓ_1 are in bold and SNRS for migration are in parentheses.

Updates via sparse inversion: Aside from interference artifacts, it is well known that gradient updates do not resolve amplitudes correctly, which may lead to slow convergence. Instead of following a Gauss-Newton-Krylov approach where the reduced Hessian is inverted with conjugate gradients (Akcelik et al., 2002; Erlangga and Herrmann, 2009; Burstedde and Ghattas, 2009), we propose a formulation, where the adjoint of the Jacobian is inverted directly. Because the Jacobian involves linearized Born scattering, contributions from internal multiple scattering are ignored in the updates (Pratt et al., 1998). We also assume \mathbf{P} to be surface-multiple free, which removes an important nonlinearity in the inversion.

In most inversion problems, people avoid inverting the Jacobian using iterative methods because it involves solutions of the forward and adjoint Helmholtz system for each iteration, which is expensive compared to inverting the Hessian iteratively, and also because this operator is ill conditioned (the Hessian is ill

compressive FWI

conditioned). However, contrary to many inverse problems, the Hessian of seismic imaging is relatively well behaved—i.e., it is near unitary for linearizations with respect to velocity models that are close to the actual velocity model. In that case, the high-frequency behavior of the Hessian can be approximated with diagonal scalings (Herrmann et al., 2008; Symes, 2008), which can lead to effective preconditioners (Herrmann et al., 2009a). In this paper, we exploit the relatively well-behaved reduced Hessian differently by using techniques from compressive sensing to reduce the dimensionality of the linearized system. This allows us to calculate ‘Newton-like’ updates by carrying out one-norm regularized linearized inversions with the reduced system—i.e., we compute $\delta\tilde{\mathbf{x}} = \mathbf{S}^H \delta\tilde{\mathbf{x}}$ by solving

$$\delta\tilde{\mathbf{x}} = \arg \min_{\delta\tilde{\mathbf{x}}} \|\delta\tilde{\mathbf{x}}\|_{\ell_1} \quad \text{subject to} \quad \|\delta\mathbf{P} - \mathbf{A}\delta\tilde{\mathbf{x}}\|_2 \leq \sigma, \quad (4)$$

where $\mathbf{A} := \mathbf{R}\mathbf{M}\mathbf{J}\mathbf{S}^* = \mathbf{J}\mathbf{S}^*$, and $\mathbf{J} = \mathbf{J}[\mathbf{m}, \mathbf{Q}]$ and where \mathbf{S}^H is the adjoint of some sparsifying transform (e.g. curvelets or wavelets). The sparsity-promoting program in Equation 4 seeks amongst all possible curvelet-coefficient vectors the vector has smallest ℓ_1 norm and that yields a fit with tolerance σ . Because of the dimensionality reduction, our algorithm should be competitive with CG on the Hessian.

Sparsity continuation with CS renewal: It is well known that FWI is plagued by local minima, which are related to oscillations that depend on the spatial wavelength. A variety of continuation methods has been proposed to mitigate this problem by allowing high-frequency components and late arriving reflections to enter into the solution only gradually. This is done through carefully designed filtering and windowing operations (Pratt et al., 1998; Burstedde and Ghattas, 2009). We follow a different and complementary strategy where we allow the one-norm of the solution to grow only gradually. In this way, we are able to preserve transform-domain sparsity on the model, and not only on the updates. Our algorithm is summarized in Table 2. The keys for the success of this algorithms lie in (i) finding a strategy for lowering the tolerance parameter σ_j for each j^{th} iteration. In the beginning, we are far from the solution and we have to avoid overfitting the data, which would result in coefficient vectors of the updates that are insufficiently sparse to sparsity on the model itself. As the algorithm progresses towards the solution, we need extract more information from the data by lowering the tolerance—i.e., $\sigma_j \rightarrow 0$ as $h \rightarrow \infty$. (ii) in renewing the CS experiment after each model update. This means we draw new data by compressively sampling the residual and by defining a new CS matrix \mathbf{A} .

EXAMPLE

To illustrate the performance of our algorithm, we study the behavior of single gradient updates by conducting a series of experiments where we vary the subsampling ratio—i.e., the aspect ratio of RM—and the frequency-to-shot subsampling ratio. All simulations are carried with 256 co-located shot and receiver positions sampled at a 29 m interval. The time sample interval is 0.016s. Comparison between the gradient updates and the updates by sparsity promotion shows remarkable high-fidelity results even for increasing subsampling ratios. As expected, the numbers in Table 1 confirm increasing recovery errors for increasing subsampling ratios. For fixed

subsampling ratios, however, we observe improved results for decreasing frequency-to-shot ratios, which suggests that simultaneous shots contribute more to the solution. These trends also hold for migrated images—i.e., standard gradient updates. Because the speedups are proportional to subsampling ratios, we can conclude that the dimensionality reductions offset the costs of the ℓ_1 recovery approximately.

We also tested the behavior of the algorithm in the setting of FWI. For this purpose, we ran an experiment with 10 simultaneous shots and 5 randomly selected frequencies starting from the initial model plotted in Figure 2(a). Because the development of proper sparsity continuation is still under development, we limited the number of iterations for the ℓ_1 solver. Our experiments showed that the algorithm only converges for an initial model that is relatively close to the actual model. Our sparsity-promoting procedure with renewals (cf. Figure 1(b) and 2(b)) clearly shows improvement because it is able to create a high-resolution FWI result from a highly dimensionality reduced FWI problem. Finally, we can also observe from the ℓ_1 -norm recovered gradient updates for the slownesses (plotted in Figure 2(c) and 2(d)) that increasingly higher frequency components are added to the solution.

CONCLUSIONS

Dimensionality-reduction strategies will be instrumental for the success of FWI. These strategies are build on the premise that whenever models exhibit structure (read transform-domain sparsity), these models can be reconstructed from randomized subsampling that are proportional to transform-domain sparsity. Because these approaches remove the ‘‘curse of dimensionality’’, which is one of the major impediments of FWI, this contribution may lead to a paradigm shift where Newton-type updates, which otherwise would have to be computed over all frequencies and all sequential sources, can now be computed by inverting the dimensionality-reduced linearized Born scattering operator. This result including the renewals will help bring FWI into the main stream of seismic data processing.

ACKNOWLEDGMENTS

FJH. would like to thank Yogi Erlangga for insightful discussions. 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).

Result: Estimate for the model $\tilde{\mathbf{m}}$

```

m ← m0 ; // initial model
j ← 0 ; // loop counter
Q ← (RM)iQ ; // Draw random sim. shot
while  $\|\mathbf{P} - \mathcal{F}[\mathbf{m}, \mathbf{Q}]\|_{2,2}^2 \geq \epsilon$  do
  j := j + 1 ; // increase counter
  A ← J[m, Q]S* ; // Calculate Jacobian
  Solve Equation 4 for  $\sigma_j$  ; // sparsity-promoting recovery
  m ← m + S*  $\delta\tilde{\mathbf{x}}$  ; // compute model update
  Q ← (RM)iQ ; // Draw random sim. shot
end

```

Algorithm 2: FWI by repeated one-norm minimization.

compressive FWI

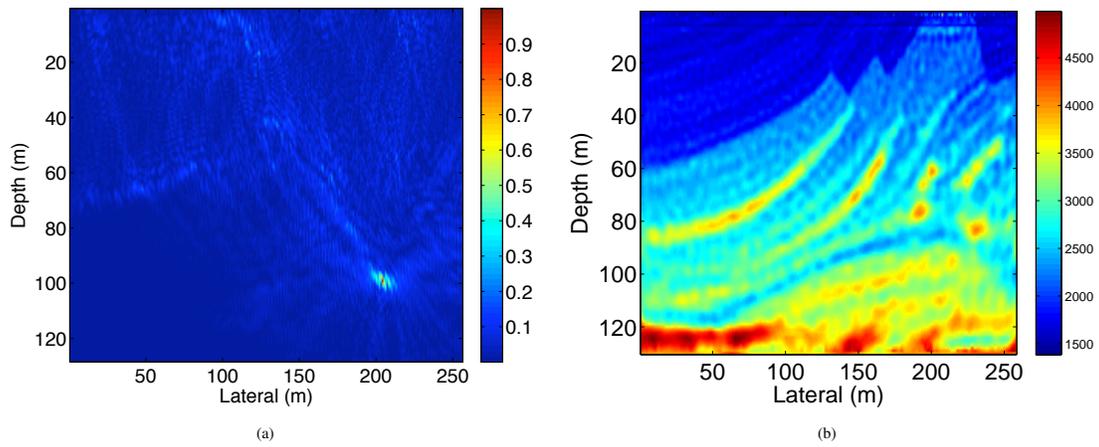


Figure 1: Sequential-shot versus simultaneous-source gradients (migration) of a single point diffractor in the target zone. **(a)** Migrated image for a single simultaneous shot. **(b)** FWI result without renewal of CS experiments.

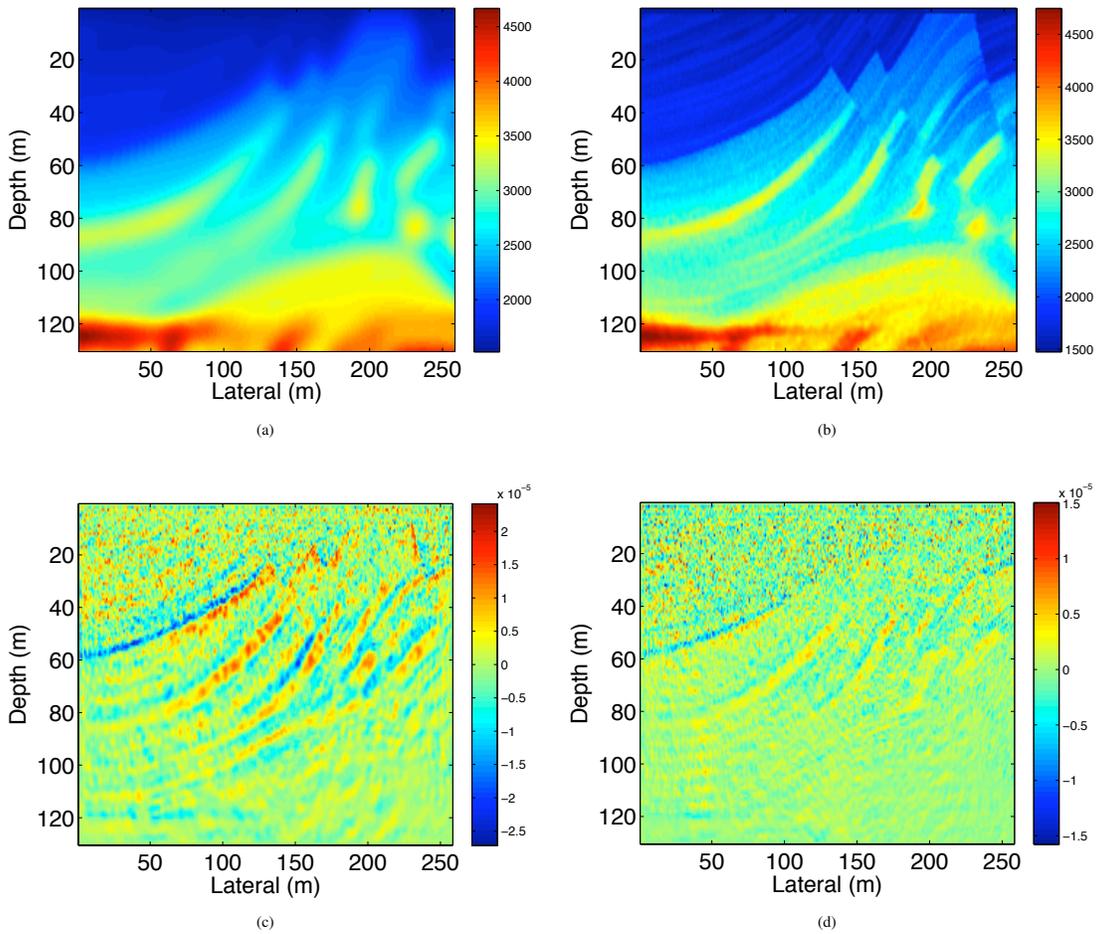


Figure 2: Proof of principle for full-waveform inversion with updates obtained by sparsity promotion **(a)** Initial smooth background model **(b)** Sparsity promoted inversion after 10 updates with $n'_f = 5$ and $n'_s = 10$ **(c)** The first update obtained by solving the ℓ_1 -norm minimization problem **(d)** The ninth update.

REFERENCES

- Akcelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Supercomputing, ACM/IEEE 2002 Conference, 41–41.
- Beasley, C. J., 2008, A new look at marine simultaneous sources: *The Leading Edge*, **27**, 914–917.
- Berg, E. v., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Technical Report 2, Department of Computer Science, University of British Columbia.
- Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: *The Leading Edge*, **27**, 924–938.
- Burstedde, C., and O. Ghattas, 2009, Algorithmic strategies for full waveform inversion: 1d experiments: *Geophysics*, **74**, WCC37–WCC46.
- 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.
- Erlangga, Y. A., and F. J. Herrmann, 2009, Seismic waveform inversion with Gauss-Newton-Krylov method: SEG Technical Program Expanded Abstracts, SEG, 2357–2361.
- Herrmann, F. J., 2001, Singularity characterization by monoscale analysis: application to seismic imaging: *Appl. Comput. Harmon. Anal.*, **11**, 64–88.
- , 2003, Multifractional splines: application to seismic imaging: *Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X*, SPIE, 240–258.
- , 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,.
- Herrmann, F. J., W. Lyons, and C. Stark, 2001, Seismic facies characterization by monoscale analysis: *Geoph. Res. Lett.*, **28**, 3781–3784.
- 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).
- 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.
- 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.
- 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).