Fast least-squares migration with multiples and source estimation

Ning Tu [1], Aleksandr Y. Aravkin[1,2], Tristan van Leeuwen[1] and Felix J. Herrmann[1]

[1] Dept. of Earth and Ocean sciences,[2] Dept. of Computer Science,

University of British Columbia, Vancouver, BC, Canada

February 4, 2013

Abstract

The advent of modern computing has made it possible to do seismic imaging using least-squares reverse-time migration. We obtain superior images by solving an optimization problem that recovers the true-amplitude images. However, its success hinges on overcoming several issues, including overwhelming problem size, unknown source wavelet, and interfering coherent events like multiples. In this abstract, we reduce the problem size by using ideas from compressive sensing, and estimate source wavelet by generalized variable projection. We also demonstrate how to invert for subsurface information encoded in surface-related multiples by incorporating the free-surface operator as an areal source in reverse-time migration. With multiples we can remove the amplitude ambiguity in wavelet estimation. We demonstrate the efficacy of the proposed method with synthetic examples.

Introduction

The goal of least-squares migration is to obtain the true-amplitude model perturbation $\delta\mathbf{m}$, by solving the following optimization problem in the frequency domain:

$$\mathbf{LS}(\mathbf{q}): \quad \underset{\delta\mathbf{m},\mathbf{q}}{\text{minimize}} \sum_{i=1}^{n_f} \|\mathbf{d}_i - \nabla\mathbf{F}_i[\mathbf{m_0}, \mathbf{Q}(q_i)]\delta\mathbf{m}\|_2^2,$$

where i indexes frequency, and $\mathbf{d}$ is the vectorized seismic data free of direct waves and multiples. The matrix $\nabla\mathbf{F}$ is the linearized Born scattering operator (demigration operator), with $\mathbf{Q}$ the monochromatic source wavefield (a function of the *unknown* source wavelet $\mathbf{q} = [q_1, q_2, \cdots, q_{n_f}]$), and $\mathbf{m_0}$ the known background velocity model. We form the operator $\nabla\mathbf{F}$ based on the two-way wave equation. In this case, applying the adjoint of $\nabla\mathbf{F}$ is widely known as the reverse-time migration (RTM). We assume that the source is omnidirectional with time-harmonic source signature $\mathbf{q}$, i.e., $\mathbf{Q}(q_i) = q_i\mathbf{I}$ in the case that the data $\mathbf{d}$ only contains primaries. When there are both primaries and multiples in the data, we follow Tu and Herrmann (2012) and solve the problem by including the total data in the source wavefield, i.e., $\mathbf{Q}(q_i) = q_i\mathbf{I} - \mathbf{D}_i$.

There are three challenges to solving this problem efficiently: prohibitive computational cost, non-convexity of the formulation, and removal of surface-related multiples. First, due to the enormous size of typical seismic data, it is necessary to use an iterative solver. During each iteration, the evaluation of the objective function of problem $\mathbf{LS}(\mathbf{q})$ involves two-way wave-equation based simulations, which requires four PDE solves for each monochromatic source experiment. Second, the formulation involves two unknowns and is non-convex, making it more difficult to solve than the standard formulation where the wavelet is known. Finally, reliable multiple removal using e.g. the SRME method remains challenging and computational expensive because they involve I/O intensive matrix multiplications (Verschuur et al., 1992).

In this abstract, we follow Herrmann and Li (2012) to reduce the computational cost, and use the generalized variable projection method by Aravkin and van Leeuwen (2012) for wavelet estimation. We investigate the possibility of including multiples in the inversion following Tu and Herrmann (2012).

Compressed imaging by sparsity promotion

Following the earlier work by Herrmann and Li (2012), we reduce the number of PDE solves that dominates the computation of solving problem $\mathbf{LS}(\mathbf{q})$, by forming randomized simultaneous sources and randomly selecting a frequency subset. Because of the subsampling, the corresponding linearized modelling operator $\nabla\mathbf{F}$ can be underdetermined, making $\mathbf{LS}(\mathbf{q})$ an ill-posed problem. This ill-posedness is exacerbated by the fact that $\nabla\mathbf{F}$ usually has a null space caused by complex overburdens. To address the ill-posedness, we look for the sparsest solution using the ℓ_1-norm. To maximally exploit the sparsity, we follow Herrmann and Li (2012) to incorporate the curvelet transform that is known to lead to fast decaying magnitude-sorted coefficients for seismic images (Candès et al., 2006). Now we formulate problem $\mathbf{LS}(\mathbf{q})$ as a basis-pursuit denoise (BPDN) type of problem:

$$\mathbf{BPDN}(\sigma,\mathbf{q}): \quad \underset{\mathbf{x},\mathbf{q}}{\text{minimize}} \|\mathbf{x}\|_1 \quad \text{subject to} \sum_{i\in\mathbb{F}} \|\underline{\mathbf{d}}_i - \nabla\mathbf{F}_i[\mathbf{m_0}, q_i\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x}\|_2^2 \leq \sigma^2,$$

where $\mathbb{F}$ is the frequency subset used for inversion, underlined variables indicate subsampled quantities (see e.g., Herrmann and Li, 2012, for more details). The matrix $\mathbf{S}^*$ is the curvelet synthesis operator where * denotes the conjugate transpose, and $\mathbf{x}$ is the curvelet coefficients.

Solving the BPDN problem using alternate formulations

The problem $\mathbf{BPDN}(\sigma,\mathbf{q})$ is not a standard optimization problem. Rather than applying an iterative optimization method directly to $\mathbf{BPDN}(\sigma,\mathbf{q})$, we would prefer to work with the following formulation:

$$\mathbf{LASSO}(\tau,\mathbf{q}): \quad \underset{\mathbf{x},\mathbf{q}}{\text{minimize}} \sum_{\mathbf{i}\in\mathbb{F}} \|\underline{\mathbf{d}}_i - \nabla\mathbf{F}_i[\mathbf{m_0}, q_i\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x}\|_2^2 \quad \text{subject to } \|\mathbf{x}\|_1 \leq \tau,$$

which can be solved using a projected gradient method. The challenge is that the sparsity parameter τ is not known, and is harder to come up ith than the noise level σ in problem $\mathbf{BPDN}(\sigma,\mathbf{q})$.

Aravkin et al. (2012) showed that for a very general class of problems, the set of minimizers of problem $\mathbf{LASSO}(\tau,\mathbf{q})$ and $\mathbf{BPDN}(\sigma,\mathbf{q})$ match, as long as any minimizer of problem $\mathbf{BPDN}(\sigma,\mathbf{q})$ $\tilde{\mathbf{x}}$ satisfies the constraint to equality, i.e., $\sum_{i\in\mathbb{F}}\|\underline{\mathbf{d}}_i-\nabla\mathbf{F}_i[\mathbf{m_0},q_i\underline{\mathbf{I}}]\mathbf{S}^*\tilde{\mathbf{x}}\|_2^2=\sigma^2$. If this holds, we also know that $\|\tilde{\mathbf{x}}\|_1=\tau$ for problem $\mathbf{LASSO}(\tau,\mathbf{q})$. This general result has made it possible to solve problem $\mathbf{BPDN}(\sigma,\mathbf{q})$ by finding the $\overline{\tau}$ such that the optimal value of $\mathbf{LASSO}(\overline{\tau},\mathbf{q})$ is precisely σ^2, which can be done using Newton's method, just as for the original problem in van den Berg and Friedlander (2008). The approach requires solving a series of $\mathbf{LASSO}(\tau,\mathbf{q})$ sub-problems, with τ values dictated by the Newton's method.

Wavelet estimation by variable projection

The structure of $\mathbf{LASSO}(\tau,\mathbf{q})$ allows us to find a closed form solution in $\mathbf{q}$ for any fixed vector $\mathbf{x}$:

$$\tilde{q}_i(\mathbf{x})=(\nabla\mathbf{F}_i[\mathbf{m_0},\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x})\mathbf{d}_i/(\|\nabla\mathbf{F}_i^*[\mathbf{m_0},\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x}\|_2^2), \tag{1}$$

by solving the optimization problem: $\text{minimize}_{q_i}\|\mathbf{d}_i-q_i\nabla\mathbf{F}_i^*[\mathbf{m_0},\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x}\|_2^2$ for each frequency, as was used in Pratt (1999) for waveform inversion applications. This allows us to eliminate $\mathbf{q}$ as a variable in the problem, and think of it as a function of $\mathbf{x}$ in $\mathbf{LASSO}(\tau,\mathbf{q})$:

$$\mathbf{LASSO}(\tau):\quad \underset{\mathbf{x}}{\text{minimize}}\sum_{\mathbf{i}\in\mathbb{F}}\|\underline{\mathbf{d}}_i-\nabla\mathbf{F}_i[\mathbf{m_0},\tilde{q}_i(\mathbf{x})\underline{\mathbf{I}}]\mathbf{S}^*\mathbf{x}\|_2^2\quad \text{subject to }\|\mathbf{x}\|_1\leq\tau.$$

At first glance this seems to complicate the problem, since one needs to be able to compute derivatives with respect to $\mathbf{x}$ to implement a projected gradient method for $\mathbf{LASSO}(\tau)$. However, it is shown in Aravkin and van Leeuwen (2012) that a stationary point of $\mathbf{LASSO}(\tau)$ remains a stationary point of problem $\mathbf{LASSO}(\tau,\mathbf{q})$ when the value (1) is used, and the gradients of the objective function in problem $\mathbf{LASSO}(\tau)$ and $\mathbf{LASSO}(\tau,\mathbf{q})$ are equal for this value of $\mathbf{q}$. Therefore we can use any gradient based method that is designed to solve problem $\mathbf{LASSO}(\tau,\mathbf{q})$ to solve $\mathbf{LASSO}(\tau)$, as long as we update (1) whenever $\mathbf{x}$ is modified. This idea is called *variable projection* in the context of nonlinear least-squares (Golub and Pereyra, 2003), and the theoretical results required to generalize it to sparse constrained case are given in Aravkin and van Leeuwen (2012).

As a joint inversion problem, $\mathbf{LASSO}(\tau,\mathbf{q})$ is still plagued by the amplitude ambiguity, i.e., a scaling of the wavelet $\mathbf{q}$ by a constant can be compensated by dividing the perturbation $\delta\mathbf{m}$ by the same constant. One interesting aspect of variable projection is that the formula (1) makes a unique choice of $\mathbf{q}$; a different choice can be obtained if necessary by normalizing the wavelet. However, the ambiguity problem in general can be solved by using formulations that include multiples, which will be discussed in the following section.

Using surface-related multiples

Conventionally, multiples are treated as noise, largely because they can result in coherent artifacts in the seismic image that hinders correct interpretation. However, Verschuur (2011); Tu and Herrmann (2012) have demonstrated that these artifacts can be removed by doing least-squares migration. We follow Tu and Herrmann (2012), and map the multiples to the right subsurface positions by including the total down-going wavefield $-\mathbf{D}_i$ in the source wavefield, i.e., $\mathbf{Q}_i=q_i\mathbf{I}-\mathbf{D}_i$. With this relatively simple operation that is derived from the SRME formulation, we are still able to speed up the inversion by subsampling over sources and frequencies (Tu and Herrmann, 2012).

Computationally, the use of the total data obviates the de-multiple procedure in data pre-processing, as long as we still have access to a reasonably accurate background model. Furthermore, the use of multiples also effectively removes the amplitude ambiguity in this blind-deconvolutional type of inversion, by having a fixed-scale term $\mathbf{D}$ in the source wavefield. However, including multiples in the inversion can lead to trivial solutions, i.e., $\mathbf{q}=\mathbf{0}$, and $\delta\mathbf{m}$ are Dirac functions. To avoid this case, we apply a mute operation to remove the perturbation in the water column after each iteration.

Synthetic examples

In this section, we demonstrate the proposed method using the synthetic Compass model (cropped, courtesy of BG group). The model is discretized onto a 341 by 417 grid with six meter spacing. We use 209 co-located sources and receivers with 12m spacing at the depth of six meters. The source wavelet is a Ricker wavelet with a peak frequency of 20Hz and a time delay of 0.1s. We make linearized data, i.e., $\mathbf{d} = \nabla\mathbf{F}\delta\mathbf{m}$. In the inversion, we use 30 simultaneous sources and 15 frequencies, which leads to a number of 94050 equations and 142197 unknown medium parameters (1.13 million curvelet coefficients). Compared with using the full data that has 160 frequencies and 209 sequential sources, we reduce the problem size by about 74 folds. For the experiments, we use our parallel framework for frequency- domain seismic inversion (van Leeuwen, 2012). The proposed algorithm is an extension of SPGℓ_1 (van den Berg and Friedlander, 2008) that includes variable projection.

We first compare three scenarios using only primaries. In the first scenario, we assume that we know the true wavelet, i.e., the one used to model the data. In the second scenario, we simply use an impulsive source excited at $t = 0$. In the third scenario, the initial guess of the source wavelet is also an impulse; however, it is updated every time we evaluate the gradient at an updated solution vector $\mathbf{x}$. The wavelet estimate is scaled to have a maximal amplitude of one after each update. In all scenarios, we choose a zero σ in equation $\mathbf{BPDN}(\sigma,\mathbf{q})$, and run 100 iterations. Finally we run the third scenario again but using data with multiples, which we synthesize with the model-space multiple predicting operator of SRME (Tu and Herrmann, 2012). For a fair comparison between with and without using multiples, we apply the mute operation in the water column for all scenarios. The inversion results are shown in Figure (1). SNRs are computed w.r.t. the true perturbation. The true and estimated source wavelets at the 15 given frequencies are plotted in Figure (2). From these figures we can see that a wrongly identified source wavelet can significantly degrade the imaging quality in least-squares migration. By updating the wavelet estimate during the inversion process using variable projection, we not only successfully reconstructed the model perturbation, we also recovered the relative amplitude and phase of the wavelet with very high accuracy. The use of multiples not only yielded an image that is virtually free of coherent artifacts, it also helped to retrieve the true amplitude of both the model perturbation and the wavelet.

Conclusions

In this abstract, we proposed a method to simultaneously invert the model perturbation as well as the source wavelet in a computationally efficient way using sparsity-promoting reverse-time migration. We demonstrated its efficacy using synthetic examples. We also showed the importance of including the wavelet estimation in our wave-equation based linearized inversion technique. In the end, we demonstrated how to include multiples in the inversion, which partially obviates the de-multiple procedure in data processing, and removes the amplitude ambiguity in wavelet estimation.

References

Aravkin, A.Y., Burke, J.V. and Friedlander, M.P. [2012] Variational properties of value functions. *submitted to SIAM J. Optimization, arXiv:1211.3724*.

Aravkin, A.Y. and van Leeuwen, T. [2012] Estimating nuisance parameters in inverse problems. *Inverse Problems*, **28**(11).

Candès, E.J., Demanet, L., Donoho, D.L. and Ying, L. [2006] Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, **5**(3), 861–899, doi:10.1137/05064182X.

Golub, G. and Pereyra, V. [2003] Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, **19**(2), R1–R26.

Herrmann, F.J. and Li, X. [2012] Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical Prospecting*, **60**(4), 696–712.

Pratt, R.G. [1999] Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, **64**(3), 888–901.

Tu, N. and Herrmann, F.J. [2012] Least-squares migration of full wavefield with source encoding. *EAGE technical program*, EAGE.

van den Berg, E. and Friedlander, M.P. [2008] Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31**(2), 890–912.

van Leeuwen, T. [2012] A parallel matrix-free framework for frequency-domain seismic modelling, imaging and inversion in matlab.

Verschuur, D.J. [2011] Seismic migration of blended shot records with surface-related multiple scattering. *Geophysics*, **76**(1), A7–A13.

Verschuur, D.J., Berkhout, A.J. and Wapenaar, C.P.A. [1992] Adaptive surface-related multiple elimination. *Geophysics*, **57**(9), 1166–1177.

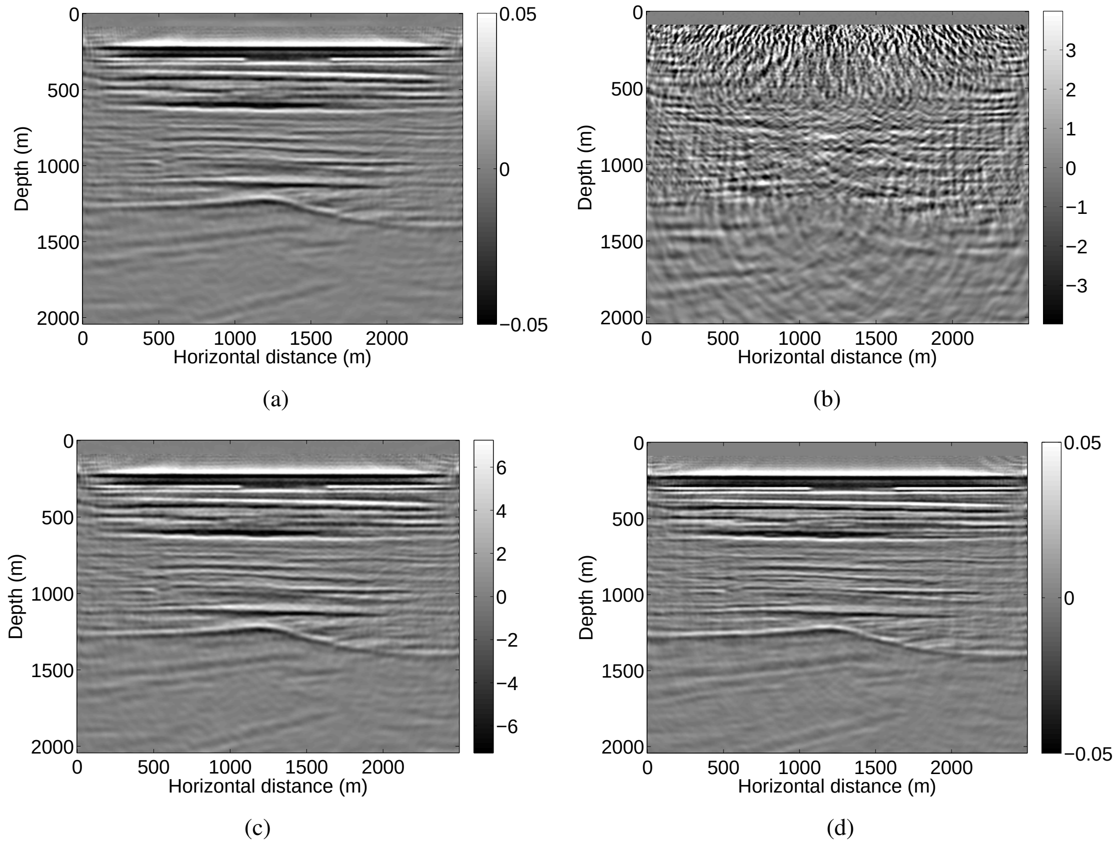

Figure 1 *The inverted model perturbations.* ***(a)*** *Using the true wavelet. SNR: 3.6dB.* ***(b)*** *Using an impulsive source excited at* $t = 0$*. SNR: -2.1dB (after scaling w.r.t.* ***(a)****).* ***(c)*** *Estimating the wavelet. SNR: 3.8dB (after scaling w.r.t.* ***(a)****).* ***(d)*** *Estimating the wavelet using multiples. SNR: 4.6dB (no scaling).*

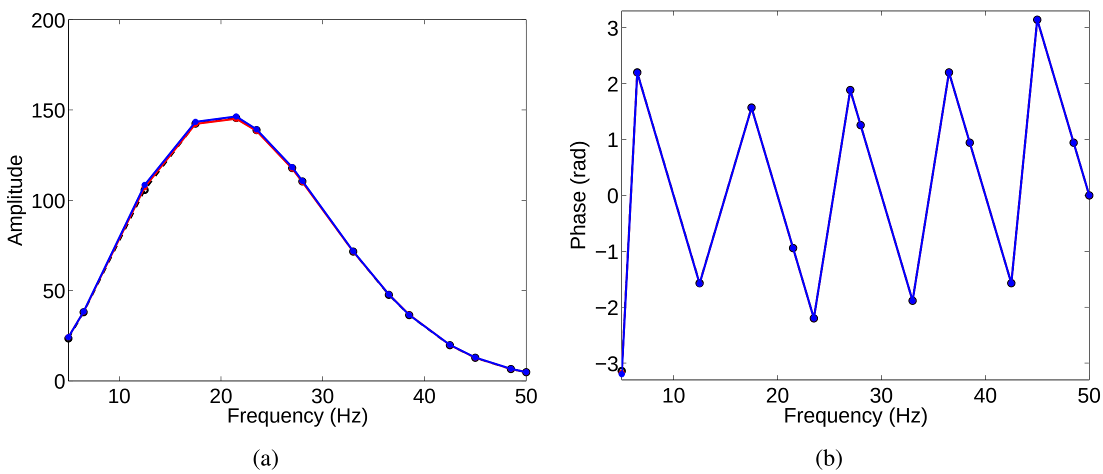

Figure 2 *Compare the true (black dashed line) and the estimated wavelets without (red solid line) and with (blue solid line) multiples.* ***(a)*** *The amplitude spectra. The wavelet estimate without using multiples are scaled w.r.t. the true wavelet.* ***(b)*** *The phase spectra.*