

Least-squares reverse-time migration for reflectivity imaging

YAO Gang¹ & WU Di^{2*}

¹*Department of Earth Science and Engineering, Imperial College London, London SW7 2BP, UK;*

²*The Unconventional Natural Gas Institute, China University of Petroleum (Beijing), Beijing 102249, China*

Received March 24, 2015; accepted May 13, 2015; published online July 27, 2015

A least-squares reverse-time migration scheme is presented for reflectivity imaging. Based on an accurate reflection modeling formula, this scheme produces amplitude-preserved stacked reflectivity images with zero phase. Spatial preconditioning, weighting and the Barzilai-Borwein method are applied to speed up the convergence of the least-squares inversion. In addition, this scheme compensates the effect of ghost waves to broaden the bandwidth of the reflectivity images. Furthermore, roughness penalty constraint is used to regularize the inversion, which in turn stabilizes inversion and removes high-wavenumber artifacts and mitigates spatial aliasing. The examples of synthetic and field datasets demonstrate the scheme can generate zero-phase reflectivity images with broader bandwidth, higher resolution, fewer artifacts and more reliable amplitudes than conventional reverse-time migration.

adjoint migration, least-squares reverse-time migration, cross-correlation imaging condition, deconvolution imaging condition, deghost

Citation: Yao G, Wu D. 2015. Least-squares reverse-time migration for reflectivity imaging. *Science China: Earth Sciences*, 58: 1982–1992, doi: 10.1007/s11430-015-5143-1

Migration is a key technique for subsurface seismic imaging by repositioning the recorded data to the location where reflection occurs. Although in principle this requires the inverse of a modeling operator, in practice the adjoint of the modeling operator is used instead. There are two main reasons for such use of adjoint migration operators. Firstly, applying adjoint operator requires significantly less calculation than inverse operators, as it is much cheaper to compute the adjoint operator than the inverse operator. Secondly, adjoint operators are unconditionally stable, since only multiplication and addition are involved in their production, whereas division is required to compute an inverse operator. Adjoint operators therefore apply to nearly all migration methods, including reverse-time migration (RTM). In cases where the data is subject to significant aliasing, truncation,

noise, or are incomplete, the adjoint operator is not a good approximation to the inverse operator, and will therefore degrade the resolution of the final migrated image. Moreover, even with perfect data, an adjoint operator still produces imperfect images (Claerbout, 1992). Hence, it is desirable to use the inverse operator to migrate seismic data.

A generalized inverse operator can be obtained by using a least-squares approach. Nemeth et al. (1999) implemented Kirchhoff migration in the least-squares inversion frame to form a least-squares migration method. The examples demonstrate improved resolution and amplitudes of the images. In addition, the least-squares migration is able to mitigate the side effects of limited aperture, gaps and coarse spatial sampling of the recorded data. To reduce the computational cost of least-squares Kirchhoff migration (Nemeth et al., 1999), Liu et al. (2005) combined least-squares inversion with wave-path migration, which is faster than Kirchhoff migration, to give a more efficient method

*Corresponding author (email: wudi800128@163.com)

titled least-squares wave-path migration. Kühl et al. (2001, 2003) presented a least-squares migration method based on the double-square-root (DSR) operator. This method can generate angle-domain common-image gathers (ADCIGs) with more accurate amplitude variations versus ray parameter (AVP) than ADCIGs using the normal DSR operator. Kaplan et al. (2010) formulated a least-squares migration method based on the shot-profile split-step migration operator. However, practical Kirchhoff migration uses an asymptotic approximation based on ray theory; thus it cannot accurately image near wavefields and has difficulties to migrate multi-arrival events in complex velocity models (Gray et al., 2001). Migration methods based on the one-way wave equation cannot properly migrate steep events or deal with lateral velocity variations and turning waves (Gray et al., 2001). As a result, the least-squares migration methods based on the Kirchhoff migration operator or one-way wave equation will inherit their respective drawbacks. To avoid such drawbacks, whilst also maintaining the advantages of least-squares migration, Dai et al. (2010, 2012) formulated least-squares reverse-time migration (LSRTM) and used a (simultaneous) multi-shots technique to reduce computation costs. The combination of least-squares inversion and conventional reverse-time migration (RTM) can retain the advantages of inverse operators and two-way wave equations for migration. However, an image relating to velocity perturbation was inverted for instead of reflectivity. In other words, the resultant images were the integral of reflectivity therefore had inaccurate phase. Dong et al. (2013) used a filter in LSRTM to match the scaled predicted data to the records, and in turn tried to address some aspects of applying LSRTM to field data, e.g., migration velocity errors and the amplitude errors of source wavelets. Zhang et al. (2015) proposed the use of zero-lag cross-correlation of predicted and observed data as the objective function to invert the reflectivity image. This method can help to mitigate the effect of the amplitude difference between predicted and observed data, and is equivalent to the least-squares function of amplitude normalized predicted and observed data.

In this work, a least-squares inversion scheme has been developed to recover the stacked reflectivity images based on a reflection modeling formula (Xu et al., 2011). Spatial preconditioning, weighting and the Barzilai-Borwein (BB) method (Barzilai et al., 1988) are used to speed up the inversion convergence significantly. Deghosting (Perz et al., 2014) has been integrated into this scheme thereby broadening the bandwidth of the reflectivity images. In cases where the data is subject to significant aliasing, truncation, noise, or is incomplete, roughness penalty constraints can be used to regularize and stabilize the inversion. This will remove high-wavenumber artifacts and mitigate the effect of spatial aliasing to generate better images. Unlike the matrix formulation of least-squares reverse-time migration (Yao et al., 2012a), this scheme does not formulate the Jacobian

matrix explicitly thus reducing the memory cost down to the same level as conventional RTM. Consequently, this scheme is ready for 3D implementation. The synthetic and field data examples successfully demonstrate that this scheme can produce images with more accurate amplitudes, broader bandwidth, higher resolution and fewer artifacts than conventional RTM.

1 Theory

Reflection data can be modeled with the expression (Xu et al., 2011):

$$d(\mathbf{x}_r, \omega) = j\omega \iint \frac{2 \cos \theta}{v(\mathbf{x})} G_r(\mathbf{x}_r | \mathbf{x}, \omega) r(\mathbf{x}, \theta) G_s(\mathbf{x} | \mathbf{x}_s, \omega) s(\omega) d\theta d\mathbf{x}, \quad (1)$$

where d is the reflection data, s is the source signature, $G_s(\mathbf{x} | \mathbf{x}_s, \omega)$ is the Green's function from the source at \mathbf{x}_s to the reflector at \mathbf{x} , $G_r(\mathbf{x}_r | \mathbf{x}, \omega)$ is the Green's function from the reflector to the receiver at \mathbf{x}_r , $r(\mathbf{x}, \theta)$ is the angle dependent reflectivity, θ is the reflection angle, v is the velocity at the reflector and j is the imaginary unit. Since the dip factor, $\frac{2 \cos \theta}{v(\mathbf{x})}$ in eq. (1) assumes the incident an-

gle equals the reflection angle, it is accurate for isotropic media. If the stacked (angle independent) reflectivity image is to be inverted, then eq. (1) can be simplified to

$$d(\mathbf{x}_r, \omega) = \int j\omega G_r(\mathbf{x}_r | \mathbf{x}, \omega) I(\mathbf{x}) G_s(\mathbf{x} | \mathbf{x}_s, \omega) s(\omega) d\mathbf{x}, \quad (2)$$

where $I(\mathbf{x})$ is the stacked reflectivity image and $j\omega$ gives a 90° phase shift to the modeling data, which is crucial for migration to produce a zero-phase image. Eq. (2) models data with the stacked reflectivity image and does not consider the dip factor; therefore it is an approximation for both isotropic and anisotropic media. Conventional adjoint migration (Claerbout, 1992) can in turn be implemented in such a form as

$$I(\mathbf{x}) = \iint \left[j\omega G_r(\mathbf{x} | \mathbf{x}_r, \omega) G_s(\mathbf{x} | \mathbf{x}_s, \omega) s(\omega) \right]^\dagger d(\mathbf{x}_r, \omega) d\mathbf{x}_r d\omega, \quad (3)$$

where \dagger denotes the complex conjugate. Eq. (3) illustrates that migration can be carried out by firstly convolving the source and receiver Green's functions with the derivative of the source wavelet, then cross-correlating with the recorded data, and finally summing over all traces and samples.

Eq. (2) describes the forward-modeling process which maps reflectivity onto the data. It can be used to solve for the reflectivity by minimizing the objective function given by

$$\varphi_d(I(\mathbf{x})) = \frac{1}{2} \sum_{\omega, \mathbf{x}_r} \|d(\mathbf{x}_r, \omega) - d_0(\mathbf{x}_r, \omega)\|_2^2, \quad (4)$$

where $d(\mathbf{x}_r, \omega)$ and $d_0(\mathbf{x}_r, \omega)$ are the predicted and recorded data respectively. This minimization can be achieved by using a variety of localized gradient methods. If only primary reflections are considered, which means the Green's functions are independent of reflectivity, differentiating eq. (4) with respect to $I(\mathbf{x})$ gives the gradient of the objective function as

$$\frac{\partial \varphi_d(I(\mathbf{x}))}{\partial I(\mathbf{x})} = \sum_{\omega, \mathbf{x}_r} \text{Re} \left\{ [j\omega G_{s,s}]^\dagger [G_r^\dagger (d(\mathbf{x}_r, \omega) - d_0(\mathbf{x}_r, \omega))] \right\}, \quad (5)$$

where Re indicates the real part of a complex number.

Equation (5) not only provides a means of calculating the gradient, but also gives it a physical meaning. In particular, $j\omega G_{s,s}$ describes the forward-propagated wavefield from the first-order derivative of the source, while $d(\mathbf{x}_r, \omega) - d_0(\mathbf{x}_r, \omega)$ is the data residual, and G_r^\dagger represents the backward propagation of the residual into the earth. The backward propagation is essentially equivalent to forward modeling the time-reversed residual as the virtual source. Multiplication of $G_r^\dagger (d(\mathbf{x}_r, \omega) - d_0(\mathbf{x}_r, \omega))$ with the complex conjugate of $j\omega G_{s,s}$ and the following summation is then equivalent to using the zero-lag cross-correlation imaging condition (Claerbout, 1971). Eq. (5) can be expressed in the time domain as

$$\frac{\partial \varphi_d(I(\mathbf{x}))}{\partial I(\mathbf{x})} = \sum_{t, \mathbf{x}_r} \left[G_s * \frac{\partial s}{\partial t} \right] [G_r \otimes (d - d_0)], \quad (6)$$

where $*$ denotes convolution, \otimes represents cross-correlation and $\frac{\partial s}{\partial t}$ is the first-order derivative of the source wavelet.

Given the gradient of the objective function, the reflectivity model can be iteratively updated as

$$\mathbf{I}_{k+1} = \mathbf{I}_k + \alpha \mathbf{q}, \quad (7)$$

where \mathbf{I} represents the vector of the image, α is an optimum step-length, and \mathbf{q} is the vector of the update direction. In practice, the step-length is determined using a linear search technique such as the secant method (Myron et al.,

1998) or the Barzilai-Borwein (BB) scheme (Barzilai et al., 1988; Wang et al., 2010), which will be described in the next section.

In theory, the operator for gradient calculation in eq. (5) is the exact adjoint of the forward modeling operator in eq. (2). However, this is difficult to implement numerically whilst also passing the ‘‘dot-product’’ test (Claerbout, 1992), which is a criterion to verify this property. As a result, the inversion may converge slowly or misconverge when linear gradient methods are used, e.g. the linear conjugate-gradient method (Scales, 1987). In our implementation, the residual and the gradient are both directly computed from the updated reflectivity model at each iteration; the step-length is calculated with a line search or the BB method. Thereby, the minimization of eq. (4) is in fact achieved with a non-linear gradient inversion scheme. This helps the convergence of the minimization process.

2 Implementation

2.1 Regularization techniques

In order to increase the convergence and therefore effectiveness of the inversion, several regularization techniques can be applied.

2.1.1 Spatial preconditioning

If the update direction in eq. (7) is opposite to the gradient, then the steepest-descent method is used. However, this method converges slowly when the update is close to the minimum. To speed up the convergence of inversion, second-order methods, e.g. Gaussian-Newton methods (Yuan et al., 1997), can be used. The update can then be implemented as

$$\mathbf{I}_{k+1} = \mathbf{I}_k - \alpha \mathbf{H}^{-1} \mathbf{g}, \quad (8)$$

where \mathbf{H}^{-1} is the inverse of the Hessian matrix and \mathbf{g} is the gradient of the objective function, given by eq. (5) or (6). If only primary reflections are considered, which means the source and receiver Green's functions are independent to the reflectivity, then differentiating the gradient with respect to \mathbf{I} gives each element of the Hessian matrix:

$$\begin{aligned} H(\mathbf{x}_i, \mathbf{x}_j) &= \frac{\partial^2 \varphi_d(I(\mathbf{x}))}{\partial I(\mathbf{x}_i) \partial I(\mathbf{x}_j)} \\ &= \sum_{\omega, \mathbf{x}_r} \text{Re} \left\{ [j\omega G_r(\mathbf{x}_i | \mathbf{x}_r) G_s(\mathbf{x}_i | \mathbf{x}_s) s]^\dagger \cdot [j\omega G_r(\mathbf{x}_j | \mathbf{x}_r) G_s(\mathbf{x}_j | \mathbf{x}_s) s] \right\}. \end{aligned} \quad (9)$$

Equation (9) shows that each element of the Hessian matrix is the zero-lag cross-correlation of $j\omega s G_s G_r$ for the two

corresponding imaging points \mathbf{x}_i and \mathbf{x}_j . Since the seismic source wavelet s is band-limited, the Hessian matrix is not a diagonal matrix. The adjoint (conventional) migration expressed by eq. (3) is just the gradient of the objective function with record as the residual. Consequently, without the Hessian matrix, it is difficult for adjoint migrations to produce images with correct amplitudes. Instead of using the full Hessian matrix, it is common practice to use the diagonal, which is trivial to invert (Virieux et al., 2009). However, the receiver Green's functions (from an imaging point to all receivers) are needed in order to calculate the diagonal elements of the Hessian matrix. This is normally considered too computationally demanding to calculate. Consequently, $G_r^\dagger G_r$ is approximated by a scalar, which is handled during the step-length calculation. Hence, the diagonal element of the Hessian matrix is approximated as

$$H(\mathbf{x}_i, \mathbf{x}_i) \approx p(\mathbf{x}_i) = \sum_{\omega} \text{Re} \left\{ \left[j\omega G_s(\mathbf{x}_i | \mathbf{x}_s) s \right]^\dagger \left[j\omega G_s(\mathbf{x}_i | \mathbf{x}_s) s \right] \right\}. \quad (10)$$

Each element in eq. (10) corresponds to the energy of the source wavefield at the corresponding imaging point. Dividing the gradient by the energy is called spatial preconditioning (Warner et al., 2013) or source illumination compensation (Kaelin et al., 2006). In practice, a small number is added into the energy to avoid the division by a zero at some imaging points. Numerical tests show this can speed up convergence significantly but at trivially minimal extra cost.

2.1.2 Weighting

Another feature of this implementation is the option to weight up the residual of events of interest, e.g. later arrivals, which thereby enhances imaging of the corresponding areas in the model. The weighting increases the contributions of these events to the objective function, and in turn affects the gradient and updates, ultimately favoring the weighted terms. By contrast, the unweighted residual becomes weaker than the weighted, and is unable to dominate the inversion. Weighting up the events of interest therefore helps improve the speed of convergence for the corresponding imaging areas. In practice, prior to migration, these events need to be determined and a weighting function is generated. During inversion, the residual of each iteration is then scaled by the weighting function.

2.1.3 Roughness penalty constraint

If the recorded data is sparse, incomplete, or the data is contaminated by noise, the inversion becomes an ill-posed problem, and may produce high-wavenumber artifacts. One way of reducing the artifacts is to use a roughness penalty constraint (Bube et al., 2008). In this implementation, this is achieved by using first-order derivatives to measure the roughness of the model. Specifically, the objective func-

tions of the roughness penalty constraint for the horizontal and vertical directions are respectively

$$\varphi_h(I(\mathbf{x})) = \frac{1}{2} \left\| \partial_h I(\mathbf{x}) \right\|_2^2, \quad (11)$$

and

$$\varphi_v(I(\mathbf{x})) = \frac{1}{2} \left\| \partial_v I(\mathbf{x}) \right\|_2^2, \quad (12)$$

where ∂_h and ∂_v are first-order derivative operators along the horizontal and vertical directions of the model. An example of the implementation of the roughness operators, ∂_h and ∂_v , can be found in eqs. (3.19) and (3.20) in Yao (2013). Combining these constraints with the original objective function, and scaling the penalty terms by the trade-off parameters, λ_h and λ_v , then gives the final objective function as

$$\varphi = \varphi_d + \lambda_h \varphi_h + \lambda_v \varphi_v, \quad (13)$$

where $\lambda_h, \lambda_v > 0$. With proper choice of regularization parameters λ_h and λ_v and incorporating a priori constraints φ_h and φ_v on the solution, minimizing eq. (13) can yield a stable solution with significantly fewer high-wavenumber artifacts. The two regularization parameters can be different. The value of them is also affected by the numerical implementation of the modeling kernel as different modeling kernels may produce different amplitudes in the data for the same model. In practice, we use the trial-and-error method to find the acceptable regularization parameters.

2.2 Deghosting

Ghost waves are generated by the reflection off a free surface. Since the free surface has negative reflectivity, ghost waves can significantly weaken the energy of low frequencies and create notches within the high frequencies in the seismic record. This reduces the resolution of seismic images (Perz et al., 2014). This effect is especially noticeable for marine seismic data because the sea surface is a perfect reflection interface and has reflectivity of almost -1 . The mechanism of ghost waves can be interpreted as a set of mirror sources and mirror receivers, which are located above the free surface and have the same distance as the true counterparts to the free surface, to respectively inject and record energy. The strength of the mirror sources and records is equal to the true counterparts scaled by the reflectivity. According to this theory, ghost waves can be simulated by using this setting. In the forward modeling process, the mirror and true sources are excited simultaneously, and final predicted data is the sum of records of the mirror and true receivers. In the back propagation step, the source side procedure is the same as that in the forward modeling process; the

residual is injected at the true receivers but simultaneously at the mirror receivers with the strength scaled by the reflectivity. In this way, LSRTM can compensate the effect of the ghost waves and give broad-band reflectivity images.

2.3 Gradient descent with BB scheme

Since only primary reflections are generated in the predicted data, the objective functions in eqs. (4) and (13) are convex. Gradient methods can be used to minimize the objective functions. The simplest method is a steepest descent. However, it converges slowly for ill-posed inversion problems. To speed up the inversion, conjugate-gradient and quasi-Newton methods can be used. But these methods may require performing a line search along update directions in a single iteration for the calculation of the optimal step-length. The line search needs extra forward modeling and/or back propagation and therefore increases computational cost significantly. To avoid the line search as well as speed up convergence by using the Hessian information, a non-monotone gradient method called the Barzilai-Borwein (BB) method (Figure 1) can be utilized (Barzilai et al., 1988; Wang et al., 2010). According to our tests, the BB method converges much faster than the steepest-descent method, but slightly slower than conjugate-gradient and quasi-Newton methods. However, the most valuable benefit is that in one iteration, the BB method only needs to calculate the gradient once and finds the step-length with negligible calculation cost; consequently the method can save the effort for more iterations.

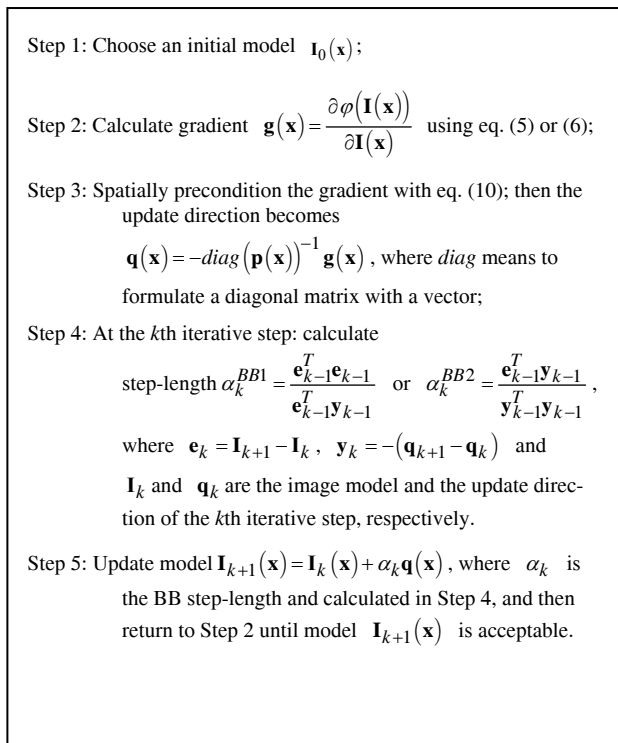


Figure 1 BB scheme for LSRTM.

3 Synthetic and field data examples

3.1 Point diffractor model

Four models are used to demonstrate the LSRTM scheme. The first model consists of nine point diffractors embedded in a medium with a constant velocity of 2000 m/s (Figure 2(a)). Figure 2(d) is one of the three shots, which have all been computed using the acoustic wave equation. A 30 Hz Ricker wavelet source is positioned at 200, 500 and 800 m on the surface with receivers spaced at 10 m intervals on the surface. The direct arrivals, which do not contribute to imaging, have been removed prior to migration. Figure 2(b) and (c) show the results of applying conventional RTM and LSRTM to the three shots, while Figure 2(e) and (f) show the modeled data corresponding to the RTM and LSRTM images. Figure 2(g) and (h) represent the LSRTM images inverted with 1950 and 2050 m/s respectively. Figure 2(i) shows the normalized residual against the percentage of velocity error, which is the ratio of the migration velocity subtracted by the true velocity and the actual true velocity.

Figure 2 shows that LSRTM has three advantages relative to RTM. Firstly, the LSRTM image has fewer artifacts than the RTM image. Eq. (5) and Figure 1 show that RTM is just the first iteration of LSRTM for an initial model with zero-value. The modeled data of the artifacts in the RTM image becomes residuals, which will be progressively diminished in the following iterations. As a result, the LSRTM generates neater images. Secondly, the diffractors in the LSRTM image have weaker sidelobes, and are significantly better resolved than those from RTM. This is because the imaging condition for RTM is based on cross-correlation, and therefore retains (actually amplifies) the imprint of the source signature. Unlike RTM, by fitting the image to recorded data, LSRTM compensates for the source signature using a deconvolution imaging condition (Yao et al., 2012a). Consequently, the LSRTM image has weaker sidelobes and higher resolution. Finally, the LSRTM image amplitudes are more accurate. For example, the deep diffractors in the LSRTM image, as expected, have almost the same amplitude as the shallow diffractors in Figure 2(c). This is due to the use of the inverse operator and the deconvolution imaging condition. Thereby, the amplitudes of the later arrivals in the modeled data of LSRTM (Figure 2(f)) are closer to those in the recorded data (Figure 2(d)) than in the case of RTM (Figure 2(e)). These advantages of LSRTM are intrinsic to the combination of least-squares inversion and reverse-time migration, and can be obtained with other implementations of LSRTM, such as that of Yao et al., (2012a). However, LSRTM is not immune to the effect of migration velocity errors. Migrations with a lower than correct velocity produces under-migrated images (Figure 2(g)) while higher velocities generate over-migrated images (Figure 2(h)). As shown by Figure 2(i),

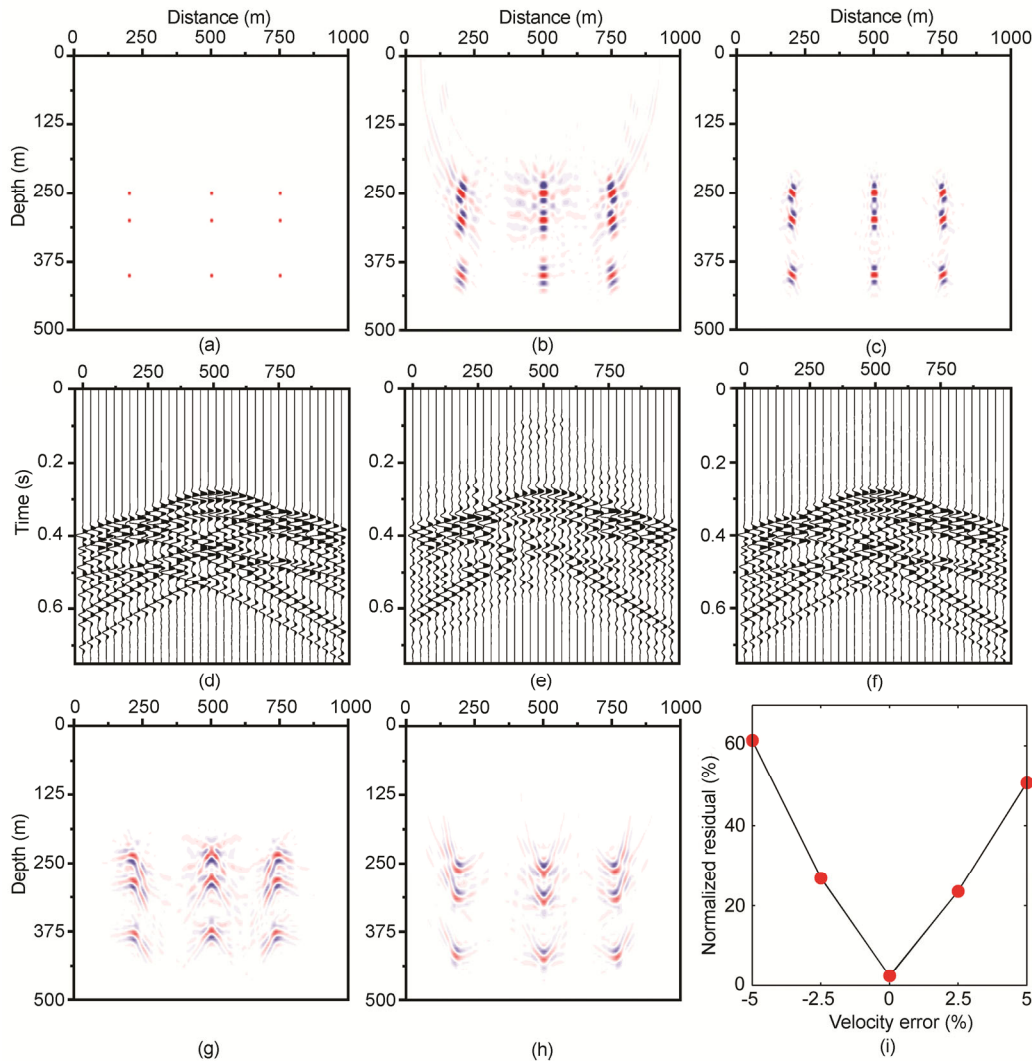


Figure 2 RTM and LSRTM images for a nine point model. (a) Model containing nine point diffractors embedded in a medium with a constant velocity of 2000 m/s; (b) RTM for three shot record from the model shown in (a); (c) the same as (b) but using LSRTM; (d) original one shot record; (e) modeled shot record for RTM; (f) modeled shot record for LSRTM. LSRTM images with a velocity of 1950 m/s (g) and 2050 m/s (h); (i) normalized residual against the percentage of velocity error,

however, only the correct migration velocity achieves the minimum residual. As can be seen, the normalized residual still has a small value when the velocity error is zero. This is because the recorded and predicted data are generated in different ways. The recorded data is produced by solving a single wave equation; the reflections are caused by the velocity contrasts and the amplitudes of which vary with reflection angle. However, the predicted data of LSRTM is modeled by solving eq. (2), which involves solving two wave equations. Here the reflections are generated by the stacked reflectivity and their amplitudes are invariant to the reflection angle. As a result, LSRTM provides a reflectivity model to achieve a least-squares fit, but not actually a full fit to the data. Nevertheless, this implies that LSRTM can be used for updating migration velocity. For the details of the theory, the reader is referred to Wang et al. (2013), Yao et al. (2014), and the references therein.

3.2 Three-layer model

The second example uses a three-layer model, the second layer of which has 200 m/s higher velocity than the first and third layers (Figure 3(a)). Three shots are generated using exactly the same sources and acquisition geometry as the first example. Figure 3(b) and (c) show the RTM and LSRTM images of the three shot data with 10 m trace spacing respectively. Compared with the RTM image, the LSRTM image also has three advantages. Firstly, the reflectors have more uniform amplitudes, which means LSRTM can better preserve the amplitudes. Secondly, the reflectors in the LSRTM image are slimmer, which implies LSRTM is able to improve the resolution of images. Finally, the LSRTM image has fewer artifacts. If the three shots are resampled to 40 m trace spacing, then RTM produces an image with noticeable arc-shaped aliasing artifacts (Figure 3(d)) and LSRTM generates an image with fewer aliasing

artifacts (Figure 3(e)). If a roughness penalty is applied in LSRTM and both λ_h and λ_v are chosen as to make the horizontal and vertical regularization terms be 0.25 and 0.125 times the data misfit term, respectively, LSRTM removes nearly all the aliasing artifacts and generates a high-quality image (Figure 3(f)). This is because large trace spacing is more likely to cause aliasing to the higher frequencies of the record and in turn aliasing artifacts have higher wavenumbers than reflector images. Hence, they can be suppressed by a roughness penalty being incorporated into LSRTM. This demonstrates LSRTM with a roughness penalty constraint can mitigate data aliasing artifacts.

As described in the section of deghosting, ghost waves exist in real seismic data and weaken low frequencies and create notches for some high frequencies. Figure 4 shows the ghost effect for LSRTM. In this test, three shots are generated in the same way as the shots in Figure 3 except a free surface is applied and source and receivers are at a depth of 10 m under surface. As a result, ghost waves are produced in the record (Figure 4(a)), which appears to have more wiggles for each event than the record without ghost waves in Figure 4(b). Figure 4(c) shows the LSRTM image by directly setting the source and receivers at the surface without applying ghost compensation, whilst Figure 4(e) shows the LSRTM image with ghost compensation. By comparison, the first noticeable difference is that the two images have opposite polarity; secondly, Figure 4(e) has less ringing; thirdly, the image with ghost compensation has

broader bandwidth, which can be seen by comparing the spectral of the images in Figure 4(d) and (f).

3.3 The Marmousi model

The third example uses the Marmousi model. The geometry of acquisition is set to be the same as the original study (Versteeg, 1993) and a 20 Hz Ricker wavelet is used as the source signature. The synthetic seismic record is modeled by using the acoustic wave equation with constant density. The source and receivers are located at a depth of 8 m. In order to generate ghost waves but no multiples, mirror sources and mirror receivers are used. The source wavelet used for migration is the same as the wavelet used for modeling. Figure 5 shows the reflectivity images obtained by migrating 16 shots and 240 shots across the Marmousi model. Figure 5(b) and (c) are the results of conventional RTM and LSRTM of 16 shots respectively without ghost compensation, which is achieved by setting the sources and receivers at surface. By contrast, Figure 5(d) is the LSRTM image after ghost compensation. Figure 5(e)–(g) are the counterparts of Figure 5(b)–(d) for 240 shots. By comparison of these results, it is obvious to see four features of LSRTM. Firstly, LSRTM can produce a thinner image of a reflector than conventional RTM, which suggests a higher resolution has been achieved. Secondly, LSRTM attenuates artifacts, including the low-wavenumber artifacts. This is especially noticeable in the results of 16 shots. Thirdly,

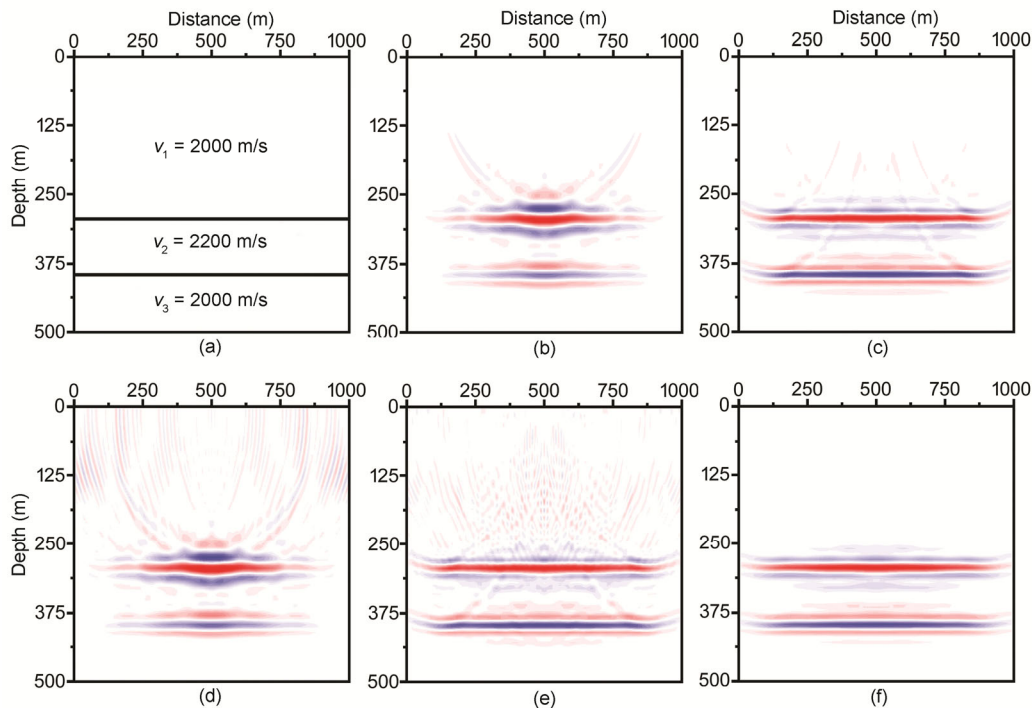


Figure 3 RTM and LSRTM images for a three-layer model. (a) The schematic diagram of the model; (b) RTM image of the three shots with sources located at 200, 500 and 800 m on the surface and 10 m trace spacing and (c) the corresponding LSRTM image; (d) RTM image of the same three shots but with 40 m trace spacing and the corresponding LSRTM image without (e) and with (f) roughness penalty.

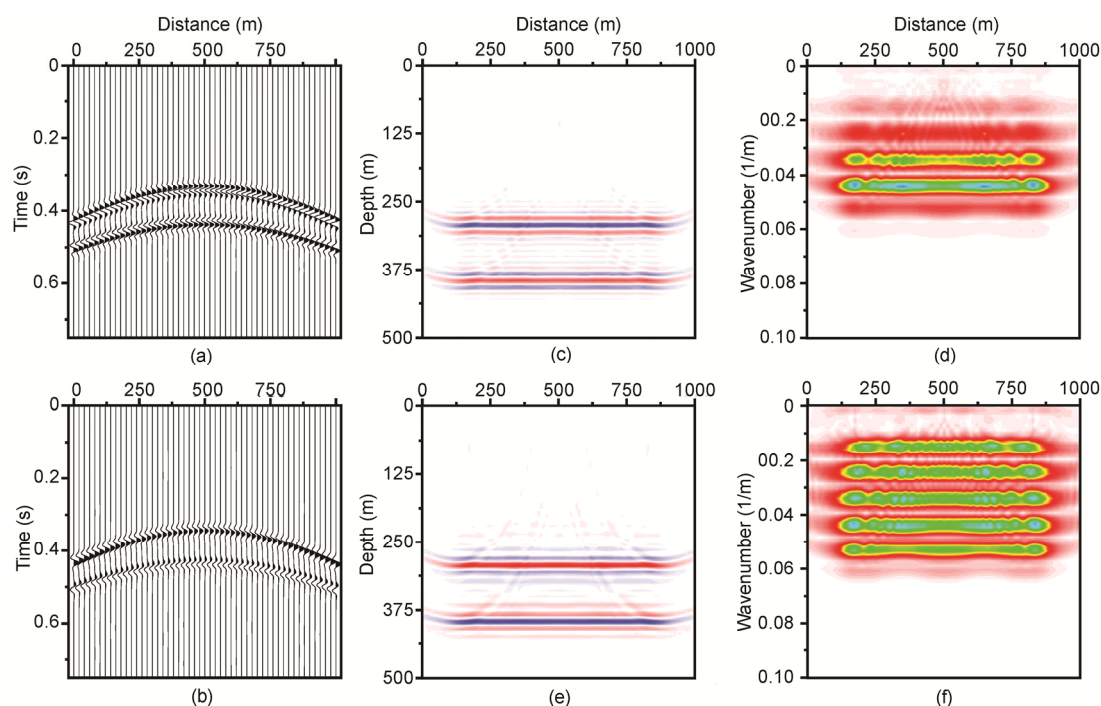


Figure 4 Ghost effect for LSRTM. Reflection data generated with free surface (a) whereas with absorbing boundary (b). LSRTM images without (c) and with (e) ghost compensation for reflection data generated with free surface; (d) and (f) are the amplitude spectral of the images in (c) and (e), respectively.

LSRTM without considering ghosting produces an incorrect polarity in the images. Finally, the reflectivity images produced by LSRTM without ghost compensation fit recorded data less well than with ghost compensation. This is demonstrated by Figure 5(h).

In this example, the BB scheme is used. In an iteration, one forward modeling step and one back propagation are needed to compute the gradient, with the cost of step-length calculation being trivial. Therefore, the computational cost of one iteration is equivalent to RTM. As a result, multi-iterative LSRTM is much more expensive than conventional RTM. However, by comparing Figure 5(d) and (g), it can be seen that even only 16 shots, which are one fifteenth of the whole set of 240 shots, can still produce a good image. To reduce the computational cost as well as use all of the information available, a random shots technique could be applied, which randomly selects a subset of the data for one or several iterations and chooses another subset for the next (Warner et al., 2013).

3.4 A field dataset

The fourth example applies the two migration methods on a field dataset provided by PGS. A dataset of 143 shots of a 2D line in a 3D towed stream survey is migrated in this example. The shot spacing is 200 m. Each of the shots has 500 traces with 12.5 m trace spacing and 6 s duration of recording with a 4 ms sample interval. Air guns and receivers were fixed at a depth of 8 and 10 m respectively. The

pre-processing, such as demultiple, deghosting and noise removal, was carried out by PGS. Due to the difference of wave propagation in 2D and 3D space, partial compensation (Wang et al., 2009) for this effect was applied on the pre-processed dataset. Afterwards, the dataset was migrated with RTM and LSRTM. Since this data was acquired with geo-streamers, which remove the receiver ghost waves, and source deghost was applied, ghost compensation is not needed for LSRTM. Prior to migration, an accurate source wavelet needs to be estimated. Otherwise, the source wavelet errors are imprinted into the image, thus reducing the image quality. In this example, the source wavelet used for migration is the processed air gun signature provided by PGS. Weighting regularization is applied to boost the convergence speed of the middle and low parts of the model (For more details please see Section 5.2 of the PhD thesis of Yao (2013)). By comparison of the two images in Figure 6, the three advantages of LSRTM shown in the previous examples can also be seen in this example. Firstly, LSRTM produces fewer artifact images. For example, the dip ringing artifacts above 2 km depth are largely removed in the LSRTM image. In addition, in the area indicated by the top dashed line box in Figure 6(a), the reflector images are contaminated by low-wavenumber artifacts, which still persist after Laplacian and further low-cut filtering. However, LSRTM suppresses these artifacts significantly (Figure 6(b)). Secondly, LSRTM achieves higher resolution images. For example, some fat reflectors in the RTM image become thinner in the LSRTM image. Finally, LSRTM generates

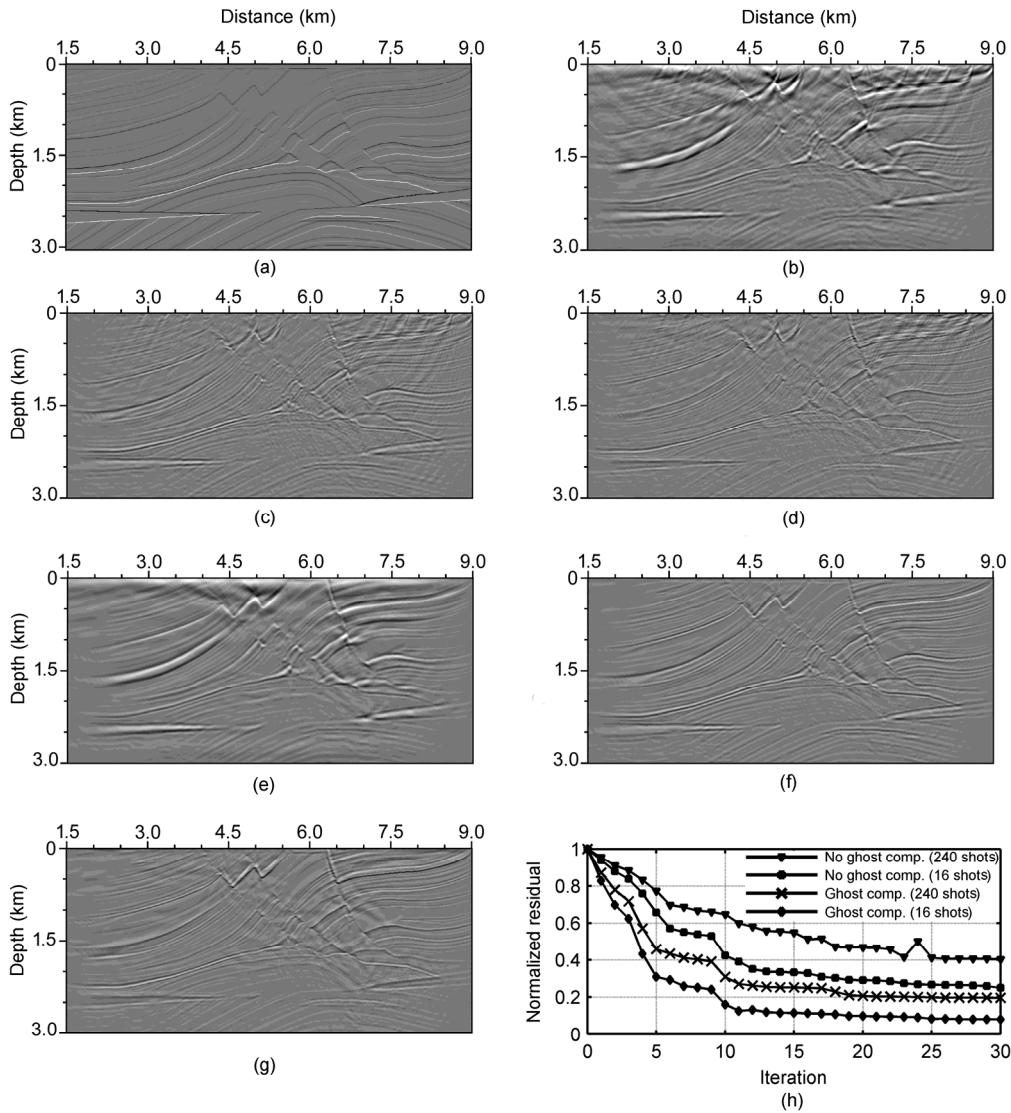


Figure 5 Different migrations applied to 16 shots and 240 shots with ghost waves from the Marmousi model. (a) True reflectivity of the Marmousi model; (b) RTM and (c) LSRTM images of 16 shots without ghost compensation whilst (d) LSRTM image with ghost compensation; (e)–(g) are the counterparts of (b)–(d) for 240 shots. Low-cut filtering wavenumbers below $4e^{-4}$ along depth is applied to (b)–(g); (h) the data residual evolution with iteration. The curves in (h) from top to bottom are corresponding to the images in (f), (c), (g) and (d).

images with more accurate amplitudes. For instance, more continuous reflectors in the middle dashed line box are shown in the LSRTM image. From this example, LSRTM improves image quality for the field dataset. In both images below 3.8 km, however, artifacts are noticeable and many reflectors are discontinuous. The main reason probably is because only one line of data of a 3D survey was used in the migration. Almost 60% residual remains throughout inversion. This may be caused by two main reasons. First and foremost, the 2D isotropic acoustic wave equation used in the LSRTM does not fully simulate the waves propagating in the real 3D world. The second reason is the noise. LSRTM is unable to find a model to fit the noise in all shots because the noise of one shot is inconsistent with that of other shots. As a result, it is kept in the residual throughout

the whole inversion.

4 Conclusions

A least-squares reverse-time migration scheme has been formulated and applied to a variety of synthetic and real datasets. It is based on an accurate modeling formula and uses an inverse of the forward modeling operator along with the two-way wave equation. Thereby, it gives correct phase in the image and inherits all the advantages of both conventional reverse-time migration and also least-squares migration. In this implementation, the spatial preconditioning, weighting and the BB scheme are all used to boost the convergence speed. With the aid of deghosting, this LSRTM

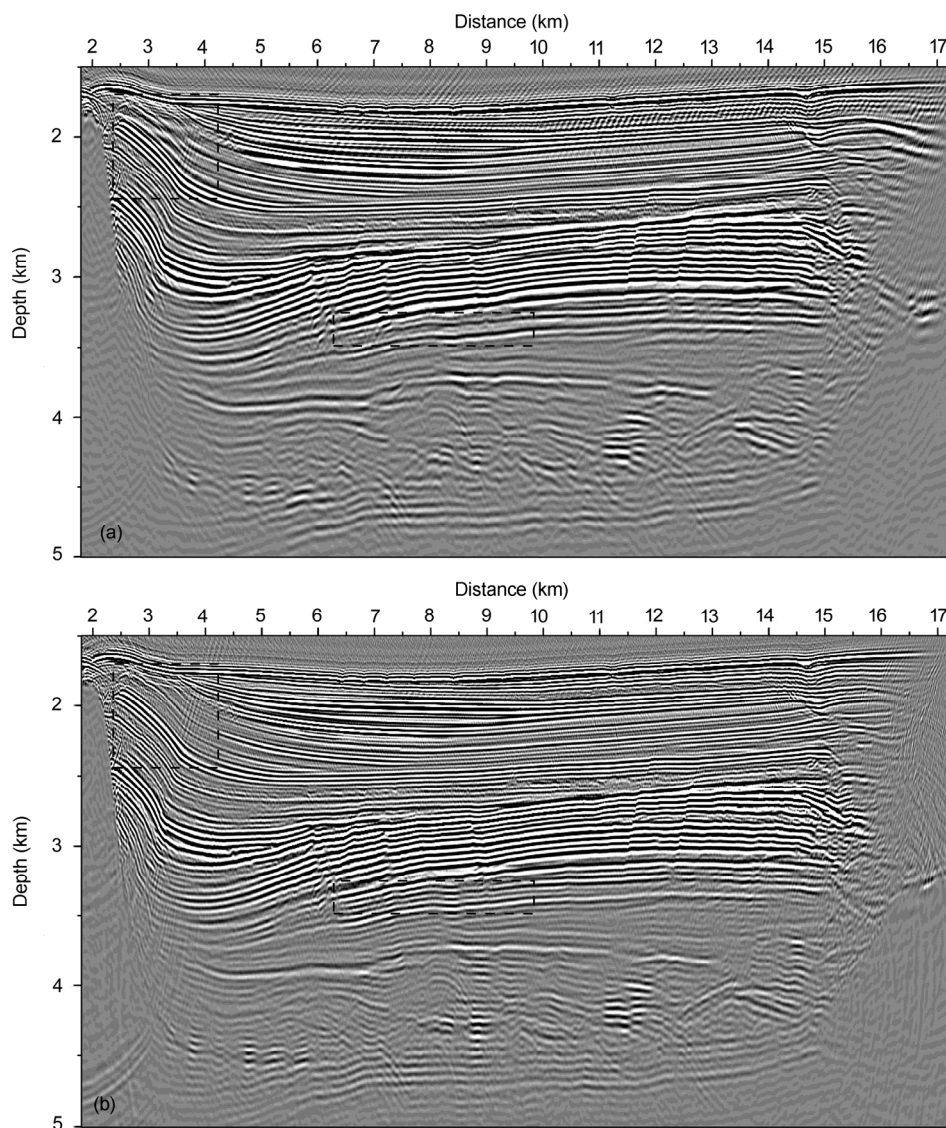


Figure 6 RTM and LSRTM applied to the field dataset. (a) RTM with Laplacian and further mild low-cut filtering; (b) LSRTM with Laplacian filtering.

scheme can produce broad bandwidth reflectivity images. Furthermore, the roughness constraint is also integrated to stabilize the inversion and suppress high-wavenumber artifacts including the spatial aliasing. All the examples demonstrate this scheme can produce stacked reflectivity images with broader bandwidth, higher resolution, fewer artifacts and more accurate amplitudes than conventional RTM. In particular, this scheme calculates the action of the Jacobian matrix instead of the explicit matrix; therefore it needs the same memory as RTM and is suitable for application in 3D.

The authors thank PGS and Sonangol for permitting to use the pre-processed field dataset and Dr. Andreas Klaedtke and Lesley Auchterlonie for preparing the field dataset and Prof. Yanghua Wang, Dr. Yu Zhang and Prof. Helmut Jakubowicz for meaningful discussions on the algorithms. The authors would like to thank our colleagues, Jack Ashley,

Oscar Agudo, Akela Silverton and Dr. Nuno Vieira Da Silva for their help to revise this paper. This work was partly supported by the National Natural Science Foundation of China (Grant No. 41272099) and the Science Foundation of China University of Petroleum, Beijing (Grant No. 2462015YJRC012). The authors greatly appreciate the two reviewers' and editors' constructive comments and suggestions, which helped to clarify and improve this manuscript significantly.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Barzilai J, Borwein J M. 1988. Two-point step size gradient methods. *IMA J Numer Anal*, 8: 141–148

Bube K P, Langan R T. 2008. A continuation approach to regularization of ill-posed problems with application to crosswell-traveltime tomography. *Geophysics*, 73: VE337–VE351

Claerbout J F. 1971. Toward a unified theory of reflector mapping. *Geo-*

- physics, 36: 467–481
- Claerbout J F. 1992. *Earth Soundings Analysis: Processing Versus Inversion*. London: Blackwell Scientific Publications
- Dai W, Schuster G T. 2010. Multi-source wave-equation least-squares migration with a deblurring filter. 72nd EAGE Conference and Exhibition, Barcelona, Spain. 276–279
- Dai W, Fowler P, Schuster G T. 2012. Multi-source least-squares reverse time migration. *Geophys Prospect*, 60: 681–695
- Dong S, Suh S, Wang B. 2013. Practical aspects of least-squares reverse time migration. 75th EAGE Conference and Exhibition, London, UK, We P09 02
- Gray S H, Etgen J, Dellinger J, Whitmore D. 2001. Seismic migration problems and solutions. *Geophysics*, 66: 1622–1640.
- Kaelin B, Guitton A. 2006. Imaging condition for reverse time migration. 76th SEG Annual Meeting, New Orleans, USA. 2594–2598
- Kaplan S T, Routh P S, Sacchi M D. 2010. Derivation of forward and adjoint operators for least-squares shot-profile split-step migration. *Geophysics*, 75: S225–S235
- Kühl H, Sacchi M D. 2001. Generalized least-squares DSR migration using a common angle imaging condition. 71st SEG Annual Meeting, San Antonio, USA. 1025–1028
- Kühl H, Sacchi M D. 2003. Least-squares wave-equation migration for AVP/AVA inversion. *Geophysics*, 68: 262–273
- Liu Y, Sun H, Chang X. 2005. Least-squares wave-path migration. *Geophys Prospect*, 53: 811–816
- Myron A B, Isaacson E L. 1998. *Numerical Analysis for Applied Science*. New York: John Wiley & Sons
- Nemeth T, Wu C, Schuster G T. 1999. Least-squares migration of incomplete reflection data. *Geophysics*, 64: 208–221
- Perz M J, Masoomzadeh H. 2014. Deterministic marine deghosting: Tutorial and recent advances. *Geoconvention 2014: Focus*, Calgary, Canada
- Scales J A. 1987. Tomographic inversion via the conjugate gradient method. *Geophysics*, 52: 179–185
- Valenciano A A, Biondi B. 2003. 2-D Deconvolution imaging condition for shot-profile migration. 73rd SEG Annual Meeting, Dallas, USA. 1059–1062
- Versteeg R J. 1993. Sensitivity of prestack depth migration to the velocity model. *Geophysics*, 58: 873–882
- Virieux J, Operto S. 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74: WCC1–WCC26
- Wang S, Chen F, Zhang H, Shen Y. 2013. Reflection-based Full Waveform Inversion (RFWI) in the frequency domain. 83rd SEG Annual Meeting, Houston, USA. 877–881
- Wang Y H, Rao Y. 2009. Reflection seismic waveform tomography. *J Geophys Res*, 114: B03304
- Wang Y, Yang C. 2010. Accelerating migration deconvolution using a nonmonotone gradient method. *Geophysics*, 75: S131–S137
- Warner M, Ratcliffe A, Nangoo T, Morgan J, Umpleby A, Shah N, Vinje V, Stekl I, Guasch L, Win C, Conroy G, Bertrand A. 2013. Anisotropic 3D full-waveform inversion. *Geophysics*, 78: R59–R80
- Xu S, Zhang Y, Tang B. 2011. 3D angle gathers from reverse time migration. *Geophysics*, 76: S77–S92
- Yao G, Jakubowicz H. 2012a. Least-squares reverse-time migration. 82nd SEG Annual Meeting, Las Vegas, USA. 1–5
- Yao G, Jakubowicz H. 2012b. Non-linear least-squares reverse-time migration. 82nd SEG Annual Meeting, Las Vegas, USA. 1–5
- Yao G. 2013. Least-squares reverse-time migration. Doctoral dissertation. London: Imperial College London, United Kingdom. <http://hdl.handle.net/10044/1/14575>
- Yao G, Warner M, Silverton A. 2014. Reflection FWI for both reflectivity and background velocity. 76th EAGE Conference and Exhibition, Amsterdam, Netherlands. We E106 10
- Yuan Y X, Sun W Y. 1997. *Optimization Theories and Methods* (in Chinese). Beijing: Science Press
- Zhang Y, Duan L, Xie Y. 2015. A stable and practical implementation of least-squares reverse time migration. *Geophysics*, 80: 23–31