Preconditioning full waveform inversion with phase-encoded Hessian

Yaxun Tang, Stanford University and Sunwoong Lee, ExxonMobil Upstream Research Company*

SUMMARY

Full waveform inversion (FWI) has received an increasing amount of attention thanks to its ability to provide a high-resolution velocity model of the subsurface. The computational cost still presents a challenge, however, and the convergence rate of the FWI problem is usually very slow without proper preconditioning of the gradient. While preconditioners based on the Gauss-Newton Hessian matrix can provide significant improvements in the convergence of FWI, computation of the Hessian matrix itself has been considered highly impractical due to its cost in computational time and storage requirements. In this paper, we design preconditioners based on an approximate Gauss-Newton Hessian matrix obtained using the phase-encoding method. The new method requires only $2N_s$ forward simulations compared to $N_s(N_r+1)$ forward simulations required in conventional approaches, where N_s and N_r are the numbers of sources and receivers, respectively. We apply the diagonal of the phase-encoded Gauss-Newton Hessian to both sequential-source FWI and encoded simultaneous-source FWI. Numerical examples using a truncated Marmousi2 model demonstrate that the phase-encoded Gauss-Newton Hessian improves the convergence of the FWI significantly.

INTRODUCTION

Velocity estimation is a challenging task in seismic exploration. In the past decade, ray-based velocity estimation methods have been widely used in practice to build the velocity model. Although the ray-based method is efficient and robust, the infinite-frequency assumption and inherent caustics in ray theory prevent it from accurately handling complex geologies. On the other hand, full-waveform inversion (FWI), which uses band-limited wavefields instead of infinite-frequency rays as carriers of information, can properly model the finite-frequency effects of seismic wavefields and is immune from multi-pathing issues. Moreover, FWI uses full waveforms in the observed data for velocity estimation. Therefore, FWI has the potential to provide a detailed velocity model even in complex geological scenarios. The theoretical advantages, as well as recent advances in seismic acquisition technologies and computer sciences, have made FWI an attractive tool for accurate and high-resolution velocity model estimation.

FWI is usually implemented within the framework of nonlinear least-squares inversion, where a data residual defined either in time or frequency domain is minimized in the least-squares sense (Tarantola, 1984; Pratt, 1999). Since this is a large-scale nonlinear optimization problem, gradient-based local optimization methods are often used in practice. Although the gradient of the least-squares misfit functional can be efficiently calculated using the adjoint state method, a gradient-only method, such as steepest descent and conjugate gradient, converges slowly without proper preconditioning. On the other hand, the Gauss-Newton method, which preconditions the gradient with a Gauss-Newton Hessian, can achieve significantly faster convergence (Pratt et al., 1998). Computing the exact Gauss-Newton Hessian, however, is prohibitively expensive for practical applications due to the size of the problem.

In this paper, we focus on the constant-density acoustic FWI problem and discuss a method based on random-phase encoding to efficiently compute the Gauss-Newton Hessian (Tang, 2009). With this new method, the cost for computing the Hessian diagonal is only two forward modeling steps for each shot. Another important feature of the randomly phase-encoded Hessian is that it converges to the true Gauss-Newton Hessian in a statistical sense, hence it provides a good approximation, the accuracy of which can even be controlled by using different numbers of realizations. The phase-encoded Hessian was first introduced by Tang (2009) and applied to least-squares wave-equation migration/inversion. Here, we show how it can be used in the FWI problem to accelerate convergence. In the next section, we review the theory of the Gauss-Newton Hessian and its approximations. In subsequent sections, we apply the hybrid time/frequency-domain preconditioned FWI algorithms to invert for a truncated Marmousi2 model (Martin, 2004).

THE GAUSS-NEWTON HESSIAN AND ITS APPROXIMATIONS

FWI is a non-linear inverse problem that tries to find an optimum velocity model $\mathbf{v}$ to minimize the difference between synthesized and observed data in a least-squares sense. The objective function to be minimized can be defined in the frequency domain as follows (Pratt et al., 1998):

$$F(\mathbf{v}) = \frac{1}{2}\sum_{\omega,\mathbf{x}_s,\mathbf{x}_r} |w(\mathbf{x}_r,\mathbf{x}_s)[d(\mathbf{x}_r,\mathbf{x}_s,\omega) - d_{\text{obs}}(\mathbf{x}_r,\mathbf{x}_s,\omega)]|^2 \quad (1)$$

where d and d_{obs} are the simulated and observed data, respectively; ω is the angular frequency, $\mathbf{x}_s$ and $\mathbf{x}_r$ are the source and receiver locations, respectively; $w(\mathbf{x}_r,\mathbf{x}_s)$ is the acquisition mask operator, it has unity value where we record data and zero where we do not. The simulated data $d(\mathbf{x}_r,\mathbf{x}_s,\omega)$ is obtained by sampling the forward modeled wavefield at the receiver locations: $d(\mathbf{x}_r,\mathbf{x}_s,\omega) = \sum_{\mathbf{x}} p(\mathbf{x},\mathbf{x}_s,\omega)\delta(\mathbf{x}-\mathbf{x}_r)$, where $p(\mathbf{x},\mathbf{x}_s,\omega)$ is the wavefield at $\mathbf{x}$ due to a point source located at $\mathbf{x}_s$. It satisfies the acoustic wave equation

$$\left[\nabla^2 + \frac{\omega^2}{v^2(\mathbf{x})}\right] p(\mathbf{x},\mathbf{x}_s,\omega) = -\delta(\mathbf{x}-\mathbf{x}_s) f(\omega), \quad (2)$$

where $v(\mathbf{x})$ is the P-wave velocity at $\mathbf{x}$ and $f(\omega)$ is the source signature in the frequency domain.

Taking the first-order partial derivatives of F with respect to

the velocity yields the gradient (Sirgue and Pratt, 2004)

$$g(\mathbf{x}) = -\sum_{\omega}\sum_{\mathbf{x}_s}\frac{2\omega^2 f^*(\omega)}{v_0^3(\mathbf{x})}G_0^*(\mathbf{x},\mathbf{x}_s,\omega) \times\sum_{\mathbf{x}_r} w^2(\mathbf{x}_r,\mathbf{x}_s)G_0^*(\mathbf{x},\mathbf{x}_r,\omega)r(\mathbf{x}_r,\mathbf{x}_s,\omega), \quad (3)$$

where * means complex conjugation; $G_0(\mathbf{x},\mathbf{x}_s,\omega)$ and $G_0(\mathbf{x},\mathbf{x}_r,\omega)$ are the Green's functions connecting the source and receiver, $\mathbf{x}_s$ and $\mathbf{x}_r$, to $\mathbf{x}$ in the subsurface, they are computed using a background velocity model $v_0(\mathbf{x})$; $r(\mathbf{x}_r,\mathbf{x}_s,\omega)$ is the data residual defined as follows:

$$r(\mathbf{x}_r,\mathbf{x}_s,\omega) = d(\mathbf{x}_r,\mathbf{x}_s,\omega) - d_{\text{obs}}(\mathbf{x}_r,\mathbf{x}_s,\omega). \quad (4)$$

The gradient $g(\mathbf{x})$ can be implemented as a reverse time migration, where the forward-propagated source wavefield is cross-correlated with the backward-propagated data residual (Tarantola, 1984).

The exact Gauss-Newton Hessian

The full Hessian operator is obtained by taking second-order derivatives of F with respect to the velocities. Since the wavefield is a nonlinear function of velocity, the full Hessian contains both first- and second-order terms (Pratt et al., 1998; Plessix and Mulder, 2004). The first-order term accounts for single-scattering events and is often known as the Gauss-Newton Hessian; the second-order term, on the other hand, accounts for multiscattering events and is often neglected due to its second-order effect (Pratt et al., 1998). In this paper, we concentrate on the computation of the Gauss-Newton Hessian and assume that the second-order term is relatively small and can be approximated using an identity operator. The Gauss-Newton Hessian takes the form (Plessix and Mulder, 2004; Tang, 2009):

$$H_0(\mathbf{x},\mathbf{y}) = \sum_{\omega}\frac{4\omega^4|f(\omega)|^2}{v_0^3(\mathbf{x})v_0^3(\mathbf{y})}\sum_{\mathbf{x}_s}G_0^*(\mathbf{x},\mathbf{x}_s,\omega)G_0(\mathbf{y},\mathbf{x}_s,\omega) \times\sum_{\mathbf{x}_r} w^2(\mathbf{x}_r,\mathbf{x}_s)G_0^*(\mathbf{x},\mathbf{x}_r,\omega)G_0(\mathbf{y},\mathbf{x}_r,\omega). \quad (5)$$

The diagonal part of the Hessian (when $\mathbf{x}=\mathbf{y}$), which contains autocorrelations of both source and receiver-side Green's functions, can be interpreted as a subsurface illumination map with contributions from both sources and receivers. The rows of the Hessian (for fixed $\mathbf{x}$'s and varying $\mathbf{y}$'s), which contain crosscorrelations of both source and receiver-side Green's functions, can be interpreted as resolution functions (Lecomte, 2008; Tang, 2009). They measure how much smearing an image can have due to a given acquisition setup.

Computing the Gauss-Newton Hessian using equation 5 is challenging even with today's most powerful computers, because it requires either saving a large number of Green's functions, which can take up hundreds of terabytes of disk space easily, or a large number of wavefield propagations (each receiver-side Green's function has to be recomputed for each shot, resulting in a cost proportional to the product of the numbers of sources and receivers).

The source-wavefield intensity approximation

The computational cost can be drastically reduced by using the source-wavefield intensity (SWI) approximation to the true Gauss-Newton Hessian, which simply neglects the receiver-side Green's functions

$$H_0^{\text{SWI}}(\mathbf{x},\mathbf{y}) = \sum_{\omega}\frac{4\omega^4|f(\omega)|^2}{v_0^3(\mathbf{x})v_0^3(\mathbf{y})}\sum_{\mathbf{x}_s}G_0^*(\mathbf{x},\mathbf{x}_s,\omega) \times G_0(\mathbf{y},\mathbf{x}_s,\omega). \quad (6)$$

Neglecting the receiver-side Green's functions makes SWI overestimate the subsurface illumination (the diagonal part), especially for acquisition geometries with short receiver spreads, such as in marine-streamer acquisition, and results in completely wrong resolution functions (the rows of the matrix) (Tang, 2009).

The phase-encoded Gauss-Newton Hessian

Instead of neglecting the receiver-side Green's functions, the phase-encoded Hessian (PEH) approximates them with phase-encoded Green's functions, which are obtained using encoded areal sources as the source function. Mathematically, the receiver-side phase-encoded Hessian takes the form

$$H_0^{\text{PEH}}(\mathbf{x},\mathbf{y}) = \sum_{\omega}\frac{4\omega^4|f(\omega)|^2}{v_0^3(\mathbf{x})v_0^3(\mathbf{y})}\sum_{\mathbf{x}_s}G_0^*(\mathbf{x},\mathbf{x}_s,\omega) \times G_0(\mathbf{y},\mathbf{x}_s,\omega)R_0^*(\mathbf{x},\mathbf{x}_s,\omega)R_0(\mathbf{y},\mathbf{x}_s,\omega), \quad (7)$$

where

$$R_0(\mathbf{x},\mathbf{x}_s,\omega) = \sum_{\mathbf{x}_r} w(\mathbf{x}_r,\mathbf{x}_s)G_0(\mathbf{x},\mathbf{x}_r,\omega)\alpha(\mathbf{x}_r,\mathbf{x}_s,\omega), \quad (8)$$

which is the encoded receiver-side Green's function obtained using an areal source: $-\sum_{\mathbf{x}_r} w(\mathbf{x}_r,\mathbf{x}_s)\delta(\mathbf{x}-\mathbf{x}_r)\alpha(\mathbf{x}_r,\mathbf{x}_s,\omega)$, where α is a weighting function, or more precisely, an encoding function, to be specified later. It is obvious from equation 7 that the computational cost for the PEH is independent of the number of receivers as opposed to the exact Gauss-Newton Hessian (equation 5), where the computational cost is proportional to the number of receivers.

However, the phase-encoded Hessian brings unwanted crosstalk. This becomes clear once we substitute equation 8 into equation 7, and expand the terms. The crosstalk can be suppressed by carefully choosing the phase-encoding function α (Tang, 2009). In this paper, we define α as follows:

$$\alpha(\mathbf{x}_r,\mathbf{x}_s,\omega) = \frac{1}{\sqrt{M}}e^{i\gamma(\mathbf{x}_r,\mathbf{x}_s,\omega)}, \quad (9)$$

where $\gamma(\mathbf{x}_r,\mathbf{x}_s,\omega)$ is a uniformly distributed random sequence from 0 to 2π and M is the number of realizations of the random sequence. With this choice of random sequence, the expectation of the crosstalk becomes zero. According to the law of large numbers, the ensemble average of a random sequence converges to its expectation (Gray and Davisson, 2003), the phase-encoded Hessian therefore converges to the true Hessian in a statisitical sense by stacking over frequencies, shots, receivers, and random realizations. For a given acquisition geometry, i.e., for a fixed number of frequencies, shots and receivers, the number of realizations M then controls the accuracy and cost of this approximation.

A simple example which compares the true Gauss-Newton Hessian with the approximated ones for a constant velocity model is shown in Figures 1 and 2. The SWI overestimates the subsurface illumination (Figures 1(c)) and produces a wrong resolution function (Figure 2(c)). The PEH, on the other hand, better estimates the subsurface illumination (Figure 1(d)) and the resolution function (Figure 2(d)). The random phase encoding function also significantly attenuates the crosstalk artifacts shown in Figures 1(b) and 2(b).

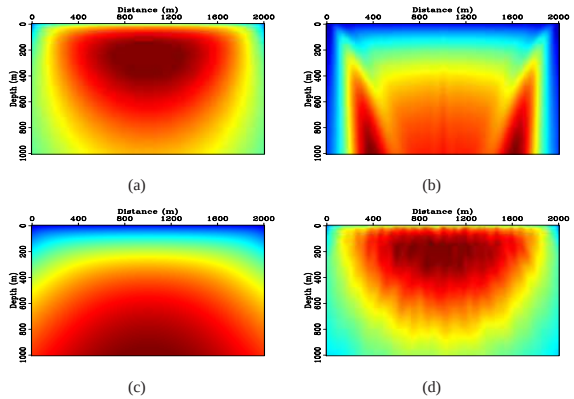


Figure 1: Diagonal of the Gauss-Newton Hessians. All panels have been scaled by depth, with (a) being the true Gauss-Newton Hessian, (b) the Hessian with crosstalk when $\alpha(\mathbf{x}_r, \mathbf{x}_s, \omega) = 1$, (c) the SWI, and (d) the PEH with 1 realization ($M = 1$).

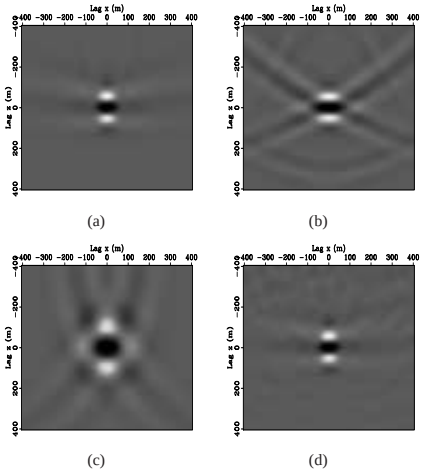


Figure 2: One row of the Gauss-Newton Hessians. View descriptions are the same as in Figure 1.

PRECONDITIONED INVERSION ALGORITHMS

Since it is easier to compute the phase-encoded Gauss-Newton Hessian in the frequency domain, we have developed a hybrid time/frequency-domain preconditioned inversion approach, where the wavefield propagation and gradient are computed in the time domain, but the phase-encoded Gauss-Newton Hessian is computed in the frequency domain with the help of the discrete Fourier transform (DFT) (Sirgue et al., 2008). Only the diagonal part of the Hessian has been used to precondition the gradient, and two different algorithms have been implemented, i.e., the preconditioned conjugate gradient (PCG) method and the pseudo Gauss-Newton (PGN) method. These two algorithms differ only in the way they compute the search direction: In PCG, the search direction is computed using

$$\mathbf{p}_k = -\mathbf{s}_k + \frac{\mathbf{s}_k^T(\mathbf{s}_k - \mathbf{s}_{k-1})}{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}} \mathbf{p}_{k-1}; \tag{10}$$

In PGN, the search direction is determined by

$$\mathbf{p}_k = -\mathbf{s}_k, \tag{11}$$

where $\mathbf{s}_k$ is the preconditioned gradient at the kth iteration

$$\mathbf{s}_k = \frac{\mathbf{g}_k}{\text{diag}\left\{\mathbf{H}_0^k\right\} + \varepsilon \mathbf{I}}, \tag{12}$$

with $\mathbf{I}$ being an identity operator and ε a damping parameter added to stabilize the division.

NUMERICAL EXAMPLES

We test our preconditioned inversion algorithms on a truncated Marmousi2 model (Martin, 2004). The acquisition geometry contains 21 shots with a 300 m spacing and 301 receivers with a 20 m spacing, the receiver spread is fixed for all shots to mimic a land survey. The source function is a minimum-phase Ricker wavelet with an 8-Hz fundamental frequency, which is assumed to be known for the FWI. The starting velocity model is a smoothed version of the true model (not shown here). We apply the PCG and PGN algorithms described in the previous section to both sequential-source FWI (SSFWI) and encoded simultaneous-source FWI (ESSFWI) (Krebs et al., 2009). In the ESSFWI, all 21 shot gathers are encoded into 1 super areal shot gather at each iteration with different binary code (Krebs et al., 2009). We run up to 100 iterations in the SSFWI and 50 iterations in the ESSFWI. Figure 3 shows the normalized gradients at the first iteration for the SSFWI. Note that preconditioning improves the amplitudes in the gradient and precoditioning with the PEH provides a more uniformed amplitude distribution than preconditioning with SWI. Figures 4 and 5 show the inverted velocity model after 20 iterations for SSFWI and ESSFWI, respectively, with different inversion methods and a variety of preconditioning approaches. Preconditioning with the PEH provides the best recovery of the velocity model. The convergence rates of both data misfit (the objective function value) and model misfit (the ℓ_2-norm of the difference between the true and inverted models) are plotted in a logarithmic scale in Figure 6. Preconditioning with the PEH gives a faster convergence rate than the other methods, especially than the one without any preconditioning.

CONCLUSIONS

We have presented a method based on phase encoding to efficiently compute the Gauss-Newton Hessian to improve the ac-

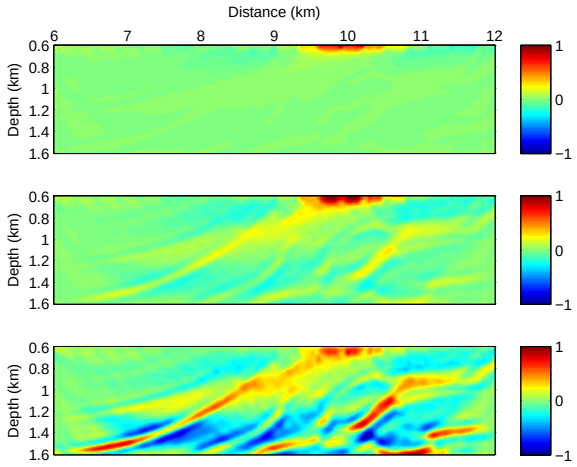


Figure 3: Normalized gradients for the first iteration. From top to bottom: no preconditioning, preconditioning with SWI and preconditioning with PEH.

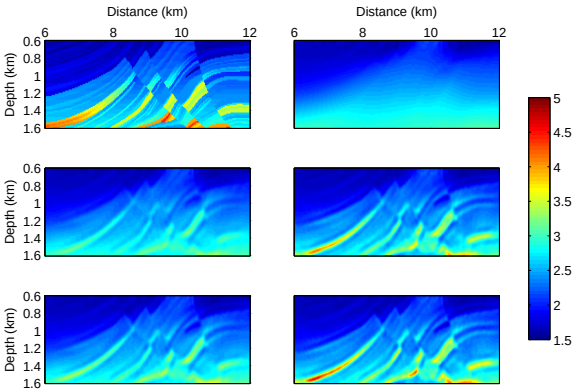


Figure 4: SSFWI after 20 iterations. From left to right and top to bottom: true model, CG without preconditioning, CG preconditioned with SWI, CG preconditioned with PEH, PGN preconditioned with SWI and PGN preconditioned with PEH.

curacy and speed of FWI problem. The phase-encoded Gauss-Newton Hessian has been used in hybrid time/frequency-domain inversion algorithms to precondition the gradient of the objective function with respect to velocity. Numerical examples demonstrate that preconditioning can significantly improve the convergence of both SSFWI and ESSFWI. Preconditioning with the PEH converges faster than preconditioning with SWI, because the PEH is more accurate than the SWI in approximating the true Gauss-Newton Hessian.

ACKNOWLEDGEMENTS

We thank Jerome Krebs for helpful discussions and we acknowledge ExxonMobil Upstream Research Company for permission to publish this work. Y. Tang also thanks the sponsors of the Stanford Exploration Project for their support.

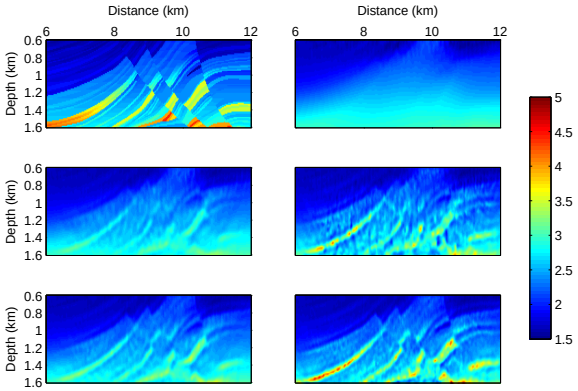
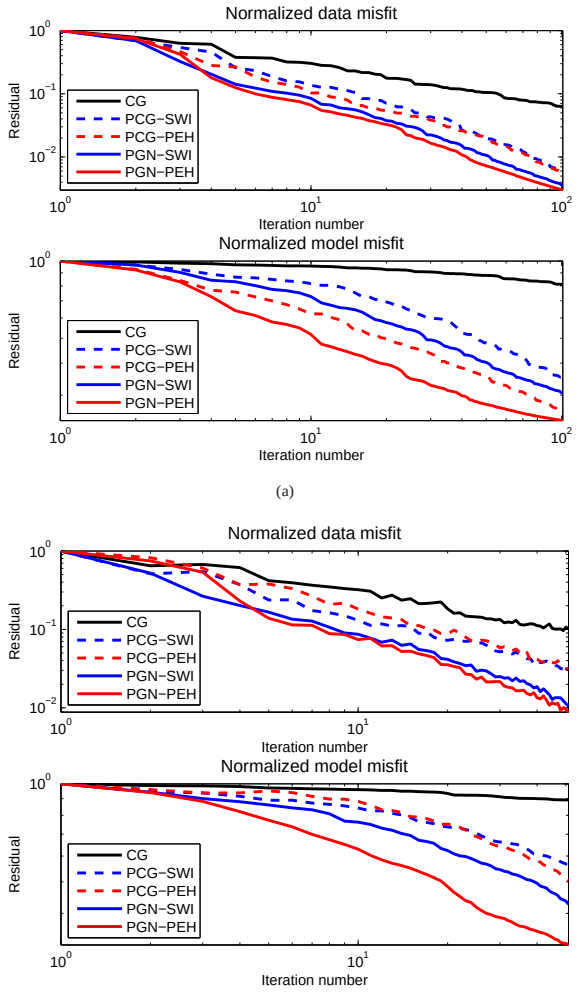


Figure 5: ESSFWI after 20 iterations. View descriptions are the same as in Figure 4.



(a)

(b)

Figure 6: Normalized data (top) and model (bottom) misfit for (a) SSFWI and (b) ESSFWI.

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2010 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Gray, R., and L. D. Davisson, 2003, Introduction to Statistical Signal Processing: Cambridge University Press.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full wavefield seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC177–WCC188, doi: 10.1190/1.3230502.

Lecomte, I., 2008, Resolution and illumination analyses in PSDM: A ray-based approach: The Leading Edge, **27**, 650–663, doi: 10.1190/1.2919584.

Martin, G. S., 2004, The Marmousi2 model, elastic synthetic data, and an analysis of imaging and AVO in a structurally complex environment: M. S. thesis, University of Houston.

Plessix, R.-E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987, doi: 10.1111/j.1365-246X.2004.02282.x.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901, doi: 10.1190/1.1444597.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362, doi: 10.1046/j.1365-246X.1998.00498.x.

Sirgue, L., J. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Conference & Technical Exhibition, EAGE, Extended Abstracts, F022.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, doi: 10.1190/1.1649391.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: Geophysics, **74**, no. 6, WCA95–WCA107, doi: 10.1190/1.3204768.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.