FAST SEISMIC IMAGING FOR MARINE DATA

Aleksandr Aravkin, Xiang Li, and Felix J. Herrmann

Seismic Laboratory for Imaging and Modeling, the University of British Columbia, Technical Report TR-2011-10

ABSTRACT

Seismic imaging can be formulated as a linear inverse problem where a medium perturbation is obtained via minimization of a least-squares misfit functional. The demand for higher resolution images in more geophysically complex areas drives the need to develop techniques that handle problems of tremendous size with limited computational resources. While seismic imaging is amenable to dimensionality reduction techniques that collapse the data volume into a smaller set of "super-shots", these techniques break down for complex acquisition geometries such as marine acquisition, where sources and receivers move during acquisition. To meet these challenges, we propose a novel method that combines sparsity-promoting (SP) solvers with random subset selection of sequential shots, yielding a SP algorithm that only ever sees a small portion of the full data, enabling its application to very large-scale problems. Application of this technique yields excellent results for a complicated synthetic, which underscores the robustness of sparsity promotion and its suitability for seismic imaging.

Index Terms— Seismic imaging, dimensionality reduction, compressive sensing, stochastic optimization

1. INTRODUCTION

Modern-day seismic wave-equation based imaging depends increasingly on computationally and data-intensive wave simulators that require solutions of partial-differential equations (PDEs) over increasingly large domains and frequency ranges. The computational costs involved with the solutions of these PDEs become prohibitively large because the linearized inversions are based on a least-squares fitting procedure carried out for a very large number of sources. This is challenging because each source corresponds to a different right-hand side of the PDE and hence the computational cost grow exponentially with the size of the domain (the sources move along the surface) and resolution.

Motivated by early work of [1, 2], we overcome this challenge of the "curse of dimensionality" by decreasing the number of source experiments. As a result, we lower the computational burden of the inversion significantly by using linearity of the PDEs with respect to the sources. This separable structure allows us to use randomized superposition, where we combine sources and their corresponding data into a smaller number of simultaneous source experiments. During these experiments, sequential sources are replaced by incoherent "beams", where each source position fires with a strength drawn from a Gaussian random distribution. As long as the number of these incoherent source experiments is smaller then the number of sequential experiments, this random superposition leads to a reduction of computations by virtue of the reduced data volume and number of sources.

However, this dimensionality reduction hinges on two conditions: cancellation of source cross talk as the number of simultaneous sources increases, and full acquisition, where each source experiment sees the same receivers. The first condition is met as long as the batch size, i.e the number of simultaneous sources, is sufficiently large. However, increasing this batch size only leads to a slow decay of the cross talk because this method corresponds to the stochastic-average approximation, which is essentially a Monte-Carlo sampling technique. Unfortunately, the alternative stochastic approximation, where randomized batches are drawn for each gradient update, does not remedy this situation because it is unstable with respect to noise.

To overcome these issues of slow convergence and noise instability, [3] introduced a dimensionality reduction approach where the source crosstalk is removed by transform domain sparsity promotion. In this approach, we turn the originally "overdetermined" problem—note that the imaging has a null space because related finite aperture and other effects—into a much smaller underdetermined system of equation. To improve the convergence of the sparsity promoting solver, [3, 4] use a stochastic-approximation type of approach that removes correlations between the solution vector and source encoding by randomized superposition.

Unfortunately, this approach relies on full acquisition (the second condition), which is generally not met during marine acquisition because receivers are towed behind the sources and therefore move. In this paper, we address this issue by replacing the randomized superposition by selecting random subsets of sequential sources instead of randomized simultaneous sources.

Supported by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (375142-08) matching support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

2. THE SEISMIC IMAGING PROBLEM

After discretization, seismic imaging requires inversion of the linearized time-harmonic Born-scattering matrix linking data $\mathbf{b}$ to the medium perturbation $\mathbf{x}$. Note that $\mathbf{b} \in \mathbb{C}^{N_f N_r N_s}$ with N_f, N_r, and N_s the number of angular frequencies, receivers, and sources, while $\mathbf{x} \in \mathbb{R}^M$, with M the number of grid points.

Because angular frequencies and sequential sources can be treated independently, the linearized inversion has the following separable form:

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 = \sum_{i=1}^{K} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2^2, \qquad (1)$$

with $K = N_f N_s$ the batch size, given by the total number of monochromatic sources. The vectors $\mathbf{b}_i \in \mathbb{C}^{N_r}$ represent the corresponding vectorized monochromatic preprocessed (free of surface-multiples and direct waves) shot records, and $\mathbf{b}$ is a stack of vectors $\mathbf{b}_i$. The matrix $\mathbf{A}_i$ represents the monochromatic linearized scattering matrix for the i^{th} source, and the matrix $\mathbf{A}$ is a stack of matrices $\mathbf{A}_i$.

Solving this problem is problematic because each iteration requires $4K$ PDE solves: two to compute the action of $\mathbf{A}_i$ and two for the action of its adjoint $\mathbf{A}_i^H$. Both actions involve solutions of the forward source and reverse-time residual wavefields [5]. Thus, the inversion costs grow linearly with the number of monochromatic sources.

To solve Equation (1) efficiently, we combine recent ideas from stochastic optimization and compressive sensing. We cast the original imaging problem into a series of much smaller subproblems that work on different subsets of data.

2.1. Solution by batching

We based our original algorithm on forming compressive seismic experiments—or to use the language of online machine learning mini batches—that consist of collections of small numbers of supershots. These supershots are made of randomized superpositions of sequential sources.

Mathematically, imaging experiments for mini batches with $K' \ll K$ monochromatic supershots, require the solution of the reduced system

$$\underset{\mathbf{x}}{\text{minimize}} \left\|\underline{\mathbf{b}} - \underline{\mathbf{A}}\mathbf{x}\right\|_2^2\} = \sum_{j=1}^{K'} \|\mathbf{b}_j - \mathbf{A}_j\mathbf{x}\|_2^2, \qquad (2)$$

which has the same structure as (1), but with much smaller dimension $K' << K$. Each $\mathbf{b}_j$ is a weighted linear combination of $\mathbf{b_i}$, and each matrix $\mathbf{A}_j$ is the same linear combination of matrices $\mathbf{A}_i$: $\mathbf{b}_j = \sum_{i=1}^{k} w_{ij}\mathbf{b}_i$, $\mathbf{A}_j = \sum_{i=1}^{k} w_{ij}\mathbf{A}_i$. To specify how sub-selection and mixing is achieved in the source-frequency space, we write the objective in (2) as $\frac{1}{2}\|\mathbf{R}\mathbf{M}\left(\mathbf{b} - \mathbf{A}\mathbf{x}\right)\|_2^2$. The restriction matrix $\mathbf{R}$ is defined by the Kronecker product: $\mathbf{R} \overset{\text{def}}{=} \mathbf{R}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}^{\Omega} \in \mathbb{R}^{(K'N_r)\times(KN_r)}$ with $\mathbf{R}^{\Sigma} \in \mathbb{R}^{n_s' \times n_s}$ selecting $n_s' \ll n_s$ rows uniform randomly amongst $[1 \cdots n_s]$ and $\mathbf{R}^{\Omega} \in \mathbb{R}^{n_f' \times n_f}$ selecting $n_f' \ll n_f$ frequencies from the seismic frequency band, and $\mathbf{I}$ the identity matrix. The measurement matrix $\mathbf{M} \in \mathbb{R}^{(KN_r)\times(KN_r)}$ is given by the Kronecker product $\mathbf{M} \overset{\text{def}}{=} \mathbf{M}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{I}$.

For the definition of the measurement matrix $\mathbf{M}^{\Sigma}$, we either use phase encoding, as in [6], which turns sequential sources into simultaneous sources, or we select random subsets of source locations by setting $\mathbf{M} = \mathbf{I}$. The latter allows us to work with marine data. It is easy to show that the number of PDE solves required for each iteration of the solution of (2) is reduced by a factor of K'/K.

2.2. Solution by sparsity promotion

Because averaging—whether via the stochastic-average approximation or via averaging amongst model iterates as part of stochastic-gradient descents—is not able to remove the source crosstalk efficiently, we rely on transform-domain sparsity promotion instead. In the noise-free case, sparsity-promoting imaging involves the solution of the following optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{b} = \mathbf{A}\mathbf{x}. \qquad (3)$$

Efficient ℓ_1 solvers for this problem are typically based on solutions of a series of relaxed subproblems, where components are allowed to enter into the solutions controllably. It is widely known that these approaches lead to a reduction in the number of iterations to reach the solution. The spectral projected-gradient algorithm [7, SPGℓ_1] uses this principle by solving a series of LASSO problems,

$$\underset{x\in\tau\mathbb{B}_1}{\text{minimize}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \,, \qquad (4)$$

where $\tau\mathbb{B}_1 = \{x \big| \|x\|_1 \leq \tau\}$. The sequence of τ's are generated using Newton root on the value function $v(\tau)$ of (4), so the final value of τ ensures $v(\tau) = 0$, so a solution of (4) with this τ also solves (3).

Each LASSO subproblem is solved using the Spectral Projected Gradient (SPG) method, so the overall algorithm uses a limited number of matrix-vector multiplies. Because the cost of the solver is determined by this number of multiplies, this approach is particularly suitable for large-scale geophysical problems [8].

Unfortunately, the degree of randomized dimensionality reduction determines the amount of cross-talk that results from the inversion, and hence we can not reduce the problem size too much. We overcome this by subsampling each LASSO problem (4), in effect solving a series of problems: $\text{minimize}_{x\in\tau\mathbb{B}_1} \|\underline{\mathbf{b}} - \underline{\mathbf{A}\mathbf{x}}\|_2^2$, where the subsampling can take

the form of either simultaneous "super-shots" or of random subsets of source locations. As in the SPGℓ_1 algorithm, each new (subsampled) LASSO problem is warm-started with the solution $\bar{\mathbf{x}}$ to the previous LASSO problem on the way to solving (3) . The algorithm can be seen as 'curve'-hopping within a random family of Pareto curves corresponding to the subsampled realizations of $\underline{\mathbf{b}}$ and $\underline{\mathbf{A}}$.

This approach dramatically decreases time per iteration, because computational effort for every iteration of the SPG algorithm is reduced by a factor of K'/K. While convergence to the true solution of (3) is still an open question, the approach works well in practice and allows sparsity promotion techniques to be used on a previously inaccessible scale.

A natural question from a practical standpoint is whether it is better to do random source encoding (random Gaussian mixing), or to select random subsets of shots and frequencies at each LASSO subproblem. This is investigated in the context of a large-scale seismic example in the next section.

3. CASE STUDY: THE BG COMPASS MODEL

To test our imaging algorithm for the two dimensionality-reduction scenarios in a realistic setting, we consider a synthetic velocity model with a large degree of complexity. To build the background-velocity model, we employ our modified Gauss-Newton method with sparse updates in 10 overlapping frequency bands on the interval $2.9 - 22.5$Hz and with initial model plotted in Figure 1. (Note that this approach reported in [6] is based on a similar dimensionality reduction technique as presented in this paper.) The output of this procedure is plotted in Figure 1 and is used as the background-velocity model for our imaging algorithm.

We parametrize the velocity perturbation on a 409×1401 grid with a grid size of $5\,\mathrm{m}$. We use a Helmholtz solver to generate data from 350 source positions sampled at an interval of $20\,\mathrm{m}$ and 701 receivers sampled with an interval of $10\,\mathrm{m}$. We use 10 random frequencies selected from $20 - 50\,\mathrm{Hz}$ and scaled by the spectrum of a $30\,\mathrm{Hz}$ Ricker wavelet. We solve 10 LASSO subproblems with and without independent redraws of $\mathbf{RM}$. The results are summarized in Figure 2 (b–c) and clearly show significant improvements from the redraws. Not only is the crosstalk removed more efficiently but the reflectors are also better imaged in particular at the deeper parts of the model where recovery without redraws is not able to image the events. It is also interesting to see that replacing the Gaussian measurement matrix by the identity matrix—i.e., replacing simultaneous shots by randomly selected sequential shots—does not seriously affect the imaging result.

4. DISCUSSION AND CONCLUSIONS

We introduced an efficient algorithm to solve the linearized imaging problem. Our method combines recent findings from the fields of stochastic optimization and compressive sensing and turns the originally "overdetermined" seismic imaging problem into a series of underdetermined dimensionality-reduced subproblems. By considering these subproblems as sparse-recovery problems, we were able to create high-fidelity images at a fraction of the computational cost for subsampling scenarios based on either randomized superposition or on random selections of subsets of sequential sources. The latter relaxes the condition of having stationary receivers, which the method conducive to marine acquisition.

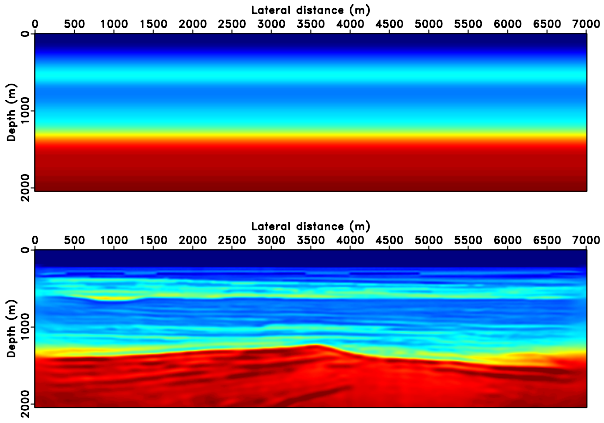

Fig. 1. Full-waveform inversion result. **(a)** Initial model. **(b)** Inverted result starting from 2.9Hz with 7 simultaneous shots and 10 frequencies in each of the 10 frequency bands.

5. REFERENCES

[1] S. A. Morton and C. C. Ober, "Faster shot-record depth migrations using phase encoding," in *SEG Technical Program Expanded Abstracts*. 1998, vol. 17, pp. 1131–1134, SEG.

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

[3] Felix J. Herrmann and Xiang Li, "Efficient least-squares migration with sparsity promotion," EAGE, 2011, EAGE Technical Program Expanded Abstracts.

[4] Felix J. Herrmann and Xiang Li, "Efficient least-squares imaging with sparsity promotion and compressive sensing.," Tech. rep., University of British Columbia, Vancouver, August 2011.

[5] R. Plessix and W. Mulder, "Frequency-domain finite difference amplitude-preserving migration," *Geoph. J. Int.*, vol. 157, pp. 975–987, 2004.

[6] Xiang Li, Aleksandr Aravkin an Tristan van Leeuwen, and Felix J. Herrmann, "Modified gauss-newton with sparse updates," SBGF, 2011.

[7] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, January 2008.

[8] G. Hennenfent, E. van den Berg, M. P. Friedlander, and F. J. Herrmann, "New insights into one-norm solvers from the Pareto curve," *Geophysics*, vol. 73, no. 4, July-August 2008.

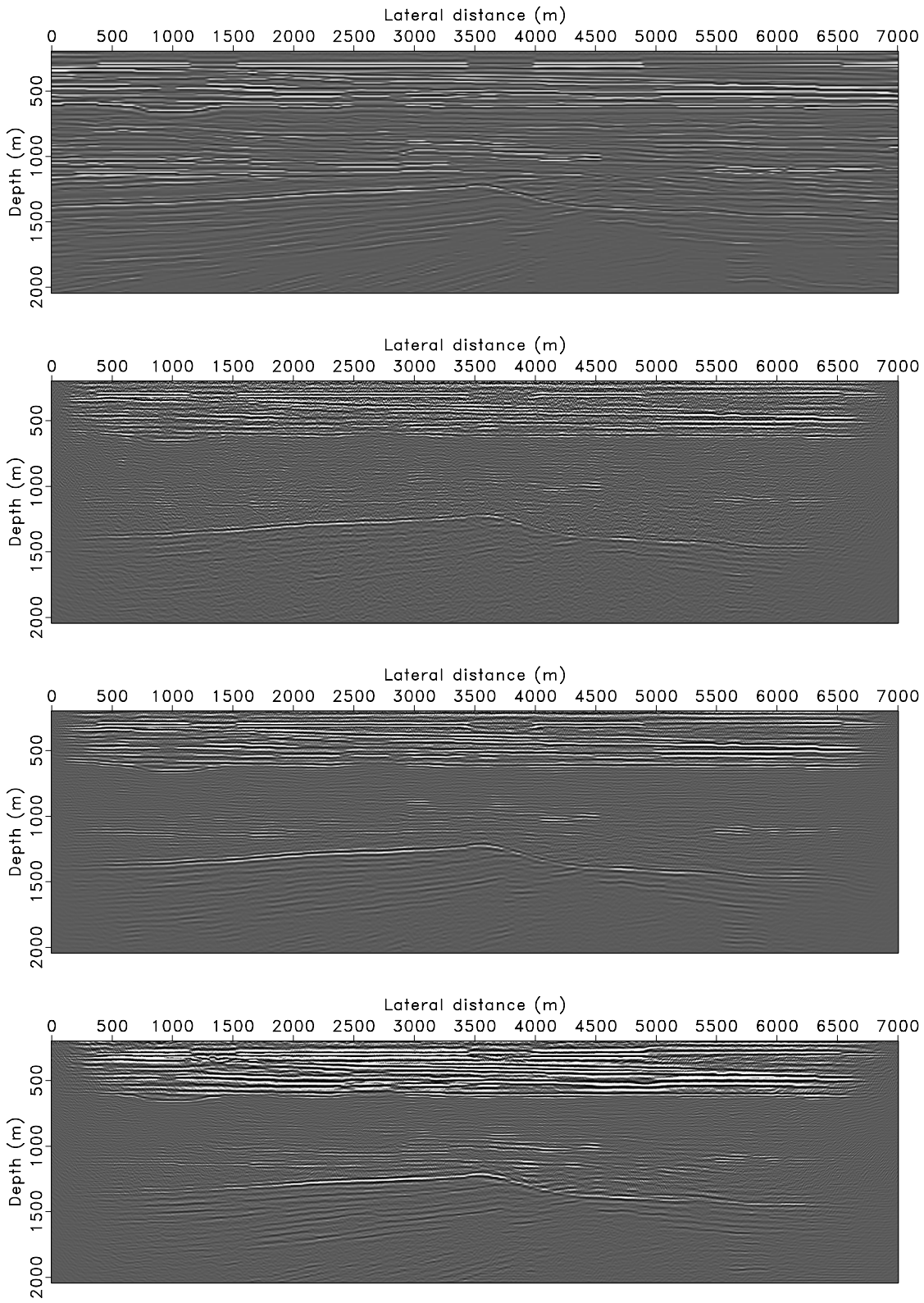

Fig. 2. Dimensionality-reduced sparsity-promoting imaging from random subsets of 17 simultaneous shots and 10 frequencies. We used the background velocity-model plotted in Figure 1 **(a)** True perturbation given by the difference between the true velocity model and the FWI result. **(b)** Imaging result without redraws for the supershots. **(c)** The same but with redraws. **(d)** Image for sequential shots with redraws. Notice the significant improvement in image quality when renewing collection of supershots after solving each LASSO subproblem. Also, notice that the result with randomized subsets of sequential shots is also very reasonable.