Comparison between least-squares reverse time migration and full-waveform inversion

Lei Yang, Daniel O. Trad and Wenyong Pan

ABSTRACT

The inverse problem in exploration geophysics usually consists of two parts: seismic imaging and velocity model constructing. In this paper, we compare the algorithms for least-squares reverse time migration (LSRTM) and full-waveform inversion (FWI) and use numerical examples to understand the differences. LSRTM uses Born approximation as the modelling method because it requires the adjoint of migration (linear inversion), while FWI uses finite-difference modelling because it does not require an adjoint-pair operator (non-linear inversion). Linearized Born modelling can update model perturbations by a linear conjugate gradient method, but may have severe inaccuracies and inversion noise if the initial model is poor. Both, FWI and LSRTM depend on the initial model largely, but FWI has a mechanism to improve the velocities and LSRTM does not. Conversely, FWI suffers from cycle skipping while LSRTM does not. For LSRTM, the long wavelength components of the gradient are considered to be noise, while for FWI they are considered to be signal. In this work we try to use a FWI algorithm to solve for reflectivity instead of using standard LSRTM.

INTRODUCTION

Reverse time migration (RTM) was proposed in 1970s, but because of its high computational cost it did not become common until after 2000. Because it is a two-way migration method, it represents a significant improvement over computational cheaper one-way migration algorithms. This characteristic gives RTM higher accuracy than other methods for imaging of steep dips. Therefore, RTM has become one of the most important migration algorithms because of the progress of computational technology. However, because RTM uses the adjoint operator, which is not the same as the inverse, there will be amplitude problems in RTM, which can be solved by applying a least-squares scheme in the migration. Also, if the velocity model has some strong impedance contrast, low-frequency artifacts will be generated which are usually eliminated by Laplacian filters.

Least-squares migration (LSM) is an alternative to migration that can, potentially, reduce migration artifacts and improve lateral resolution. It was first proposed by Schuster (1993), and applied for the first time to a real data set by Nemeth et al. (1999). Since then much research has been done on this topic although its application to real data by industry remains elusive. This problem is often solved by considering migration as a linear problem, which is a common approximation in seismic processing if multiple reflections are not considered. LSM uses iterative methods to match the observed data for every iteration and can solve the amplitude inaccuracy of RTM. Therefore, RTM can be implemented by a least-squares scheme, which is LSRTM.

Generally, linear operators are used in geophysical modelling calculations that predict data from models. A common task is to find the inverse of these calculations, i.e., given the observed data, find a model that can predict these data. For linear operators, it is

possible to use iterative algorithms which require forward and adjoint operators to solve the inverse problem (Claerbout, 1992). Although in general these adjoint operators are different from the inverse operators, often they approximate the inverse quite well and are more stable and faster. In fact, most commonly used migration algorithms are very similar to the respective adjoint operators but not exactly the same, because they are good approximations to the inverse of the modeling operator. In some cases, the exact adjoint operators are not easy to design, and that is the particular case of RTM. The importance of knowing the exact adjoint operator for RTM is that provides a practical method to calculate LSRTM. Without the adjoint, more computational algorithms for linear search are required, as is often applied for related techniques like FWI. In those cases, the forward problem of data modeling has to be applied many times to calculate the magnitude of corrections applied at each iteration. The adjoint, on the other hand, permits to calculate these step sizes with the cost of only one forward and one adjoint operation.

The methods mentioned above are all linear problems. However, the observed data depend nonlinearly on the parameters of the earth in reality. FWI is a nonlinear inversion method which uses all information in the seismogram to get the earth model. Classical FWI involves the minimization of square objective function for the observed and synthetic data and then moving from the initial model to the minimum of that objective function. For the synthetic data, FWI uses finite-difference method rather than Born modelling to calculate wavefield and may update the initial model in every iteration. The gradient and Hessian operator are calculated in the conjugate gradient algorithm to decide how much the model should move in a certain direction. In the following section, we review the theories of LSRTM and FWI and compare them to find the relationship and use numerical examples to illustrate the merits and weakness of each method.

THEORY

Least-squares Reverse Time Migration

The 2D acoustic wave equation with constant density in time domain is:

$$\frac{1}{v^2(x)}\frac{\partial^2 u}{\partial t^2}(x,t) - \nabla^2 u(x,t) = f(t,x,x_s), \tag{1}$$

where x is the is the position of the earth, $v(x)$ is the acoustic wave velocity, $u(x,t)$ is the acoustic wave potential, ∇^2 is the Laplacian operator and $f(t,x,x_s)$ is the source term. In frequency domain, this equation turns into:

$$\left(\frac{\omega^2}{v^2(x)} + \nabla^2\right) u(x,\omega,v) = f(\omega,x,x_s), \tag{2}$$

where ω is the angular frequency. The conventional imaging condition of RTM can be described as the following equation:

$$I(x) = \int dx_s dx_r dt d\tau G(x_s,x,\tau)\, G(x,x_r,t-\tau) d\left(x_s,x_r,t\right). \tag{3}$$

This is the most typical forms of RTM and the image can be formed at every location where the wavefields coexist in space and time. $G(x_s,x,\tau)$ and $G(x,x_r,t-\tau)$ are the source and receiver Green's function respectively and $d\left(x_s,x_r,t\right)$ is the seismic reflection data in common RTM imaging. However, in LSRTM, $d\left(x_s,x_r,t\right)$ is usually the data residual for

every iteration. As can be seen, this imaging condition is very stable because it only contains multiplication and summation. Unfortunately, this algorithm will provide amplitude inaccuracies and low-frequency artifacts, especially when the velocity model has some strong impedance contrasts. Therefore, the Laplacian filter is usually applied to the migration results to attenuate the low-frequency artifacts. The mathematical form of the Laplacian filter is given by:

$$I_L(x,z) = -\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial z^2}\right), \tag{4}$$

where $I_L(x,z)$ is the filtered image and x,z is the 2D spatial directions respectively. For the imaging condition in equation (3), the source and receiver wavefields are correlated in the space. The imaging result maybe separated into illuminated and unilluminated areas. The illuminated area contains the imaging points where the correlation is non-zero, while in the unilluminated area the source energy did not pass through or did not reach receivers. Between these two areas, the correlation values vary slowly and smoothly. In the homogeneous area, there is only transmission and the wavefield is not perturbed. Therefore, the correlation value of the imaging is a constant in the homogeneous area. In the boundary, this value changes from one constant to another constant, which is corresponding to the velocity and density properties of the medium. To highlight the impedance boundaries, one need to zero-out all constant-valued homogeneous layers and only map impedance boundaries. A divergence operator should be able to carry this out. However, the correlation values vary from zero in the unilluminated area to a constant in the illuminated area smoothly and this will contaminate the divergence result. To handle this problem a Laplacian operator is often used. The problem with the Laplacian operator is some changes in phase and amplitude which can be solved by applying phase-shift filter and LSRTM respectively.

RTM uses the adjoint operators to calculate the phase correctly, but not the amplitude because adjoint operators are not the same as the inverse operators. LSM calculates an inverse operator and can solve the amplitude problem and produce image with fewer artifacts and more accurate amplitudes. Therefore, the combination of RTM and LSM is necessary, which is LSRTM. The linear geophysical modelling process is present as follows:

$$d = Am, \tag{5}$$

where d is the observed data, m is the geophysical model and A is the linear operator which maps m to d. In LSRTM, we use Born modelling (BM, see Appendix) to calculate the data residual for every iteration, which is the operator A in equation (5). In Born approximation, we assume the velocity of the earth model can be split into a smooth part m_0 and a singular part Δm:

$$m = m_0 + \Delta m. \tag{6}$$

As a result, the equation can be re-written in:

$$\Delta d = A\Delta m, \tag{7}$$

Therefore, the objective function is:

$$J(\Delta m) = \left|\left|\Delta d - A\Delta m\right|\right|^2. \quad (8)$$

To get the m that produces the best predictions, the objective function should be minimum, which happens when:

$$\frac{\partial J(\Delta m)}{\partial \Delta m} = 0. \quad (9)$$

This condition implies

$$A^\dagger A\Delta m - A^\dagger \Delta d = 0, \quad (10)$$

where † is the conjugate transpose. Therefore, we may get the least-squares solution below:

$$\Delta m = (A^\dagger A)^{-1} A^\dagger \Delta d, \quad (11)$$

which is usually not stable and expensive to calculate. Fortunately, we have iterative methods, such as steepest-descent (SD) or conjugate gradient (CG) to solve this problem. Table 1 shows the specific preconditioned conjugate gradient algorithm to solve for the best Δm.

Table 1. Preconditioned Conjugate Gradient Method

Given Δm_0, preconditioner M;

Set $r_0 \leftarrow A\Delta m_0 - \Delta d$;

Solve $My_0 = r_0$ for y_0;

Set $p_0 = -y_0, k \leftarrow 0$;

While $r_k \neq 0$

$$\alpha_k \leftarrow \frac{r_k^T y_k}{p_k^T A p_k};$$

$$\Delta m_{k+1} \leftarrow \Delta m_k + \alpha_k p_k;$$

$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$

Solve $My_{k+1} = r_{k+1}$;

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T y_{k+1}}{r_k^T y_k};$$

$$p_{k+1} \leftarrow -y_{k+1} + \beta_{k+1} p_k;$$

$$k \leftarrow k + 1;$$

End (while)

When given the initial smoothed model m_0 and the observed data d, the initial data d_0 can be calculated by the forward modelling operators and thus the data residual Δd is obtained. Inject the data residual into the CG algorithm. When the iterations are finished, one can get the model perturbation Δm from the output. Figure 1 shows the workflow for

the LSRTM. One can use BM to get data residuals for every iteration, but they will depend not only on the reflector amplitudes but also on the initial model m_0. Because RTM focuses on the high frequency components and treat the low frequency components as the noise, the quality of the initial model would have significant influence on the inversion result. This is the limitation of LSRTM comparing with FWI.

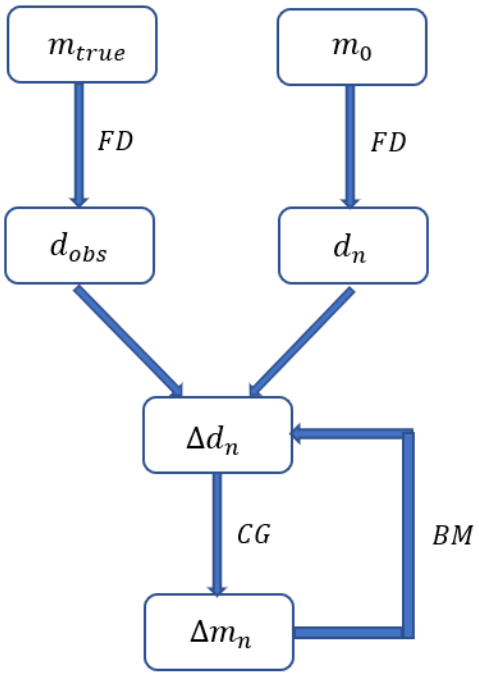

FIG. 1. A flowchart for LSRTM

Full Waveform Inversion

Full waveform inversion is a nonlinear inversion method. Similar to LSRTM, it starts from a least-squares problem:

$$J(m) = \left|\left|d_{obs} - d_{syn}\right|\right|^2, \tag{12}$$

where d_{obs} are the observed data generated by the true model and d_{syn} are the synthetic data obtained by the current model for every iteration. In the frequency domain, the Green's function can be separated into two parts:

$$G(m_0 + \delta m) = G_0(m_0) + \delta G(m_0, \delta m), \tag{13}$$

where m_0 represents the long wavelength components of the model and δm indicates the short wavelength components or model perturbations. $\delta G(m_0, \delta m)$ is the wavefield perturbation which has a linear relationship with δm:

$$\delta G(m_0, \delta m)(x_r, \omega, x_s) = \omega^2 \int dx G_0(x_r, \omega, x)\, G(x, \omega, x_s) \delta m(x). \quad (14)$$

This is the conventional kernel of the Born operator (also see in Appendix). However, FWI is usually a nonlinear method for the inversion problem. Figure 2 shows a basic workflow for the FWI. First, we have the true model and the initial model, which is formulated by the finite-difference method to get the data. By subtracting the synthetic data from the observed data, one can get the data residuals, and from them the model perturbations by a mapping (migration) from data to model space. Once the perturbation is obtained, the initial model can be updated and used to generate new synthetic data in the next iteration.

Back to the objective function, we have

$$\frac{\partial J(m+\Delta m)}{\partial m} = 0, \quad (15)$$

when the minimum value of the misfit function is reached. Ignoring the high-order terms,

$$\frac{\partial J(m+\Delta m)}{\partial m} \approx \frac{\partial J(m)}{\partial m} + \frac{\partial^2 J(m)}{\partial m^2} \Delta m = 0, \quad (16)$$

and

$$\Delta m = -\left(\frac{\partial^2 J(m)}{\partial m^2}\right)^{-1} \frac{\partial J(m)}{\partial m}. \quad (17)$$

The first order derivative is the gradient and the second order derivative is the Hessian. Unfortunately, the Hessian is an extremely large and dense matrix for large scale inverse problems and its inverse has a large computation cost. If the model dimension is $M * N$, the Hessian will be a $MN * MN$ symmetric square matrix. Therefore, there are many approximations of the Hessian. The nonlinearity of FWI occurs because the initial model is updated for every iteration to generate synthetic data. Also, the Hessian changes from point to point in model space. In comparison, LSRTM uses the data residuals to find the best model perturbation by the CG method and does not change the initial model while the FWI updates the initial model before doing forward modeling at every iteration. LSRTM focuses on the short wavelength components to get the reflectivity image but it depends on the initial model largely. If the initial model is not accurate, the migration result will be wrong. FWI can correct for model errors but it also has problems if the initial model is inaccurate. To address this issue, FWI starts from low frequency components of the data which are less sensitive to cycle-skipping and corrects large wavelengths components of the background velocity model before using the higher frequency part of the data.

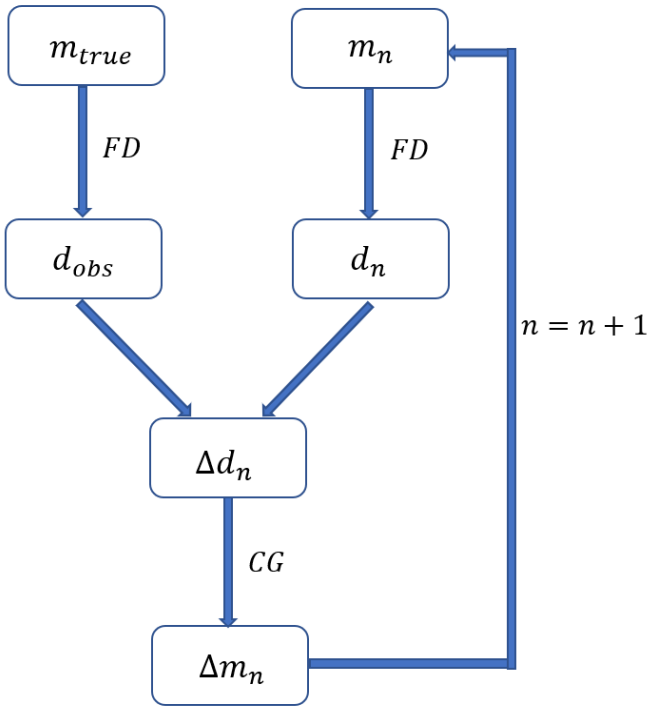

FIG. 2. A flowchart for basic FWI algorithm

NUMERICAL EXAMPLES

In this section we show some simple numerical examples for LSRTM and FWI using an inaccurate version of the Marmousi velocity model, resampled in a grid with 126*288 cells. The horizontal and vertical sampling are both equal to 8m. We use a constant density in this example. There are 285 shots and 287 receivers on the surface and the frequency is 5-25 Hz for RTM and LSRTM and 3-12 Hz for FWI.

Figure 3 shows the true velocity model and the initial model, which was obtained by applying a Gaussian smoother on the true velocity model. In Figure 4 we see the results when the smooth initial model is used for RTM and LSRTM. The RTM image is distorted by the low-frequency components severely. LSRTM image has better result which recovers the amplitude problem but still have poor resolution. This is because of the poor quality of the initial model. RTM and LSRTM are designed to focus on the reflections to generate the reflectivity image of the model, but they require a good initial model. Now we use FWI from the same initial model and see in Figure 5 the results. We use this result as the starting model and get RTM and LSRTM results which are showed in Figure 6. The accuracy of the model is improved. Because FWI starts the inversion from low frequency components first and build the background model before using the high frequencies, it can deal with the inaccuracies in the velocities. This simple example suggests that using the non-linear approach of FWI for LSRTM could be useful to provide better images. Our goal in future work is to explore how a LSRTM based on the ideas of FWI would compare with the cascade approach FWI+LSRTM shown in this example

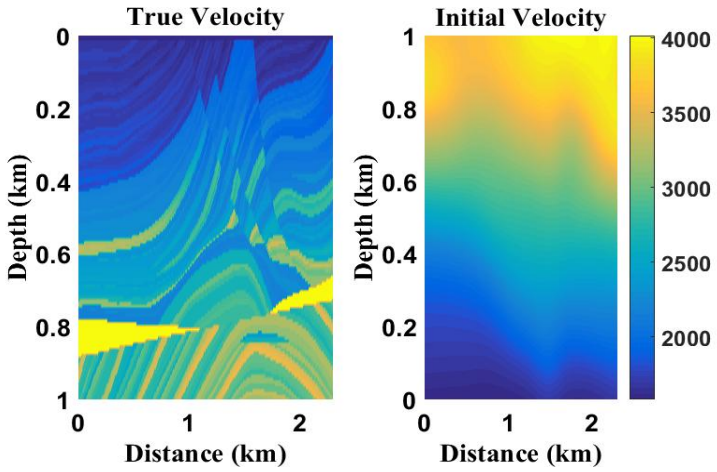

FIG. 3. The true velocity model of Marmousi and the initial model applied Gaussian smoother

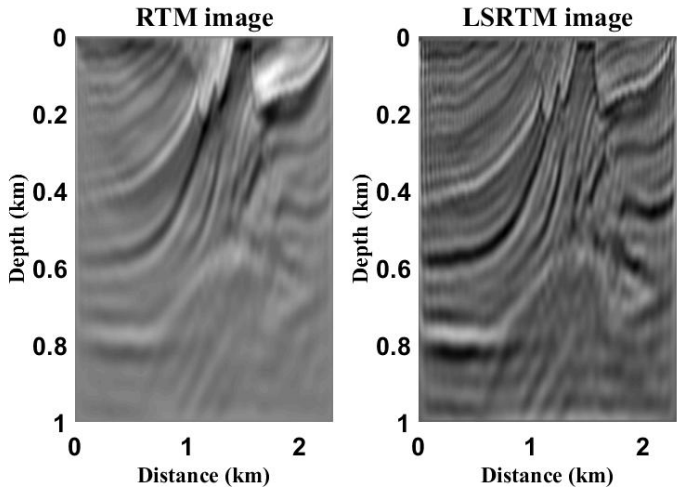

FIG. 4. RTM image and LSRTM image of the initial model

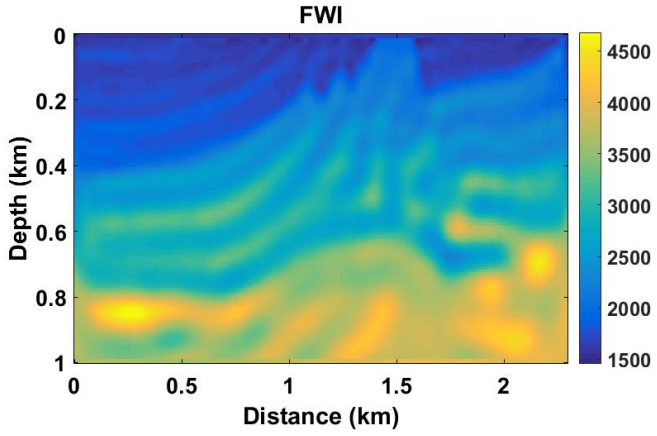

FIG. 5. FWI result from the smoothed model

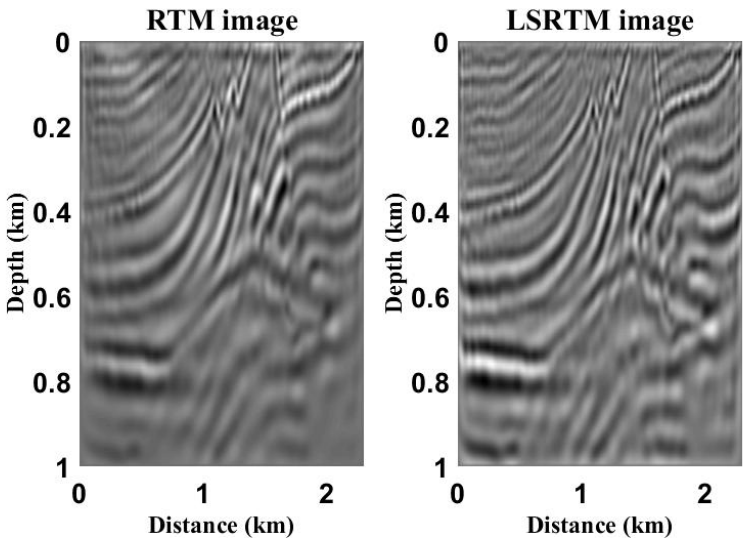

FIG. 6. RTM and LSRTM image using FWI result

SUMMARY

We have discussed differences between LSRTM and FWI. LSRTM uses a linearized wave equation based on the Born approximation, which allows one to use a linear inversion method with parameterized step size calculation. FWI uses the finite-difference method which is more precise than Born modeling because it produces first and multiple scattering waves, but it requires to use a non-linear inversion algorithm to correct for the model updates at each iteration. In LSRTM the high frequency components are emphasized to produce a model of reflectivities. FWI focuses on the low frequency components to correct for the background velocity model. In LSRTM the inversion result will be correct only if the background velocity model is accurate, but not otherwise. FWI will update the background velocity model if it can avoid local minima produced by cycle-skipping. If LSRTM can be implemented by using non-linear inversion with finite difference modeling, it may be possible to overcome its sensitivity to inaccuracies in the model.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for the support. This work was funded by CREWES and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. We also thank Bernie Law and Junxiao Li for their valuable discussions on this paper.

APPENDIX

Born Approximation

In the Born approximation, we assume the velocity of the earth model can be split into a smooth part v_0 and a singular part δv :

$$v(x) = v_0(x) + \delta v(x). \qquad \text{(A-1)}$$

$v_0(x)$ represents the long-wavelength smooth background velocity model and $\delta v(x)$ is the short-wavelength velocity model, which contains the singular features. This short-

wavelength component will produce reflections and contains high resolution features, which is the main object of the LSRTM.

In time domain, the first order wavefield perturbation δu corresponding to δv may be expressed in the form of the Green's function and the background velocity v_0. The Born modeling is formulated as followed:

$$(F[v_0]\delta v)(x_s, x_r, t) = \frac{\partial^2}{\partial t^2}\int dx d\tau G(x_s, x, \tau)\frac{2\delta v(x)}{v_0(x)^3}G(x, x_r, t-\tau) \qquad \text{(A-2)}$$

where $F[v_0]$ is the Born modelling operator, $G(x_s, x, \tau)$ and $G(x, x_r, t-\tau)$ is the source wavefield and receiver wavefield respectively. This equation gives us the wavefield perturbation δu by a simple linear relationship. Correspondingly, the inverse Born modelling operator is given by the $F[v_0]^*$, which is adjoint to $F[v_0]$ and applied to the wavefield perturbation δu:

$$(F^*[v_0]\delta d)(x) = \frac{2}{v_0(x)^3}\int dx_s dx_r dt d\tau G(x_s, x, \tau)$$
$$G(x, x_r, t-\tau)\frac{\partial^2}{\partial t^2}\delta d(x_s, x_r, t). \qquad \text{(A-3)}$$

In the frequency domain, the Green's function can also be separated into the background velocity component and perturbation as:

$$G(x, x_s, \omega|v) = G(x, x_s, \omega|v_0) + \delta G(x, x_s, \omega|v_0, \delta v), \qquad \text{(A-4)}$$

where the perturbation term can be expressed as:

$$\delta G(x, x_s, \omega|v_0, \delta v) = \omega^2\int dx' G\left(x, x', \omega\middle|v_0\right)\delta v\left(x'\right)G\left(x', x_s, \omega\middle|v\right)$$
$$= \omega^2\int dx' G\left(x, x', \omega\middle|v_0\right)\delta v\left(x'\right)G\left(x', x_s, \omega\middle|v_0\right)$$
$$+\ \omega^4\int dx' G\left(x, x', \omega\middle|v_0\right)\delta v\left(x'\right)\int dx'' G\left(x', x'', \omega\middle|v_0\right)\delta v\left(x''\right)G(x'', x_s, \omega|v_0)$$
$$+\cdots \qquad \text{(A-5)}$$

The Born approximation can be obtained by discarding the high order terms in equation (A-5) as

$$\delta G_{Born}(x, x_s, \omega|v_0, \delta v) = \omega^2\int dx' G\left(x, x', \omega\middle|v_0\right)\delta v\left(x'\right)G\left(x', x_s, \omega\middle|v_0\right), \qquad \text{(A-6)}$$

$$G_{Born}(x, x_s, \omega|v) = G(x, x_s, \omega|v_0)$$
$$+\omega^2\int dx' G\left(x, x', \omega\middle|v_0\right)\delta v\left(x'\right)G\left(x', x_s, \omega\middle|v_0\right). \qquad \text{(A-7)}$$

In this formulation, the nonlinear relationship between the model perturbation δv and the scattered wavefield $\delta G(x, x_s, \omega|v_0, \delta v)$ is linearized by the Born approximation, which is that the wavefield perturbation can be calculated by the integration between the Green's function and model perturbation δv. This approximation is often used in LSRTM iteratively to get the best δv in the inverse problem. In addition, the background and perturbation Green's function can be expressed in the following equation:

$$[\omega^2 v_0(x) + \nabla^2]G(x, x_s, \omega|v_0) = -\delta(x - x_s), \quad \text{(A-8)}$$

$$[\omega^2 v_0(x) + \nabla^2]\delta G_{Born}(x, x_s, \omega|v_0, \delta s) = -\omega^2 \delta v\left(x^{'}\right) G\left(x^{'}, x_s, \omega\middle|v_0\right). \quad \text{(A-9)}$$

REFERENCES

Chen, K. and Sacchi, M.D., 2017, Elastic least-squares reverse time migration via linearized elastic full waveform inversion with pseudo-Hessian preconditioning: Geophysics, **82**, No.5, 1-89

Dai, W., 2012, Multisource least-squares migration and prism wave reverse time migration: PhD thesis, Univ. of Utah

Dong, S., Cai, J., Guo, M., Suh, S., Zhang, Z., Wang, B. and Li, Z., 2012, Least-squares reverse time migration: towards true amplitude imaging and improving the resolution: SEG Expanded Abstracts, 1-5

Douma, H., Yingst, D., Vasconcelos, I. and Tromp, J., 2010, On the connection between artifact filtering in reverse-time migration and adjoint tomography: Geophysics, **75**, No.6, S219–S223

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

Geng, Y., and Innanen, K.A., 2016, Frequency domain nonlinear full-waveform inversion: CREWES Annual Report, **28**

Hou, J., 2016, Accelerating seismic imaging and velocity model building with approximate extended Born inversion: PhD thesis, Rice Univ.

Innanen, K.A., 2009, Born series forward modelling of seismic primary and multiple reflections: an inverse scattering shortcut: Geophysical Journal International, 177, No.3, 1197-1204

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

Luo, S. and Hale, D., 2014, Least-squares migration in the presence of velocity errors: Geophysics, **79**, No. 4, S153-S161

Margrave, G.F., Yedlin, M. and Innanen, K.A., 2011, Full waveform inversion and the inverse Hessian: CREWES Annual Report, **23**

Mulder, W.A. and Plessix, R.E., 2004, A comparison between one-way and two-way wave-equation migration: Geophysics, **69**, No.6, 1491-1504

Pan, W., 2017, Waveform inversion for estimating subsurface properties: phase-encoding strategies, optimization methods, interparameter tradeoffs quatification and reduction: PhD thesis, Univ. of Calgary

Pica, A., Diet, J.P. and Tarantola, A., 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, No.3, 284-292

Plessix, R.E. and Mulder, W.A., 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, No.3, 975-987

Rocha, D., Tanushev, N. and Sava, P., 2016, Acoustic wavefield imaging using the energy norm: Geophysics, **81**, No.4, S151–S163

Virieux, J. and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1-WCC26

Whitmore, N.D. and Crawley, S., 2012, Applications of RTM inverse scattering imaging conditions: SEG Expanded Abstracts, 1-6

Xu, L. and Sacchi, M.D., 2017, Preconditioned acoustic least-squares two-way wave equation migration with exact adjoint operator: Geophysics, 83, No.1, 1-49.

Yang, P., Gao, J. and Wang, B., 2015, A graphics processing unit implementation of time-domain full-waveform inversion: Geophysics, **80**, No.3, F31-F39

Yao, G., 2013, Least-squares reverse time migration: PhD thesis, Imperial College London