

Least-squares reverse-time migration in a matrix-based formulation

Gang Yao* and Helmut Jakubowicz

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

Received November 2013, revision accepted July 2014

ABSTRACT

This paper describes least-squares reverse-time migration. The method provides the exact adjoint operator pair for solving the linear inverse problem, thereby enhancing the convergence of gradient-based iterative linear inversion methods. In this formulation, modified source wavelets are used to correct the source signature imprint in the predicted data. Moreover, a roughness constraint is applied to stabilise the inversion and reduce high-wavenumber artefacts. It is also shown that least-squares migration implicitly applies a deconvolution imaging condition. Three numerical experiments illustrate that this method is able to produce seismic reflectivity images with higher resolution, more accurate amplitudes, and fewer artefacts than conventional reverse-time migration. The methodology is currently feasible in 2-D and can naturally be extended to 3-D when computational resources become more powerful.

Key words: Adjoint migration, Least-squares reverse-time migration, Cross-correlation imaging condition, Deconvolution imaging condition.

INTRODUCTION

Migration attempts to produce an image of the subsurface by reversing the effects of propagation in seismic data. Although, in principle, this requires the inverse of a modelling operator, in practice, the adjoint of the modelling operator is used instead. There are two main reasons for using adjoint migration operators. First, adjoint migration requires significantly less calculation than inverse migration because it is much cheaper to compute the adjoint than the inverse operator. Second, adjoint operators are unconditionally stable because only multiplication and addition are involved in their calculation, whereas inverse operators require division. In cases where the data are subject to significant aliasing, truncation, and noise or are incomplete, the adjoint operator is not a good approximation to the inverse operator (Claerbout 1992) and will degrade the resolution of the final migrated image. Even with perfect data, an adjoint operator still produces imperfect images (Claerbout 1992). Consequently, it is desirable to

use the inverse operator to migrate seismic data; this is most commonly performed using least-squares approximations.

Nemeth, Wu, and Schuster (1999) provided an implementation of least-squares migration (LSM) based on the Kirchhoff integral, demonstrating improved resolution and treatment of amplitudes. Moreover, LSM is able to heal the effects of gaps and coarse spatial sampling in recorded data. To reduce the computational cost of least-squares Kirchhoff migration (Nemeth *et al.* 1999), Liu, Sun, and Chang (2005) combined least-squares inversion and wave-path migration, which together are faster than least-squares Kirchhoff migration. Kühl and Sacchi (2001, 2003) also suggested an LSM method based on the double-square-root (DSR) operator. This generates angle-domain common-image gathers with more accurate amplitudes than are obtained using the normal DSR operator. Kaplan, Routh, and Sacchi (2010) formulated an LSM method using the shot-profile split-step migration operator. These three LSM methods use either the Kirchhoff migration operator or the one-way wave equation, both of which have inherent drawbacks. The limitations were addressed by Dai and Schuster (2010) and Dai, Fowler, and Schuster (2012), who formulated a least-squares reverse-time

*E-mail: g.yao09@imperial.ac.uk

migration (LSRTM) scheme using a multi-shot technique to reduce computational cost.

In this work, we present a formulation of matrix-based LSRTM (MLSRTM) that is based upon an explicit matrix representation of generalised diffraction-stack migration (Schuster 2002). This uses a modified source wavelet to perform forward modelling at each iteration and is regularised using a roughness penalty constraint. Experimental results demonstrate that the method produces images with increased resolution, improved amplitude recovery, and fewer artefacts than conventional reverse-time migration (RTM).

THEORY

Seismic reflection data can be modelled using the following expression:

$$d(\mathbf{x}_r, \omega) = \int G_r(\mathbf{x}_r | \mathbf{x}, \omega) I(\mathbf{x}) G_s(\mathbf{x} | \mathbf{x}_s, \omega) s(\omega) d\mathbf{x}, \quad (1)$$

where d is the reflection data, $G_s(\mathbf{x} | \mathbf{x}_s, \omega)$ 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 , $I(\mathbf{x})$ is the stacked zero-phase reflectivity image, and $s(\omega)$ is the source wavelet. Conventional adjoint migration (Claerbout 1992), in turn, can be implemented as

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

where \dagger denotes complex conjugate. Equation (2) can be expressed in the time domain as

$$I(\mathbf{x}) = \int \int [G_r(\mathbf{x} | \mathbf{x}_r, t) * G_s(\mathbf{x} | \mathbf{x}_s, t) * s(t)] \otimes d(\mathbf{x}_r, t) d\mathbf{x}_r dt, \quad (3)$$

where $*$ denotes convolution, and \otimes represents cross-correlation. Equation (3) illustrates that migration can be carried out by firstly convolving the source with the source and receiver Green's functions, cross-correlating with the recorded data, and then summing over all traces and samples. If a two-way wave equation is used to calculate the convolution of the Green's functions with the source, then equation (3) is equivalent to generalised diffraction-stack migration (Schuster 2002). Equation (2) can also be expressed as

$$I(\mathbf{x}) = \int \int [G_s(\mathbf{x} | \mathbf{x}_s, \omega) s(\omega)]^\dagger [G_r(\mathbf{x} | \mathbf{x}_r, \omega)]^\dagger d(\mathbf{x}_r, \omega) d\mathbf{x}_r d\omega. \quad (4)$$

The term in the second square bracket of equation (4) is difficult to compute using finite-difference (FD) methods because the receiver Green's function G_r has an infinite

bandwidth. Fortunately, band-limited seismic data modelling avoids this so that the problem can be solved by decomposing the source wavelet as the product of two terms given by

$$s(\omega) = f_1(\omega) f_2(\omega), \quad (5)$$

where f_1 and f_2 are defined in the next section. Equation (5) then becomes

$$I(\mathbf{x}) = \int \int [G_s(\mathbf{x} | \mathbf{x}_s, \omega) f_1(\omega)]^\dagger [G_r(\mathbf{x} | \mathbf{x}_r, \omega) f_2(\omega)]^\dagger d(\mathbf{x}_r, \omega) d\mathbf{x}_r d\omega. \quad (6)$$

Equation (6) can be implemented as a three-stage calculation as follows:

- (i) solving the two-way wave equation with two source signatures $f_1(\omega)$ and $f_2(\omega)$ to generate the product of the Green's functions and source terms in the two square brackets;
- (ii) evaluating the kernel, which corresponds to the terms before the data term $d(\mathbf{x}_r, \omega)$ in equation (6), by conjugating and multiplying (i.e., time-reversing and convolving) the terms in square brackets obtained from the first step, and then multiplying (convolving) the result with the recorded data;
- (iii) summing all the traces and frequencies.

The above procedure is equivalent to RTM (Schuster 2002). However, both methods use adjoints of the forward-modelling operators rather than inverse operators and therefore only provide an estimate of the subsurface reflectivity.

The adjoint modelling of equation (6) can be described by

$$d(\mathbf{x}_r, \omega) = \int [G_s(\mathbf{x} | \mathbf{x}_s, \omega) f_1(\omega)] [G_r(\mathbf{x} | \mathbf{x}_r, \omega) f_2(\omega)] I(\mathbf{x}) d\mathbf{x}. \quad (7)$$

As can be seen from equation (7), the other advantage of source decomposition is that only one source signature is imprinted in the forward-modelled data.

Equations (6) and (7) can be written in explicit matrix form (see Appendix A), which guarantees that the adjoint migration described by equation (6) is the exact adjoint operator of the forward modelling given by equation (7). This is a requirement of gradient-based iterative linear least-squares schemes such as the conjugate gradient method (Scales 1987). Otherwise, the inversion may converge slowly or not at all. This frequently occurs in other formulations of LSRTM, which do not use matrices (Nemeth *et al.* 1999; Kühl and Sacchi 2003; Dai *et al.* 2012) because the adjoint migration operator implemented in those algorithms does not act as a

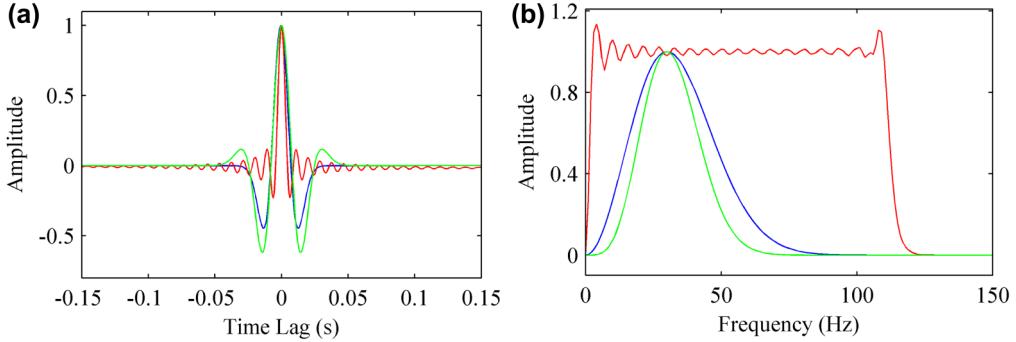


Figure 1 Comparison of least-squares migration (LSM) and conventional migration in the (a) time and (b) frequency domains. The blue curve represents the incident and reflected wavefields. The red curve represents the LSM image, whereas the green curve is the conventional migration image.

true adjoint. In contrast, the matrix formulation guarantees this and, in turn, can be used to formulate an iterative linear least-squares scheme that produces an optimum estimate of the reflectivity that is consistent with the recorded data. This is done by minimizing the objective function

$$\varphi_d(I) = \|\langle d \rangle - d\|^2, \quad (8)$$

where $\langle d \rangle$ represents the modelled data obtained from equation (7) (corresponding to the estimated reflectivity), and d describes the recorded data. This process is similar to full-waveform inversion (FWI), which also aims to find the best model to fit seismic data (Tarantola 1984). Indeed, if the input data for FWI are dominated by reflection data, the behaviour of FWI is very similar to LSRTM (Warner *et al.* 2013). However, LSRTM and FWI are essentially different. First, whereas FWI recovers the velocity of the medium, LSRTM only recovers the reflectivity. Second, FWI is able to invert both reflected and refracted arrivals, including multiples (Warner *et al.* 2013), whereas LSRTM builds its models only using primary reflection data. Finally, whereas the background velocity is updated in each iteration of FWI, this is fixed for LSRTM. Minimization of equation (8) forms the basis of the matrix-based least-squares reverse-time migration (MLSRTM) method presented here and is equivalent to finding the least-squares solution to equation (7).

Rewriting equation (7) in matrix notation as

$$\mathbf{d} = s\mathbf{G}_r\mathbf{G}_s\mathbf{I}, \quad (9)$$

where \mathbf{G}_r and \mathbf{G}_s are matrices, \mathbf{d} and \mathbf{I} are column vectors, s is a scalar, and $s\mathbf{G}_r\mathbf{G}_s$ represents the kernel of the integral in equation (7), the standard least-squares solution for \mathbf{I} is given by

$$\mathbf{I} = [(s\mathbf{G}_r\mathbf{G}_s)^\dagger (s\mathbf{G}_r\mathbf{G}_s)]^{-1} (s\mathbf{G}_r\mathbf{G}_s)^\dagger \mathbf{d}, \quad (10)$$

each element of which, in turn, can be shown to be equal to

$$I(\mathbf{x}_j, \omega) = \frac{u_r(\mathbf{x}_j, \omega)}{u_s(\mathbf{x}_j, \omega)}, \quad (11)$$

where u_s is the source wavefield, and u_r is the reflected wavefield (see Appendix B). The wavefields in equation (11) are those pertaining at a single point and at a single frequency, so that this equation corresponds to the piecewise division of two wavefields.

Equation (11) is equivalent to the deconvolution imaging condition of Valenciano and Biondi (2002, 2003), describing the reflectivity as the ratio of the reflected and incident wavefields rather than their cross-correlation. It follows that, in providing an optimum match between the data and reflectivity model, MLSRTM also implicitly implements a deconvolution imaging condition. Thus, MLSRTM has the additional benefit of compensating for the effects of the source signature and increasing the resolution of the resulting reflectivity estimate relative to conventional RTM.

Figure 1 illustrates how the deconvolution imaging condition implicit in MLSRTM produces higher resolution than conventional RTM. In this case, the amplitude of reflectivity is one. Reflection only modifies the amplitudes of incident wavefields, and otherwise maintains the shapes of wavefronts. Consequently, the reflected wavefield in this example is identical to the incident wavefield represented by the blue curve. The red curve corresponds to the MLSRTM image, whereas the green curve is the corresponding image for conventional migration using a cross-correlation imaging condition. It is clear from Fig. 1b that the MLSRTM can produce an image with broader temporal bandwidth than conventional migration, resulting in higher resolution. However, MLSRTM does not produce unlimited resolution. This is because the high-frequency components (in this case, higher than 125 Hz)

have low amplitudes to which the least-squares inversion is insensitive. This also helps make least-squares inversion highly stable. Although this example is an idealised problem, in cases with the same dominant frequency, MLSRTM will achieve almost as much of an improvement in resolution, even for a complex structural model.

IMPLEMENTATION

Source wavelet decomposition

The two brackets in equations (6) and (7) are calculated by solving the two-way wave equation using source signatures $f_1(\omega)$ and $f_2(\omega)$, which satisfy equation (5). If the wave equation is solved in the frequency domain (Xu, Zhou, and McMechan 2010; Kim, Min, and Shin 2011), then the two source signatures can simply be chosen to be the same:

$$f_1(\omega) = f_2(\omega) = \sqrt{s(\omega)}. \quad (12)$$

However, if the wave equation is solved in the time domain, then the source signatures are expected to have a duration period similar to $s(\omega)$. To achieve this, the source signatures can be chosen to be

$$f_1(\omega) = \text{abs}(\sqrt{s(\omega)}) \quad (13)$$

and

$$f_2(\omega) = \text{abs}(\sqrt{s(\omega)}) \exp(i\phi(s(\omega))), \quad (14)$$

where $\text{abs}()$ extracts the amplitude of a complex number, and $\phi()$ denotes the phase of a number. As can be seen, $f_1(\omega)$ has the square-root amplitude of the original wavelet with zero phase, whereas $f_2(\omega)$ has the same amplitude as $f_1(\omega)$ but with the phase of the original wavelet.

Memory requirements

Equations (6) and (7) can be reformulated in matrix form (see Appendix A). The elements of the matrices are then the product in the frequency domain, or convolution in the time domain, of the source and receiver Green's functions with the two source signatures $f_1(\omega)$ and $f_2(\omega)$. The product or convolution of the Green's functions with the source signatures can be obtained by solving the acoustic wave equation with the two source signatures. In this work, the wave equation is solved in the time domain using FD methods. To reduce the numerical dispersion intrinsic to FD methods, the sample rates in time and space are much higher than those required by the Nyquist theorem. However, modelling and adjoint migration

using the matrix formulation (equations (A1) and (A3)) do not induce dispersion. Memory requirements can therefore be significantly reduced by sampling the Green's functions sparsely. Considering imaging condition aliasing (Zhang, Sun, and Gray 2003), the two Green's functions can be sampled as sparsely as almost two samples per period in time and four samples per wavelength in space. Furthermore, if the same location is recorded in multiple shots, then the receiver Green's function only needs to be stored once.

Conjugate gradient inversion

The objective function given by equation (8) can be minimised using a gradient-based iterative inversion scheme. Since equations (6) and (7) are implemented here in matrix form, the modelling and adjoint migration operators are exactly conjugate transposes of each other. A modified linear conjugate gradient method (Scales 1987) can hence be used to achieve improved convergence rates. Moreover, these matrices are pre-computed ahead of the inversion; the iterative inversion process is thereby more efficient than implementations of LSRTM, which repeatedly solve the wave equation and therefore require much more computation than the simple matrix–vector multiplication used in MLSRTM during the inversion.

Roughness constraint

Similar to conventional RTM, the use of a two-way wave equation can produce low-wavenumber artefacts that can be removed using a Laplacian filter after migration (Youn and Zhou 2001) or least-squares attenuation (Guitton, Kaelin, and Biondi 2006). However, whereas the adjoint operator used in RTM is unconditionally stable, the inverse required for LSRTM can still be ill-posed without adequate data. This could generate additional high-wavenumber noise but can be controlled using a roughness constraint (Bube and Langan 2008). As a result, the objective function with the constrained terms can be written as

$$\varphi(I) = \|d - I\|^2 + \lambda_b^2 \|\partial_b I\|^2 + \lambda_v^2 \|\partial_v I\|^2, \quad (15)$$

where ∂_b and ∂_v represent first-order FD operators that describe roughness; the first $L2$ -norm term measures the misfit between predicted and recorded data; the second and third $L2$ -norms measure roughness of the image along the horizontal and vertical axes, respectively; and λ_b^2 and λ_v^2 are two positive regularization parameters controlling the contributions of each $L2$ -norm term to the overall objective function. A simple

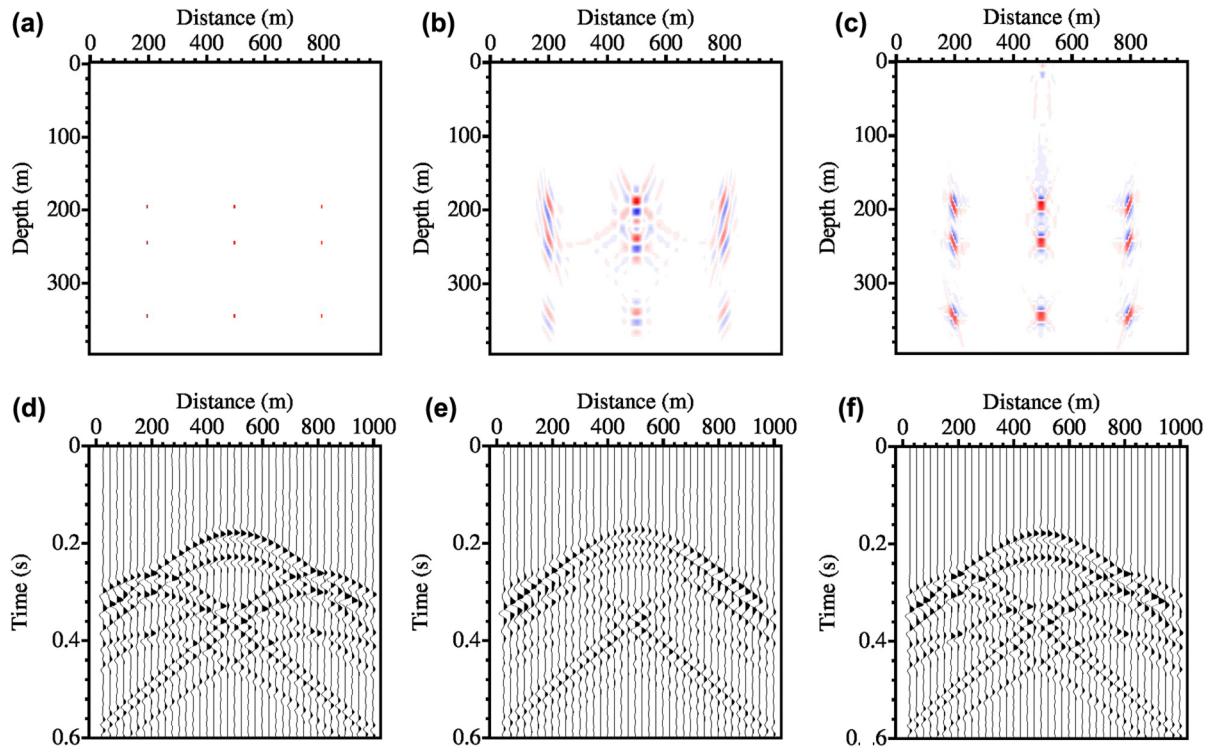


Figure 2 (a) Model containing nine point diffractors embedded in a medium with a constant background velocity of 2000 m/s. Migration of true data (b) using reverse-time migration (RTM) and (c) using matrix-based least-squares reverse-time migration (MLSRTM). (d) Original shot record and (e) modelled shot record for RTM and (f) for MLSRTM.

method to determine the two regularization parameters is the L-curve criterion (Aster, Brian, and Thurber 2005).

NUMERICAL EXAMPLES

The implementation of the matrix-based least-squares reverse-time migration (MLSRTM) is demonstrated using three models. The first model has a size of 500 m by 1000 m and consists of nine point diffractors embedded in a medium with a constant background velocity of 2000 m/s (Fig. 2a). Figure 2d shows a shot record from this model for a surface source at 500 m, with receivers located along the surface every 25 m; the source signature is a 30 Hz Ricker wavelet, and the direct arrivals, which do not contribute to imaging, have been removed. Figure 2b and c shows the results of applying conventional reverse-time migration (RTM) and MLSRTM to the data in Fig. 2d, whereas Fig. 2e and f shows the modelled data corresponding to the RTM and MLSRTM images. In this case, since there are only nine diffractors in the model, the available data are sufficient to construct the image using a least-squares approach. Noise and instabilities are therefore not an issue, and Laplacian filtering is not required;

the MLSRTM result also does not incorporate any roughness constraint.

Figure 2 demonstrates that MLSRTM has two major advantages relative to RTM. First, the artificial sidelobes of the diffractors in the MLSRTM image are weaker, and the diffractors are better resolved than those produced by RTM. This is primarily because the imaging condition for RTM is based on cross-correlation and therefore retains and in fact amplifies the imprint of the source signature. By contrast, MLSRTM effectively employs the deconvolution imaging condition given by equation (11) and thereby compensates for the source signature. Second, the images of the deep diffractors have more accurate amplitudes in the MLSRTM result than in the case of RTM. This is because of the deconvolution imaging condition and inverse operator used. RTM uses adjoint, rather than inverse, operators, and these fail to accurately reconstruct the reflected wavefield when used to back-propagate the recorded data. By contrast, by using an approximate inverse, MLSRTM provides a better representation of the amplitude effects related to back propagation. This helps preserve the amplitudes inherent in the reflectivity and leads to a more accurate image. As a result, the amplitudes of the later arrivals in the modelled

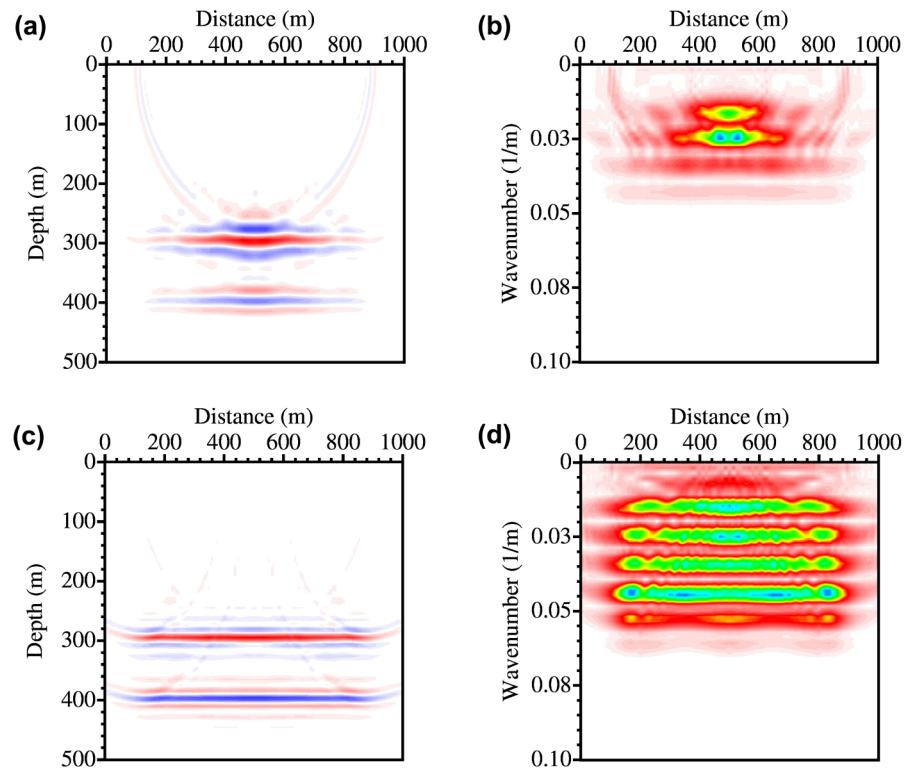


Figure 3 RTM and MLSRTM images of three shots and their amplitude spectra. (a) RTM image. (c) MLSRTM image after 30 iterations. (b) and (d) Amplitude spectra corresponding to the images to their immediate left.

MLSRTM data are closer to those in the recorded data than in the case of RTM.

The second example investigates the migration of a three-layer model, which has the same size as the first example, and comprises a 2200 m/s layer embedded from the depth of 300 m–400 m in a 2000-m/s constant background. Three shot records are acquired at 200 m, 500 m, and 800 m on the surface with the same acquisition geometry as the diffractor model. Similarly, a 30 Hz Ricker wavelet is again used as the source, with the direct arrivals also muted from the records prior to migration.

Figure 3 compares the results of migrating the three shots using RTM and MLSRTM. Figure 3a shows the RTM image, whereas Fig. 3c shows the MLSRTM image after 30 iterations. It can be seen that the RTM image has several flaws. First, the reflectors in the RTM image are very rough, and the amplitudes in the middle of each reflector are stronger than at the sides. The ends of reflectors imaged by each shot with an adjoint migration operator do not stop at the end of the effective illumination range inherent in the limited recording aperture but sweep upwards and outwards; as a result, when the images from the different shots are stacked, the reflector appears

uneven and is distorted compared with the true model. The strong amplitudes at the middle of each reflector are caused by near-critical reflections in the shots located at 200 m and 800 m. Second, the top reflector is stronger than the second because the cross-correlation imaging condition and adjoint operator do not preserve amplitudes. Finally, there are two arc-shaped artefacts above the top reflector. These may be caused by the limited recording aperture. By comparison, the image in Fig. 3c, produced by MLSRTM after 30 iterations, is much improved relative to the RTM image. The artefacts in the RTM image are almost completely removed, and the amplitudes are much more uniform both along and between the two reflectors. However, there are still faint remnants of the arc-shaped artefacts. This is because only three shots are used in the inversion process, and this prevents the artefacts from a single shot being fully suppressed by the others. Figures 3b and d shows the spatial amplitude spectra of the images to their left and demonstrate that MLSRTM also improves the spatial resolution.

The third example uses the Marmousi model (Versteeg 1993). Figure 4 shows the velocity and reflectivity models for the Marmousi data, whereas Fig. 5a shows the shot record

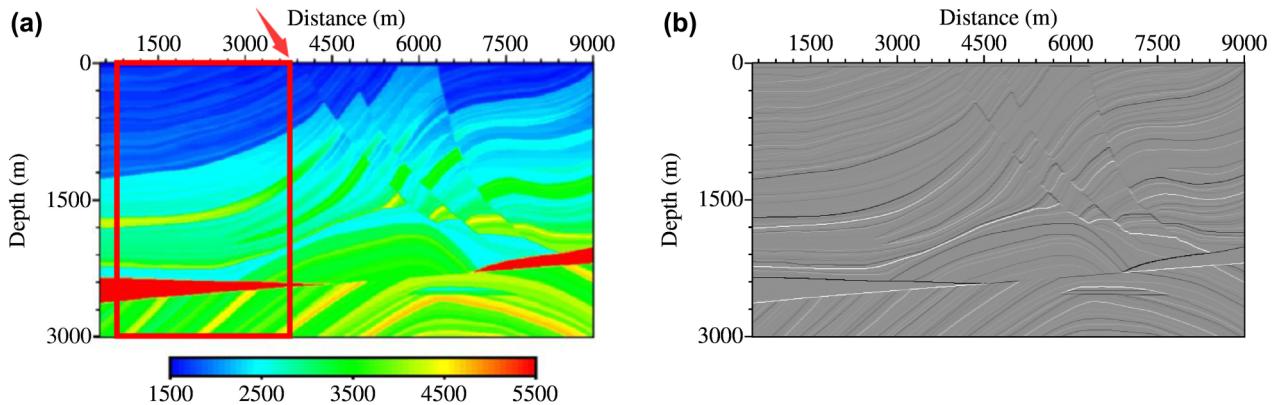


Figure 4 (a) Velocity model for the synthetic Marmousi dataset, with its associated colour scale in meters per second and (b) its corresponding normal incident reflectivity. The red box in (a) indicates the area covered by the migrations in Fig. 5.

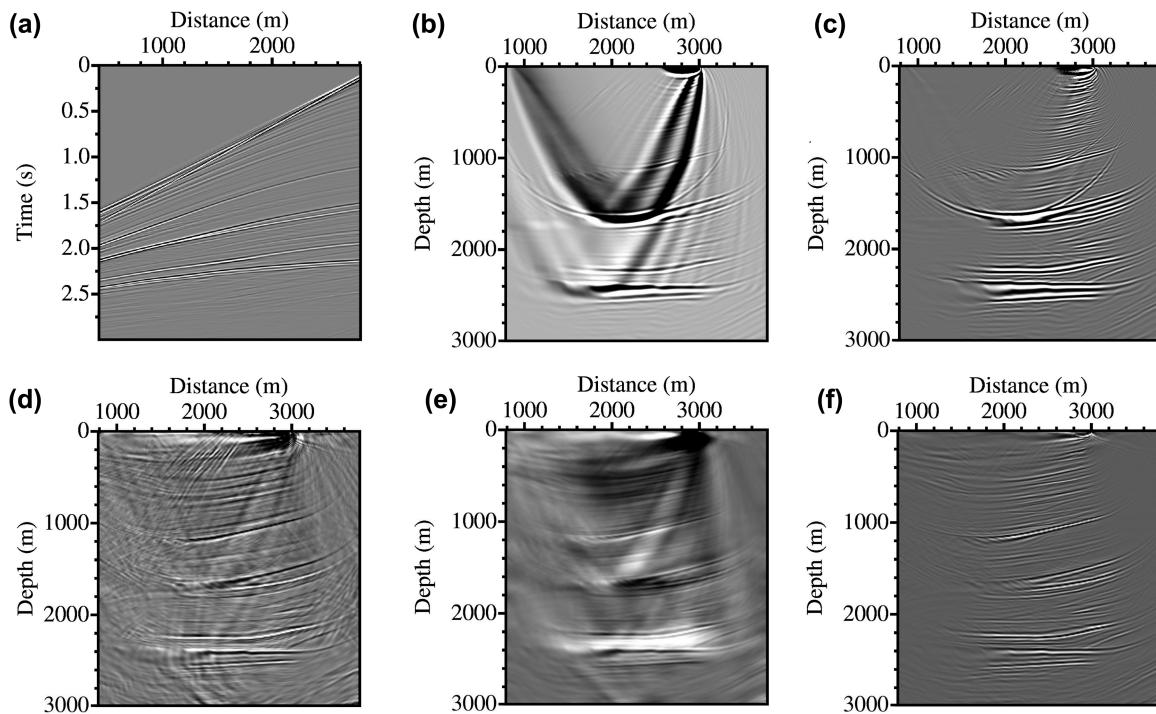


Figure 5 RTM and MLSRTM images for a single-shot record of the Marmousi model. (a) Original shot record. (b) RTM image. (c) RTM image after Laplacian filtering. (d) MLSRTM image without a roughness constraint. (e) MLSRTM image with a roughness constraint. (f) MLSRTM image with a roughness constraint after application of a Laplacian filter.

from source location at 3000 m on the surface for receivers located between 425 m and 2800 m at intervals of 12.5 m. The data were synthesised using a high-order FD modelling algorithm and perfectly-matched-layer absorbing boundaries around the model. The RTM image from this shot record is shown in Fig. 5b and corresponds to the area enclosed within the red box in Fig. 4a. In this case, low-wavenumber artefacts

arising from the use of the two-way wave equation are an issue but can be removed using a Laplacian filter (Fig. 5c). However, even after Laplacian filtering, the final RTM image retains high-wavenumber artefacts.

Figure 5d shows the image produced by MLSRTM without a roughness constraint. Relative to conventional RTM, the MLSRTM implementation suppresses high-wavenumber

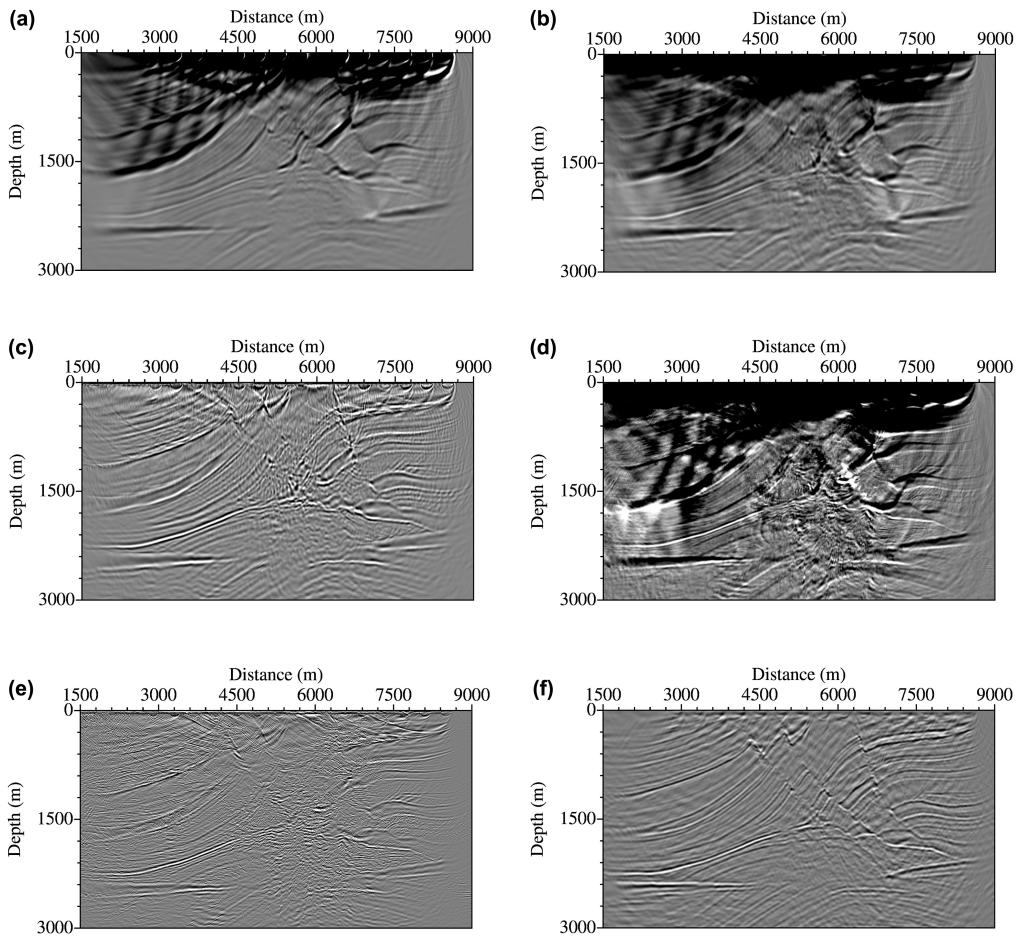


Figure 6 RTM images for 15 shots from the Marmousi model. (a)–(c) RTM image with zero-lag cross-correlation imaging condition: (a) raw, (b) with illumination compensation, and (c) with illumination compensation plus Laplacian filtering. (d)–(e) RTM image with deconvolution imaging condition: (d) raw and (e) with Laplacian filtering. (f) MLSRTM image with a roughness constraint and mild low-cut filtering only.

artefacts and increases the resolution in the final image; however, it also produces a significant amount of high-wavenumber noise. Figure 5e shows the corresponding MLSRTM result but using a roughness constraint. This successfully eliminates the high-wavenumber noise but retains low-wavenumber artefacts that are inherent to the numerical solution of the two-way wave equation. Finally, Fig. 5f shows the result of applying a Laplacian filter to the data in Fig. 5e. This image has fewer artefacts, is more highly resolved than the RTM image (Fig. 5c), and also has more consistent amplitudes over a wider area.

Figure 6 compares images obtained by migrating 15 shots from across the Marmousi model using conventional RTM and MLSRTM. Figure 6a shows the conventional RTM image with the zero-lag cross-correlation imaging condition. As can be seen, strong low-wavenumber noise distorts the shallow

reflectors. In addition, the amplitudes noticeably decrease with depth because the cross-correlation imaging condition could not correct for amplitude decline. This can be improved by using source illumination compensation (Kaelin and Guittot 2006), as shown in Fig. 6b where amplitudes are more balanced between shallow and deep reflectors. However, the low-wavenumber noise remains. This is effectively and efficiently removed by Laplacian filtering (Fig. 6c). Unfortunately, Laplacian filtering also artificially boosts the high wavenumbers, and although this increases the apparent resolution, it also boosts high-wavenumber noise.

A better way to improve resolution is to use a deconvolution imaging condition (Valenciano and Biondi 2002, 2003) as shown in Fig. 6d. During this process, frequencies ranging from 2 Hz to 58 Hz are imaged, and 1% of the source wavefield energy is used for stabilization. As expected,

the resolution in Fig. 6d is dramatically improved but the low-wavenumber noise is still prevalent. Laplacian filtering is thereby applied; however, the high-wavenumber artefacts again become very strong (Fig. 6e).

Figure 6a–d together shows that the source illumination compensation, Laplacian filtering, and deconvolution imaging condition do not fully solve the low-wavenumber noise, high-wavenumber artefacts, image amplitude imbalance, and resolution problems inherent in RTM. Fortunately, MLSRTM provides an improved solution to all of these problems. The image in Fig. 6f, generated using MLSRTM with a roughness constraint and mild low-cut filtering, is better than all of the images in Fig. 6a–d in terms of resolution, amplitudes, and noise attenuation. There are three likely reasons for this. First, MLSRTM implicitly employs a deconvolution imaging condition that improves resolution. Second, MLSRTM applies an inverse operator that corrects amplitudes. Third, for a multi-shot dataset, MLSRTM iteratively suppresses artefacts. The artefacts in any given iteration generate artificial events in the modelled data of some shots. These events, in turn, become residuals, which are eliminated by subsequent iterations of the least-squares inversion. By contrast, conventional RTM only removes artefacts by stacking and is therefore a much less effective technique.

CONCLUSIONS

We have presented a matrix-based least-squares implementation of reverse-time migration (MLSRTM). In this approach, the forward-modelling operator is expressed by a matrix, which, in turn, allows the adjoint migration operator to be calculated exactly as the conjugate transpose of this matrix. This then allows the modified linear least-squares conjugate gradient scheme (Scales 1987) to be applied. In order to provide the correct source imprint in the predicted data, a modified source wavelet is used to compute the matrices. Furthermore, it is shown that the method implicitly uses a deconvolution imaging condition. MLSRTM has several advantages over conventional RTM. First, MLSRTM uses the inverse operator to backward-propagate the recorded data and reconstruct the reflected wavefield more accurately. Second, since MLSRTM implicitly applies a deconvolution imaging condition, the source signature is properly removed and the amplitude and resolution of the image are improved. In addition, roughness regularization can be incorporated into MLSRTM to suppress high-wavenumber noise in the image. All these advantages are demonstrated by the examples. Unfortunately, MLSRTM requires a relatively large amount of memory and

storage and is currently impractical for application in 3-D. An alternative non-linear formulation of LSRTM, which retains the benefits of least-squares imaging but is more suitable for 3-D data, is described by Yao and Jakubowicz (2012).

ACKNOWLEDGEMENTS

The authors would like to thank Prof. Yanghua Wang and Dr. Yu Zhang for useful discussions on the theory in this work. They would also like to thank the two reviewers and editors for their comments and suggestions that helped improve and clarify this manuscript significantly.

REFERENCES

- Aster R.C., Brian B. and Thurber C.H. 2005. *Parameter Estimation and Inverse Problems*. Elsevier. ISBN: 0-12-065604-3.
- Bube K.P. and 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. *Geophysics* 36, 467–481.
- Claerbout J.F. 1992. *Earth Soundings Analysis: Processing Versus Inversion*. Blackwell Scientific Publications. ISBN: 9780865422100.
- Dai W., Fowler P. and Schuster G.T. 2012. Multi-source least-squares reverse time migration. *Geophysical Prospecting* 60, 681–695.
- Dai W. and Schuster G.T. 2010. Multi-source wave-equation least-squares migration with a deblurring filter. 72nd EAGE meeting, Barcelona, Spain, Expanded Abstracts, 276.
- Guitton A., Kaelin B. and Biondi B. 2006. Least-squares attenuation of reverse-time migration artefacts. *Geophysics* 72, S19–S23.
- Kaelin B. and Guitton A. 2006. Imaging condition for reverse time migration. 76th SEG meeting, New Orleans, USA, Expanded Abstracts, 2594–2598.
- Kaplan S.T., Routh P.S. and Sacchi M.D. 2010. Derivation of forward and adjoint operators for least-squares shot-profile split-step migration. *Geophysics* 75, S225–S235.
- Kim Y., Min D.J. and Shin C. 2011. Frequency-domain reverse-time migration with source estimation. *Geophysics* 76, S41–S49.
- Kühl H. and Sacchi M.D. 2001. Generalised least-squares DSR migration using a common angle imaging condition. 71st SEG meeting, San Antonio, USA, Expanded Abstracts, 1025–1028.
- Kühl H. and Sacchi M.D. 2003. Least-squares wave-equation migration for AVP/AVA inversion. *Geophysics* 68, 262–273.
- Liu Y., Sun H. and Chang X. 2005. Least-squares wave-path migration. *Geophysical Prospecting* 53, 811–816.
- Nemeth T., Wu C. and Schuster G.T. 1999. Least-squares migration of incomplete reflection data. *Geophysics* 64, 208–221.
- Scales J.A. 1987. Tomographic inversion via the conjugate gradient method. *Geophysics* 52, 179–185.
- Schuster G.T. 2002. Reverse-time migration = generalised diffraction stack migration. 72nd SEG meeting, Salt Lake City, USA, Expanded Abstracts, 1280–1283.

- Tarantola A. 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **49**, 1259–1266.
- Valenciano A.A. and Biondi B. 2002. Deconvolution imaging condition for reverse-time migration. *Stanford Exploration Project, Report* **112**, 83–96.
- Valenciano A.A. and Biondi B. 2003. 2-D deconvolution imaging condition for shot-profile migration. 73rd SEG meeting, Dallas, USA, Expanded Abstracts, 1059–1062.
- Versteeg R.J. 1993. Sensitivity of prestack depth migration to the velocity model. *Geophysics* **58**, 873–882.
- Warner M., Ratcliffe A., Nangoo T., Morgan J., Umpleby A., Shah N. et al. 2013. Anisotropic 3D full-waveform inversion. *Geophysics* **78**, R59–R80.
- Xu K., Zhou B. and McMechan G.A. 2010. Implementation of prestack reverse time migration using frequency-domain extrapolation. *Geophysics* **75**, S61–S72.
- Yao G. and Jakubowicz H. 2012. Non-linear least-squares reverse-time migration. 82nd SEG meeting, Las Vegas, USA, Expanded Abstracts, 1–5.
- Youn O.K. and Zhou H. 2001. Depth imaging with multiples. *Geophysics* **66**, 246–255.
- Zhang Y., Sun J.C. and Gray S.H. 2003. Aliasing in wavefield extrapolation prestack migration. *Geophysics* **68**, 629–633.

APPENDIX A: MATRIX FORMULATION OF MODELLING AND ADJOINT MIGRATION

Equations (6) and (7) can be written explicitly in matrix form. For the single-shot situation, adjoint migration can be expressed as

$$\begin{aligned} & \begin{bmatrix} I(\mathbf{x}_1) \\ I(\mathbf{x}_2) \\ \vdots \\ I(\mathbf{x}_n) \end{bmatrix} = \\ & \begin{pmatrix} G_1^\dagger(\mathbf{x}_1, \omega_1) & \cdots & G_1^\dagger(\mathbf{x}_1, \omega_m) & \cdots & G_k^\dagger(\mathbf{x}_1, \omega_1) & \cdots & G_k^\dagger(\mathbf{x}_1, \omega_m) \\ G_1^\dagger(\mathbf{x}_2, \omega_1) & \cdots & G_1^\dagger(\mathbf{x}_2, \omega_m) & \cdots & G_k^\dagger(\mathbf{x}_2, \omega_1) & \cdots & G_k^\dagger(\mathbf{x}_2, \omega_m) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ G_1^\dagger(\mathbf{x}_n, \omega_1) & \cdots & G_1^\dagger(\mathbf{x}_n, \omega_m) & \cdots & G_k^\dagger(\mathbf{x}_n, \omega_1) & \cdots & G_k^\dagger(\mathbf{x}_n, \omega_m) \end{pmatrix} \\ & \times \begin{bmatrix} d_1(\omega_1) \\ \vdots \\ d_1(\omega_m) \\ \vdots \\ d_k(\omega_1) \\ \vdots \\ d_k(\omega_m) \end{bmatrix}, \end{aligned} \quad (\text{A1})$$

where

$$G_l^\dagger(\mathbf{x}_j, \omega_i) = [G_s(\mathbf{x}_j | \mathbf{x}_s, \omega_i) f_1(\omega_i)]^\dagger [G_r(\mathbf{x}_j | \mathbf{x}_{r_l}, \omega_i) f_2(\omega_i)]^\dagger \quad (\text{A2})$$

whereas the modelling process can be written as

$$\begin{bmatrix} d_1(\omega_1) \\ \vdots \\ d_1(\omega_m) \\ \vdots \\ d_k(\omega_1) \\ \vdots \\ d_k(\omega_m) \end{bmatrix} = \begin{pmatrix} G_1(\mathbf{x}_1, \omega_1) & G_1(\mathbf{x}_2, \omega_1) & \cdots & G_1(\mathbf{x}_n, \omega_1) \\ \vdots & \vdots & \ddots & \vdots \\ G_1(\mathbf{x}_1, \omega_m) & G_1(\mathbf{x}_2, \omega_m) & \cdots & G_1(\mathbf{x}_n, \omega_m) \\ \vdots & \vdots & \ddots & \vdots \\ G_k(\mathbf{x}_1, \omega_1) & G_k(\mathbf{x}_2, \omega_1) & \cdots & G_k(\mathbf{x}_n, \omega_1) \\ \vdots & \vdots & \ddots & \vdots \\ G_k(\mathbf{x}_1, \omega_m) & G_k(\mathbf{x}_2, \omega_m) & \cdots & G_k(\mathbf{x}_n, \omega_m) \end{pmatrix} \times \begin{bmatrix} I(\mathbf{x}_1) \\ I(\mathbf{x}_2) \\ \vdots \\ I(\mathbf{x}_n) \end{bmatrix}, \quad (\text{A3})$$

where

$$G_l(\mathbf{x}_j, \omega_i) = [G_s(\mathbf{x}_j | \mathbf{x}_s, \omega_i) f_1(\omega_i)][G_r(\mathbf{x}_j | \mathbf{x}_{r_l}, \omega_i) f_2(\omega_i)]. \quad (\text{A4})$$

For multi-shot migration, the equation system for each shot (equations (A1) and (A3)) is formulated before simply combining them in a larger system of equations.

APPENDIX B: DERIVATION OF THE DECONVOLUTION IMAGING CONDITION USED IN LEAST-SQUARES MIGRATION

For the sake of simplicity, only a single frequency for one shot is considered, with the corresponding amplitude of this mono-frequency source set to one. The source Green's function from the source at \mathbf{x}_s to an imaging point at \mathbf{x}_j is denoted as $G_s(\mathbf{x}_j | \mathbf{x}_s)$; therefore, the incident source wavefield at this imaging point can be expressed as

$$u_s(\mathbf{x}_j) = G_s(\mathbf{x}_j | \mathbf{x}_s), \quad (\text{B1})$$

where the source term is one and thus is ignored. Assuming that the reflectivity at \mathbf{x}_j is $I(\mathbf{x}_j)$, then the reflected wavefield is

$$u_r(\mathbf{x}_j) = G_s(\mathbf{x}_j | \mathbf{x}_s) I(\mathbf{x}_j). \quad (\text{B2})$$

The matrix formulation for equation (B2) can, in turn, be expressed as

$$\begin{bmatrix} u_r(\mathbf{x}_1) \\ \vdots \\ u_r(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} G_s(\mathbf{x}_1 | \mathbf{x}_s) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & G_s(\mathbf{x}_n | \mathbf{x}_s) \end{bmatrix} \begin{bmatrix} I(\mathbf{x}_1) \\ \vdots \\ I(\mathbf{x}_n) \end{bmatrix}, \quad (\text{B3})$$

which can also be denoted as

$$\mathbf{u}_r = \mathbf{G}_s \mathbf{I}, \quad (\text{B4})$$

where \mathbf{G}_s is a diagonal matrix, and \mathbf{u}_r and \mathbf{I} are both column vectors.

According to equation (2), the record of the receiver at \mathbf{x}_k can be expressed as

$$d(\mathbf{x}_k) = \sum_j G_r(\mathbf{x}_k | \mathbf{x}_j) u_r(\mathbf{x}_j), \quad (\text{B5})$$

where $G_r(\mathbf{x}_k | \mathbf{x}_j)$ is the receiver Green's function from the imaging point at \mathbf{x}_j to the receiver at \mathbf{x}_k . Equation (B5) hence can be written in matrix formation as

$$\begin{bmatrix} d(\mathbf{x}_1) \\ \vdots \\ d(\mathbf{x}_k) \end{bmatrix} = \begin{bmatrix} G_r(\mathbf{x}_1 | \mathbf{x}_1) & \cdots & G_r(\mathbf{x}_1 | \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ G_r(\mathbf{x}_k | \mathbf{x}_1) & \cdots & G_r(\mathbf{x}_k | \mathbf{x}_n) \end{bmatrix} \begin{bmatrix} u_r(\mathbf{x}_1) \\ \vdots \\ u_r(\mathbf{x}_n) \end{bmatrix}, \quad (\text{B6})$$

which can also be denoted as

$$\mathbf{d} = \mathbf{G}_r \mathbf{u}_r, \quad (\text{B7})$$

where \mathbf{G}_r is a matrix, and \mathbf{u}_r and \mathbf{d} are both column vectors. The least-squares solution of equation (B7) is, in turn, given by

$$\mathbf{u}_r = (\mathbf{G}_r^\dagger \mathbf{G}_r)^{-1} \mathbf{G}_r^\dagger \mathbf{d} \quad (\text{B8})$$

where \mathbf{G}_r is a matrix, and \mathbf{u}_r and \mathbf{d} are both column vectors.

Substituting equation (B7) with equation (B4) gives

$$\mathbf{d} = \mathbf{G}_r \mathbf{G}_s \mathbf{I}. \quad (\text{B9})$$

The least-squares solution of the image can thereby be expressed as

$$\mathbf{I} = [(\mathbf{G}_r \mathbf{G}_s)^\dagger (\mathbf{G}_r \mathbf{G}_s)]^{-1} (\mathbf{G}_r \mathbf{G}_s)^\dagger \mathbf{d}. \quad (\text{B10})$$

Using simple matrix–vector algebra, equation (B10) can be simplified to

$$\mathbf{I} = \mathbf{G}_s^{-1} (\mathbf{G}_r^\dagger \mathbf{G}_r)^{-1} \mathbf{G}_r^\dagger \mathbf{d}. \quad (\text{B11})$$

According to equation (B8), equation (B11) can be further simplified to

$$\mathbf{I} = \mathbf{G}_s^{-1} \mathbf{u}_r. \quad (\text{B12})$$

Since \mathbf{G}_s is a diagonal matrix where, according to equations (B1) and (B3), each element on the diagonal is the source wavefield, it can be concluded that each element in equation (B12) is given by

$$I(\mathbf{x}_j, \omega) = \frac{u_r(\mathbf{x}_j, \omega)}{u_s(\mathbf{x}_j, \omega)}, \quad (\text{B13})$$

which is equivalent to the deconvolution imaging condition (Claerbout 1971; Valenciano and Biondi 2002, 2003). For multi-frequency and multi-shot solutions, the image is simply the stack of all the individual images expressed by equation (B13).