GEOPHYSICS, VOL. 81, NO. 4 (JULY-AUGUST 2016); P. S221–S238, 20 FIGS.
10.1190/GEO2015-0529.1

An efficient step-length formula for correlative least-squares reverse time migration

Youshan Liu[1], Jiwen Teng[1], Tao Xu[2], Zhiming Bai[1], Haiqiang Lan[1], and José Badal[3]

ABSTRACT

In correlative least-squares reverse time migration (CLSRTM), the estimation of the optimal step size is usually determined by fitting a parabola and finding its minimum; it involves at least two times extra reading of all seismic records, which significantly lowers the efficiency of the algorithm. To improve the efficiency of the CLSRTM algorithm, we have deduced an analytical step-length (ASL) formula based on the linear property of the demigration operator. Numerical examples performed with the data synthetized by the Marmousi and Sigsbee2A models were used to test its validity. In complex models with imperfect migration velocity, such as the Sigabee2A model, our formula makes the value of the objective function converges to a much smaller minimum. Additional numerical tests performed with the data either acquired irregularly or contaminated by different noise levels verify the robustness of the ASL formula. Compared with the commonly used parabolic search method, the ASL formula is much more efficient because it is free from an extra estimation of the value of the objective function.

INTRODUCTION

In seismic exploration, data migration has been playing an important role in imaging the subsurface structures of the earth. Conventional imaging methods are based on prestack depth migration: from Kirchhoff migration (Schneider, 1978) and Beam migration (Hill, 1990, 2001) to one-way wave-equation migration (Claerbout, 1971; Claerbout and Doherty, 1972; Gazdag, 1978; Stoffa et al., 1990; Huang and Fehler, 1998) and reverse time migration (RTM) (Hemon, 1978; Baysal et al., 1983; McMechan, 1983; Whitmore, 1983; Liu et al., 2011; Lan et al., 2014), each migration procedure uses an extrapolation operator to adapt to complicated subsurface structures. Nowadays, RTM is recognized as the state-of-the-art of imaging technology to seek increasingly complicated subsurface structures. However, in practice, the above conventional depth imaging methods cannot achieve a perfect image because of irrgular data sampling, aliased seismic data, and uneven subsurface illumination from acquisition geometry (Etgen et al., 2009; Wong et al., 2015).

To remove the data acquisition footprint and improve the image quality, the imaging problem can be posed as a linear inverse problem. This inversion-based imaging method is also widely known as least-squares migration (LSM) (Lambaré et al., 1992; Nemeth et al., 1999) or linearized wavefield inversion (Tarantola, 1984; Clapp, 2005; Valenciano, 2008). LSM was originally proposed in Kirchhoff migration (Schuster, 1993; Nemeth et al., 1999; Duquet et al., 2000), and then it was introduced into one-way wave-equation migration (Gazdag, 1978; Kuehl and Sacchi, 1999, 2001) and applied to RTM (Tang, 2009; Dai et al., 2011, 2012; Wong et al., 2011, 2015; Dong et al., 2012). The idea behind conventional LSM is that it generates simulated data to perfectly match the amplitude of the observed seismic data.

In general, conventional LSM can boost the signal for the true reflector, balance the relative amplitude, and suppress noise. However, various factors hamper it when applied to field data. On the one hand, the earth is viscoelastic or even anisotropic. The acoustic-wave equation is not adequately accurate to model the propagation of seismic waves across earth structures. On the other hand, the source strength may vary from experiment to experiment, and an accurate estimation of the source strength is not always possible. Consequently, it is difficult to effectively match the amplitude. Thus, conventional (i.e., amplitude-matching-based) least-squares

Manuscript received by the Editor 4 October 2015; revised manuscript received 16 January 2016; published online 30 May 2016.
[1]Chinese Academy of Sciences, Institute of Geology and Geophysics, State Key Laboratory of Lithospheric Evolution, Beijing, China. E-mail: ysliu@mail.iggcas.ac.cn; jwteng@mail.iggcas.ac.cn; bbzzmm@mail.iggcas.ac.cn; lanhq@mail.iggcas.ac.cn.
[2]Chinese Academy of Sciences, Institute of Geology and Geophysics, State Key Laboratory of Lithospheric Evolution, Beijing, China and CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing, China. E-mail: xutao@mail.iggcas.ac.cn.
[3]University of Zaragoza, Physics of the Earth, Sciences B, Zaragoza, Spain. E-mail: badal@unizar.es.

RTM (LSRTM) faces huge challenges when applied to field seismic data. Zhang et al. (2015) propose a new general framework for LSRTM, which they call correlative LSRTM (hereafter CLSRTM). CLSRTM is able to generate a high-resolution image through minimizing the negative crosscorrelation coefficient of the simulated and observed data at zero lag. It relaxes the amplitude matching and uses phase information to measure the closeness between simulated and observed seismic data. Even if differences in amplitude between the simulated and observed seismic data exist, the involved objective function value is unchanged (Zhang et al., 2015). Thus, differences in amplitude have a smaller contribution than phase differences. Therefore, CLSRTM has good performance and is revealed as a useful and stable tool to work with field seismic data (Zhang et al., 2015).

Although Zhang et al. (2015) construct a complete framework for CLSRTM, they do not give an efficient step-length formula in their flowchart for inversion. For gradient-based inversion methods, only if the descent direction is scaled by a proper scalar (i.e., a step length) will the value of the objective function decrease iteratively. In this paper, we derive an analytical step-length (ASL) formula for CLSRTM based on the linear characteristic of the demigration operator. Compared with the commonly used linear search method, such as the parabolic search method (PSM) (Vigh et al., 2009), this new step-length formula involves little computational overhead because the optimal step length can be incidentally calculated during the data predicition process (demigration). In contrast, the PSM involves at least twice as much reading of the simulated, perturbed, and observed seismic records of all shots to estimate the objective function value, which degrades the efficiency of CLSRTM algorithm.

In this study, first, we review the migration and demigration operators of CLSRTM. Then, we analytically derive a step-length formula based on the linear property of the demigration operator. After that, we present two numerical examples with noise-free data to demonstrate the effectiveness of the ASL when compared with the commonly used PSM. Last, we adopt CLSRTM with PSM and ASL to migrate irregularly acquired data to verify the robustness of ASL. In addition, we also add different noise levels into the data to check the robustness of the proposed procedure further.

THEORY

In LSM, the data are migrated (migration) and the resulting reflectivity image is used to generate simulated data (demigration). These simulated data are substracted from the field data, and the residuals are migrated to correct for the estimated reflectivty model at each iteration. Therefore, a typical LSM scheme involves two core parts, namely, the migration and demigration processes. The former computes the gradient of the cost function with respect to reflectivity (or stacked image), and the latter computes the predicted data using the reflectivity (or stacked image). Once the gradient of the cost function and the initial stacked image are available, one can iteratively update the stacked image to reduce migration artifacts and to obtain a more balanced amplitude. For completeness, we briefly review the theoretical background of CLSRTM. The detailed algorithm can be seen in Zhang et al. (2015) and Figure 1.

Correlative LSRTM

In 2015, Zhang et al. (2015) construct the following crosscorrelation-based objective function to overcome the problem of imperfect amplitude-matching characteristic in the conventional LSRTM (i.e., the L_2 norm objective function). The crosscorrelation-based objective function can be written as

$$E(\mathbf{r}(\boldsymbol{x})) = -\int\int \frac{\int \mathbf{d}(\boldsymbol{x}_r, t|\boldsymbol{x}_s)\cdot\mathbf{D}(\boldsymbol{x}_r, t|\boldsymbol{x}_s)dt}{\sqrt{\int \mathbf{d}^2(\boldsymbol{x}_r, t|\boldsymbol{x}_s)dt}\sqrt{\int \mathbf{D}^2(\boldsymbol{x}_r, t|\boldsymbol{x}_s)dt}} d\boldsymbol{x}_r d\boldsymbol{x}_s, \tag{1}$$

where $\mathbf{r}(\boldsymbol{x})$ is the stacked image at position $\boldsymbol{x}$ rather than physical angle-dependent reflectivity; the middle dot denotes the inner product of the observed and simulated seismic records at each receiver; $\mathbf{d}(\boldsymbol{x}_r, t|\boldsymbol{x}_s)$ is the simulated or predicted data at the time instant t and the receiver $\boldsymbol{x}_r$, which is excited by a source located at position $\boldsymbol{x}_s$; $\mathbf{D}(\boldsymbol{x}_r, t|\boldsymbol{x}_s)$ is the observed data at the time instant t and the receiver $\boldsymbol{x}_r$, which is excited by a source located at position $\boldsymbol{x}_s$. The negative sign on right side means that the negative crosscorrelation coefficient between the observed and simulated data will be minimized. In the best-case scenario, in which the two data sets are identical or with a constant scaling difference, the objective function value reaches its minimum -1 (Zhang et al., 2015). Mathematically, this LSM can be ascribed to obtain the solution of the optimization problem 1. Many methods can be adopted to solve it. Although Monte Carlo-based inversion methods tend to search for a global optimal solution, they are rarely used for seismic inversion in modern computing architecture due to the huge and unaffordable computing cost. At present, the gradient-based optimization methods are generally recognized as more practical choices.

In this paper, we adopt the limited-memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS) method as our inversion method (Nocedal, 1980). Using the L-BFGS method, the descent direction can be approximated as

$$\boldsymbol{\delta}\mathbf{r}_i = -\mathbf{B}_i\mathbf{g}_i, \tag{2}$$

where $\mathbf{B}_i$ is the approximate inverse Hessian matrix, $\mathbf{g}_i$ is the gradient of the objective function, $\boldsymbol{\delta}\mathbf{r}_i$ is the descent direction, and i is the iteration index. The product of the approximated or quasi-inverse Hessian matrix and the gradient can be calculated using a rescurive formula with information from the last m iterations, where m is any number supplied by the user. The application details can be seen in Nocedal (1980). At each iteration, we use the following initial inverse Hessian matrix given by Nocedal (1980)

$$\mathbf{B}_i^0 = \mathbf{s}_i^T\mathbf{y}_i/\mathbf{y}_i^T\mathbf{y}_i, \tag{3}$$

where $\mathbf{r}_i$ is the stacked image, $\mathbf{s}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$ is the stacked image change, and $\mathbf{y}_i = \mathbf{g}_i - \mathbf{g}_{i-1}$ is the gradient change. In all the following experiments, the number of the stored y and s for corrections used in the L-BFGS is set to 10. Usually, an effective descent direction generated by the L-BFGS method must be well-behaved, which is ensured by using the Wolfe linear search (Wu et al., 2015). Here, we also consider this special case. In this case, we use the negative gradient as the descent direction when the sufficient descent condition $\mathbf{g}_i^T\delta\mathbf{r}_i < 0$ is not satisfied (where superscript T means the transpose operator; Hu and Wang, 2014). In this case, the L-BFGS method fails to generate an effective descent direction to decrease the objective function value (i.e., a nonpositive definite Hessian matrix).

In relation to this optimization method, we need the gradient (migration process) of the objective function to update the initial stacked image and need the simulated data (demigration process) to compute the value of the objective function at each iteration. In the following subsections, we briefly review the migration and demigration operators of CLSRTM. After, an ASL formula is deduced based on the linear characteristic of the demigration operator.

Reverse time migration

RTM is an advanced migration tool and is suitable to use as a migration engine in LSM. Zhang et al. (2015) use the following RTM operator M^T:

$$\begin{cases} \left(\frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2}-\nabla^2\right)p_s(\boldsymbol{x},t|\boldsymbol{x}_s)=\delta(\boldsymbol{x}-\boldsymbol{x}_s)f(t), \\ \left(\frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2}-\nabla^2\right)p_r(\boldsymbol{x},t|\boldsymbol{x}_s)=-\frac{\partial}{\partial t}\Delta\mathbf{d}(\boldsymbol{x}_r,t|\boldsymbol{x}_s)\delta(\boldsymbol{x}-\boldsymbol{x}_r), \end{cases} \tag{4}$$

where $v(\boldsymbol{x})$ is the velocity of the medium at position $\boldsymbol{x}$, $p_s(\boldsymbol{x},t|\boldsymbol{x}_s)$ is the forward-propagated source wavefield, $p_r(\boldsymbol{x},t|\boldsymbol{x}_s)$ is the backward-propagated receiver wavefield, δ is the Dirac-delta function, ∇^2 is the Laplacian operator, and $f(t)$ is the source signature. The reweighted residual between the rescaled simulated data and observed data $\Delta\mathbf{d}(\boldsymbol{x}_r,t|\boldsymbol{x}_s)$ is expressed as

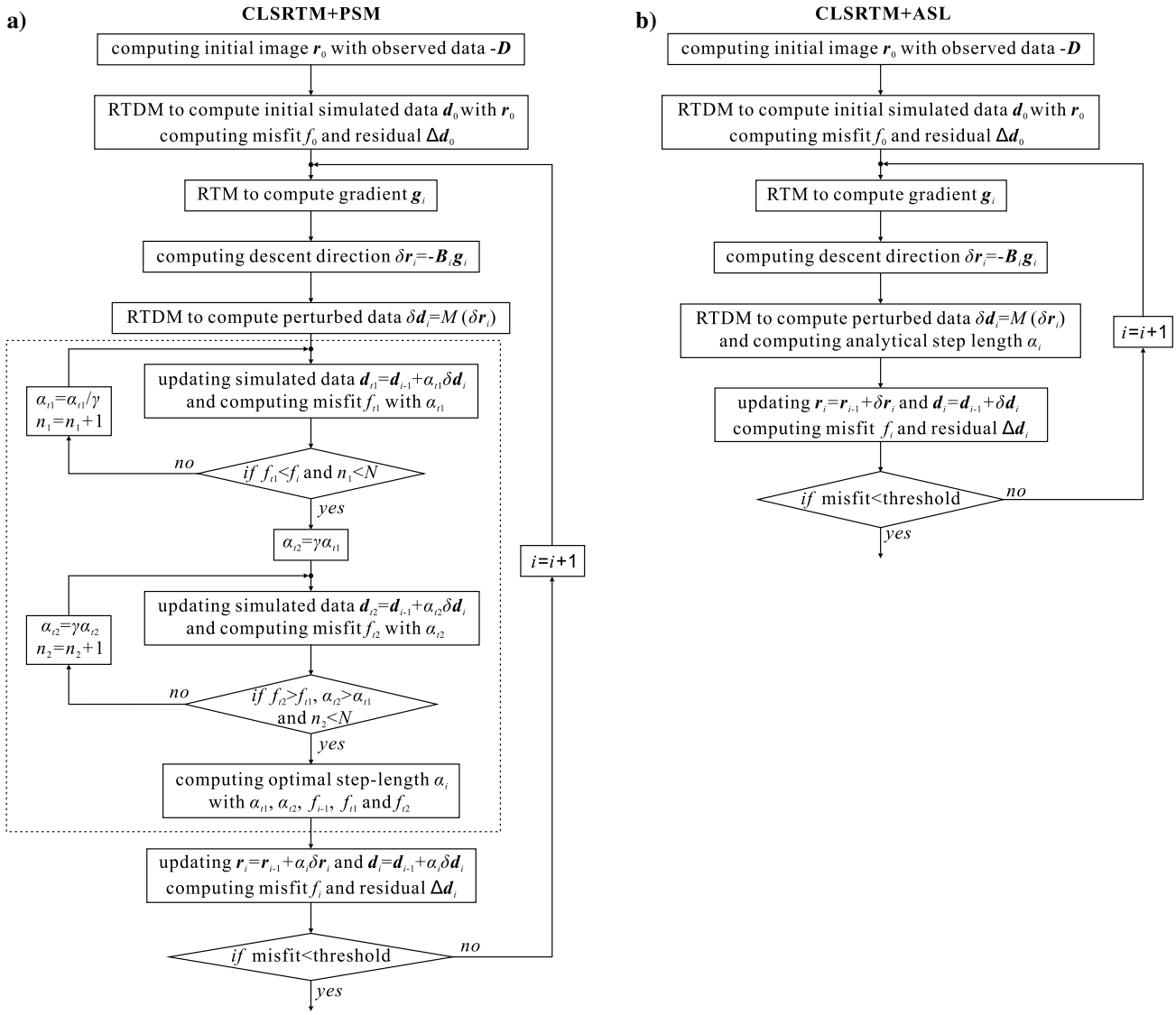

Figure 1. Flowcharts for the implementation of CLSRTM with the (a) PSM and (b) ASL; $\mathbf{B}_i$ is the approximate inverse Hessian matrix; $\mathbf{g}_i$ is the gradient; $\delta\mathbf{r}_i$ is the descent direction; $\delta\mathbf{d}_i$ is perturbed data; $\mathbf{d}_i$ is the simulated data; i is the iteration index, its maximum value is taken as 100; f_i is the value of objective function at the current iteration; f_{t1} and f_{t2} are the first and second trial objective function values, respectively; α_i is the step length at the current iteration; α_{t1} and α_{t2} are the step lengths used to estimate the first and second trial objective function values, respectively; $\mathbf{d}_{t1}$ and $\mathbf{d}_{t2}$ are the updated data with the trial step lengths, respectively; n_1 and n_2 are the iteration indexes of the first and second trial estimation; and N is the maximum iteration number of the first and second trial estimation.

$$\Delta\mathbf{d}(\boldsymbol{x}_r, t|\boldsymbol{x}_s) = \frac{1}{\sqrt{\int \mathbf{d}^2 dt}\sqrt{\int \mathbf{D}^2 dt}}\left[\frac{\int \mathbf{d}\cdot\mathbf{D}dt}{\int \mathbf{d}^2 dt}\mathbf{d} - \mathbf{D}\right], \quad (5)$$

where $\mathbf{d}$ denotes the predicted or simulated data and $\mathbf{D}$ is the observed data. Here, $(\int \mathbf{d}\cdot\mathbf{D}dt/\int \mathbf{d}^2 dt)\times\mathbf{d}$ represents the rescaled simulated data to correct the differences in amplitude between the observed and simulated data. The weight before the square brackets aims to remove the amplitude strength effects from the observed and predicted data.

The migrated image can be produced by applying the following imaging condition:

$$\mathbf{r}(\boldsymbol{x}) = \iint p_s(\boldsymbol{x}, t|\boldsymbol{x}_s) p_r(\boldsymbol{x}, t|\boldsymbol{x}_s) dt d\boldsymbol{x}_s. \quad (6)$$

In CLSRTM, the computation of the gradient (equation 4) is used to update the stacked image (equation 6) to obtain a stacked image with higher resolution and balanced amplitude. This process is accomplished by RTM with data residuals (equation 5), but the first iteration of CLSRTM is equivalent to RTM. Therefore, RTM plays the role of the "engine" in the inversion of CLSRTM. Then, the simulated data are computed with the updated-stacked image. The fitness between the observed and simulated data measures whether the model can well interpret the observed data. Correspondingly, a data prediction process, i.e., reverse time demigration (RTDM), is naturally involved in CLSRTM.

Reverse time demigration

RTDM is a powerful tool to predict seismic data from the stacked image r obtained with CLSRTM. Zhang et al. (2015) use the following RTDM operator M:

$$\begin{cases} \left(\frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right) p_s(\boldsymbol{x}, t|\boldsymbol{x}_s) = \delta(\boldsymbol{x}-\boldsymbol{x}_s) f(t), \\ \left(\frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right) p_r(\boldsymbol{x}, t|\boldsymbol{x}_s) = \mathbf{r}(\boldsymbol{x})\frac{\partial}{\partial t} p_s(\boldsymbol{x}, t|\boldsymbol{x}_s), \\ \mathbf{d}(\boldsymbol{x}_r, t|\boldsymbol{x}_s) = p_r(\boldsymbol{x}_r, t|\boldsymbol{x}_s). \end{cases} \quad (7)$$

Theoretically, the migration operator (4) and the demigration operator (7) form an exact adjoint operator pair in the time-domain extrapolation. Zhang et al. (2015) analytically prove that this adjoint operator pair satisfies the dot-product test that an exact adjoint operator pair should obey.

The adjoint of a linear operator is its complex-conjugated transpose. In the time domain, the operator is real-valued and the adjoint is the transpose of the forward modeling operator (Ji, 2009). An analytically exact adjoint operator pair is just a theoretical presupposition. To ensure that the adjoint pair is numerically exact, as pointed by Ji (2009), a correct implementation of the transpose operator must be applied in the extrapolation of the backward-propagated receiver wavfields. Instead of using wavefields from the two previous time steps together for one time-step extrapolation, the exact adjoint operator uses the wavefields of the two previous time steps separately (Ji, 2009). Althougth a correct implementation of exact adjoint pair can drive the objective function value to converge to a smaller minimum (the Figure 8 of Ji, 2009), in this paper, we still use the identical wavefield extrapolator for the source and receiver wavefields as in most of the RTM-related papers. Nevertheless, this never impedes us from verifying the effectiveness of our proposed ASL in the following subsection, because we use a completely identical algorithms except for different step-length formulas.

In all experiments, we adopt a central finite-difference stencil of the 16th order of accuracy in space and the second-order of accuracy in time to extrapolate the source wavefield (equation 4) and the receiver wavefield (equation 7). The second-order acoustic wave equation is solved and a second-order PML boundary conditions (Liu et al., 2012) are applied to four edges to suppress spurious reflections from the artificial boundaries. Because the source and receiver wavefields should be synchronously available at the same time step, the source wavefield is either saved on disk or recomputed at each time instant. In this paper, the source wavefield history is reconstructed in the reverse time direction by storing the wavefileds history at PMLs with the thickness of a half-number of the finite-difference stencil as the boundary condition and by storing wavefields at the last two slices as initial condition (Dussaud et al., 2008; Liu et al., 2015; Nguyen and McMechan, 2015). In addition, at each iteration, we adopt a Laplacian filter as a preconditioner to remove low-frequency artifacts throughout all experiments (Pratt, 1978; Youn and Zhou, 2001; Guitton and Kaelin, 2006; Zhang and James, 2009).

ASL formula

As is well-known, the gradient-based optimization methods always need a step length to scale the gradient of the objective function, which ensures that the objective function value will be reduced iteratively. Here, we derive an ASL formula for CLSRTM based on the linear characteristic of the RTDM operator.

At the current state (i.e., stacked image $\mathbf{r}$ at ith iteration), by applying the Taylor series expansion of α to the second order, the objective function one can be approximated as follows:

$$\begin{aligned} E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) &\approx E(\mathbf{r}) + \alpha\left.\frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\right|_{\alpha=0} \\ &\quad + \frac{\alpha^2}{2}\left.\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\right|_{\alpha=0} \\ &= c + b\alpha + a\alpha^2, \end{aligned} \quad (8)$$

where α is the optimal step length. After comparing the terms of the right sides of the last expressions, we can obtain the following relationships:

$$\begin{cases} c = E(\mathbf{r}), \\ b = \left.\frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\right|_{\alpha=0}, \\ a = \frac{1}{2}\left.\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\right|_{\alpha=0}. \end{cases} \quad (9)$$

If the value of the objective function approaches a local or global minimum, expression 8 satisfies the following condition:

$$\frac{\partial}{\partial\alpha} E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) = 0. \quad (10)$$

Substituting the relationship 9 into the formula 10, we obtain the general form of the optimal step-length formula:

$$\alpha_{\text{opt}} = -\frac{b}{2a} = -\frac{\left.\frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\right|_{\alpha=0}}{\left.\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\right|_{\alpha=0}}. \tag{11}$$

After derivation (see Appendix A), the specific form of the optimal step-length formula can be written as follows:

$$\alpha_{\text{opt}} = -\frac{b}{2a}, \tag{12}$$

where (omitting $(\boldsymbol{x}, t|\boldsymbol{x}_s)$ for clarity)

$$a = \frac{1}{2}\iint \frac{1}{\sqrt{\int \mathbf{D}^2 dt}\sqrt{\int \mathbf{d}^2 dt}}\left[2\frac{\int \mathbf{d}\cdot\boldsymbol{\delta}\mathbf{d}dt}{\int \mathbf{d}^2 dt}\int \boldsymbol{\delta}\mathbf{d}\cdot\mathbf{D}dt + \frac{\int(\boldsymbol{\delta}\mathbf{d})^2 dt}{\int \mathbf{d}^2 dt}\int \mathbf{d}\cdot\mathbf{D}dt - 3\left(\frac{\int \mathbf{d}\cdot\boldsymbol{\delta}\mathbf{d}dt}{\int \mathbf{d}^2 dt}\right)^2\int \mathbf{d}\cdot\mathbf{D}dt\right]d\boldsymbol{x}_r d\boldsymbol{x}_s \tag{13}$$

and

$$b = \iint \frac{1}{\sqrt{\int \mathbf{D}^2 dt}\sqrt{\int \mathbf{d}^2 dt}} \times \left[\frac{\int \mathbf{d}\cdot\mathbf{D}dt\int \mathbf{d}\cdot\boldsymbol{\delta}\mathbf{d}dt}{\int \mathbf{d}^2 dt} - \int \boldsymbol{\delta}\mathbf{d}\cdot\mathbf{D}dt\right]d\boldsymbol{x}_r d\boldsymbol{x}_s. \tag{14}$$

To verify the effectiveness of our ASL, we make a comparative study with the PSM. Figure 1 shows the flowcharts for the detailed implementation of CLSRTM with the PSM (Figure 1a) and ASL (Figure 1b). Compared with the latter, the former needs to update the simulated data by a trial step length (highlighted by the dotted line rectangle in Figure 1a) at least two times, which involves extra I/O and computational operations among the simulated, perturbed, and observed data of all shots. Consequently, it degrades the efficiency of CLSRTM algorithm. In this study, in terms of the PSM, the maximum estimation of the objective function value is set to 10 (at each iteration). The first trial step length should obey the condition: the maximum updated value of the stacked image is less than 1% of the migrated image, i.e., $\max(|\alpha_t \delta r|_i) \le 0.01\ \max(|r_i|)$ (Pica et al., 1990).

NUMERICAL EXAMPLES

To demonstrate the effectiveness of the ASL for CLSRTM versus the PSM, we develop two synthetic examples and compare the results. First, the Marmousi and Sigabee2A models are adopted to verify the validity and efficiency of the ASL. Then, the data acquired irregularly and also with different noise levels are migrated to verify its robustness.

In all experiments, we give an adequate maximum number of iterations (i.e., 100) and the identical stopping criterion. The inversion process will exit if the objective function value increases or the number of iterations reaches its maximum value or the relative change of the objective function value is less than 0.0002. Besides, the implementation algorithms of CLSRTM with the PSM and ASL are identical, except different step-length formulas.

Marmousi model

The Marmousi consists of 1151×376 grid cells in the horizontal and vertical directions, respectively. The horizontal and vertical grid spacing is 10 m. The synthetic data acquired from 58 shots separated by an interval of 200 m are used as shot gathers, which are generated by a Ricker wavelet with a dominant frequency of 20 Hz at a depth of 50 m. Consequently, the maximum frequency of the seismic source is 60 Hz (for a Ricker wavelet, typically $f_{\max} = 3f_0$, where f_0 is the dominant frequency). The sampling interval is 0.8 ms, and the recording length is 4 s. The first shot locates at the sixth grid cell. At most, 356 receivers are deployed at both sides with split-spread acquisition geometry to record each shot. The separation of receivers is 10 m. The minimum and maximum offsets are 0 and 3560 m, respectively. The number of receivers on either left side or right side of some shotpoints is less than 356 when the shot location is close to two ends of the model. In such cases, the receiver arrangement on either the left side or the right side is truncated by the end of the model. Figure 2a shows the model used to synthetize the observed data, whereas Figure 2b shows a Gaussian smoothed version of the Marmousi velocity model as our migration velocity model. In Figure 2a, the gray pentagram denotes the location of the 29th shot. Two gray triangles denote the first and last receivers of the 29th shot experiment, respectively.

Figure 3 shows the migrated images for the Marmousi model. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses represent the corresponding iteration numbers. The RTM image (i.e., the image at the first iteration) is shown in Figure 3a. The inverted images after nine iterations using CLSRTM with the PSM and ASL are shown in Figure 3b and 3c, respectively. Black colors represent positive values, whereas white colors represent negative values. Compared with the initial RTM image (Figure 3a), the images obtained with the PSM (Figure 3b) and ASL (Figure 3c) show more balanced amplitude (as can be appreciated in the small area enclosed by dotted line ellipses in Figure 3a) and overall high resolution. In the bottom left corner, the resolution has been significantly improved as the number of iterations increases. In addition, the top/bottom boundaries of the oil/gas cap beneath the anticline become sharper.

To compare the resolution of the migrated results, Figure 4a–4c shows the wavenumber spectra of the migrated images presented in Figure 3a–3c, respectively. To improve the visualization of these wavenumber spectra, the range of the wavenumber axes are reduced from 0.05 (in the original plot) to 0.025. Compared with the wavenumber spectrum of the initial RTM image (Figure 4a), both the wavenumber spectra of the images after nine iterations with the PSM (Figure 4b) and ASL (Figure 4c) clearly contain higher wavenumber components, especially along the k_z-axis, which further proves that Figure 3b and 3c obtains higher resolution than Figure 3a.

Figure 5 shows common shot gathers of the 29th shot with direct waves muted. Figure 5a shows the synthetic (or observed) data using the Marmousi velocity model (Figure 2a), and Figure 5b shows the simulated data using the initial stacked image $\mathbf{r}_0$. Figure 5c and 5d shows the simulated data after nine iterations using CLSRTM with the PSM and ASL, respectively. The numbers in parentheses represent corresponding iteration numbers. At more iterations, CLSRTM gradually makes the simulated data approach the synthetic or observed data. As can be seen in the shot records supplied

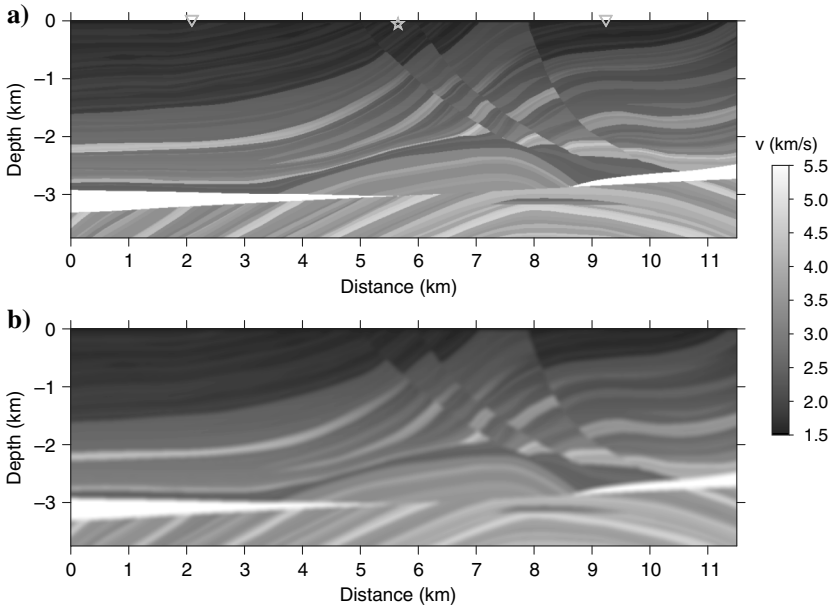

Figure 2. (a) The Marmousi velocity model. (b) A Gaussian smoothed version of the Marmousi model as migration velocity model. The gray pentagram denotes the source location of the 29th shot, and two gray triangles denote the first and last receivers of the 29th shot experiment.

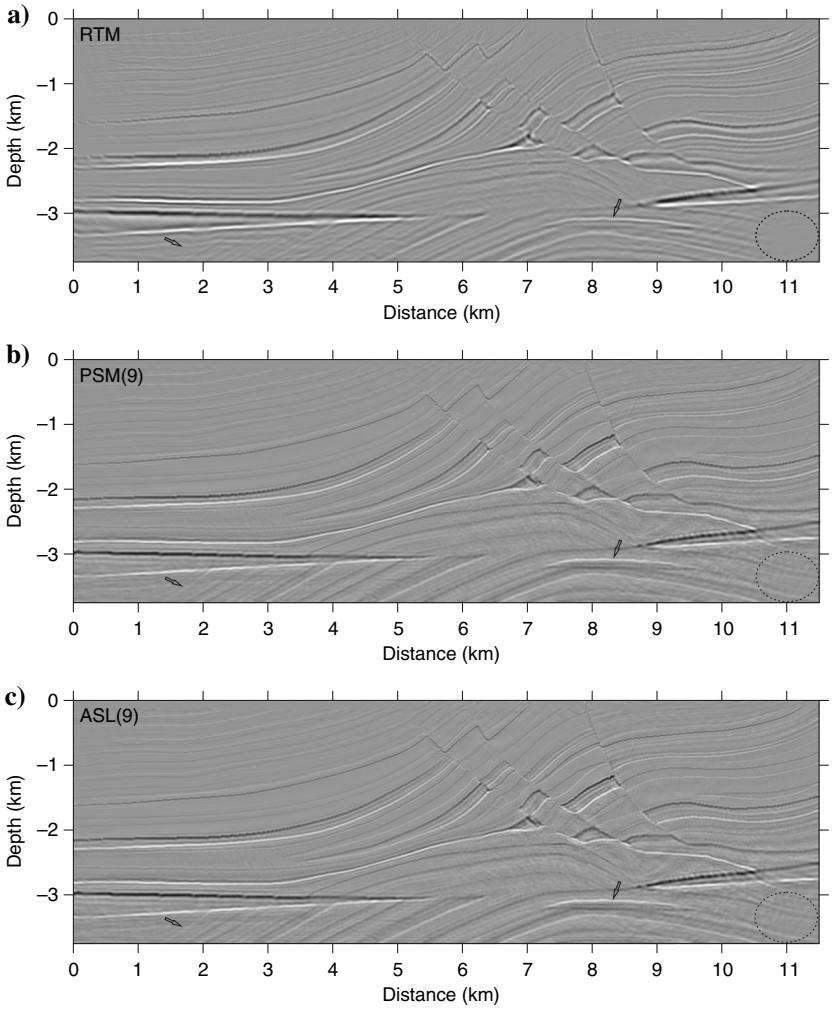

Figure 3. Migrated images of the Marmousi model. (a) The initial RTM image using CLSRTM. (b) The inverted image after nine iterations using CLSRTM with the PSM. (c) The inverted image after nine iterations using CLSRTM with the ASL. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses denote the corresponding iteration numbers.

by the PSM and ASL, reconstructed data match very well with the synthetic shot record. Although some events are weak and discontinuous on the initial simulated data (indicated by arrows) due to the limited acquisition aperture and imperfect migration operator, the simulated data become more continuous and balanced with the increase in the number of iterations. These results demonstrate that the ASL for CLSRTM is valid when compared with the conventional PSM.

The crosscorrelation coefficients at receivers of the 29th common shot are also plotted in Figure 6. The numbers in parentheses represent the corresponding iteration numbers. At the two ends of the receiver arrangement of the 29th shot, the crosscorrelation coefficient of the simulated data with the initial RTM image (red line) is significantly lower than that in the middle part. At a higher number of iterations, the crosscorrelation coefficient increases. In particular, the crosscorrelation coefficient at the two ends becomes more balanced after nine iterations. The crosscorrelation profiles after nine iterations with the PSM (solid blue line) and ASL (dashed green line) are almost identical, which further verifies that the migrated images after nine iterations can better interpret the observed

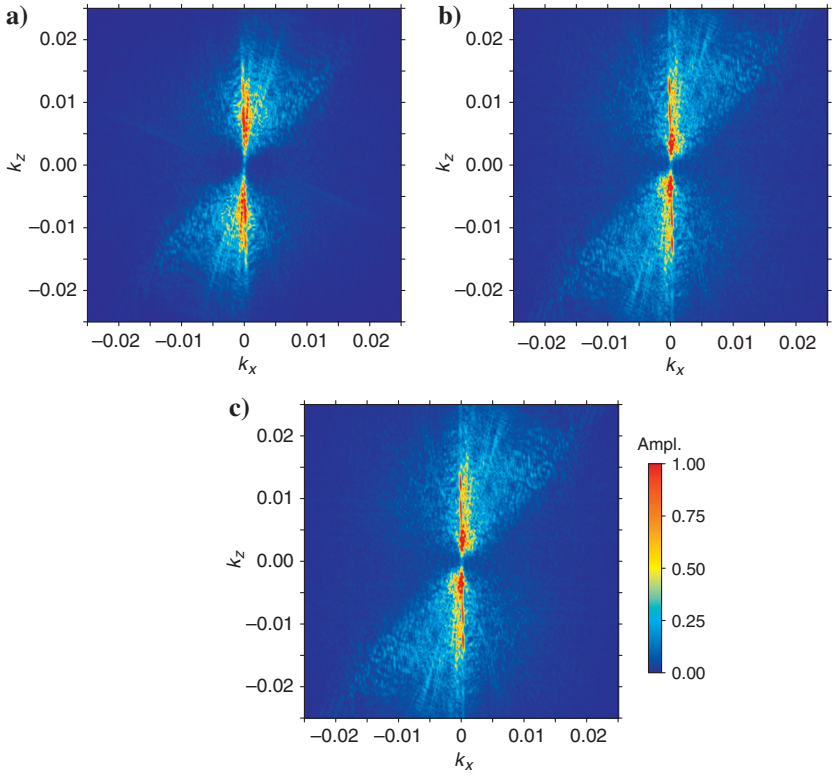

Figure 4. Wavenumber spectra of migrated images of the Marmousi model. (a) The wavenumber spectrum of the initial RTM image. (b) The wavenumber spectrum of the inverted image after nine iterations using CLSRTM with the PSM. (c) The wavenumber spectrum of the inverted image after nine iterations using CLSRTM with the ASL. To improve the visualization of the wavenumber spectra, the range of the wavenumber axes was reduced from 0.05 (in the original plot) to 0.025 (as can be seen here). Compared to the initial RTM image, the inverted images using CLSRTM with the PSM and ASL generate increased amplitude at a higher wavenumber.

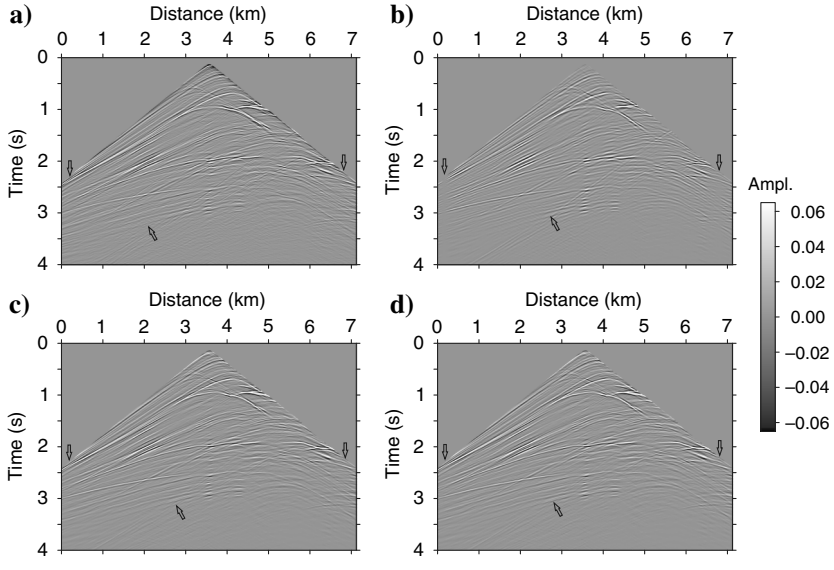

Figure 5. Common shot records of the 29th shot (direct waves are muted) in the case of the Marmousi model. (a) Synthetic data, (b) initial simulated data, (c) simulated data after nine iterations using CLSRTM with the PSM, and (d) simulated data after nine iterations using CLSRTM with the ASL.

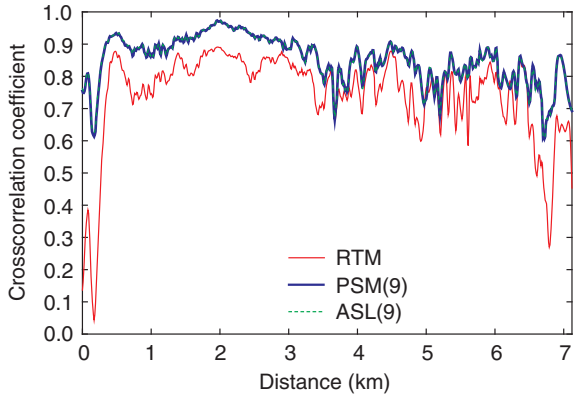

Figure 6. The crosscorrelation coefficients of the 29th shot in the case of the Marmousi model. Solid red, solid blue, and dashed green lines denote the crosscorrelation coefficients of the initial simulated data, simulated data after nine iterations using CLSRTM with the PSM, and simulated data after nine iterations using CLSRTM with the ASL, respectively. The numbers in parentheses denote the corresponding iteration numbers. The crosscorrelation coefficients of the simulated data after nine iterations using CLSRTM with the PSM and ASL are approximately identical. The PSM and ASL are better because the crosscorrelation coefficient is closer to one.

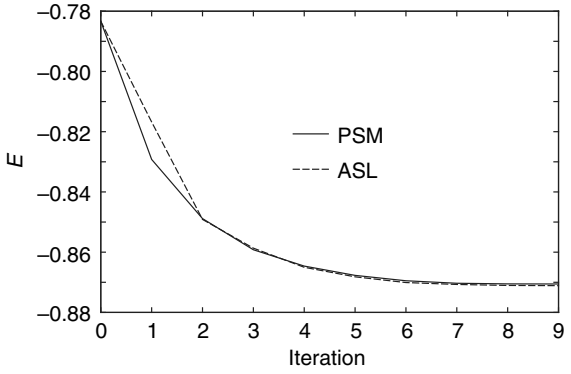

Figure 7. Values of the objective function E versus number of iterations. The results make reference to the Marmousi model and CLSRTM, either with the PSM (solid line) or the ASL (dotted line).

data than the initial RTM image and verifies the effectiveness of the ASL.

The objective function values E versus the number of iterations with the PSM (solid line) and ASL (dotted line) are shown in Figure 7. It can be seen that in the two cases, these values converge to approximately -0.87 and they decrease very rapidly after the first three iterations. Besides, the final convergence value of the objective function using the ASL is slightly smaller than that of the PSM.

Sigsbee2A model

In the exploration community, the subsalt imaging problem is always one of the most challenging topics due to the poor illumination beneath salt bodies. Liu et al. (2011) use multiples to provide better subsalt images because the different propagation wavepaths of multiples can complement the illuminations where the primary reflections beneath the salt are not acquired. Because LSM can boost up the true reflector and improve the image quality, we use the challenging Sigsbee2A model (Paffenholz et al., 2002) to check the imaging ability and the effectiveness of ASL for CLSRTM.

The model we used is a modified version of the released Sigsbee2A model. To save computational amount, we mute the top 120 water layers in the released Sigsbee2A model, and the model is sparsely resampled in the x- and z-directions by a factor of two. The stratigraphy and migration velocity models are shown in Figure 8. The modified Sigsbee2A model consists of 1601 and 481 grid cells in the x- and z-directions, respectively. The grid spacing in the x- and z-axes is reset to 10 m. The synthetic data acquired from 80 shots separated by an interval of 200 m are used as shot gathers, which are generated by a Ricker wavelet with dominant frequency of 20 Hz at the depth of 50 m. Consequently, the maximum frequency of the seismic source is 60 Hz. The sampling interval is 1 ms, and the record length is 6 s. The first shot locates at the 10th grid cell. At most, 340 receivers are deployed at both sides with split-spread acquisition geometry to record each shot. The separation of receivers is 10 m. The minimum and maximum offsets are 0 and 3400 m, respectively. When the location of shotpoint approaches to either end of the model, only one side has 340 receivers, whereas another side is truncated by the end of the model.

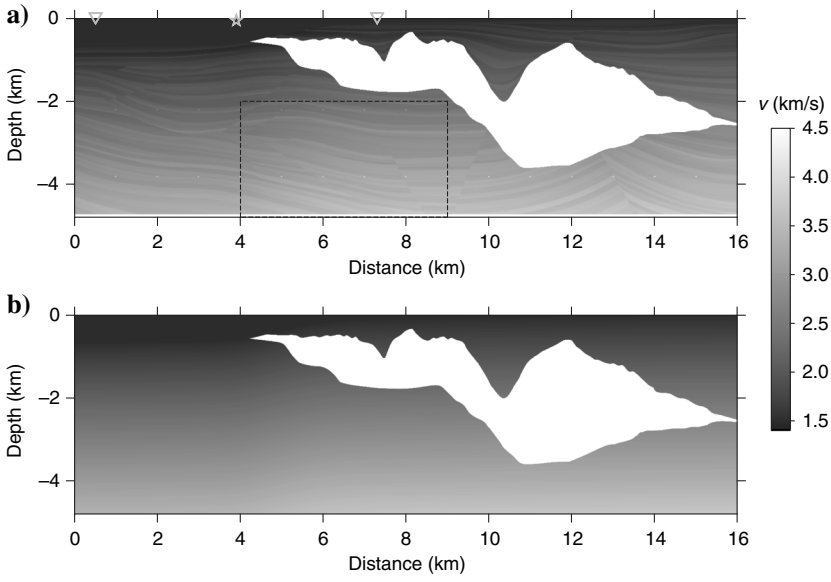

Figure 8. (a) The Sigsbee2A velocity model and (b) the Sigsbee2A migration velocity model. The gray pentagram denotes the source location of the 20th shot; two gray triangles denote the first and last receivers of the 20th shot.

In Figure 8a, the gray pentagram denotes the shot location of the 20th shot. In addition, two gray triangles denote the first and last receivers of the 20th shot experiment, respectively.

The initial RTM image (the image at the first iteration) of the Sigsbee2A model is shown in Figure 9a. Figure 9b is the inverted image after 10 iterations using CLSRTM with the PSM, whereas Figure 9c and 9d is the inverted images after 10 and 20 iterations using CLSRTM with the ASL, respectively. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses represent the corresponding iteration numbers. Black colors represent positive values, whereas white colors represent negative values. As can be seen, the stacked images are improved with much more balanced amplitudes and high resolution when compared with the initial RTM image. In particular, the amplitudes beneath subsalt areas (delimited by rectangles in Figure 9) are enhanced by CLSRTM. However, some artifacts still exist on migrated images (marked by arrows), which may due to the imperfect migration velocity (Figure 8b). It can be seen clearly that CLSRTM with the ASL at 10 iterations (Figure 9c) generates much higher resolution image than the initial RTM image (Figure 9a) and also fewer artifacts than CLSRTM with the PSM at 10 iterations (Figure 9b). At 20 iterations, the migration artifacts for images constructed using CLSRTM with the ASL (Figure 9d) are further removed (see arrows in rectangles). These fea-

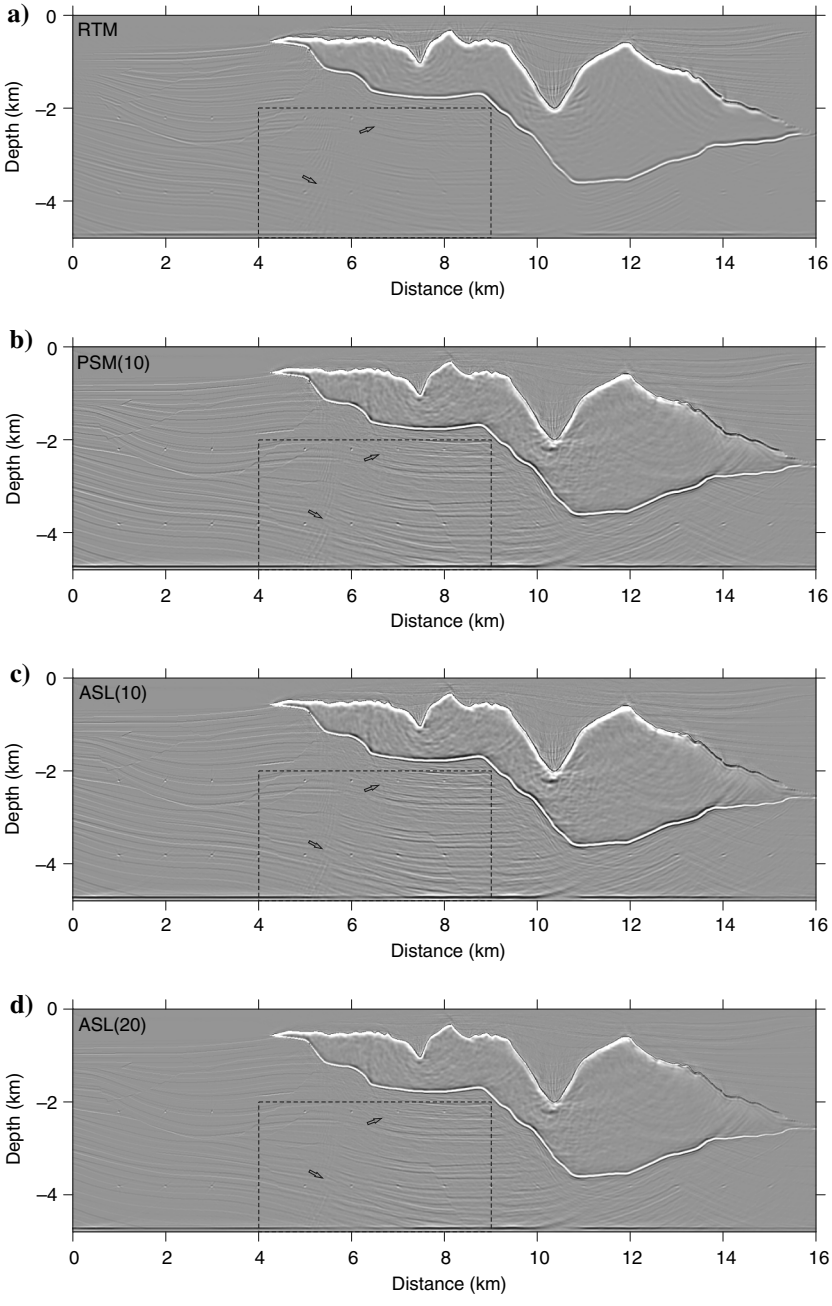

Figure 9. Migrated images of the Sigsbee2A model. (a) The initial RTM image using CLSRTM. (b) The inverted image after 10 iterations using CLSRTM with the PSM. (c and d) The inverted imaged after 10 and 20 iterations using CLSRTM with the ASL, respectively. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses denote the corresponding iteration numbers.

tures support the validity of the proposed computational procedure based on CLSRTM and the analytically determined step-length formula.

Figure 10 is a zoomed view of the partial images delimited by rectangles in Figure 9. Figure 10a–10d allows us to observe the subsalt image quality with CLSRTM. Compared with the corresponding stratigraphy of the zoomed zone (Figure 10e), the stratum structures are well-delineated. It is also appreciated that the artifacts (see arrows on images) on the migrated image obtained using CLSRTM with the ASL are weaker than those given by CLSRTM with the PSM, which demonstrates that the ASL provides better results than the PSM.

For the Sigsbee2A model, the salt reflection in migrated images has very high amplitude. The wavenumber spectrum for the entire model essentially represents the amplitude of the salt boundary. To view an improvement in the resolution of the migrated images, we only plot the wavenumber spectra of the zoomed zone instead of the entire images (Figure 10). Figure 11a–11d shows the wavenumber spectra of the migrated images plotted in Figure 10a–10d, respectively. To improve the visualization of these wavenumber spectra, the range of wavenumber axes are cut from the original 0.05 to 0.025. Figure 11a is the wavenumber spectrum of the initial RTM image, and Figure 11b is the wavenumber spectrum of the inverted image after 10 iterations using CLSRTM with the PSM, and Figure 11c and 11d is the wavenumber spectra of the inverted image after 10 and 20 iterations using CLSRTM with the ASL, respectively. Compared with the wavenumber spectrum of the initial RTM image (Figure 11a), the wavenumber spectra of the images with the PSM (Figure 11b) and ASL (Figure 11c and 11d) clearly contain higher wavenumber components. As before in the case of the Marmousi model, it can be observed that the wavenumber spectra of the images with the PSM and ASL exhibit wider wavenumber range than that of the initial RTM image, which means that a much better resolution is obtained with CLSRTM than that of RTM. Compared with Figure 11c, the wavenumber range of the Figure 11d is slightly wider, which demonstrates that a resolution improvement has been obtained at more iterations.

The synthetic and simulated data acquired from the 20th shot with direct waves muted in the cases of the Sigsbee2A model are shown in Figure 12. Figure 12a shows the synthetic (or observed) data, Figure 12b is the initial simulated data, Figure 12c is the simulated data after 10 iterations using CLSRTM with the PSM, and Figure 12d is the simulated data after 20 iterations using CLSRTM with the ASL. Compared with the initial simulated data generated using RTM image (Figure 12b), the simulated data obtained with the inverted images (Figure 12c and 12d) match well with the synthetic data (Figure 12a). Although some events are weak on the initial simulated data (indicated by arrows in Figure 12b), the simulated data demigrated with the inverted results demonstrate that CLSRTM with the ASL works well compared with the conventional PSM even when using imperfect migration velocity model (Figure 8b).

The crosscorrelation coefficients at receivers of the 20th shot are shown in Figure 13. The numbers in parentheses denote the corresponding iteration numbers. In the right gray-filled area, the quasi-horizontal strata are intruded by a high-velocity salt body (see Figure 8a). In this salt-body-intruded area, the initial cross-correlation coefficient is negative because of the very strong re-

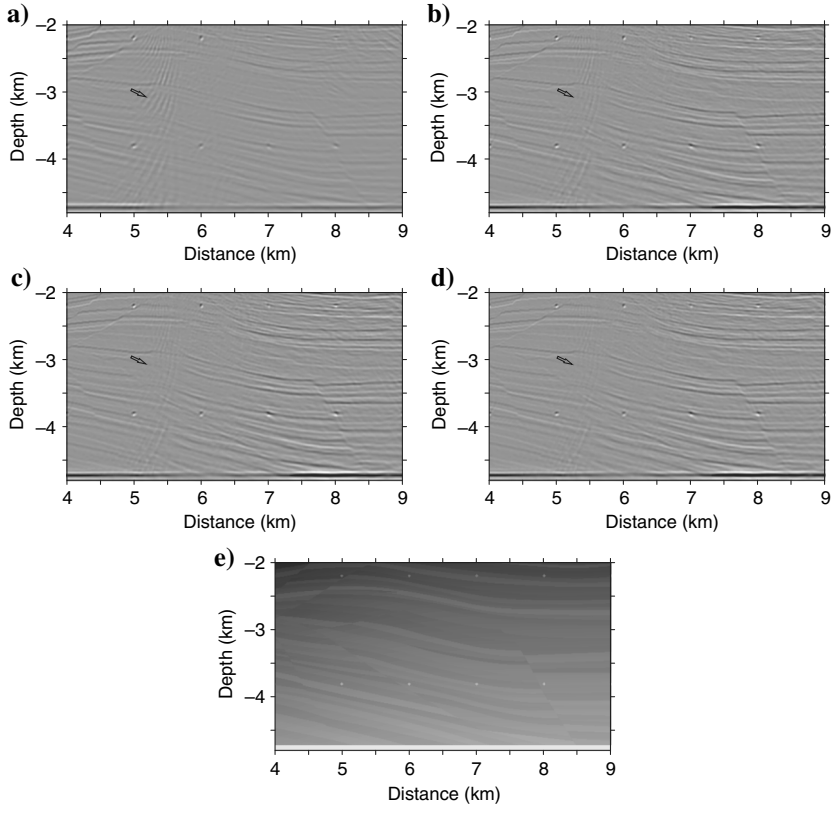

Figure 10. Zoomed view (a-d) of the partial migrated images enclosed by the rectangles shown in Figure 9a–9d, respectively. The panel in the bottom (e) shows the corresponding stratigraphy of the reference zone. The annotation shows that the ASL has a clearer image when compared with the PSM.

flection at salt boundaries. Generally, the crosscorrelation profile obtained using the initial RTM image (red line) is remarkably lower than those obtained using the migrated images with the PSM (blue line) after 10 iterations and with the ASL (green line) after 20 iterations. At more iterations, the crosscorrelation coefficient becomes more balanced and larger. In addition, the crosscorrelation coefficient obtained by the ASL is slightly larger than that obtained by the PSM, which illustrates that the ASL is more efficient than the PSM.

The objective function values E versus the number of iterations with the PSM (solid line) and ASL (dotted line) are shown in Figure 14. It can be seen that in the second case (ASL), these values converge to a much smaller value. In particular, after the 10th iteration, the objective function values with the ASL are clearly smaller than those of the PSM, whereas the convergence curve of the PSM becomes flat (meet the stopping criterion) and no longer decreases. This further demonstrates that the ASL is very effective, and it can drive the objective function value converges to a smaller minimum.

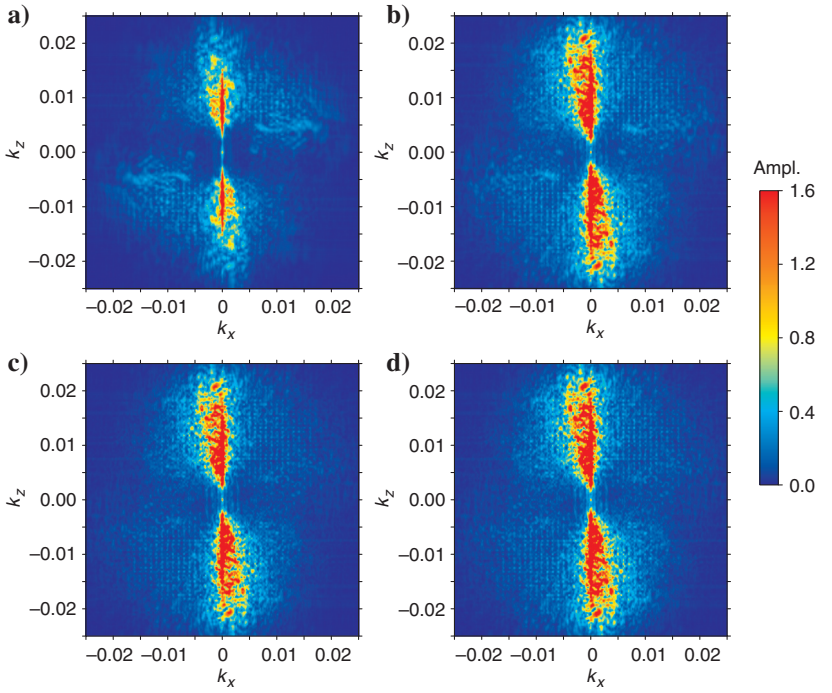

Figure 11. Wavenumber spectra of migrated images of the Sigsbee2A model. (a) The wavenumber spectrum of the initial RTM image. (b) The wavenumber spectrum of the inverted image after 10 iterations using CLSRTM with the PSM. (c and d) The wavenumber spectra of the inverted image after 10 and 20 iterations using CLSRTM with the ASL, respectively. To improve the visualization of the wavenumber spectra, the range of the wavenumber axes was reduced from 0.05 (in the original plot) to 0.025 (as can be seen here). Compared with the initial RTM image, the inverted images using CLSRTM with the PSM and ASL generate increased amplitude at a higher wavenumber.

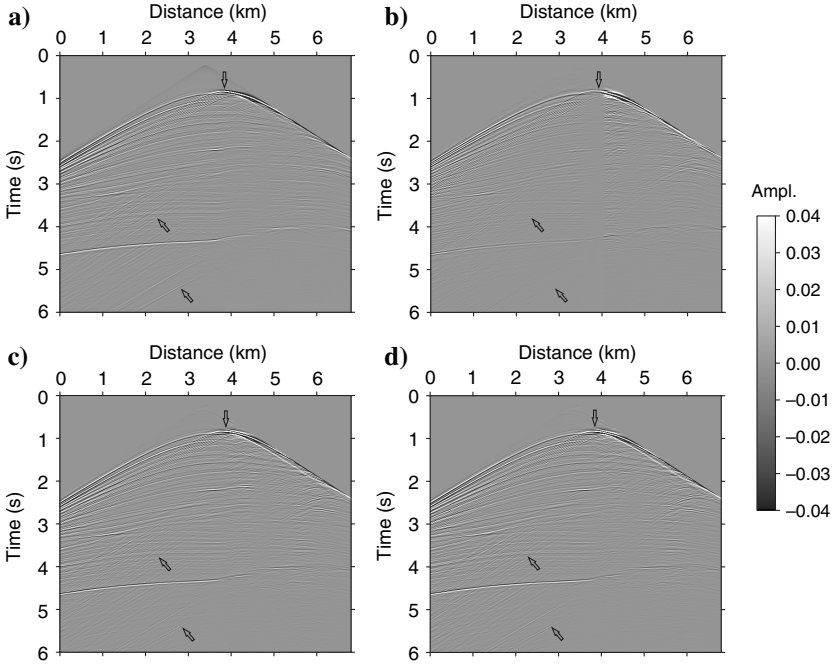

Figure 12. Common shot records of the 20th shot (direct waves are muted) in the case of the Sigsbee2A model. (a) Synthetic data, (b) initial simulated data, (c) simulated data after 10 iterations using CLSRTM with the PSM, and (d) simulated data after 20 iterations using CLSRTM with the ASL.

Robustness tests

In the above tests, the data are regularly acquired. In this subsection, we use the irregularly acquired data and the noisy data to check the robustness of the ASL.

Irregularly acquired data

We randomly delete some seismic traces from the Marmousi data set to simulate an irregularly acquired data set. Whether a trace is deleted is determined by a pseudorandom number drawn from the standard uniform distribution on the open interval (0, 1). This random number varies from one trace to another. The seismic traces will be deleted to simulate a missing trace wherever the random number is greater than 0.9. The common shot records of the 29th shot is shown in Figure 15a (hereafter Marmousi-irregular), whereas Figure 15b shows the miss rate for all 58 shots. Generally, the miss rate range from 4% to 12%. The miss rate is slightly low when the shotpoint is close to both ends of the model, whereas the missing trace is approximate to 10% when the shotpoint is located in the middle of the model.

The migrated images are shown in Figure 16. Figure 16a is the initial RTM image; Figure 16b and 16c is the inverted images after 10 iterations using CLSRTM with the PSM and ASL, respectively. Black colors represent positive values, whereas white colors represent negative values. Generally, much higher resolutions (marked by arrows) and balanced amplitudes (enclosed by the dotted line ellipses) are obtained using CLSRTM with the PSM (Figure 16b) and the ASL (Figure 16c) when compared to the initial RTM image (Figure 16a). This illustrates the robustness of the ASL for irregular acquisition geometry. However, some random artifacts at deeper layers are obvious when compared with Figure 3, which may be due to the inexact gradient caused by the missing seismic traces.

Tests with noisy data

In the previous examples, the data are free of noise. Unfortunately, in real cases, the data are always contaminated by stochastic and/or coherent noise. To verify the effectiveness of our deduced step-length formula, different levels of Gaussian white noise are added into the data set generated from the Marmousi model to achieve the decreasing signal-to-noise ratio (S/N) of 40 and 20 dB, respectively. To purely investigate the robustness of the ASL, an impractical high noise level (20 dB) is also considered. Figure 17 shows common shot records of the 29th

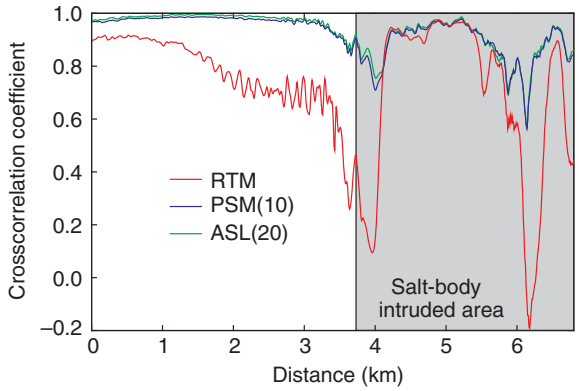

Figure 13. The crosscorrelation coefficients of the 20th shot in the case of the Sigsbee2A model. Red, blue, and green lines denote the crosscorrelation coefficients of the initial simulated data, simulated data after 10 iterations using CLSRTM with the PSM, simulated data after 20 iterations using CLSRTM with the ASL, respectively. The numbers in parentheses denote the corresponding iteration numbers. The PSM and ASL are better as the crosscorrelation coefficient is closer to one.

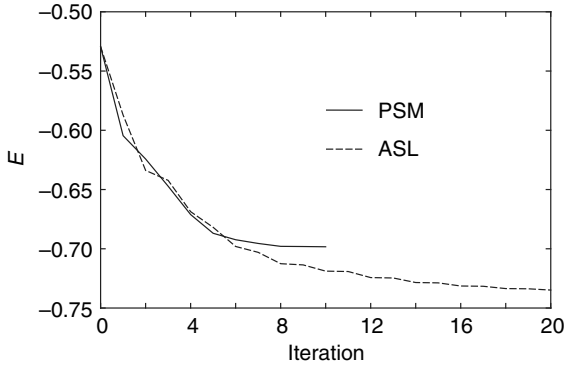

Figure 14. Values of the objective function E versus the number of iterations. The results make reference to the Sigsbee2A model and CLSRTM, either with the PSM (solid line) or the ASL (dotted line). After 10 iterations, the convergence curve of CLSRTM with the PSM becomes flat and no longer decreases.

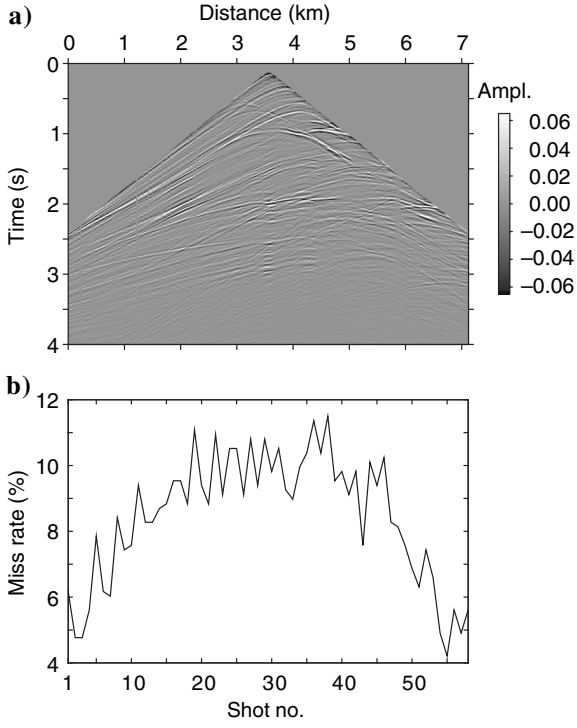

Figure 15. (a) Data acquired irregularly from the common shot records of the 29th shot in the case of the Marmousi model. (b) Miss rate for all 58 shots.

shot with an S/N of 40 (Figure 17a) and 20 dB (Figure 17b). Compared with Figure 17a, the amplitude of the data contaminated by 20 dB is larger. This illustrates that the amplitude of noise is comparable with that of signal. Then, the noisy data are migrated.

The migrated images from data with S/N of 40 (hereafter Marmousi-40 dB) and 20 dB (hereafter Marmousi-20 dB) are shown in Figures 18 and 19, respectively. Black colors represent positive values, whereas white colors represent negative values. The numbers in parentheses denote corresponding iteration numbers. Figures 18a and 19a shows the respective initial RTM images. In Figure 18b and 18c, we present the migrated images after eight iterations using the PSM and ASL with the data set shown in Figure 17a, respectively. Analogously, in Figure 19b and 19c, we present the migrated images after seven iterations using the PSM and ASL with the data set shown in Figure 17b, respectively. The migrated images (Figure 18) obtained using data shown in Figure 17a still have high resolution and definition, whereas the migrated images in Figure 19 look smeared. It can be seen that migration artifacts become more and more obvious (see the locations marked by arrows in Figure 19) as the S/N decreases, which may be due to the inaccurate gradient caused by severe noise. In deeper layers, the traveltime of seismic events maybe is not correct because of the large cumulated error. The number of iterations for CLSRTM to meet a predefined stopping threshold decreases with the increase in the noise level, which may be accounted for noise hampering the inversion to continue. Although the seismic record with an S/N of 20 dB (Figure 13b) hardly allows the identification of seismic events, the structures of the Marmousi model are well-imaged and can be recognized with a certain detail. This example proves that the ASL also works well and is robust when using a data set with severe white Gaussian noise contamination.

Computational efficiency

For the above numerical examples, we run the same code with MPI but different versions of step-length formulas. On each node, the number of CPU cores is six and the dominant frequency of CPU is 2.4 GHz (Six-Core AMD Opteron™ Processor 2431). In all cases, each CPU core deals with tasks of one shot. We allocate five computing nodes for the Marmousi model and seven computing nodes for the Sigsbee2A model. The stopping criteria are the same for all numerical examples namely (1) the increase in the objective function value, (2) or that the number of iterations reaches its maximum value, and (3) or that the relative change rate of the objective function values is less than 0.0002. Thus, the differences in the

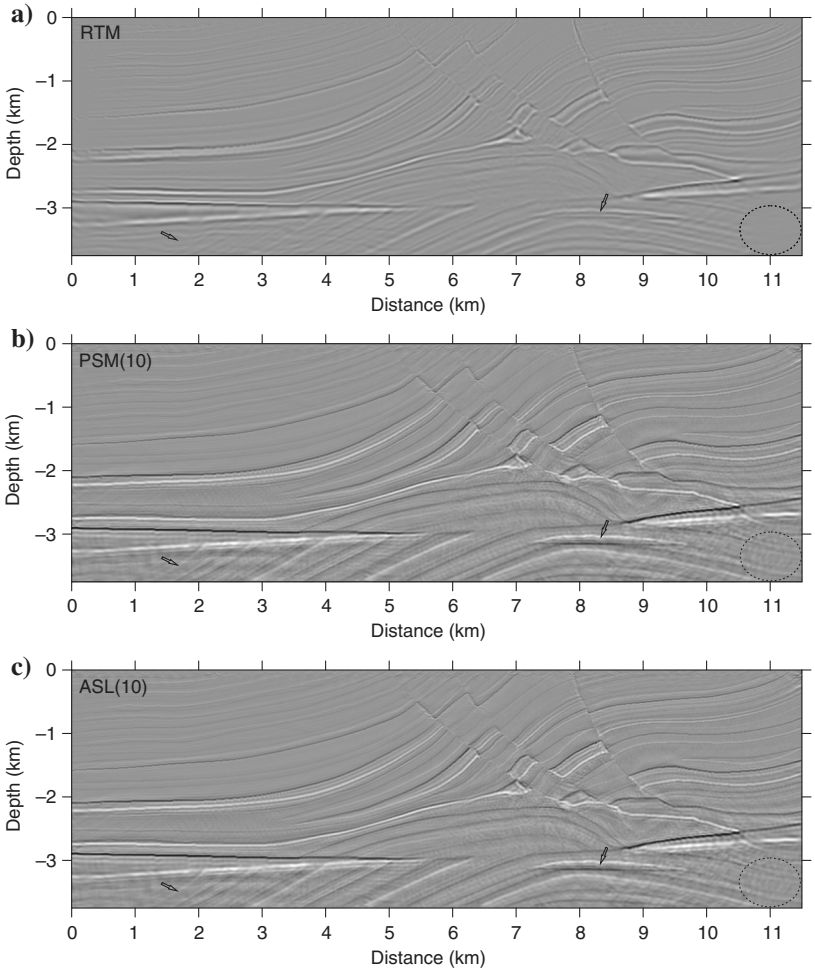

Figure 16. Migrated images of the Marmousi model with irregular acquisition geometry. Panel (a) shows the initial RTM image by using the CLSRTM, (b) the inverted image after 10 iterations using CLSRTM with the PSM, and (c) the inverted image after 10 iterations using CLSRTM with the ASL. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses denote the corresponding iteration numbers.

results are mainly due to the differences between the step-length formula of the algorithms and not extraneous issues.

At each iteration, CLSRTM involves the migration (for gradient) and demigration (for objective function) processes once, respectively. Therefore, four times the modeling are required, i.e., modeling twice each for RTM and RTDM. For gradient-based CLSRTM, an optimal step length is always necessary to iteratively decrease the value of the objective function. To estimate this optimal step length, the PSM always needs to read the simulated, perturbed, and observed data of all shots, whereas, for the ASL, it is incidentally calculated in the RTDM process. As is well-known, the I/O operations are always slow especially for MPI communication, which significantly degrades the efficiency of CLSRTM algorithm. Figure 20 shows the number of the extra estimations of the objective function value for the PSM in the above experiments. This metric is independent of hardware differences and implementations details. It is also unbiased to evaluate the efficiency of both step-length formulas. In Figure 20, Marmousi and Sigsbee2A represent experiments with the noise-free data; Marmousi-irregular represents the experiment with the irregular acquisition geometry; Marmousi-40 dB and Marmousi-20 dB indicate data sets with S/N of 40 and 20 dB, respectively. It can be seen that the PSM needs several times extra estimation of the objective function value in experiments with noise-free data (Marmousi and Sigsbee2A), with irregularly acquired data (Marmousi-irregular), and with noisy data (Marmousi–40 dB and Marmousi–20 dB). Compared with the ASL, the PSM is not always efficient due to at least twice reading of the simulated, perturbed, and observed seismic records of all shots. The detailed differences between CLSRTM with the PSM and ASL are listed in Figure 1.

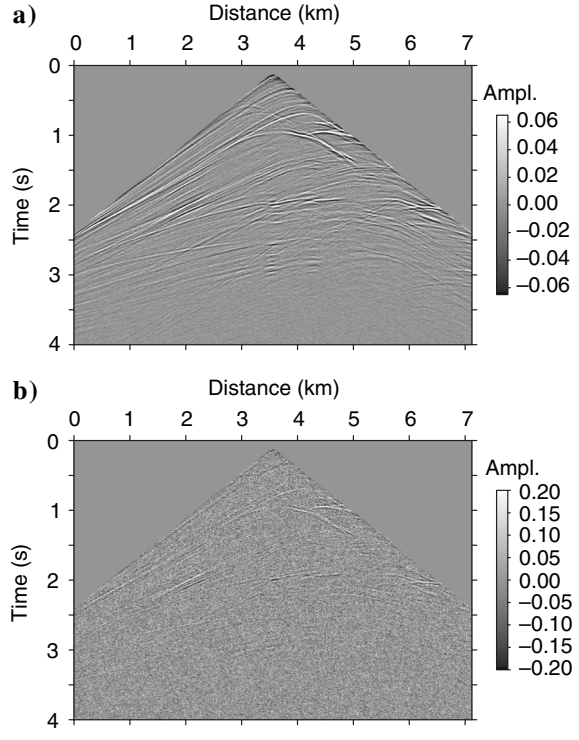

Figure 17. Common shot records of the 29th shot by adding Gaussian white noise with the S/N of (a) 40 and (b) 20 dB.

CONCLUSIONS

Based on the linear characteristic of the demigration operator, we derive an ASL formula for CLSRTM. The validity of the ASL is verified using the Marmousi and Sigsbee2A models. Numerical examples prove that the PSM is not effective, especially in complex models with imperfect migration velocity (such as the Sigsbee2A model test), whereas the ASL can make the value of the objective function converges to a smaller minimum. Although some artifacts are remained at deep layers in the test with the data acquired irregularly, the PSM and ASL can obtain high-resolution and balanced amplitude results. This experiment verifies the robustness of the ASL for irregular acquisition geometry. Tests performed with data contaminated by different levels of noises further demonstrate the robustness of the proposed ASL. An impractically high noise level shows that the migration artifacts become more obvious at deep layers with increase in the number of iterations, which may be responsible for the large cumulated traveltime errors caused by high-level noise. In terms of computational efficiency, the PSM always needs some extra I/O operations (at least twice reading of the simulated, perturbed, and observed seismic records of all shots), whereas the ASL for CLSRTM requires no extra operations. Compared with the commonly used PSM, the ASL is clearly more efficient.

ACKNOWLEDGMENTS

We want to honor the memory of Z. Zhang, who was the leader of the research group in the Institute of Geology and Geophysics of Chinese Academy of Sciences (IGGCAS) and also our dear friend, who passed away unexpectedly on 6 September 2013 at the young age of 49 years. We thank the Computer Simulation Laboratory at IGGCAS for the allocation of computing time.

The authors would like to thank editor M. Sacchi and assistant editor J. Shragge, reviewers M. Wong, and another two anonymous reviewers for their very attentive reviews and constructive comments, which led to a significant improvement of the early manuscript. Fruitful discussions with Y. Zhang, L. Duan, Q. Liu, and X. Liu are also greatly appreciated. We gratefully acknowledge the financial support for this work contributed by the China Earthquake Administration (grant no. 201408023) and the National Natural Science Foundation of China (grant nos. 41522401, 41374062, 41274070, and 41404073).

APPENDIX A

ANALYTICAL STEP-LENGTH FORMULA

In this appendix, we give a detailed deduction of the proposed ASL formula based on the linear characteristic of the demigration operator.

At the current state (i.e., stacked image $\mathbf{r}$), by applying the Taylor series expansion of α to the second order, the objective function (1) can be approximated as

$$E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) = -\iint \frac{\int M(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})\cdot\mathbf{D}dt}{\sqrt{\int[M(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})]^2dt}\sqrt{\int\mathbf{D}^2dt}}d\boldsymbol{x}_s d\boldsymbol{x}_r, \tag{A-1}$$

so that the objective function depends only on step length α when the stacked image $\mathbf{r}$ and perturbed-stacked image $\boldsymbol{\delta}\mathbf{r}$ are fixed (i.e., at the current state). Consequently, equation A-1 can be expanded by a Taylor series of step length α

$$\begin{aligned} E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) &\approx E(\mathbf{r}) + \alpha[\nabla E(\mathbf{r})]^T\boldsymbol{\delta}\mathbf{r} + \frac{\alpha^2}{2}(\boldsymbol{\delta}\mathbf{r})^T[\nabla^2 E(\mathbf{r})]\boldsymbol{\delta}\mathbf{r} \\ &= c + b\alpha + a\alpha^2 \\ &= E(\mathbf{r}) + \frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\bigg|_{\alpha=0}\alpha + \frac{1}{2}\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\bigg|_{\alpha=0}\alpha^2. \end{aligned} \tag{A-2}$$

Considering the linear relationship between the data and parameters (the stacked image $\mathbf{r}$ and perturbed-stacked image $\boldsymbol{\delta}\mathbf{r}$) for this problem, the data in the LSM problem are regarded as linear with respect to the reflectivity or image,

$$M(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) = M(\mathbf{r}) + \alpha M(\boldsymbol{\delta}\mathbf{r}). \tag{A-3}$$

Substituting equation A-3 into A-1, we obtain

$$\begin{aligned} E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r}) &= -\int\int \frac{\int M(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})\cdot\mathbf{D}dt}{\sqrt{\int[M(\mathbf{r})+\alpha M(\boldsymbol{\delta}\mathbf{r})]^2dt}\sqrt{\int\mathbf{D}^2dt}}d\boldsymbol{x}_r d\boldsymbol{x}_s \\ &= -\int\int \frac{\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})\cdot\mathbf{D}dt}{\sqrt{\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})^2dt}\sqrt{\int\mathbf{D}^2dt}}d\boldsymbol{x}_r d\boldsymbol{x}_s, \end{aligned} \tag{A-4}$$

where $\boldsymbol{\delta}\mathbf{d} = M(\boldsymbol{\delta}\mathbf{r})$ is the perturbed seismic data generated by the perturbed-stacked image or descent direction $\boldsymbol{\delta}\mathbf{r}$. After comparing the terms of the right sides of equation A-2, we obtain the following relationships:

$$\begin{cases} c = E(\mathbf{r}), \\ b = \frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\Big|_{\alpha=0}, \\ a = \frac{1}{2}\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\Big|_{\alpha=0}. \end{cases} \tag{A-5}$$

If the objective function approaches its local or global minimum, the expression satisfies the following condition:

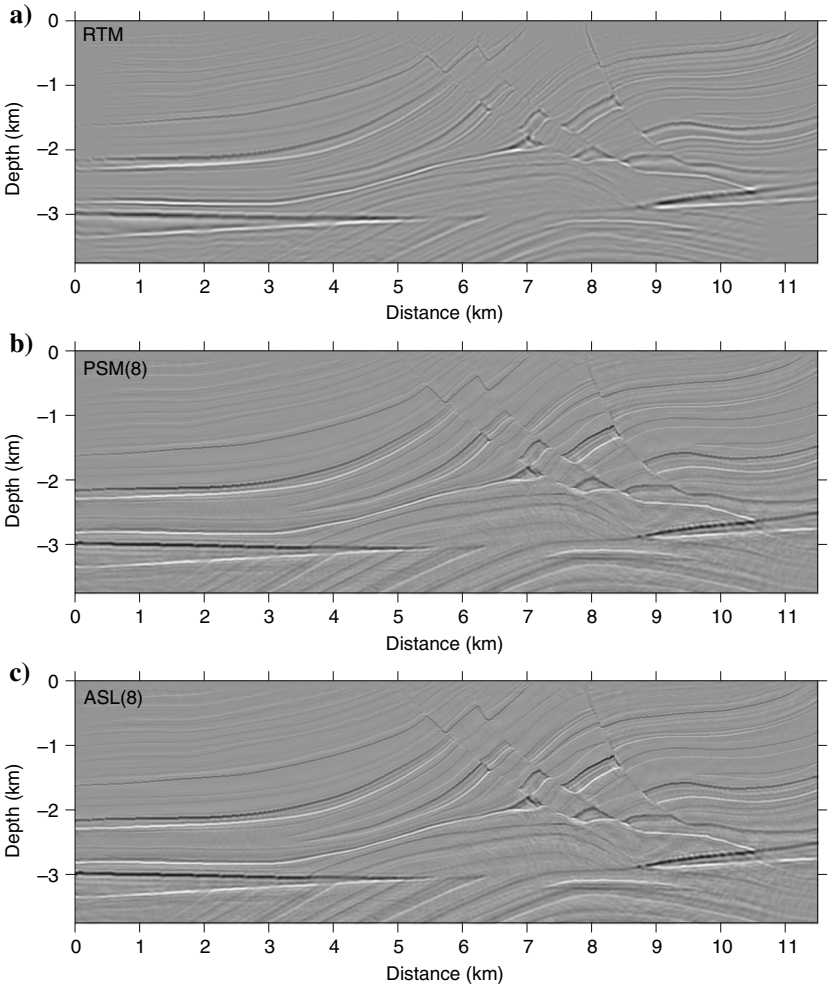

Figure 18. Migrated images of the Marmousi model using noisy data contaminated with white Gaussian noise and S/N of 40 dB. Panel (a) shows the initial RTM image using CLSRTM, (b) the inverted image after eight iterations using CLSRTM with the PSM, and (c) the inverted image after eight iterations using CLSRTM with the ASL. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses denote the corresponding iteration numbers.

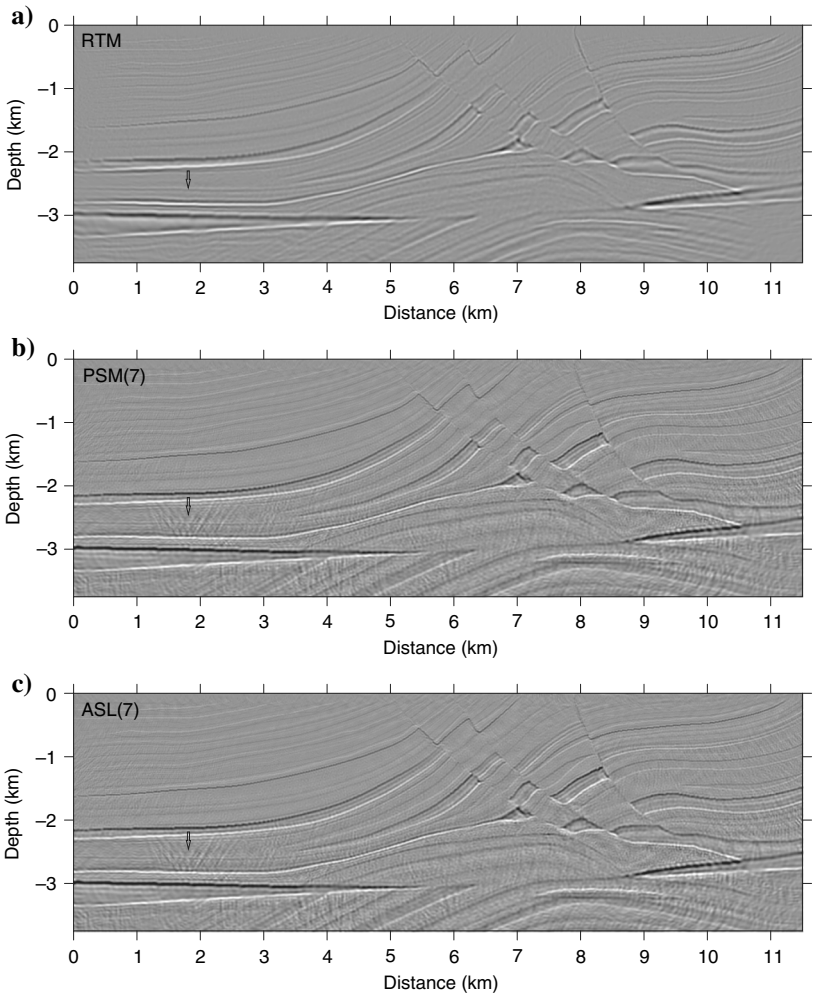

Figure 19. Migrated images of the Marmousi model using noisy data contaminated with white Gaussian noise and S/N of 20 dB. Panel (a) shows the initial RTM image using CLSRTM, (b) the inverted image after nine iterations using CLSRTM with the PSM, and (c) the inverted image after seven iterations using CLSRTM with the ASL. The acronym on the top left corner of each image refers to the inversion method used for computation. The numbers in parentheses denote the corresponding iteration numbers.

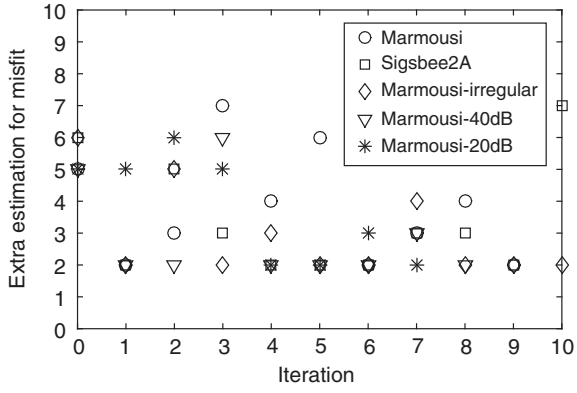

Figure 20. The number of the extra estimation of objective function. Marmousi and Sigsbee2A denote the experiments with noise-free data as those shown in Figures 2a and 8a, respectively. Marmousi-irregular (Figure 15a) denotes the experiment with data acquired irregularly. Marmousi-40 dB and Marmousi-20 dB denote the experiments using noisy data with S/Ns of 40 (Figure 17a) and 20 dB (Figure 17b), respectively.

$$\frac{\partial}{\partial\alpha}E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})=0. \tag{A-6}$$

Substituting the relationship A-6 into the formula A-2, we obtain the general form of the optimal step-length formula:

$$\alpha_{\text{opt}}=-\frac{b}{2a}=-\frac{\left.\frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}\right|_{\alpha=0}}{\left.\frac{\partial^2 E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha^2}\right|_{\alpha=0}}. \tag{A-7}$$

Taking the first-order derivatives of formula A-4 with respect to step length α, we obtain the following expression:

$$\frac{\partial E(\mathbf{r}+\alpha\boldsymbol{\delta}\mathbf{r})}{\partial\alpha}=\iint\frac{1}{\sqrt{\int\mathbf{D}^2dt}}\left[-\frac{\int\boldsymbol{\delta}\mathbf{d}\cdot\mathbf{D}dt}{\sqrt{\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})^2dt}}+\frac{\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})\cdot\mathbf{D}dt}{\left(\sqrt{\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})^2dt}\right)^3}\int(\mathbf{d}+\alpha\boldsymbol{\delta}\mathbf{d})\cdot\boldsymbol{\delta}\mathbf{d}dt\right]d\boldsymbol{x}_r d\boldsymbol{x}_s. \tag{A-8}$$

And taking the first-order derivatives of formula A-8 with respect to step length α, we obtain the following expression:

$$\frac{\partial^2 E(\mathbf{r}+\alpha\delta\mathbf{r})}{\partial\alpha^2} = \iint \frac{1}{\sqrt{\int \mathbf{D}^2 dt}} \left\{ \frac{\int \delta\mathbf{d}\cdot\mathbf{D}\,dt \int(\mathbf{d}+\alpha\delta\mathbf{d})\cdot\delta\mathbf{d}dt}{\left(\sqrt{\int(\mathbf{d}+\alpha\delta\mathbf{d})^2 dt}\right)^3} + \frac{\int \delta\mathbf{d}\cdot\mathbf{D}\,dt \int(\mathbf{d}+\alpha\delta\mathbf{d})\cdot\delta\mathbf{d}\,dt + \int(\mathbf{d}+\alpha\mathbf{d})\cdot\mathbf{D}\,dt \int(\delta\mathbf{d})^2 dt}{\left(\sqrt{\int(\mathbf{d}+\alpha\delta\mathbf{d})^2 dt}\right)^3} - 3\frac{\int(\mathbf{d}+\alpha\delta\mathbf{d})\cdot\mathbf{D}\,dt[\int(\mathbf{d}+\alpha\delta\mathbf{d})\cdot\delta\mathbf{d}\,dt]^2}{\left(\sqrt{\int(\mathbf{d}+\alpha\delta\mathbf{d})^2 dt}\right)^5} \right\} d\boldsymbol{x}_r d\boldsymbol{x}_s. \quad \text{(A-9)}$$

Substituting equations A-8 and A-9 into equation A-7, we can obtain the optimal step length:

$$\alpha_{opt} = -\frac{b}{2a} \quad \text{(A-10)}$$

being

$$a = \frac{1}{2}\iint \frac{1}{\sqrt{\int \mathbf{D}^2 dt}\sqrt{\int \mathbf{d}^2 dt}} \left[2\frac{\int \mathbf{d}\cdot\delta\mathbf{d}dt}{\int \mathbf{d}^2 dt}\int \delta\mathbf{d}\cdot\mathbf{D}dt + \frac{\int(\delta\mathbf{d})^2 dt}{\int \mathbf{d}^2 dt}\int \mathbf{d}\cdot\mathbf{D}dt - 3\left(\frac{\int \mathbf{d}\cdot\delta\mathbf{d}dt}{\int \mathbf{d}^2 dt}\right)^2 \int \mathbf{d}\cdot\mathbf{D}dt \right] d\boldsymbol{x}_r d\boldsymbol{x}_s \quad \text{(A-11)}$$

and

$$b = \iint \frac{1}{\sqrt{\int \mathbf{D}^2 dt}\sqrt{\int \mathbf{d}^2 dt}} \left[\frac{\int \mathbf{d}\cdot\mathbf{D}dt \int \mathbf{d}\cdot\delta\mathbf{d}dt}{\int \mathbf{d}^2 dt} - \int \delta\mathbf{d}\cdot\mathbf{D}dt \right] d\boldsymbol{x}_r d\boldsymbol{x}_s. \quad \text{(A-12)}$$

REFERENCES

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524, doi: 10.1190/1.1441434.

Claerbout, J., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481, doi: 10.1190/1.1440185.

Claerbout, J., and S. Doherty, 1972, Downward continuation of moveout corrected seismograms: Geophysics, **37**, 741–768, doi: 10.1190/1.1440298.

Clapp, M. L., 2005, Imaging under salt: Illumination compensation by regularized inversion: Ph.D. thesis, Stanford University.

Dai, W., P. Fowler, and G. T. Schuster, 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **60**, 681–695, doi: 10.1111/j.1365-2478.2012.01092.x.

Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: Geophysics, **76**, no. 5, R135–R146, doi: 10.1190/geo2010-0159.1.

Dong, S., J. Cai, M. Guo, S. Suh, Z. Zhang, B. Wang, and Z. Li, 2012, Least-squares reverse time migration: Towards true amplitude imaging and improving the resolution: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-1488.1.

Duquet, B., K. J. Marfurt, and J. A. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209, doi: 10.1190/1.1444812.

Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271.

Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, no. 6, WCA5–WCA17, doi: 10.1190/1.3223188.

Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, **43**, 1342–1351, doi: 10.1190/1.1440899.

Guitton, A., and B. Kaelin, 2006, Least-squares attenuation of reverse-time migration artifacts: 76th Annual International Meeting, SEG, Expanded Abstracts, 2348–2352.

Hemon, C., 1978, Equations d'onde et modeles: Geophysical Prospecting, **26**, 790–821, doi: 10.1111/j.1365-2478.1978.tb01634.x.

Hill, R., 1990, Gaussian beam migration: Geophysics, **55**, 1416–1428, doi: 10.1190/1.1442788.

Hill, R., 2001, Prestack Gaussian-beam depth migration: Geophysics, **66**, 1240–1250, doi: 10.1190/1.1487071.

Hu, P., and Z. Wang, 2014, A non-monotone line search combination technique for unconstrained optimization: The Open Electrical and Electronic Engineering Journal, **8**, 218–221, doi: 10.2174/1874129001408010218.

Huang, L. J., and M. C. Fehler, 1998, Accuracy analysis of the split-step Fourier propagator: Implications for seismic modeling and migration: Bulletin of the Seismological Society of America, **88**, 18–29.

Ji, J., 2009, An exact adjoint operation pair in time extrapolation and its application in least-squares reverse-time migration: Geophysics, **74**, no. 5, H27–H33, doi: 10.1190/1.3173894.

Kuehl, H., and M. D. Sacchi, 1999, Least-squares split-step migration using the Hartley transform: 69th Annual International Meeting, SEG, Expanded Abstracts, 1548–1551.

Kuehl, H., and M. D. Sacchi, 2001, Split-step WKBJ least-squares migration/inversion of incomplete data: Proceedings of the 5th SEG International Symposium on Imaging Technology.

Lambaré, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154, doi: 10.1190/1.1443328.

Lan, H., Z. Zhang, J. Chen, and Y. Liu, 2014, Reverse time migration from irregular surface by flattening surface topography: Tectonophysics, **627**, 26–37, doi: 10.1016/j.tecto.2014.04.015.

Liu, S., X. Li, W. Wang, and T. Zhu, 2015, Source wavefield reconstruction using a linear combination of the boundary wavefield in reverse time migration: Geophysics, **80**, no. 6, S203–S212, doi: 10.1190/geo2015-0109.1.

Liu, Y., X. Chang, D. Jin, R. He, H. Sun, and Y. Zheng, 2011, Reverse time migration of multiples for subsalt imaging: Geophysics, **76**, no. 5, WB209–WB216, doi: 10.1190/geo2010-0312.1.

Liu, Y., S. Liu, M. Zhang, and D. Ma, 2012, An improved perfectly matched layer absorbing boundary condition for second order elastic wave equation (in Chinese): Progress in Geophysics, **27**, 2113–2122.

McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophysical Prospecting, **31**, 413–420, doi: 10.1111/j.1365-2478.1983.tb01060.x.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221, doi: 10.1190/1.1444517.

Nguyen, B. D., and G. A. McMechan, 2015, Five ways to avoid storing source wavefield snapshots in 2D elastic prestack reverse time migration: Geophysics, **80**, no. 1, S1–S18, doi: 10.1190/geo2014-0014.1.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782, doi: 10.1090/S0025-5718-1980-0572855-7.

Paffenholz, J., B. McLain, J. Zaske, and P. J. Keliher, 2002, Subsalt multiple attenuation and imaging: Observations from the Sigsbee2B synthetic dataset: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2122–2125.

Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292, doi: 10.1190/1.1442836.

Pratt, W. K., 1978, Digital image processing: Wiley Interscience.

Schneider, W., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, 49–76, doi: 10.1190/1.1440828.

Schuster, G., 1993, Least-squares crosswell migration: 63rd Annual International Meeting, SEG, Expanded Abstracts, 110–113.

Stoffa, P. L., J. T. Fokkema, R. M. D. Freire, and W. P. Kessinger, 1990, Split-step Fourier migration: Geophysics, **55**, 410–421, doi: 10.1190/1.1442850.

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.

Valenciano, A., 2008, Imaging by wave-equation inversion: Ph.D. thesis, Stanford University.
Vigh, D., W. Starr, and J. Kapoor, 2009, Developing earth models with full waveform inversion: The Leading Edge, **28**, 432–435, doi: 10.1190/1.3112760.
Whitmore, D., 1983, Iterative depth migration by backward time propagation: 53rd Annual International Meeting, SEG, Expanded Abstracts, 382–385.
Wong, M., B. L. 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, doi: 10.1190/geo2015-0093.1.
Wong, M., S. Ronen, and B. Biondi, 2011, Least-squares reverse time migration/inversion for ocean bottom data: A case study: 81st Annual International Meeting, SEG, Expanded Abstracts, 2369–2373.
Wu, S., Y. Wang, Y. Zheng, and X. Chang, 2015, Limited-memory BFGS based least-squares pre-stack Kirchhoff depth migration: Geophysical Journal International, **202**, 738–747, doi: 10.1093/gji/ggv156.
Youn, O. K., and H. Zhou, 2001, Depth imaging with multiples: Geophysics, **66**, 246–255, doi: 10.1190/1.1444901.
Zhang, Y., L. Duan, and Y. Xie, 2015, A stable and practical implementation of least-squares reverse time migration: Geophysics, **80**, no. 1, V23–V31, doi: 10.1190/geo2013-0461.1.
Zhang, Y., and S. James, 2009, Practical issues of reverse time migration: True amplitude gathers, noise removal and harmonic-source encoding: Presented at the CPS/SEG Beijing 2009 International Geophysical Conference and Exposition.