Full-waveform inversion with randomized L1 recovery for the model updates

Xiang Li, Aleksandr Aravkin, Tristan van Leeuwen, Felix Herrmann

March 5, 2011

Abstract

Full-waveform inversion (FWI) is a data fitting procedure that relies on the collection of seismic data volumes and sophisticated computing to create high-resolution results. With the advent of FWI, the improvements in acquisition and inversion have been substantial, but these improvements come at a high cost because FWI involves extremely large multi-experiment data volumes. The main obstacle is the 'curse of dimensionality' exemplified by Nyquist's sampling criterion, which puts a disproportionate strain on current acquisition and processing systems as the size and desired resolution increases. In this paper, we address the 'curse of dimensionality' by randomized dimensionality reduction of the FWI problem adapted from the field of CS. We invert for model updates by replacing the Gauss-Newton linearized subproblem for subsampled FWI with a sparsity promoting formulation, and solve this formulation using the SPGl1 algorithm. We speed up the algorithm and avoid overfitting the data by solving for the linearized updates only approximately. Our approach is successful because it reduces the size of seismic data volumes without loss of information. With this reduction, we can compute a Newton-like update with the reduced data volume at the cost of roughly one gradient update for the fully sampled wavefield.

Introduction

As we reported in earlier work (Li and Herrmann, 2010), the cost of computing gradient and Newton updates is one of the major impediments preventing the successful application of FWI to industry-size data volumes. The cost of computing the gradient depends on the size of the data and on the discretization of the Helmholz operator, while Newton updates are difficult because the Hessian of FWI is dense and possibly indefinite (negative eigenvalues). Finally, FWI is ill-posed, and so requires use of prior information. It is often formulated as a nonlinear least squares optimization problem, and regularized by an attractor term with an initial model guess or by total variation terms (Virieux and Operto (2009)). In contrast to these methods we do not append any extra terms to the FWI objective, but instead regularize the model updates to be sparse in the curvelet frame (Herrmann et al., 2008), which is known to provide efficient representations of models that are smooth except for discontinuities along smooth curves. It is known that velocity parameters in the earth contain singularities (zero-, first, and fractional-order discontinuities (Herrmann, 2003; Herrmann et al., 2001)) that trace curved interfaces.
Our approach to computing model updates uses a sparsity-promoting program for the solution of linear systems of equations, and connects to recent ideas from CS (CS in short throughout the paper, Candès et al., 2006; Donoho, 2006), where it is shown that compressible signals can be recovered from severely sub-Nyquist sampling rates by solving a sparsity promoting (ℓ_1) program. These developments, in conjunction with the recent resurgence of simultaneous sources (Beasley, 2008; Berkhout, 2008; Krebs et al., 2009; Herrmann et al., 2009; 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 in FWI (migration). Randomly phase-encoded sources significanlty reduce crosstalk by turning crosstalk artifacts into incoherent noise, and provide significant speedup in gradient computation (Li and Herrmann, 2010). The Jacobian adjoint inversion for the recovery of the gradient required by sparsity-promoting optimization is prohibitive for our application, as it is comparable to a Gauss-Newton update. We overcome this problem by using simultaneous sources as a dimensionality reduction technique, resampling the sources at each linearization.
Rather than solving the sparsity promoting programs fully, we solve them only approximately, exiting the SPGℓ_1 solver when we have achieved the maximum possible error reduction for a particular level of sparsity. In addition to speeding up the update calculation, this reduces the risk of overfitting the data. Limiting solver iterations is another common approach in regularization of ill-conditioned inverse problems; for example the reduced Hessian may be approximately inverted with a limited number of iterations of the conjugate gradient (CG) solver (Akcelik et al., 2002; Erlangga and Herrmann, 2009; Burstedde and Ghattas, 2009). Finally, we use a custom technique that relies on sampling the Pareto curve (SPGℓ_1 - Berg and Friedlander, 2008) for warm-starting the SPGℓ_1 solver between update calculations. This approach decreases the number of matrix-vector multiplies that are necesssary for the solver to proceed, which makes our approach particularly suitable for large-scale geophysical problems.

Dimensionality reduction by compressive sampling

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

$$\min_{\mathbf{m}} \frac{1}{2K} \sum_{i=1}^{K} \|\mathbf{b}_i - \boldsymbol{\mathcal{F}}_i[\mathbf{m}, \mathbf{q}_i]\|_2^2, \tag{1}$$

with $K = n_f \times n_s$ the batch size, given by the total number of monochromatic sources $\mathbf{q}_i$ (n_f, n_s are the total number of frequencies and source experiments). The vectors $\mathbf{b}_i$ represent the corresponding vectorized monochromatic shot records and the $\boldsymbol{\mathcal{F}}_i[\mathbf{m}, \mathbf{q}_i] = \mathbf{P}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{q}_i$ represents the monochromatic forward operator for the i^{th} source, with $\mathbf{P}$ the detection operator, resticting the data to the receiver positions. 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. Unfortunately, the solution of the above minimization problem with (quasi)Newton techniques 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 replacing Eq. 1 with

$$\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{RM}(\mathbf{b}-\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{q}])\|_2^2 = \min_{\mathbf{m}} \frac{1}{2}\|\underline{\mathbf{b}}-\underline{\boldsymbol{\mathcal{F}}}[\mathbf{m},\underline{\mathbf{q}}]\|_2^2, \tag{2}$$

where the vector $\mathbf{b}$, $\mathbf{q}$, and matrix $\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{q}]$ represent the data, source, and modeling operator for the full batch, respectively. The corresponding dimensionality-reduced quantities $\underline{\mathbf{b}} = \mathbf{RMb}$, $\underline{\mathbf{q}} = \mathbf{RMq}$, and matrix $\underline{\boldsymbol{\mathcal{F}}}[\mathbf{m},\underline{\mathbf{q}}]$ are underlined. Each simultaneous-source experiment in the collection of supershots is given by a different restriction. The i^{th} block of the restriction matrix $\mathbf{R}$ is defined by the Kronecker product: $\mathbf{R}_i := \mathbf{R}_i^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_i^{\Omega}$ for $i = 1 \cdots n_s'$. The measurement matrix $\mathbf{M}$ is given by the Kronecker product $\mathbf{M} := \mathbf{M}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{I}$. As in Lin and Herrmann (2007), we follow Romberg (2009) to phase encode sequential shots.

The above identity follows from linearity of randomized subsampling by $\mathbf{RM}$ in combination with linearity and separability of monochromatic sources. Consequently, the number of PDE solves required per iteration of Eq. 2 is reduced by a factor of K'/K (see also Herrmann et al., 2009, for details). However, this speed up comes at the expense of leaking energy from imaged reflectors to incoherent artifacts. Hence, the key question is to find a solver that mitigates these artifacts.

Replacing Gauss-Newton subproblem with sparsity promoting formulation: A common approach to solving the FWI problem (or in our case, the subsampled FWI problem) is by applying the Gauss-Newton method, where at each iteration an update $\delta\mathbf{m}$ is obtained by solving the following least-squares subproblem

$$\min_{\delta\mathbf{m}} \|\underline{\delta\mathbf{b}} - \underline{\nabla\boldsymbol{\mathcal{F}}}[\mathbf{m},\underline{\mathbf{q}}]\delta\mathbf{m}\|_2^2, \tag{3}$$

with $\underline{\delta\mathbf{b}}$ and $\underline{\nabla\boldsymbol{\mathcal{F}}}$ the dimensionality reduced data residue and Jacobian, respectively. Because the Jacobian involves linearized Born scattering, contributions from internal multiple scattering are ignored in the updates. We also assume our observation to be surface-multiple free, which removes an important nonlinearity from the inversion.

Our approach is to use techniques from CS to recover sparse updates for the linearized subsampled FWI problem Eq. (3). Specifically, given a sparsifying transform $\mathbf{S}$ with adjoint $\mathbf{S}^H$ (wavelets or curvelets), we compute the update $\delta\widetilde{\mathbf{m}} = \mathbf{S}^H\delta\widetilde{\mathbf{x}}$ by solving the corresponding basis pursuit de-noise (BPDN) problem

$$\delta\widetilde{\mathbf{x}} = \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_1} \quad \text{subject to} \quad \|\underline{\delta\mathbf{b}} - \nabla\underline{\boldsymbol{\mathcal{F}}}[\mathbf{m},\underline{\mathbf{q}}]\mathbf{S}^*\delta\mathbf{x}\|_2 \le \sigma, \tag{4}$$

with σ an appropriately chosen tolerance. The BPDN problem is harder to solve than the least square subproblem Eq. (3). However, because of the dimensionality reduction, we can afford more iterations. Comparing this approach with Eq. (3), it is clear that at each iteration the linearlized subsampled misfit is reduced (now to a tolerance σ); but the key idea is that the sparsity-promoting program in Eq. (4) finds amongst all possible updates $\delta\mathbf{m}$ the vector that has smallest ℓ_1 norm when represented in curvelet frames.

Sparsity continuation with CS renewals: Efficient ℓ_1 solvers are based on solutions of series of relaxed subproblems, where the ℓ_1-norms of the solutions are increased intelligently. For linear problems, the spectral projected-gradient algorithm SPGℓ_1 developed by (Berg and Friedlander, 2008) exploits the Pareto boundary—the trade-off curve delineating feasible and infeasible solutions as a function the ℓ_2-norm of the data misfit and the model's ℓ_1-norm—to compute the relaxations by root finding. This allows components to enter into solutions controllably with a minimal number of matrix-vector multiplies and this makes this solver suitable for large-scale geophysical problems (Hennenfent et al., 2008). We use the SPGℓ_1 algorithm as our computational kernel to compute the updates (see Algorithm 1). Since our formulation decreases the linearized subsampled FWI residual, we expect that the recovered update is a descent direction of the full problem. The key point is that in our formulation, the solution of each of

Algorithm 1: FWI with modified updates

initialize $\mathbf{m}$
for ω in freq. range **do**
  while not converged **do**
    sample to obtain $\underline{\mathbf{b}}$, $\underline{\mathbf{q}}$, compute $\delta\underline{\mathbf{b}}$
    $\delta\widetilde{\mathbf{x}} \leftarrow \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_1}$ subject to $\|\delta\underline{\mathbf{b}} - \nabla\underline{\mathcal{F}}[\mathbf{m}, \underline{\mathbf{q}}]\mathbf{S}^*\delta\mathbf{x}\|_2 \leq \sigma$ (SPGℓ_1)
    $\mathbf{m} \leftarrow \mathbf{m} + \mathbf{S}^H\delta\widetilde{\mathbf{x}}$
  end while
end for

these subproblems is maximally sparse in the curvelet frame. This helps reduce the effects of randomized subsampling. Furthermore, we resample at each linearization to further reduce these effects. Note that we only solve the BPDN supbroblems approximately, exiting each time the SPGℓ_1 algorithm reaches the Pareto curve. Moreover, at each new iteration the solver is warm-started using data from the previous iteration.

Example

To demonstrate the performance of our algorithm, we run a series of experiments on the Marmousi model summarized in Fig. 1. All simulations are carried out with 384 co-located shots and receivers sampled at a 24m interval, yielding an maximum offset of 9.2km. We used a Ricker wavelet with a central frequency of 15Hz. The time record has a duration of 3$\,\mathrm{s}$ and is sampled with a sample interval of 16ms. To improve convergence, the inversions are carried out sequentially in 10 overlapping frequency windows on the interval $1.7 - 22.5\mathrm{Hz}$, each using 15 different simultaneous shots and 10 randomly selected frequencies. For each subproblem, we use roughly 20 iterations of SPGℓ_1. Hence, we obtain the Gauss-Newton update at a cost roughly equivalent to one tenth of the cost of a gradient calculation with all of the sources. As a starting model, we use a velocity profile obtained by smoothing the original model, followed by horizontal averaging.
The result after ten Gauss-Newton iterations for each frequency band is depicted in bottom panel of Fig. 1. Since we do not use a line search, we never evaluate the misfit for all the sources. The cost of ten Gauss-Newton iterations is then roughly equivalent to one evaluation of the full misfit. This gives us an order of magnitude speed up without loss of quality.

Conclusions

We introduced an efficient algorithm to solve linearized subproblems of full-waveform inversion. Our method combines recent findings from stochastic optimization (Haber et al., 2010) and CS and turns the 'overdetermined' Gauss-Newton subproblems into a series of underdetermined dimensionality-reduced experiments. By considering these subproblems as compressive-sampling experiments, we were able to reduce the computational costs significantly without loss of quality.

Acknowledgments

We would like to thank the sponsors of the SINBAD project. This work was partly supported by a NSERC Discovery Grant (22R81254) and CRD Grant DNOISE II (375142-08).

References

Akcelik, V., Biros, G. and Ghattas, O. [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**(7), 914–917, doi:10.1190/1.2954033.
Berg, E.v. and Friedlander, M.P. [2008] Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31**(2), 890–912, doi:10.1137/080714488.
Berkhout, A.J. [2008] Changing the mindset in seismic data acquisition. *The Leading Edge*, **27**(7), 924–938, ISSN 7, doi:10.1190/1.2954035.
Burstedde, C. and Ghattas, O. [2009] Algorithmic strategies for full waveform inversion: 1d experiments. *Geophysics*, **74**(6), WCC37–WCC46, doi:10.1190/1.3237116.
Candès, E., Romberg, J. and Tao, T. [2006] Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, **59**(8), 1207–1223.

Donoho, D.L. [2006] Compressed sensing. *IEEE Transactions on Information Theory*, **52**(4), 1289–1306.
Erlangga, Y.A. and Herrmann, F.J. [2009] Seismic waveform inversion with Gauss-Newton-Krylov method. *SEG Technical Program Expanded Abstracts*, SEG, vol. 28, 2357–2361.
Haber, E., Chung, M. and Herrmann, F.J. [2010] An effective method for parameter estimation with pde constraints with multiple right hand sides. Tech. Rep. TR-2010-4, UBC-Earth and Ocean Sciences Department.
Hennenfent, G., van den Berg, E., Friedlander, M.P. and Herrmann, F.J. [2008] New insights into one-norm solvers from the Pareto curve. *Geophysics*, **73**(4).
Herrmann, F.J. [2009] Compressive imaging by wavefield inversion with group sparsity. *SEG Technical Program Expanded Abstracts*, SEG, SEG, vol. 28, 2337–2341.
Herrmann, F.J., Erlangga, Y.A. and Lin, T. [2009] Compressive simultaneous full-waveform simulation. *Geophysics*, **74**, A35.
Herrmann, F.J. [2003] Multifractional splines: application to seismic imaging. *Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X*, SPIE, vol. 5207, 240–258.
Herrmann, F.J., Lyons, W. and Stark, C. [2001] Seismic facies characterization by monoscale analysis. **28**(19), 3781–3784.
Herrmann, F.J., Moghaddam, P.P. and Stolk, C.C. [2008] Sparsity- and continuity-promoting seismic imaging with curvelet frames. *Journal of Applied and Computational Harmonic Analysis*, **24**(2), 150–173, doi:10.1016/j.acha.2007.06.007.
Krebs, J.R. et al. [2009] Fast full wave seismic inversion using source encoding. SEG, vol. 28, 2273–2277, doi: 10.1190/1.3255314.
Li, X. and Herrmann, F.J. [2010] Full-waveform inversion from compressively recovered model updates. SEG, vol. 29, 1029–1033, doi:10.1190/1.3513022.
Lin, T.T.Y. and Herrmann, F.J. [2007] Compressed wavefield extrapolation. *Geophysics*, **72**(5), SM77–SM93, doi:10.1190/1.2750716.
Romberg, J. [2009] Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, **2**(4), 1098–1128, doi:10.1137/08072975X.
Virieux, J. and Operto, S. [2009] An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC1–WCC26, doi:10.1190/1.3238367.

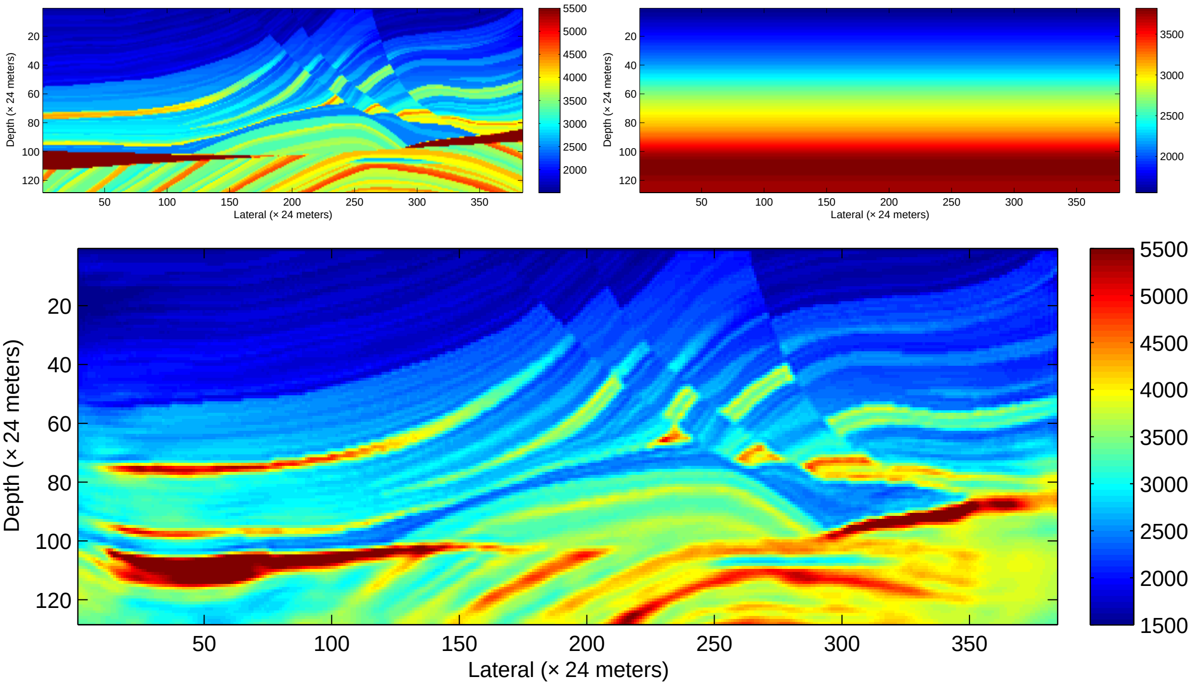

Figure 1: Marmousi model, (velocity unit: $\mathrm{m/s}$). **(Upper left:)** True model. **(Upper right:)** Starting model. **(Lower:)** Inversion results with renewals.