Linearized waveform inversion with (small) velocity updating

Alejandro Cabrales-Vargas, Biondo Biondi, and Robert Clapp

ABSTRACT

Current implementations of linearized waveform inversion rely on an optimum background model, and only allow updating the high wavenumber component, *a.k.a.* reflectivity. We attempt to take one step further allowing controlled perturbations in the background model. We propose constraining such perturbations in a way that maximizes the stacking power, therefore improving the estimated reflectivity even further. We introduce theoretical insight about what we have called *linearized waveform inversion with velocity updating*.

INTRODUCTION

Reverse-time migration (RTM) (Baysal et al., 1983; Kosloff and Baysal, 1983; Gazdag and Carrizo, 1986; McMechan, 1983) constitutes the best available technique to image the subsurface for petroleum exploration purposes. RTM performance is superior than sophisticated implementations of one-way wave equation migration that flourished since the early 90s (e.g. Stoffa et al., 1990; Ristow and Rühl, 1994; Biondi, 2002). Such solutions aimed at solving steep dipping events. RTM is based on the two-way wave equation solution, which accounts not only for 90° dipping events and beyond, but also for wavepath trajectories that are difficult, if not impossible, to recover with one-way wave equation migration, such as prismatic waves, and even multiples (e.g. Liu et al., 2011, 2015; Wong et al., 2015).

There is an intrinsical limitation of RTM and seismic migration in general. It constitutes the first approximation to the inverse of the seismic modeling experiment, the so-called *adjoint* operator (Claerbout, 2014). As a consequence, the migration image typically suffers from degradation in resolution and incorrect seismic amplitudes, becoming a "blurred" version of the reflectivity image. The main reason is the limited, irregular, and/or sparse acquisition coverage. Intense research has been devoted to producing realistic estimations of the subsurface reflectivity. One product of such research is *linearized waveform inversion* (LWI), more commonly known as *least-squares migration* (e.g. Nemeth et al., 1999; Duquet et al., 2000; Ronen and Liner, 2000; Jiang and Schuster, 2003). This procedure consists of minimizing a scalar misfit function that quantifies the mismatch between synthetic data and recorded data, in the least-squares sense. Synthetic data are produced by applying the *Born modeling operator* to reflectivity models of the subsurface. We aim at finding the reflectivity

model that minimizes the misfit function. We usually perform a gradient-based optimization scheme (Hestenes and Stiefel, 1952) in the data space or in the model space, iteratively updating the reflectivity. In the data space we need the modeling operator (Born operator) and the adjoint (RTM). In the model space we need the *Hessian* operator, which can be constructed by applying the modeling operator followed by the adjoint operator to "spiky" perturbations in the model space. It is often necessary to include constraints in both schemes, in order to reject solutions from either the model null space, or from the overfitting of undesired events of the data (e.g. unphysical events such as noise, or propagation modes not accounted for by our modeling and adjoint operators). In the data space every iteration of LWI costs somewhat more than two conventional migrations, and in the model space every iteration cost one computation of the Hessian. Not surprisingly, the first implementations of LWI used comparatively cheap Kirchhoff-based algorithms. Only after RTM itself became affordable did least-squares RTM (LSRTM) become subject of research (e.g. Ji, 2009; Dai et al., 2010; Wong et al., 2011, among many others), although the original idea can be traced back many years earlier (Ji, 1992). Some techniques have been proposed to tackle the intense computational burden demanded by LWI, such as source blending and target oriented methods (Dai et al., 2013).

So far, LWI methods aim at improving the reflectivity estimation assuming that the model parameters (particularly velocity) are optimum. Therefore, the motivation for this report is envisioning an algorithm capable of performing LWI including controlled velocity perturbations. Such velocity perturbations are expected to be rather small to deserve their incorporation into the velocity model, but significant enough to promote the improvement of the image by maximizing the stacking power. The method is conceptualized as a linear optimization scheme that updates two aspects of the model: a perturbation on the low-wavenumber component (related to background velocity) and a perturbation on the high-wavenumber component (related to reflectivity). In contrast, full-waveform inversion (FWI) is built upon a non-linear optimization scheme that updates the model parameters as a single entity.

This report is organized in three sections. We first make a brief review about basic concepts of LWI and FWI. We then introduce an algorithm to implement LWI with velocity updating using the conjugate gradient method. Finally, we analyze strategies for practical implementation, and discuss potential issues and challenges.

LINEARIZED WAVEFORM INVERSION VS. FULL WAVEFORM INVERSION

In this section we offer a brief discussion about the differences between LWI and FWI.

Following the notation convention proposed by Barnier and Almomin (2014), we characterize the subsurface using model parameters (e.g. slowness, density) encompassed by the real variable vector $\mathbf{m}$. We can split such a variable into low-

wavenumber and high-wavenumber components,

$$\mathbf{m} = \mathbf{b} + \mathbf{r},$$

which can be individually perturbed (Barnier and Almomin, 2014),

$$\mathbf{m} = \mathbf{b_0} + \mathbf{\Delta b} + \mathbf{r_0} + \mathbf{\Delta r}.$$

Full waveform inversion performs the minimization of a misfit function Φ_{FWI} that quantifies the mismatch between the recorded data, $\mathbf{d}_r$, and the modeled data, $\mathbf{d} = \mathcal{L}(\mathbf{m})$, in the least-squares sense:

$$\Phi_{FWI}(\mathbf{m}) = \|\mathcal{L}(\mathbf{m}) - \mathbf{d_r}\|_{\mathbf{2}}^{\mathbf{2}}. \tag{1}$$

Here $\mathcal{L}(\mathbf{m})$ represents the full wave propagation operator. In the case of a constant-density acoustic medium, $\mathbf{m}$ usually represents slowness squared. Therefore, $\mathcal{L}(\mathbf{m})$ is given by

$$\begin{cases} \left[\mathbf{m}\dfrac{\partial^2}{\partial t^2} - \nabla^2\right]\mathbf{u}(\mathbf{x},t) = \mathbf{s}(\mathbf{x},t) \\ \mathcal{L}(\mathbf{m}) = \mathbf{u}(\mathbf{x} = \mathbf{x}_r, t) \end{cases} \tag{2}$$

The first term of Equation 2 represents the solution of the acoustic wave equation in a subsurface medium with slowness squared $\mathbf{m}$, excited by the source function $\mathbf{s}(\mathbf{x},t)$. The second term samples the wavefield at the receivers positions, $\mathbf{x}_r$.

The operator $\mathcal{L}(\mathbf{m})$ is non-linear with respect to the model parameters, although it is linear with respect to the source (Barnier and Almomin, 2014). The non-linearity of FWI with respect $\mathbf{m}$ makes the misfit function non-quadratic, hence it generally has a global minimum and several local minima. For such reason, current implementations of FWI rely on initial models that are assumed to be close enough to the global minimum. FWI is popularly implemented using non-linear conjugate gradient methods and Newton-Raphson methods.

On the other hand, in linearized waveform inversion we keep the background model fixed: $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{\Delta b} + \mathbf{r}_0$. Then we invert for the perturbation of the reflectivity, $\mathbf{\Delta r}$, using the Born approximation, which consists of the linearization of the modeling operator around $\mathbf{r}_0$ [See Barnier and Almomin (2014) for mathematical details]. Setting $\mathbf{r}_0 = 0$ is equivalent to smoothing the background model (Barnier and Almomin, 2014). In the data domain, such smoothing corresponds to removal of direct arrivals and diving waves from the recorded data. Thus, the LWI misfit function is giving by

$$\Phi_{LWI}(\mathbf{m}) = \|L(\mathbf{m}_0)\mathbf{\Delta r} - \left[\mathbf{d}_r - \mathcal{L}(\mathbf{m}_0)\right]\|_2^2 \tag{3}$$

where $\mathcal{L}(\mathbf{m}_0)$ represents the synthetic data obtained using the background model, $\mathbf{m}_0$. These data contains no reflections, just direct arrivals and diving waves. They are subtracted from the recorded data, $\mathbf{d}_r$ (in practice, "surgical" filters are designed

to remove such seismic events). $L(\mathbf{m}_0)$ constitutes the Born modeling operator, which operates on the reflectivity, $\mathbf{\Delta r}$, as follows:

$$\begin{cases} \left[\mathbf{m}_0 \dfrac{\partial^2}{\partial t^2} - \nabla^2\right] \mathbf{u}_0(\mathbf{x},t) = \mathbf{s}(\mathbf{x},t) \\ \left[\mathbf{m}_0 \dfrac{\partial^2}{\partial t^2} - \nabla^2\right] \mathbf{u}(\mathbf{x},t) = \mathbf{\Delta r}(\mathbf{x},t) \dfrac{\partial^2 \mathbf{u}_0(\mathbf{x},t)}{\partial t^2} \\ L(\mathbf{m}_0)\mathbf{\Delta r}(\mathbf{x},t) = \mathbf{u}(\mathbf{x}=\mathbf{x}_r,t). \end{cases} \tag{4}$$

$\mathbf{u}_0$ represents the source wavefield, $\mathbf{u}$ represents the *scattered wavefield*, which uses the source wavefield interacting with the perturbation in reflectivity as a source term. This operator is linear with respect to $\mathbf{\Delta r}$(*a.k.a* the seismic image), so the LWI misfit function (Equation 3) is quadratic. There is only one global minimum, although not necessarily a unique solution. The reason is because the limited bandwidth of the data implies the presence of a model null space, that we normally reduce with model regularization. Using gradient-based iterative solvers one could theoretically reach the minimum after enough iterations. In practice, seismic data are huge, allowing just few tens of iterations. Even if we could afford more iterations, the data null space would prevent fitting exactly the data (Aster et al., 2013). Fortunately, we do not have to fit the data exactly in order to get useful results.

The premise behind LWI with velocity updating is that the full FWI Hessian can be expressed as the sum of two components (Biondi et al., 2015): the Gauss-Newton Hessian, $\mathbf{H}_{GN}$, and the "wave-equation migration velocity analysis" (WEMVA) Hessian, $\mathbf{H}_W$, or simply WEMVA operator, $\mathbf{W}$. In FWI the Gauss-Newton Hessian constitutes the product of a *Jacobian matrix*, $\mathbf{J}(\mathbf{m})$, pre-multiplied by its adjoint. The Jacobian matrix is obtained by deriving the synthetic wavefield with respect to the model parameters, and evaluating the matrix at the current model.

In the case of LWI, from Equation 3 the Jacobian matrix becomes simply

$$\mathbf{J}(\mathbf{\Delta r}) = \mathbf{L},$$

where $\mathbf{L}$ is the matrix representation of the linear Born modeling operator, $L(\mathbf{m}_0)$. Hence, the corresponding Gaussian-Newton Hessian is given by

$$\mathbf{H}_{GN} = \mathbf{L}^T\mathbf{L}. \tag{5}$$

The WEMVA component of the full Hessian, $\mathbf{W}$, requires the derivative of Jacobian with respect to the model parameters, so it is zero in the case of LWI. For such reason, conventional LWI works exclusively with the Gauss-Newton Hessian (Equation 5). We propose the inclusion of model (velocity) updates in the process by incorporating WEMVA to the LWI misfit function (Equation 3). In this report we do not derive LWI with velocity updating from FWI. Such derivation is an objective for future research.

ALGORITHM FOR LWI WITH VELOCITY UPDATING

In this section we present an algorithm to perform LWI that updates the reflectivity model, $\mathbf{\Delta r}$, and allows controlled updates of the background model, $\mathbf{\Delta b}$.

Let us consider the following optimization problem:

$$\Phi(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) = \frac{1}{2}\|\mathbf{H}_{GN}(\mathbf{b}_0)\mathbf{\Delta r}+\mathbf{H}_W(\mathbf{b}_0)\mathbf{\Delta b}-\mathbf{\Delta r}_{mig}\|_2^2-\frac{\lambda}{2}\|\mathbf{H}_W(\mathbf{b}_0)\mathbf{\Delta b}+\mathbf{\Delta r}_{mig}\|_2^2. \tag{6}$$

$\mathbf{\Delta r}_{mig}$ represents the RTM image. Minimizing Equation 6 is interpreted as the search of the optimal perturbation of reflectivity, $\mathbf{\Delta r}$, and perturbation in the background slowness, $\mathbf{\Delta b}$. The smooth background slowness, $\mathbf{b}_0$, is fixed respect to the optimization, but spatially variable. The reflectivity obtained from $\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{H}_W\mathbf{\Delta b}$ is expected to fit the RTM image. The optimization is subject to the condition that the perturbation in the image contributed by $\mathbf{H}_W\mathbf{\Delta b}$, when added to the migrated image, maximizes the stacking power (second term). This is enforced by minimizing the negative value of the second term of Φ. We control this constraint with the parameter λ.

Let us simplify Equation 6 by dropping the explicit dependence on $\mathbf{b}_0$, and substitute $\mathbf{W}$ for $\mathbf{H}_W$:

$$\Phi(\mathbf{\Delta r}, \mathbf{\Delta b}) = \frac{1}{2}\|\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{W}\mathbf{\Delta b} - \mathbf{\Delta r}_{mig}\|_2^2 - \frac{\lambda}{2}\|\mathbf{W}\mathbf{\Delta b} + \mathbf{\Delta r}_{mig}\|_2^2. \tag{7}$$

In the first term we can express the Hessian components as a matrix product,

$$\Phi(\mathbf{\Delta r}, \mathbf{\Delta b}) = \frac{1}{2}\| \begin{bmatrix}\mathbf{H}_{GN} & \mathbf{W}\end{bmatrix} \begin{bmatrix}\mathbf{\Delta r} \\ \mathbf{\Delta b}\end{bmatrix} - \mathbf{\Delta r}_{mig}\|_2^2 - \frac{\lambda}{2}\|\mathbf{W}\mathbf{\Delta b} + \mathbf{\Delta r}_{mig}\|_2^2. \tag{8}$$

The corresponding fitting functions (Claerbout, 2014) are

$$\begin{aligned}\begin{bmatrix}\mathbf{H}_{GN} & \mathbf{W}\end{bmatrix} \begin{bmatrix}\mathbf{\Delta r} \\ \mathbf{\Delta b}\end{bmatrix} - \mathbf{\Delta r}_{mig} &\approx \mathbf{0} \\ -\lambda\mathbf{W}\mathbf{\Delta b} - \lambda\mathbf{\Delta r}_{mig} &\approx \mathbf{0}\end{aligned}$$

We can re-cast the fitting functions as a matrix operation,

$$\begin{bmatrix}\mathbf{H}_{GN} & \mathbf{W} \\ \mathbf{0} & -\lambda\mathbf{W}\end{bmatrix} \begin{bmatrix}\mathbf{\Delta r} \\ \mathbf{\Delta b}\end{bmatrix} \approx \begin{bmatrix}\mathbf{\Delta r}_{mig} \\ \lambda\mathbf{\Delta r}_{mig}\end{bmatrix}, \tag{9}$$

which defines our forward modeling operator. The corresponding adjoint operator is defined by taking the transpose of the big matrix in Equation 9,

$$\begin{bmatrix}\mathbf{\Delta \tilde{r}} \\ \mathbf{\Delta \tilde{b}}\end{bmatrix} = \begin{bmatrix}\mathbf{H}_{GN}^T & \mathbf{0} \\ \mathbf{W}^T & -\lambda\mathbf{W}^T\end{bmatrix} \begin{bmatrix}\mathbf{\Delta r}_{mig} \\ \lambda\mathbf{\Delta r}_{mig}\end{bmatrix}. \tag{10}$$

Now $[\mathbf{\Delta \tilde{r}} \quad \mathbf{\Delta \tilde{b}}]^T$ constitutes the first estimation of the perturbations in reflectivity and background model. Equations 9 and 10 constitute the main elements to set an iterative solution scheme.

DISCUSSION: IMPLEMENTATION AND POTENTIAL ISSUES

Figure 1 shows a flowchart with the implementation of LWI with velocity updating using the conjugate gradient method. It follows the classic implementation of an augmented system (e.g. Nemeth et al., 1999). The objective is to analyze which elements could be recycled or precomputed (in trapezoid boxes), and which ones have to be computed "on the fly" (rectangular boxes). The former are simply supplied to processes (thick arrows), whereas the later are calculated just before they get into processes (thin arrows). Quantities in the model space (gradient, $\vec{\mathbf{g}}$, and corresponding search direction, $\vec{\mathbf{h}}$) have one component related to $\mathbf{\Delta r}$ and another related to $\mathbf{\Delta b}$. Quantities in the data space (residual, $\vec{\mathbf{f}}$, and corresponding search direction, $\vec{\mathbf{q}}$) have two components related to perturbations in reflectivity, but the first one in turn relates to the Gauss-Newton Hessian (I will henceforth refer to it as "the Hessian"), and the second relates to WEMVA. Such components are distinguished throughout using corresponding superindices. The data space is comprised of the "migration" space, or to be more precise, the "blurred image" space (Hu et al., 2001). The model space consists of the estimated reflectivity domain, and the estimated perturbation in the background model domain. Finally, although the Hessian and the WEMVA operators are symmetric, I distinguish between "forward" and "adjoint" operations for the sake of convenience.

The flowchart shows that every iteration of the conjugate gradient method demands two applications of the Hessian and two applications of WEMVA. For big-scale problems it is convenient to precompute the Hessian (or some of its elements) and then supply it as needed. In the case of WEMVA, the source wavefield, and the receiver wavefield could be in principle precomputed and stored, for they only act upon the background model, $\mathbf{b}_0$. However, the source and receiver scattered wavefields operate either on $\mathbf{\Delta b}$ ($\delta\mathbf{S}$ and $\delta\mathbf{R}$ for $\mathbf{W}$) or $\mathbf{\Delta r}$ ($\delta\mathbf{S}^T$ and $\delta\mathbf{R}^T$ for $\mathbf{W}^T$), which are updated every iteration. As a consequence, such scattered wavefields must be recalculated twice every iteration.

In practice, the precomputation and storage of the Hessian is difficult because of the huge number of elements it contains. The simplest approach is the computation of the diagonal elements only. Inverting the diagonal of the Hessian has the property of correcting the amplitudes of the image. The inclusion of the off-diagonal elements additionally corrects for the smearing effect related to sparse and/or irregular data acquisition. Totally neglecting the off-diagonal elements of the Hessian contradicts our original intention to recover a more accurate and crisp image by including velocity updates. Two possible solutions can be 1) computing only some off-diagonal elements, and 2) computing point spread functions sparsely around the space domain, and interpolate as required (Fletcher et al., 2016).

The precomputation of the source and the receiver wavefields has practical issues. The storage of the entire set of time frames is computationally overwhelming at the industrial scale. A much better alternative is to implement random boundary

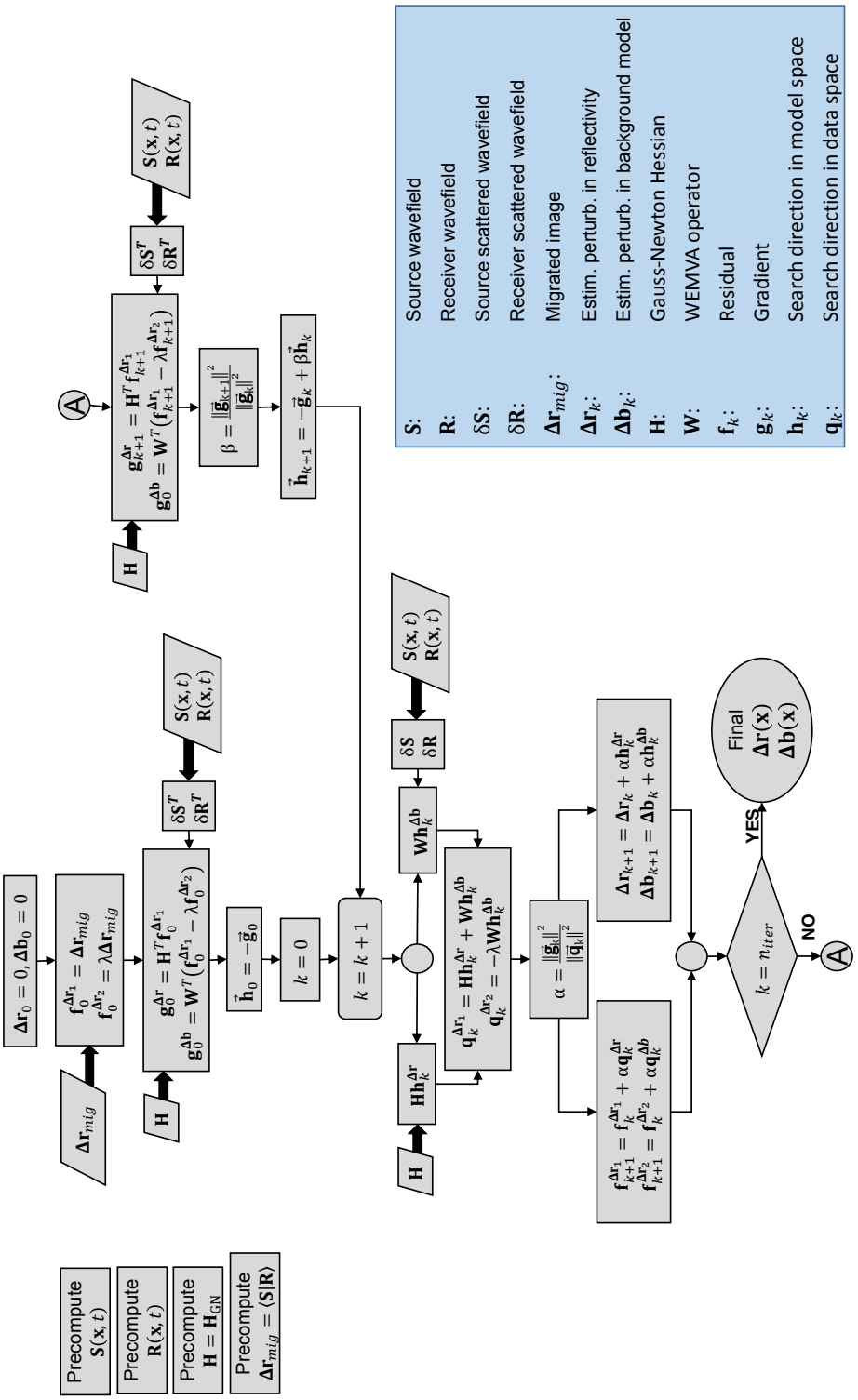

Figure 1: Flowchart of the conjugate gradient method applied to LWI with velocity updating. [**NR**]

conditions (Clapp, 2009, 2010), thus reducing the storage to two time frames per wavefield. This is in fact a tremendous reduction in storage. However, one needs to recompute the source and receiver wavefields for each iteration.

FUTURE WORK

So far, we have proposed LWI with velocity updating as an *ad hoc* solution. We still need a formal mathematical derivation from a more general case, such as FWI. Numerical implementation of the algorithm in simple, comprenhensible models constitutes a must. Finally, we need to find an optimal convergence criterion, which would very likely contemplate both the perturbation on the reflectivity and the perturbation on the background model.

ACKNOWLEDGEMENTS

We would like to thank the SEP sponsors for their continuos support. Special thanks to Guillaume Barnier and Taylor Dahlke for their merciless review and useful discussions, which highly improved this report. Alejandro Cabrales would like to thank Petróleos Mexicanos for their financial support.

REFERENCES

Aster, R., B. Borchers, and C. Thurber, 2013, Parameter estimation and inverse problems: Elsevier Inc.

Barnier, G. and A. Almomin, 2014, Tutorial on two-way wave equation operators for acoustic, isotropic, constant-density media: SEP-Report, **155**, 23–56.

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

Biondi, B., 2002, Stable wide-angle Fourier finite-difference downward extrapolation of 3-D wavefields: Geophysics, **67**, 872–882.

Biondi, B., E. Biondi, M. Maharramov, and Y. Ma, 2015, Dissection of the full-waveform inversion hessian: SEP-Report, **160**, 19–38.

Claerbout, J. F., 2014, Geophysical image estimation by example: Jon Claerbout.

Clapp, R. G., 2009, Reverse time migration with random boundaries: SEP-Report, **138**, 29–38.

———, 2010, More fun with random boundaries: SEP-Report, **140**, 95–102.

Dai, W., C. Boonyasiriwat, and G. Schuster, 2010, 3D multi-source least-squares reverse time migration: SEG Technical Program Expanded Abstracts, 3120–3124.

Dai, W., Y. Huang, and G. Schuster, 2013, Least-squares reverse time migration of marine data with frequency-selection encoding: SEG Technical Program Expanded Abstracts, 3231–3236.

Duquet, B., K. Marfurt, and J. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209.

Fletcher, R., D. Nichols, R. Bloor, and R. Coates, 2016, Least-squares migation - data domain versus image domain using point spread functions: The Leading Edge, **35**, no. 2, 157–162.

Gazdag, J. and E. Carrizo, 1986, On reverse-time migration: Geophysical Prospecting, **34**, 822–832.

Hestenes, M. and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, no. 6, 409–436.

Hu, J., G. Schuster, and P. Valasek, 2001, Poststack migration deconvolution: Geophysics, **66**, 939–952.

Ji, J., 1992, Reverse-time migration as the conjugate of forward modeling: SEP-Report, **75**, 135–143.

———, 2009, An exact adjoint operation pair in time extrapolation and its application in least-squares reverse-time migration: Geophysics, **74**, no. 5, H27–H33.

Jiang, Z. and S. Schuster, 2003, Target-oriented least squares migration: SEG Technical Program Expanded Abstracts, 1043–1046.

Kosloff, D. and E. Baysal, 1983, Migration with the full acoustic wave equation: Geophysics, **48**, 677–687.

Liu, F., Z. Guanquan, S. Morton, and J. Leveille, 2011, An effective imaging condition for reverse-time migration using wavefield decomposition: Geophysics, **76**, no. 1, S29–S39.

Liu, Y., H. Hu, X. Xie, Y. Zheng, and P. Li, 2015, Reverse time migration of internal multiples for subsalt imaging: Geophysics, **80**, no. 5, S175–S185.

McMechan, G., 1983, Migration by extrapolation of time-dependent boundary values: Geophysical Prospecting, **31**, 413–420.

Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Ristow, D. and T. Rühl, 1994, Fourier finite-difference migration: Geophysics, **59**, 1882–1893.

Ronen, S. and C. Liner, 2000, Least-squares DMO and migration: Geophysics, **65**, 1364–1371.

Stoffa, P., J. Fokkema, de Luna Freire R., and W. Kessinger, 1990, Split-step Fourier migration: Geophysics, **56**, 410–421.

Wong, M., B. Biondi, and S. Ronen, 2015, Imaging with primaries and free-surface multiples by joint least-squares reverse time migration: Geophysics, **80**, no. 6, S223–S235.

Wong, M., S. Ronen, and B. Biondi, 2011, Least-squares reverse time migration/inversion for ocean bottom data: a case study: SEG Technical Program Expanded Abstracts, 2369–2373.