OXFORD

Journal of Geophysics and Engineering

Journal of Geophysics and Engineering (2021) **18**, 304–316 doi:10.1093/jge/gxab015

Least-squares reverse-time migration with sparsity constraints

Di Wu[1], **Yanghua Wang**[2], **Jingjie Cao**[3], **Nuno V. da Silva**[4] **and Gang Yao**[1,*]

[1] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China
[2] Resource Geophysics Academy, Imperial College London, London, SW7 2BP, UK
[3] College of Earth Sciences, Hebei GEO University, Shijiazhuang,050031, China
[4] Department of Earth Science and Engineering, Imperial College London, London, SW7 2BP, UK

[*]Corresponding author: Gang Yao. E-mail: yaogang@cup.edu.cn

Received 10 January 2021, revised 14 February 2021
Accepted for publication 31 March 2021

Abstract

Least-squares reverse-time migration (RTM) works with an inverse operation, rather than an adjoint operation in a conventional RTM, and thus produces an image with a higher resolution and more balanced amplitude than the conventional RTM image. However, least-squares RTM introduces two side effects: sidelobes around reflectors and high-wavenumber migration artifacts. These side effects are caused mainly by the limited bandwidth of seismic data, the limited coverage of receiver arrays and the inaccuracy of the modeling kernel. To mitigate these side effects and to further boost resolution, we employed two sparsity constraints in the least-squares inversion operation, namely the Cauchy and L1-norm constraints. For solving the Cauchy-constrained least-squares RTM, we used a preconditioned nonlinear conjugate-gradient method. For solving the L1-norm constrained least-squares RTM, we modified the iterative soft thresholding method. While adopting these two solution methods, the Cauchy-constrained least-squares RTM converged faster than the L1-norm constrained least-squares RTM. Application examples with synthetic data and laboratory modeling data demonstrated that the constrained least-squares RTM methods can mitigate the side effects and promote image resolution.

Keywords: Cauchy constraint, least-squares RTM, L1-norm constraint, reverse-time migration, sparse optimization

1. Introduction

Seismic migration is a key technique in imaging the subsurface of the Earth. It produces subsurface reflectivity images by shifting the recorded seismic reflections to where they originated from. Reverse-time migration (RTM) is an advanced seismic migration technology for imaging complex models (McMechan 1983; Whitmore 1983) when compared to either a ray-based or one-way wave equation-based method. The merit of the RTM image is fundamentally because RTM uses a two-way wave equation, which produces a seismic wavefield with high accuracy (Zhang *et al.* 2011; Shi *et al.* 2019a) when compared to either a ray-based method (Červený 2001; Shi *et al.* 2019b) or an one-way wave equation-based method (Claerbout & Doherty 1972; Stolt 1978).

However, all these three methods, including the ray-based method, the one-way wave equation-based method and the conventional RTM method, employ an adjoint operator, instead of the inverse, of a linearized modeling operator to produce the migration image (Claerbout 1992). The adoption of the adjoint operator involves an assumption that the inverse

© The Author(s) 2021. Published by Oxford University Press on behalf of the Sinopec Geophysical Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of the Hessian matrix is an identity operator. But this assumption is valid only if the seismic data are acquired around the entire model and meanwhile the seismic data have full bandwidth (Xu *et al.* 2011). Unfortunately, these two conditions are impractical in seismic acquisition and, consequently, the resolution and accuracy of the migration image are compromised. The least-squares migration method recasts a migration method into a least-spares inverse problem, so as to improve the resolution and accuracy (Nemeth *et al.* 1999; Kuhl & Sacchi 2003). The least-squares RTM method is a combination between the least-squares inversion and RTM (Dai *et al.* 2012; Yao *et al.* 2018a; Yang *et al.* 2019; Guo & Wang 2020; Liu *et al.* 2020; Song *et al.* 2020; Sun *et al.* 2020). Basically, least-squares RTM applies the inverse of the Hessian matrix as the RTM adjoint operator. This application will result in the removal of source signature blurring and uneven illumination.

The Hessian matrix is ill-conditioned, because the acquired seismic data generally are incomplete due to the recording limitation. The application of this ill-conditioned Hessian will cause some side effects in the least-squares RTM, including sidelobes around reflectors (Yao & Jakubowicz 2016) and high-wavenumber migration artifacts. The sidelobes are caused because the least-squares RTM in effect applies a deconvolution imaging condition on bandlimited seismic data, and the deconvolution condition generates oscillations in the image. The high-wavenumber artifacts such as some arc-shaped artifacts in the image are generated because of the limited spatial coverage of the acquisition geometry. The inaccuracy of the modeling kernel used in the least-squares RTM also causes some migration artifacts. The kernel often uses a Born linearization of the full-wave equation, but the actual seismic wavefields comply with the full-wave equation. This discrepancy between the predicted seismic data and the recorded seismic data generates migration artifacts in the least-squares RTM image.

Suppressing these artifacts and boosting resolution requires improvement to the least-squares inversion, for instance, incorporating a sparsity constraint in the inversion. One of the widely used sparsity constraints in the least-squares RTM is the L1 norm. Wu *et al.* (2016) formulated L1-norm-constrained least-squares RTM as a basis pursuit denoising (BPDN) problem and demonstrated that, although it converges much slower than the unconstrained least-squares RTM, the L1-norm constraint in least-squares RTM can boost the resolution of the image. A total-variation (TV) constraint is another method that could be used to remove the migration artifacts and meanwhile to promote the sparsity of the image (Lin & Huang 2015). However, because the TV constraint enforces piece-wise smoothness, it is suitable for the velocity inversion or the impedance inversion (Feng *et al.* 2019), rather than for the reflectivity inversion with the least-squares RTM. The sparsity promotion can also be achieved by other non-convex constraints, such as the Cauchy constraint (Wang 2016). In contrast to the L1-norm constraint, the Cauchy constraint is differentiable at the origin. Therefore, the problem can be solved efficiently using a gradient-based method.

In this paper, we compare the Cauchy and L1-norm constraints in the context of least-squares RTM for suppressing the sidelobe artifacts and high-wavenumber artifacts. We solve the Cauchy-constrained least-squares RTM using a nonlinear preconditioned conjugate-gradient method and solve the L1-norm-constrained least-squares RTM using a modified iterative soft thresholding (IST) method. In the latter, we use a preconditioned conjugate-gradient direction to replace the gradient direction in the standard IST, so as to improve the efficiency. In addition, we also present effective strategies for selecting the regularization parameter and other parameters for both constrained least-squares RTM methods. Applications to synthetic and laboratory modeling data demonstrate that the Cauchy constraint is more efficient than the L1-norm constraint in the least-squares RTM.

2. Least-squares RTM with the Cauchy and L1-norm constraint

The objective function of the Cauchy-constrained least-squares RTM is expressed as (Wang 2016):

$$\phi_{\text{Cauchy}} = \frac{1}{2}\left\|\mathbf{d}-\mathbf{d}_0\right\|_2^2 + \lambda\sum_{k=1}^{N}\frac{1}{2}\ln\left(1+\frac{m_k^2}{\mu_k^2}\right), \qquad (1)$$

where $\mathbf{d}_0$ is the vector of the recorded data, $\mathbf{d}$ represents the vector of the predicted data, $\mathbf{m}$ is the vector of the reflectivity model with elements m_k, λ represents the regularization parameter and μ_k is the Cauchy hyper-parameter associated with the kth model elements.

Since the Cauchy constraint term in the objective function is nonquadratic, a nonlinear preconditioned conjugate-gradient method with Polak-Ribière is used (Algorithm 1). The gradient of the objective function (1) is given by

$$\frac{\partial\phi_{\text{Cauchy}}}{\partial\mathbf{m}} = \mathbf{L}^T(\mathbf{L}\mathbf{m}-\mathbf{d}_0) + \lambda\left[\frac{m_1}{\mu_1^2+m_1^2}\ \cdots\ \frac{m_k}{\mu_k^2+m_k^2}\ \cdots\ \frac{m_N}{\mu_N^2+m_N^2}\right]^T, \qquad (2)$$

where $\mathbf{L}$ denotes the linearized wave equation operator that maps the reflectivity model $\mathbf{m}$ to predicted data $\mathbf{d}$ (Yao *et al.* 2018b) and the superscript T represents transpose. We used a preconditioner $\mathbf{M}$ to speed up convergence. The preconditioner $\mathbf{M}$ is a diagonal matrix, one diagonal element of which is computed by summing the square of all shots' source wavefields S along time t:

$$M_{kk} = \sum_{\mathbf{x}_s}\sum^{T_{\max}} S(\mathbf{x}_k, t; \mathbf{x}_s)\, S(\mathbf{x}_k, t; \mathbf{x}_s), \qquad (3)$$

Algorithm 1. *The preconditioned nonlinear conjugate-gradient method with Polak-Ribière for Cauchy-constrained least-squares RTM (adapted from Shewchuk (1994, p. 53)).*

Input initial $\mathbf{m}$, λ, μ, $n_{\max}$
$n = 0$
$\mathbf{r} = -\frac{\partial \phi}{\partial \mathbf{m}}$
Compute the preconditioner $\mathbf{M}$
$\mathbf{s} = \mathbf{M}^{-1}\mathbf{r}$, $\mathbf{p} = \mathbf{s}$, $\delta_{\text{new}} = \mathbf{r}^T\mathbf{p}$
While $n < n_{\max}$ *do*
$\alpha = \arg\min_{\alpha} \phi(\mathbf{m} + \alpha\delta\mathbf{m})$, *where* $\delta\mathbf{m}$ *is a small perturbation along* $\mathbf{p}$ *and* α *is the step length obtained through line search*
$\mathbf{m} \leftarrow \mathbf{m} + \alpha\delta\mathbf{m}$
$\mathbf{r} = -\frac{\partial \phi}{\partial \mathbf{m}}$, $\delta_{\text{old}} = \delta_{\text{new}}$, $\delta_{\text{mid}} = \mathbf{r}^T\mathbf{s}$, $\mathbf{s} = \mathbf{M}^{-1}\mathbf{r}$, $\delta_{\text{new}} = \mathbf{r}^T\mathbf{s}$
$\beta = \max(\frac{\delta_{\text{new}} - \delta_{\text{mid}}}{\delta_{\text{old}}}, 0)$, $\mathbf{p} \leftarrow \mathbf{s} + \beta\mathbf{p}$, $n \leftarrow n + 1$
End

Algorithm 2. *Improved iterative soft thresholding method.*

1: Input initial $\mathbf{m}$, λ, N;
2: For $i = 0, 1, \cdots, N-1$ *do*
i. $\hat{\mathbf{m}} = \mathbf{m} - \alpha\delta\mathbf{m}$, *where* $\delta\mathbf{m}$ *is a small perturbation along the preconditioned conjugate-gradient direction and*α*is the step length*
ii. $\mathbf{m} = T_\lambda(\hat{\mathbf{m}})$, *where* $T_\lambda(\hat{m}_k) = \text{sgn}(\hat{m}_k)\max(0, |\hat{m}_k| - \lambda W_{kk})$, *and* $\hat{m}_k$ *is an element of the vector* $\hat{\mathbf{m}}$ *and* W_{kk} *a diagonal element of matrix* $\mathbf{W}$.
End for

where $\mathbf{x}_k$ indicates the position of the kth element and $\mathbf{x}_s$ denotes the source position, $T_{\max}$ is the data recording length.

The objective function of the L1-norm-constrained least-squares RTM is expressed as

$$\phi_{\text{L1}} = \frac{1}{2}\left\|\mathbf{d} - \mathbf{d}_0\right\|_2^2 + \lambda\|\mathbf{W}\mathbf{m}\|_1, \tag{4}$$

where $\mathbf{W}$ is a diagonal matrix consisted of weighting factors. The setting of these weighting factors will reduce the penalty on weak reflectors by scaling down their contribution to the objective function.

The IST method is one of the effective approaches to minimize the objective function in equation (4) (Daubechies *et al.* 2004). The IST method is a special case of the proximal method (Parikh & Boyd 2014), which solves a general optimization problem

$$\mathbf{min}\,\{g(\boldsymbol{m}) + h(\boldsymbol{m})\}, \tag{5}$$

iteratively with

$$\mathbf{m}^i = \text{prox}_{\alpha_i h}\left\{\mathbf{m}^{i-1} - \alpha_i \nabla g\left(\mathbf{m}^{i-1}\right)\right\}, \tag{6}$$

where $g(\mathbf{m})$ is a smooth function (i.e. $\frac{1}{2}\|\mathbf{d} - \mathbf{d}_0\|_2^2$ in this paper), $h(\mathbf{m})$ is a convex function but not differentiable everywhere (i.e. $\lambda\|\mathbf{W}\mathbf{m}\|_1$ in this paper), α is the step length, i is the iteration number and prox$(\cdot)$ is called the proximity operator, the soft thresholding operator for IST.

Conventionally, the proximal method operates along the gradient direction (i.e. $-\mathbf{L}^T(\mathbf{L}\mathbf{m} - \mathbf{d})$). We found that the Polak-Ribière conjugate-gradient direction (Shewchuk 1994) preconditioned by the stack of all shots' source wavefield energy (equation (3)) can be used in IST for fast convergence. Then, we have an improved IST method, as described in Algorithm 2. Step (i) in Algorithm 2 is achieved by using Algorithm 1. Step (ii) is the soft thresholding operation, in which 'sgn' takes the sign of a number and 'max' means a non-negative value of the updated model element $|\hat{m}_k| - \lambda W_{kk}$. The regularization parameter λ, in general, should be smaller than the magnitude of weak reflectors.

3. Application to an overthrust model

We apply the two constrained least-squares RTM methods to a synthetic dataset generated from an overthrust model, and compare them to the result of a conventional RTM and an unconstrained least-squares RTM.

Figure 1a shows a slice of the 3D overthrust model. A dataset comprised of 183 shots is generated by an acoustic wave equation. The source signature is a 15-Hz Ricker wavelet (Wang 2015). The shot interval is 100 m. The first shot is located at a distance of 0.9 km. A split-spread geometry with a maximum offset of 5.25 km is used to record the data. All sources and receivers are deployed at a depth of 50 m. Figure 1b shows the migration velocity model generated by smoothing the slowness of the true velocity model with the SeismicUnix command smooth2, and 1c is the ideal reflectivity image.

Figure 2 displays one shot gather. The direct arrivals and refractions are muted before migration because direct arrivals and refractions are transmitted waves, which do not contain the information of reflectivity but the information of background velocity (Yao *et al.* 2020). The mute position is indicated by the dashed black curve in the figure.

Figure 3a displays the conventional RTM image, which does not distinguish the top and bottom boundaries of thin layers (blue arrows in figure 1c) and also gives unbalanced amplitude. For instance, the amplitude of the top interface (the red arrow in figure 3a) is very weak. However, the amplitude of some other reflectors is over strong (blue arrows in figure 3a). Thus, this result shows that the conventional RTM yields an image with low resolution and inaccurate amplitude.

A least-squares RTM that applies a deconvolution imaging condition can improve resolution and balance amplitude. The least-squares RTM image with 50 iterations (figure 3b) clearly shows much thinner reflectors, as well as the top and bottom boundaries of the thin layers. The amplitude of the reflectors is also more balanced than that in the RTM image.

However, least-squares RTM also introduces some side effects (figure 3b): first, sidelobes appeared around the reflectors, and second, pronounced vertical and arc-shape high-wavenumber artifacts exist in the image. To mitigate

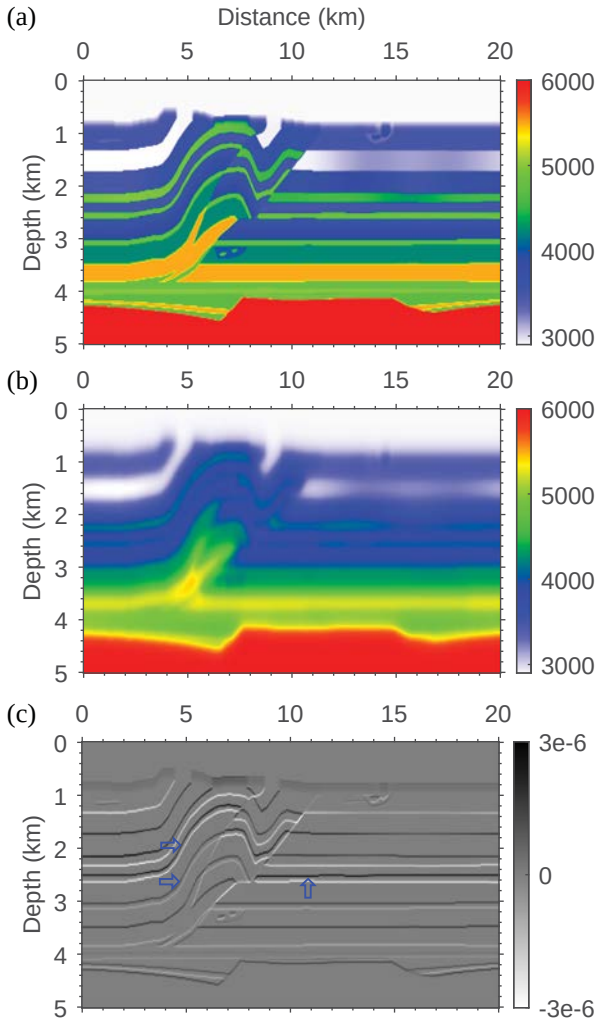

Figure 1. (a) One slice of the 3D overthrust velocity model. (b) The migration velocity, which is a smoothed version of (a). (c) The ideal reflectivity model for reference. The unit of velocity is m s^{-1} while the reflectivity is dimensionless.

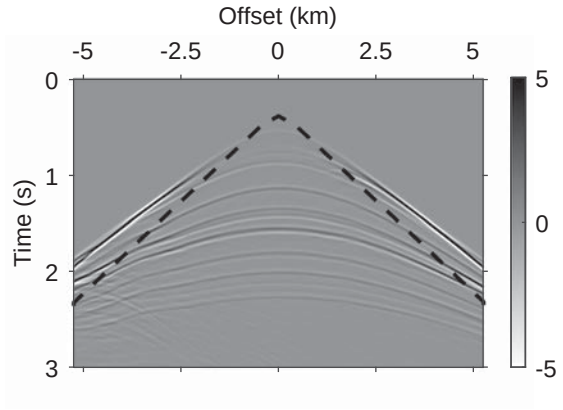

Figure 2. One shot record for migration. The direct arrivals are removed. The dashed black curve indicates the mute position. The data above the mute are not used in migration.

these side effects and further improve the resolution, we use sparsity constraints in least-squares RTM.

In the Cauchy-constrained least-squares RTM, the RTM image is used as the initial model. The regularization parameter is λ =1.84, which is the auto-determined regularization parameter at the 50th iteration by using the multiplicative objective function (da Silva & Yao 2018). Theoretic analysis and numerical tests show that the auto-determined regularization parameter is a good choice for the regularization parameter in the additive objective function.

Figure 4a shows the image of the Cauchy-constrained least-squares RTM after 20 iterations. Comparing with the unconstrained least-squares RTM image in figure 3b, it is obvious that the sidelobes and other migration artifacts are mitigated significantly. In addition, the reflectors are even thinner than those in the least-squares RTM image, which means higher resolution. Figure 4b shows the image after 50 iterations. More iterations improve resolution and mitigate artifacts further.

Then, we applied the L1-norm constraint to promote sparsity and suppress the artifacts. Again, we use the conventional RTM image as the initial model. The revised IST is used to minimize the L1-norm constrained objective function. We set the regularization parameter to $\lambda = 5 \times 10^{-8}$, which is smaller than the amplitude of weak reflectors. Step (ii) of algorithm 2 shows that the elements that have a smaller magnitude than the threshold λ shrink to zero; therefore, the threshold should be decreased for weak reflectors. Otherwise, these weak reflectors will be over penalized.

Figure 5a is the recovered reflectivity image after 50 iterations. By comparison, the L1-norm constrained image has much fewer sidelobes and migration artifacts, and has slimmer reflectors than the unconstrained least-squares RTM image shown in figure 3b.

We notice that the L1-norm-constrained image has faint sidelobes around some reflectors but these faint sidelobes are removed in the Cauchy-constrained image. This is because the thresholding step in IST breaks the conjugacy of conjugate-gradient methods partially; consequently, the L1-norm constrained least-squares RTM converges slower than the Cauchy-constrained least-squares RTM with a preconditioned nonlinear conjugate-gradient method.

As indicated by the blue arrow in figure 5b, IST with the steepest-descent direction leaves some vertical artifacts in the image. Moreover, the better convergence is also demonstrated by a wider range of reflectors recovered with the revised IST, which is indicated by the vertical dashed lines in figure 5.

Comparing figures 4a and 5a, the Cauchy-constrained least-squares RTM with 20 iterations delivers a similar result as the L1-norm constrained least-squares RTM with 50 iterations. This implies that the Cauchy constraint is more

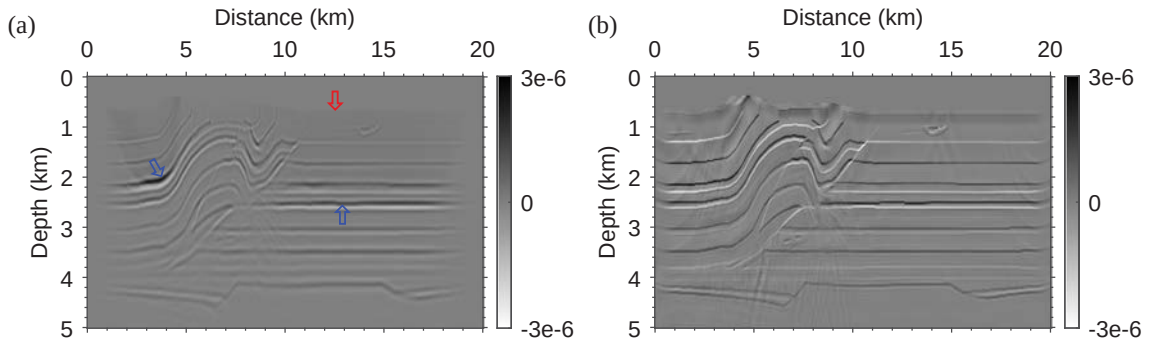

Figure 3. Migration images. (a) RTM. (b) Unconstrained least-squares RTM.

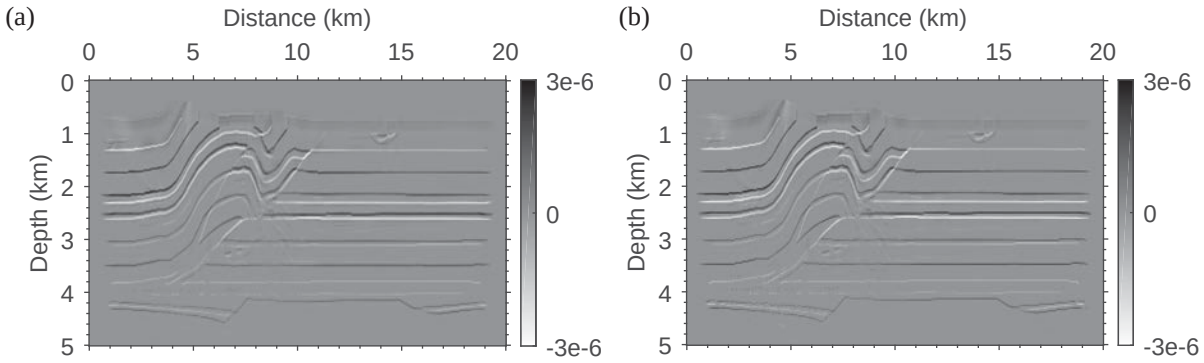

Figure 4. Migration images obtained by the Cauchy-constrained least-squares RTM. (a) The image after 20 iterations. (b) The image after 50 iterations.

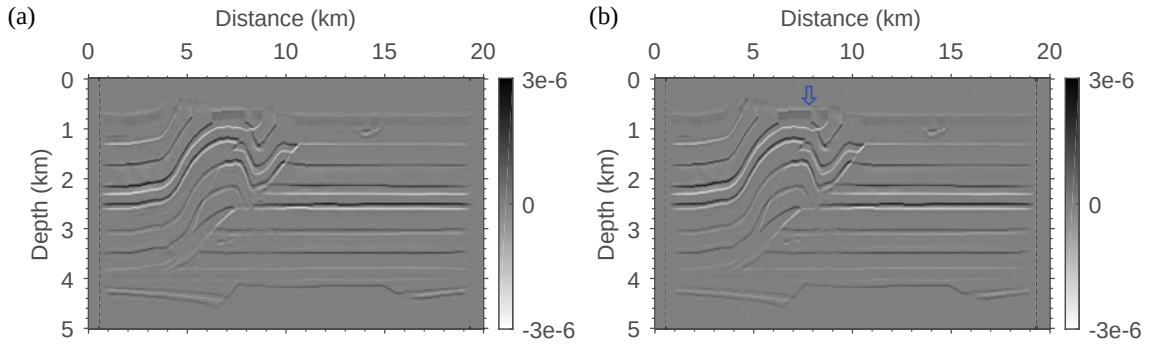

Figure 5. Migration images with L1-norm constrained least-squares RTM solved using (a) revised IST along a preconditioned conjugate-gradient direction and (b) original IST along a steepest-descent direction. The blue arrow indicates the artifact residual. The vertical dashed lines indicate the reflector range recovered by using the revised IST.

efficient than the L1-norm constraint thanks to the preconditioned nonlinear conjugate-gradient method. However, one minor imperfectness of the Cauchy constraint is that it leaves some spike-like artifact residuals in the images (figure 4b), which can be removed using further filtering after migration. A more detailed comparison can be found in an enlarged view of migration results and a trace comparison in figure 6.

We also notice that some weak reflectors at a depth of about 4 km (figure 1c) are recovered by unconstraint least-squares RTM but not by RTM. However, the faint images of these reflectors are suppressed by both constrained least-squares RTM. One way to mitigate this issue is to relax constraint at the cost of keeping more artifacts in the image or to relax the constraint locally. An alternative approach is to carry out joint interpretation of images obtained from RTM, unconstrained least-squares RTM and constrained least-squares RTM side by side for interpretation.

We noticed that the reflector of the top interface in the least-squares RTM image is much weaker than that of the other reflectors below. Therefore, we applied a spatially varying hyper-parameter shown in figure 7a. The

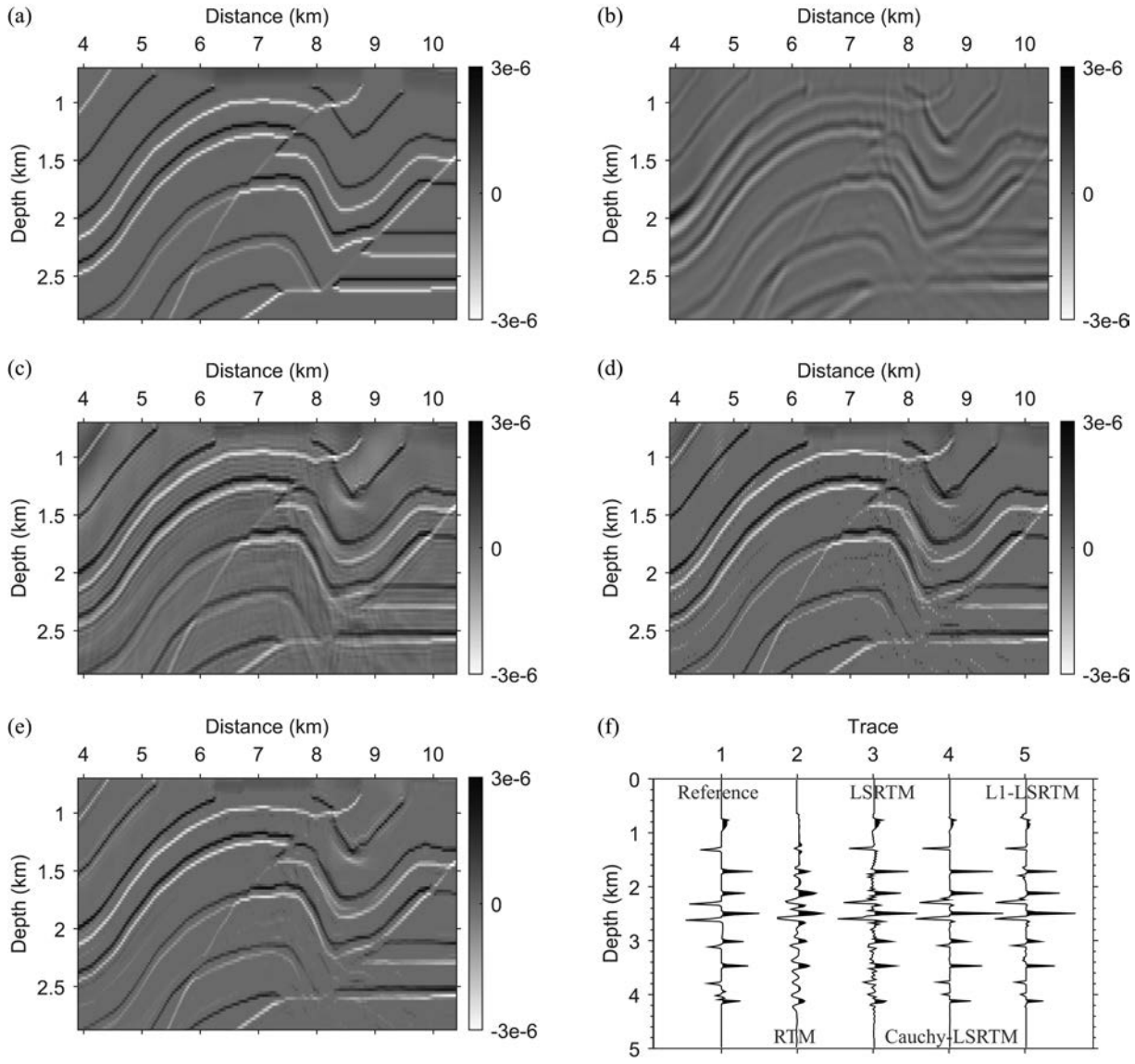

Figure 6. Enlarged view of migration results. Panels (a)–(e) correspond to figure 1c (the ideal reflectivity model), figure 3a (RTM), figure 3b (unconstrained least-squares RTM), figure 4b (Cauchy-least-squares RTM) and figure 5a (L1-least-squares RTM), respectively. (f) Wiggle display of the trace at the distance of 12 km.

hyper-parameter has a bigger value around the top interface to avoid the Cauchy constraint wiping out the weak reflector.

Figure 8 parts c and d compare shot gathers of the predicted data from the Cauchy and L1-norm constrained least-squares RTM. Figure 9 parts c and d show the corresponding residuals. Although these gathers are very similar to the gather of the unconstraint least-squares RTM (figure 8b), the migration images are very different. This implies that null space exists in the inversion due to incomplete data. The sparsity constraints are effective ways to reduce the null space.

Figure 10 shows the convergence rate with different migration methods applied to the synthetic data example. The Cauchy-constrained least-squares RTM converges faster than the L1-norm constrained least-squares RTM but slower than the unconstrained least-squares RTM. The unconstrained least-squares RTM leaves smaller residuals. This is because that the unconstrained least-squares RTM generates artifacts to fit the recorded data further. However, the constraints remove the artifacts resulting in less fitting. In addition, updating the model along the preconditioned conjugate direction in IST gives faster convergence than along the steepest-descent direction (the magenta curve in figure 10). Consequently, the migration image with the preconditioned conjugate-gradient direction is better than the image (figure 5b) with the steepest-descent direction.

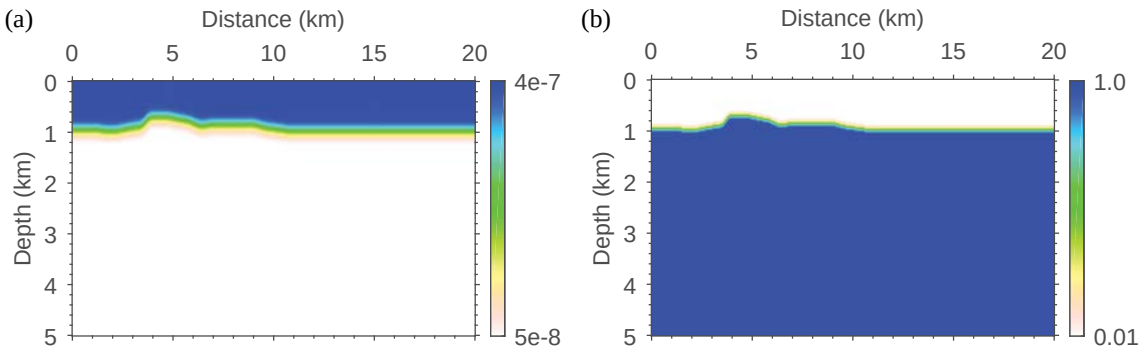

Figure 7. (a) The hyper-parameter used in the Cauchy-constrained least-squares RTM. (b) The weighting factors in the L1-norm constrained least-squares RTM.

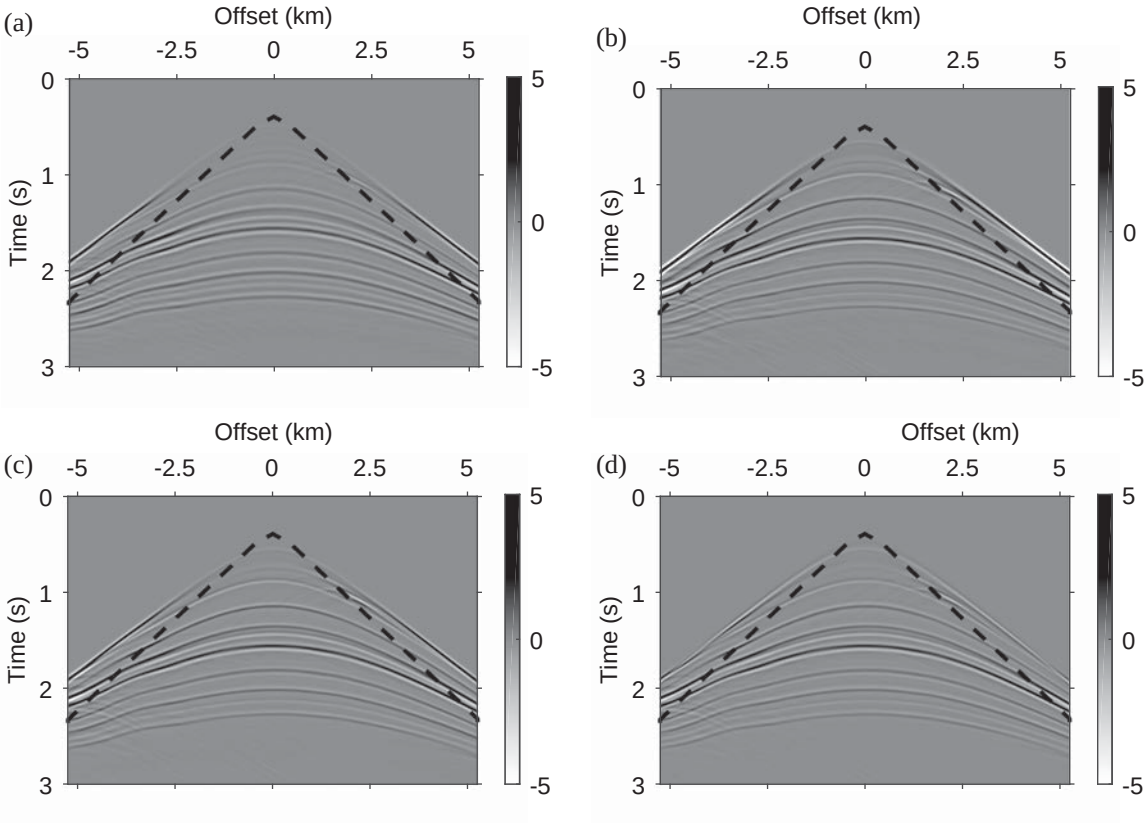

Figure 8. The predicted shot gathers from four reflectivity models. (a) The model generated by RTM (figure 3a). (b) The model generated by an unconstrained least-squares RTM (figures 3b). (c) The model generated by the Cauchy-constrained least-squares RTM (figure 4b). (d) The model generated by the L1-norm constrained least-squares RTM (figure 5a).

4. Application to laboratory modeling dataset

We also apply the constrained least-squares RTM methods to an ultrasonic dataset acquired from a physical scale model.

Figure 11a displays the first shot of the ultrasonic dataset, which was acquired along a 2D line on the surface of a 3D model. The line contains 41 shots and the shot interval is 8 mm. The minimal offset is 20 mm and the trace interval is 2 mm. The first 20 shots contain 96 traces per shot. Each shot of the rest has four traces fewer than its previous shot. Piezoelectric ultrasonic transducers were used as the sources and receivers.

In this acquisition, the P-wave mode was used for source generation and data recording. Therefore, the dataset is dominated by P-waves but it also contains energy from S-waves resulting mainly from PS conversions. For instance, the steeper direct arrival in figure 11a is S-wave energy. In addition, there are multiples in the record. Since we only migrate primary P-wave energy, the S-wave energy and multiples need to be removed before migration. In this test, we applied *F-K* filtering

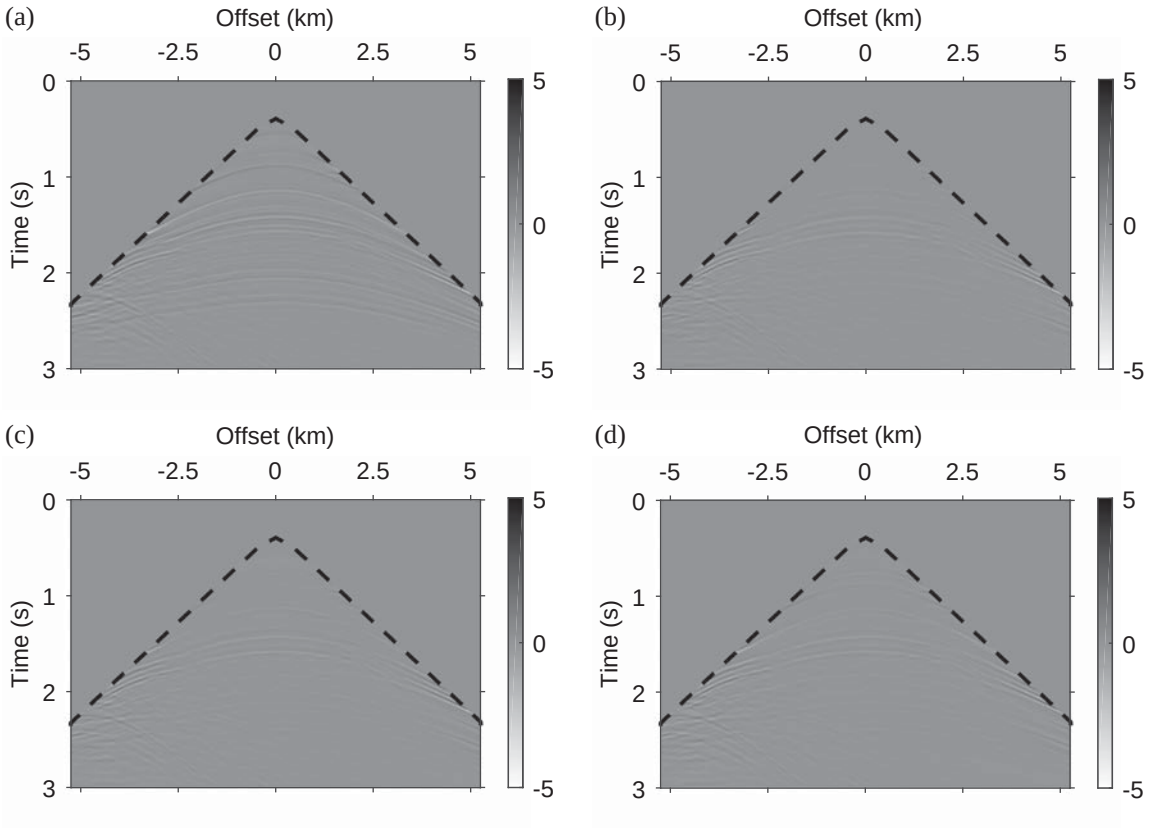

Figure 9. The residuals corresponding to the four shot gathers shown in figure 8. (a) The model generated by RTM (figure 3a). (b) The model generated by an unconstrained least-squares RTM (figure 3b). (c) The model generated by the Cauchy-constrained least-squares RTM (figure 4b). (d) The model generated by the L1-norm constrained least-squares RTM (figure 5a).

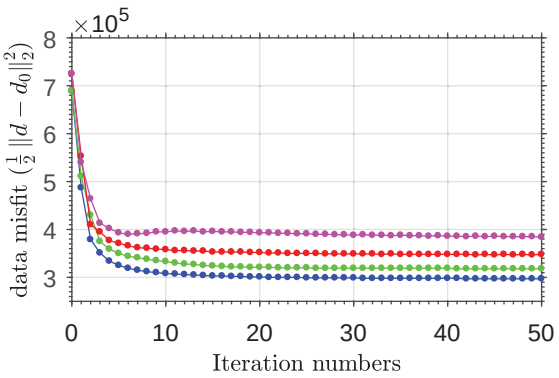

Figure 10. The convergence of the synthetic data example. The data misfits in an unconstrained least-squares RTM (blue), the Cauchy-constrained least-squares RTM (green) and the L1-norm-constrained least-squares RTM with modified IST (red) and standard IST (magenta).

to remove these events. We also scaled the data by a factor of $t^{0.5}$, where t is the travel-time, for geometrical spreading compensation between 2D and 3D propagation. Figure 11b is the first shot of the preprocessed dataset.

This survey also acquired a dataset (figure 12a) by using a single-receiver-per-shot geometry. The offset is 2 mm and, therefore, this acquisition mimics a zero-offset profile. As seen from the profile, this model contains several layers and two structures between 125 and 150 μs. We extracted the wavelet from the third event recorded at about 90 μs. Figure 12 parts b and c depict the source wavelet and its frequency spectrum, respectively. The peak frequency is about 0.3 MHz. The final wavelet for migration is generated by further applying half derivative on this wavelet to correct phase difference between 2D and 3D propagation (Wang & Rao 2009).

Next, we built the velocity model. We performed migration velocity analysis to generate a smooth velocity model (figure 13a), and built a velocity model with sharp boundaries based on the smooth velocity model and its migration image. The final sharp-boundary velocity model (figure 13b) is obtained by further tuning the velocity and interface depth of each layer according to the travel-time misfit between the modeled data and the recorded data. The model contains a fault-like structure and a syncline structure. The velocity model was smoothed using a Gaussian smoother with a standard deviation of 13 before migrating the data.

In the migration step, we carried conventional RTM first. Figure 14 parts a and b are the RTM images by migrating

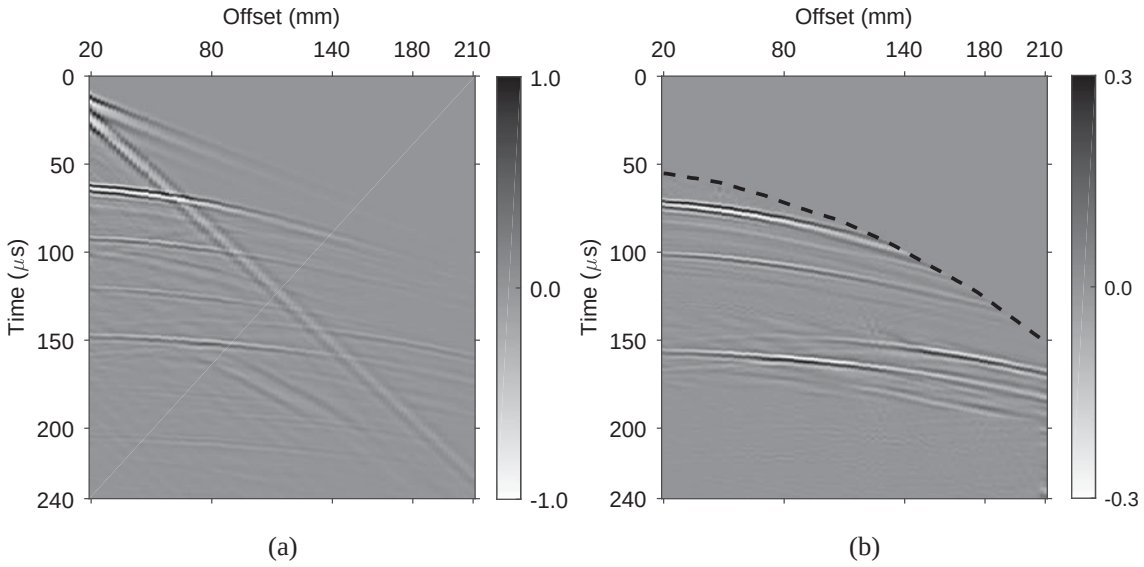

Figure 11. The first shot of the ultrasonic dataset: (a) the raw data and (b) the preprocessed data. The preprocess removes the direct arrivals, other linear noise, S-wave energy and the multiples. The dashed curve indicates the top mute, the data above which are not used in the migration.

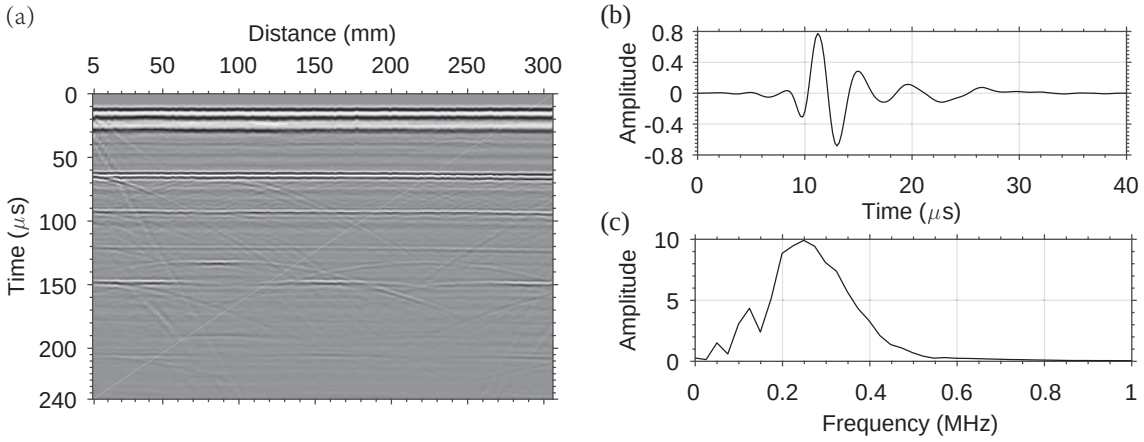

Figure 12. (a) The dataset acquired by using a single-receiver-per-shot geometry. The dataset contains 151 shots/traces, the first shot of which is fired at a distance of 5 mm. The offset of each trace is 2 mm. (b) The wavelet for inversion extracted by stacking the traces of 81–150 between the times of 80 and 126 μs. (c) The frequency spectrum of the extracted wavelet.

the raw record (figure 11a) and the preprocessed record (figure 11b), respectively. By comparison of the images, we have some interesting observations. First, preprocess can clean the image by removing the noise but it also removes some true details of the model, for example, the tiny diffractors indicated by the red circles in figure 14a. These diffractors may be air bubbles created when forging this model. Second, migration is a powerful tool for eliminating noise, including random noise and S-wave energy; even the multiples are attenuated significantly.

Then, we carried out unconstrained least-squares RTM with 20 iterations, and display the image in Figure 14c. By comparison, the least-squares RTM method improves the resolution and balances the amplitude of the image because it uses the inverse operator. This is more clearly observed from the wiggle display of the RTM and the least-squares RTM images shown in figure 14d: the central peak of a reflector is compressed, the sidelobes are suppressed and the amplitude of the bottom reflector is increased.

However, the least-squares RTM method also generated high-wavenumber noise, i.e. the strips and ringing sidelobes, in the image. These artifacts are suppressed with the Cauchy and L1-norm constraint, as depicted in figure 14 parts e and f, respectively. The regularization parameter of the Cauchy constraint is set to $\lambda = 0.01$ and set to $\lambda = 8 \times 10^{-5}$ for the L1-norm constraint.

Figure 15 plots the hyper-parameter μ and the weighting factors in the diagonal matrix $\mathbf{W}$. By comparison, the

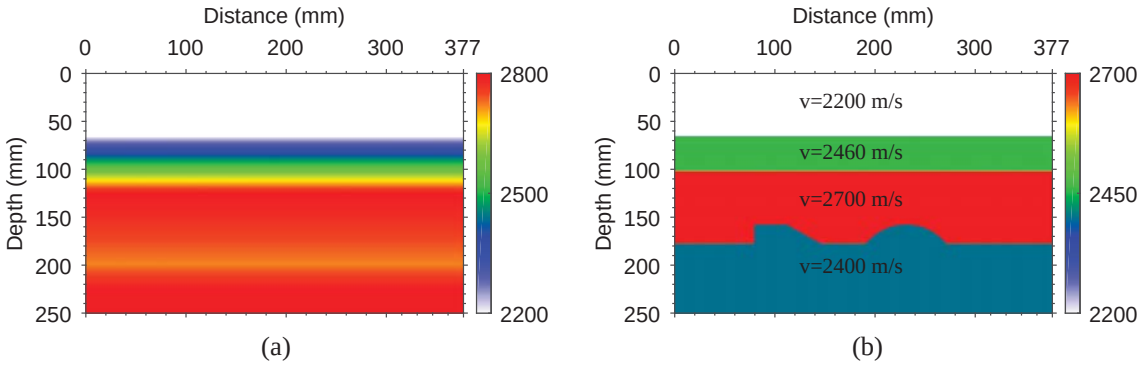

Figure 13. Velocity models. (a) The velocity model built with migration velocity analysis. (b) The final velocity model built based on (a), the migration image and travel-time misfit between the synthetic data and the recorded data.

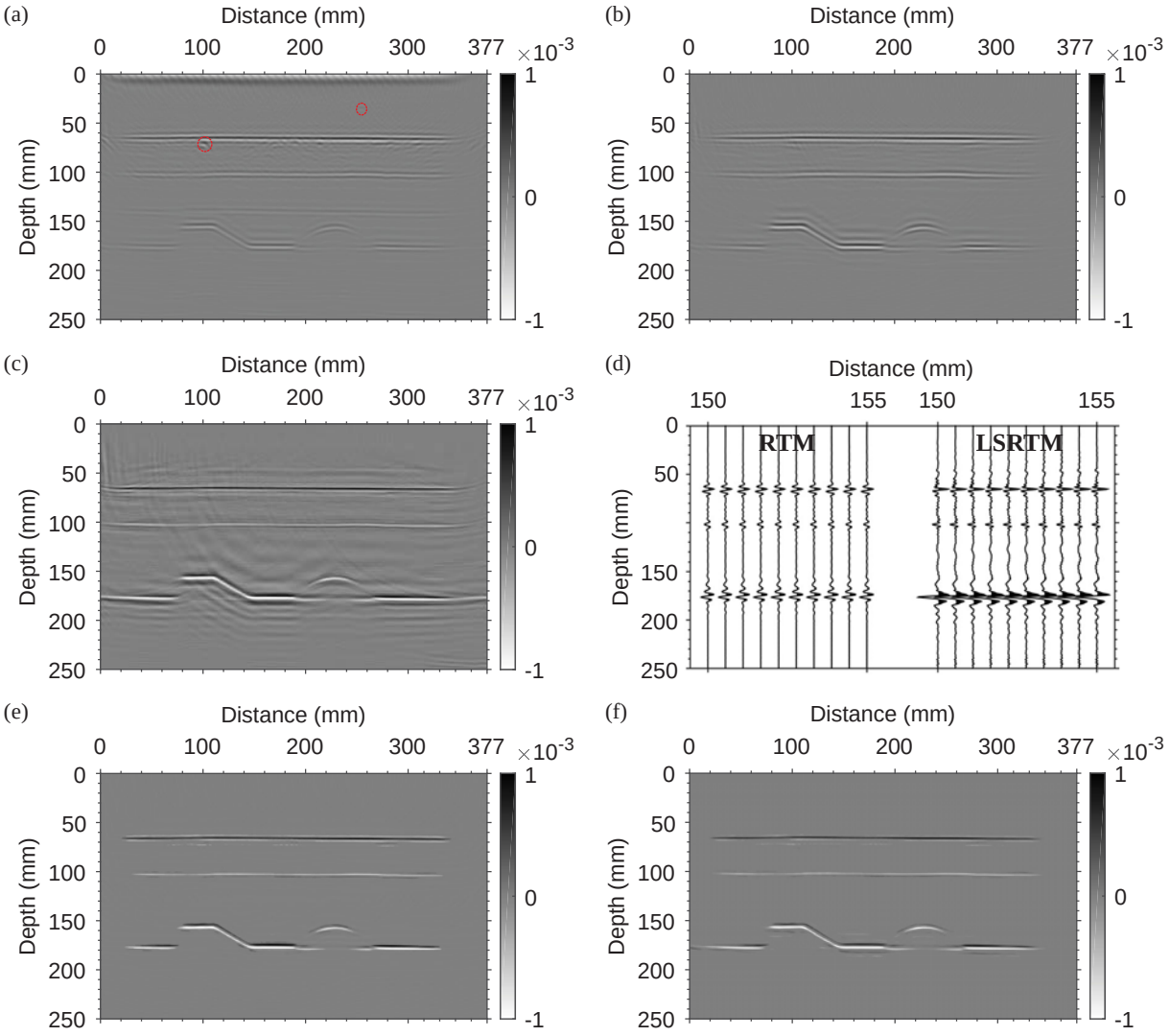

Figure 14. Migration images. (a, b) The conventional RTM images by using the raw dataset and the preprocessed dataset. The red circles in (a) indicate the diffractors in the model. (c) The image of the unconstrained least-squares RTM after 20 iterations. (d) The wiggle display of the conventional RTM image and the unconstrained least-squares RTM image. (e) The image of the Cauchy-constrained least-squares RTM. (f) The image of the L1-norm-constrained least-squares RTM.

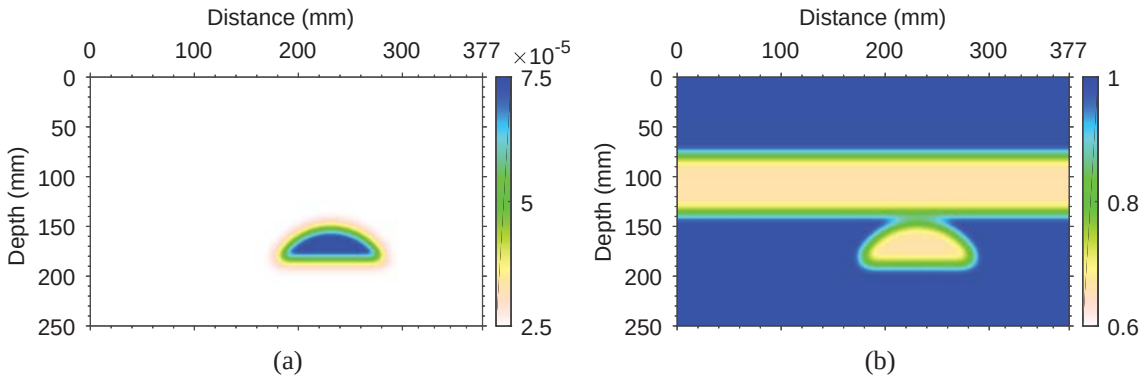

Figure 15. (a) The hyper-parameter used in the Cauchy-constrained least-squares RTM. (b) The weighting factors used in the L1-norm constrained least-squares RTM.

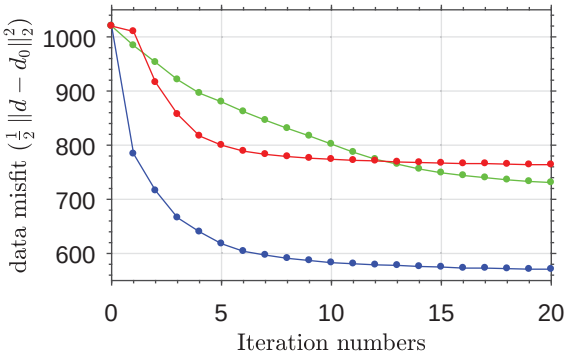

Figure 16. The convergence of the laboratory model example. The data misfits in an unconstrained least-squares RTM (blue), the Cauchy-constrained least-squares RTM (green) and in the L1-norm-constrained least-squares RTM (red).

high-wavenumber artifacts are mitigated significantly by using both constraints and the resolution of the image is further improved.

Figure 16 shows that, since the constraint is enforced in the laboratory data example, the data misfit of both constrained least-squares RTM becomes relatively larger than that of unconstrained least-squares RTM. The constraint can be relaxed to allow reducing the data misfit at the expense of obtaining an image with a higher level of noise. Hence, a trade-off between the level of noise and the level of data fitting is inevitable and effectively depends on some level of trial-and-error and expectations.

The data residuals obtained with two least-squares RTM method are relatively large in the laboratory data example because of two factors. The first one is that the piezoelectric source has a directional feature: the energy becomes weak as the direction is away from the pointing direction of the transducer, while the piezoelectric receiver has similar directional sensitivity (Buddensiek *et al.* 2009). As a result, the amplitude of the events decreases as the offset increases. By contrast, the modeling kernel of migration assumes an isotropic point source. The second factor is that the AVO characteristic of the record follows that of an elastic medium, while the AVO characteristic of the synthetic data is that of acoustic media as the data are modeled using an acoustic wave equation.

During this laboratory data test, we also notice that one simple way to mitigate the strip noise in figure 14c, which is quite steep, is high-pass filtering along the depth direction. However, the high-pass filtering should not be included in constrained least-squares RTM because filtering out low frequencies boosts the sidelobes of a reflector resulting in many artificial sharp reflectors around the true reflector in the migration images.

5. Discussion

Both the Cauchy and L1-norm constrained systems, i.e. equations (1) and (4), include a regularization parameter λ and a hyper-parameter μ or weighting factors in the diagonal matrix $\mathbf{W}$. Choosing the right value for them is key for successful inversion in least-squares RTM.

To determine the regularization parameter λ automatically, we first carried out Cauchy-constrained least-squares RTM with a multiplicative objective function (da Silva & Yao 2018). We then chose the regularization parameter determined in the last iteration, e.g. 50 in the first test, as the regularization parameter of the additive objective function (1). The tests show that this regularization parameter produces even better results than the multiplicative objective function.

The guideline for choosing the hyper-parameter μ is that an appropriate hyper-parameter should give small values of the gradient of the Cauchy term at reflectors but relatively large values at the sidelobes and other migration artifacts. As the maximum value of the gradient of $\ln(1+m^2/\mu^2)$ happens at $m=\mu$, the hyper-parameter μ should be set to around the noise level. However, if the noise level is close to the magnitude of reflectors or even stronger, μ should be

decreased, resulting in less penalization for noise. Otherwise, the constrained inversion removes the reflectors from the estimated image. If the model includes reflectors that are much weaker than the majority reflectors, the hyper-parameter μ should be increased (not decreased) to reduce their penalization. The hyper-parameter μ is then refined by trial-and-error. In this study, we used an improved IST to minimize the L1-norm constrained least-squares RTM. The thresholding parameter λ should be around the noise level but smaller than the weak reflectors. The final optimal value of λ comes by further trial-and-error.

Different from the Cauchy-constrained least-squares RTM, λ in the L1-norm constrained least-squares RTM should be decreased for weak reflectors for avoiding over penalization. Thus, we use a diagonal weighting matrix $\mathbf{W}$ in the objective function (4). The general guideline is that λW_{kk} should be smaller than the magnitude of the weak reflectors. Its optimal value is also determined by trial-and-error.

In a large-scale application, a cost-effective practice is testing the constrained migration methods on a subset of the whole data set (e.g. a 2D line or a section) to find the optimal parameters, and then applying them to the whole data set.

6. Conclusion

Least-squares RTM refines the reflectivity model iteratively, which equivalently applies the inverse of the Hessian to the RTM image. Consequently, the least-squares RTM method produces migration images with higher resolution and more balanced amplitudes than the RTM images. However, least-squares RTM presents two side effects: sidelobes around the reflectors and high-wavenumber artifacts. Including either the Cauchy or the L1-norm constraint in the least-squares RTM suppresses these side effects and boosts resolution with sparseness. In general, least-squares RTM with these two constraints produces similar high-quality results with appropriate parameters. However, application examples with synthetic data and laboratory modeling data demonstrated that the Cauchy constraint is more efficient than the L1-norm constraint in least-squares RTM but the Cauchy constraint can leave some spike-like artifact residuals in the images.

Both constrained least-squares RTM methods need to determine the parameters for a successful inversion. A cost-effective strategy is to test the parameters in a subset of the whole dataset to identify the optimal parameters. Finding effective methods for automatically determining the parameters is a valuable research direction. More importantly, exploring more regularization methods for various geologic models than just mathematically sparse constraints is paramount for improving imaging of seismic data.

Acknowledgements

This work was partially supported by National Key R&D Program of China (grant no. 2018YFA0702502), the National Natural Science Foundation of China (grant nos. 41974142, 42074129 and 41674114), Science Foundation of China University of Petroleum (Beijing) (grant nos. 2462019YJRC005 and 2462020YXZZ005) and the Strategic Cooperation Technology Projects of CNPC and CUPB (ZLZX2020-03).

Conflict of interest statement

Authors do not have any conflicts of interest to declare.

References

Buddensiek, M.L., Krawczyk, C.M., Kukowski, N. & Oncken, O., 2009. Performance of piezoelectric transducers in terms of amplitude and waveform, *Geophysics*, **74**, T33–T45.

Červený, V., 2001. *Seismic Ray Theory*, Cambridge University Press.

Claerbout, J.F., 1992. *Earth Soundings Analysis: Processing Versus Inversion*, Blackwell Scientific Publications.

Claerbout, J.F. & Doherty, S.M., 1972. Downward continuation of moveout-corrected seismograms, *Geophysics*, **37**, 741–768.

Dai, W., Fowler, P. & Schuster, G.T., 2012. Multi-source least-squares reverse time migration, *Geophysical Prospecting*, **60**, 681–695.

Daubechies, I., Defrise, M. & De Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on Pure & Applied Mathematics*, **57**, 1413–1457.

Feng, D., Cao, C. & Wang, X., 2019. Multiscale full-waveform dual-parameter inversion based on Total Variation regularization to on-ground GPR data, *IEEE Transactions on Geoscience and Remote Sensing*, **57**, 9450–9465.

Guo, S. & Wang, H., 2020. Image domain least-squares migration with a Hessian matrix estimated by non-stationary matching filters, *Journal of Geophysics and Engineering*, **17**, 148–159.

Kuhl, H. & Sacchi, M.D., 2003. Least-squares wave-equation migration for AVP/AVA inversion, *Geophysics*, **68**, 262–273.

Lin, Y. & Huang, L., 2015. Least-squares reverse-time migration with modified total-variation regularization, in *SEG Technical Program Expanded Abstracts*, pp. 4264–4269.

Liu, Q., Lu, Y. & Zhang, H., 2020. Fast single-step least-squares reverse-time imaging via adaptive matching filters in beams, *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 1913–1919.

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

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

Parikh, N. & Boyd, S., 2014. *Proximal Algorithms*. Now Publishers Inc.

da Silva, N. V. & Yao, G., 2018. Wavefield reconstruction inversion with a multiplicative cost function, *Inverse Problem*, **34**, 015004.

Shewchuk, J. R., 1994. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, Carnegie Mellon University.

Shi, X., Mao, W. & Li, X., 2019b. Elastic Gaussian-beam migration for four-component ocean-bottom seismic data, *Geophysics*, **85**, S11–S19.

Shi, Y., Zhang, W. & Wang, Y., 2019a. Seismic elastic RTM with vector-wavefield decomposition, *Journal of Geophysics and Engineering*, **16**, 509–524.

Song, M., Zhang, J. & Zhang, J., 2020. Viscoacoustic least-squares migration with a blockwise Hessian matrix: an effective Q approach, *Journal of Geophysics and Engineering*, **17**, 618–634.

Stolt, R. H., 1978. Migration by Fourier transform, *Geophysics*, **43**, 3–48.

Sun, Y., Kim, Y. S., Qu, S. & Verschuur, E., 2020. Joint migration inversion: features and challenges, *Journal of Geophysics and Engineering*, **17**, 525–538.

Wang, Y., 2015. Frequencies of the Ricker wavelet, *Geophysics*, **80**, A31–A37.

Wang, Y., 2016. *Seismic Inversion: Theory and Applications*, Wiley-Blackwell, Oxford.

Wang, Y. & Rao, Y., 2009. Reflection seismic waveform tomography, *Journal of Geophysics Research-Solid Earth*, **114**, B03304.

Whitmore, N.D., 1983. Iterative depth migration by backward time propagation, in *Proc. SEG Tech. Program Expanded Abstract*, pp. 382–385.

Wu, D., Yao, G., Cao, J. & Wang, Y., 2016. Least-squares RTM with L1 norm regularisation, *Journal of Geophysics and Engineering*, **13**, 666–673.

Xu, S., Zhang, Y. & Tang, B., 2011. 3D angle gathers from reverse time migration, *Geophysics*, **76**, S77–S92.

Yang, J., Li, Y., Cheng, A., Liu, Y. & Dong, L., 2019. Least-squares reverse time migration in the presence of velocity errors, *Geophysics*, **84**, S567–S580.

Yao, G. & Jakubowicz, H., 2016. Least-squares reverse-time migration in a matrix-based formulation, *Geophysical Prospecting*, **64**, 611–621.

Yao, G., da Silva, N.V. & Wu, D., 2018a. An effective absorbing layer for the boundary condition in acoustic seismic wave simulation, *Journal of Geophysics and Engineering*, **15**, 495–511.

Yao, G., da Silva, N.V. & Wu, D., 2018b. Forward modelling formulas for least-squares reverse-time migration, *Exploration Geophysics*, **49**, 506–518.

Yao, G., Wu, D. & Wang, S.X., 2020. A review on reflection-waveform inversion, *Petroleum Science*, **17**, 334–351.

Zhang, Y., Zhang, H. & Zhang, G., 2011. A stable TTI reverse time migration and its implementation, *Geophysics*, **76**, WA3–WA11.