

# Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion

Zedong Wu<sup>1</sup> and Tariq Alkhalifah<sup>1</sup>

## ABSTRACT

The gradient of standard full-waveform inversion (FWI) attempts to map the residuals in the data to perturbations in the model. Such perturbations may include smooth background updates from the transmission components and high wavenumber updates from the reflection components. However, if we fix the reflection components using imaging, the gradient of what is referred to as *reflected-waveform inversion* (RWI) admits mainly transmission background-type updates. The drawback of existing RWI methods is that they lack an optimal image capable of producing reflections within the convex region of the optimization. Because the influence of velocity on the data was given mainly by its background (propagator) and perturbed (reflectivity) components, we have optimized both components simultaneously using a modified objective function. Specifically, we used an objective function that combined the data generated from a source using the background velocity, and that

by the perturbed velocity through Born modeling, to fit the observed data. When the initial velocity was smooth, the data modeled from the source using the background velocity will mainly be reflection free, and most of the reflections were obtained from the image (perturbed velocity). As the background velocity becomes more accurate and can produce reflections, the role of the image will slowly diminish, and the update will be dominated by the standard FWI gradient to obtain high resolution. Because the objective function was quadratic with respect to the image, the inversion for the image was fast. To update the background velocity smoothly, we have combined different components of the gradient linearly through solving a small optimization problem. Application to the Marmousi model found that this method converged starting with a linearly increasing velocity, and with data free of frequencies below 4 Hz. Application to the 2014 Chevron Gulf of Mexico imaging challenge data set demonstrated the potential of the proposed method.

## INTRODUCTION

Conventional full-waveform inversion (FWI) is based on minimizing the misfit between the predicted wavefield and the observed one (Virieux and Operto, 2009) at the sensor locations. After recognizing that the calculation of the gradient is of the same computational order as the forward modeling, and that the gradient is closely related to the seismic migration (Lailly, 1983; Tarantola, 1984), it became practical to finely invert for parameterized models, enabling high-resolution inversion results when the initial velocity is close enough to the exact one. However, the conventional objective function based on the least-squares misfit is not convex and has a lot of local minima, especially with high-frequency data and reflections. Thus, to converge to an acceptable result, we need a good starting

velocity model that is able to produce events within a half-cycle of the observed ones (Virieux and Operto, 2009). We pursue such models using migration-based traveltimes tomography (Clément et al., 2001) or migration velocity analysis (MVA) methods (Biondi and Symes, 2004; Sava et al., 2005; Symes, 2008). This requirement comes from the separation of the velocity to those components responsible for scattering and those responsible for wave propagation. To allow both components to update the velocity in FWI, Symes (2008) introduces an extended velocity model to include a nonphysical dimension such as subsurface offset. Biondi and Almomin (2014) use a similar idea to introduce an extended velocity along the time axis. The extension, however, has deemed the approach impractical, which forces Almomin and Biondi (2013) to do the extension only to the perturbed components of the velocity

Manuscript received by the Editor 8 August 2014; revised manuscript received 14 June 2015; published online 2 September 2015.

<sup>1</sup>King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering Division, Thuwal, Saudi Arabia. E-mail: zedong.wu@kaust.edu.sa; tariq.alkhalifah@kaust.edu.sa.

© 2015 Society of Exploration Geophysicists. All rights reserved.

model. Thus, they invert the extended image in a nested fashion. Albertin et al. (2013) propose a similar idea through a gradient orthogonalization to update the smooth parts of the velocity.

Xu et al. (2012a) and Zhou et al. (2012) develop a method based mainly on the work of Plessix et al. (1995) to invert for smooth velocity models using modeled reflected energy from an image. They refer to the method as *reflected-waveform inversion* (RWI). The idea is based on the migration followed by demigration to predict the reflections. Because the demigration is obtained from the image, for an imperfect velocity, the modeled data from the image will have residuals, hopefully at far offsets. As a result, RWI inverts mainly the propagator (smooth) part of the model, like MVA, without the need for extended images or angle gathers. Wang et al. (2013) implement the same approach in the frequency domain to use a sequential frequency implementation, which they think is necessary to avoid the nonlinearity caused by an incorrect image.

Zhou et al. (2015) propose a similar idea to invert the velocity and impedance. Wu and Alkhalifah (2014c) implement RWI with the spectral method as a wavefield extrapolator (Wu and Alkhalifah, 2014b). They find a new term that has the opposite direction of the gradient of standard RWI, and they linearly combine the new term with the original terms to update the velocity with the smoothest function possible while guaranteeing the ascending direction of the approximate gradient. A similar idea is used by Wu and Alkhalifah (2014a) for combining the tomographic and reflected part of the standard FWI gradient. However, a drawback of the RWI method is the need of a true amplitude migration (image) in each iteration. From another prospective, standard RWI tries to approximate the observed data with that generated by Born modeling. The diving wave, a major source of background velocity information, is considered as noise and neglected (Alkhalifah and Choi, 2012; Tang et al., 2013).

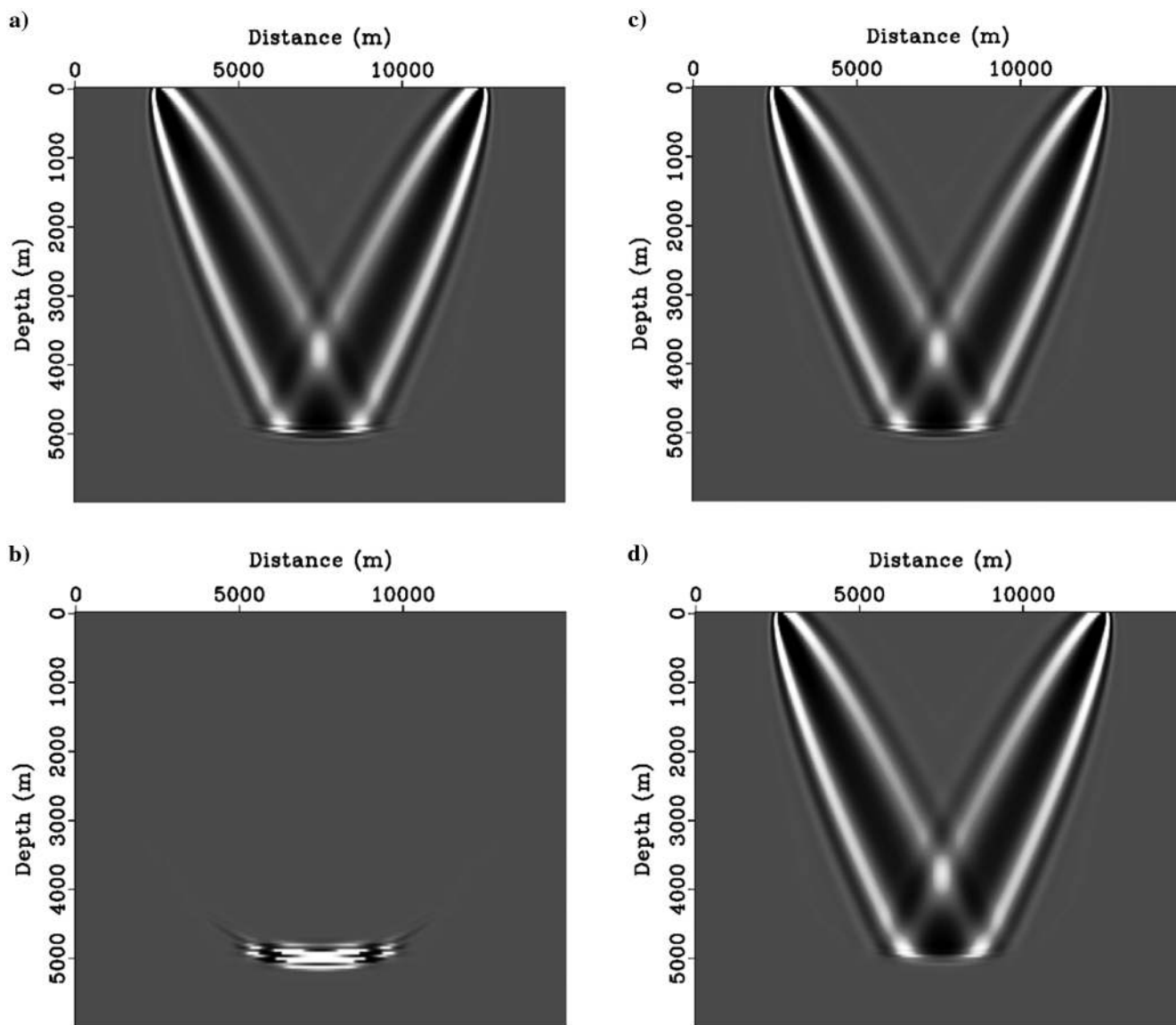


Figure 1. (a) Part II + IV of the gradient, (b) part III of the gradient, (c) RWI part of the gradient (II + III + IV), and (d) the optimized summation gradient (II + IV + 3.905III).

In this article, we present a method that does not require computing the true amplitude image as part of RWI because the image itself is a part of the inversion. Thus, we formulate an optimization problem that will invert for the background velocity  $v$  and perturbed velocity  $w$  simultaneously. We use a new objective function introduced by Alkhalifah and Wu (2014), which allows us to mix RWI and FWI in a natural way. Finally, we test our approach on the Marmousi velocity model using data without low frequencies, starting with a linearly increasing velocity model. The application to the 2014 Chevron Gulf of Mexico imaging challenge synthetic data set shows that the method can provide a reasonably good inversion result.

### OPTIMIZATION PROBLEM

Wave propagation in isotropic acoustic media with constant density is governed by the following wave equation:

$$\frac{\partial^2 p}{\partial t^2} - v^2(\mathbf{x})\Delta p = 0, \quad (1)$$

where  $p(\mathbf{x}, t)$  is the pressure wavefield described in a given domain with space coordinates given by  $\mathbf{x} = \{x, y, z\}$ ,  $t$  is the time,  $v$  is the velocity, and  $\Delta$  is the Laplacian operator.

Here, we use the optimized expansion-based low-rank method (Wu and Alkhalifah, 2014b) to solve equation 1 with the initial and boundary conditions. We give a short description of the method in Appendix A. Let us denote the discretized wavefield as  $\mathbf{p} = (p_0; p_1; \dots; p_n)$ , where  $p_n = p(x, t_n)$ , then the discrete formulation can be written in matrix form as

$$\mathbf{A}\mathbf{p} = \mathbf{B}(v)\mathbf{p} + \mathbf{f}, \quad (2)$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ \dots & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}, \quad (3)$$

$$\mathbf{B}(v) = \begin{pmatrix} 0 & b(v) & 0 & \dots & 0 & 0 \\ 0 & 0 & b(v) & \dots & 0 & 0 \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & 0 & b(v) \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (4)$$

and

$$b(v)p_m = \sum_{i=1}^M \left( \sum_{j=1}^N a_{i,j} 2 \cos(vk_j \Delta t) \right) \times \int \hat{p}_m(\mathbf{k}, t) 2 \cos(v_i k \Delta t) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k}. \quad (5)$$

The standard FWI minimizes the misfit between the predicted and observed wavefields. Such an optimization using the least-squares norm is given by

$$\min_{v_{\text{all}}} J(\mathbf{p}_{\text{all}}) = \frac{1}{2} \|\mathbf{C}\mathbf{p}_{\text{all}} - \mathbf{g}\|_2^2 \quad (6)$$

$$\text{s.t. } \mathbf{A}\mathbf{p}_{\text{all}} = \mathbf{B}(v_{\text{all}})\mathbf{p}_{\text{all}} + \mathbf{f}, \quad (7)$$

where  $\mathbf{g}$  is the recorded data at the receivers,  $\mathbf{C}$  is the operator that projects the data onto the receiver positions, and  $\mathbf{f}$  is the source wavelet. For simplicity, we use the two-norm, which is defined as  $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ . Alternatively, we can use other objective functions, such as the inner product (Choi and Alkhalifah, 2012; Wang et al., 2012), the Huber norm (Guitten and Symes, 2003), and the envelope function (Bozda et al., 2011; Wu et al., 2014).

Similar to Plessix et al. (1995), Clément et al. (2001), Xu et al. (2012a), and Biondi and Almomin (2014), we split the velocity into a mainly background component and a perturbed component. Let us assume that  $v_{\text{all}} = v + w$ , where  $v$  is the background component and  $w$  is the perturbed component. We also define the corresponding wavefield as  $\mathbf{p}_{\text{all}} = \mathbf{p} + \mathbf{q}$ , where  $\mathbf{p}$  is the wavefield modeled from the source using the background velocity and  $\mathbf{q}$  is the wavefield modeled from the perturbation  $w$ , and  $\mathbf{p}_{\text{all}}$  is the total wavefield. Thus,  $\mathbf{p}_{\text{all}}$  satisfies the wave equation for  $v + w$ , given by

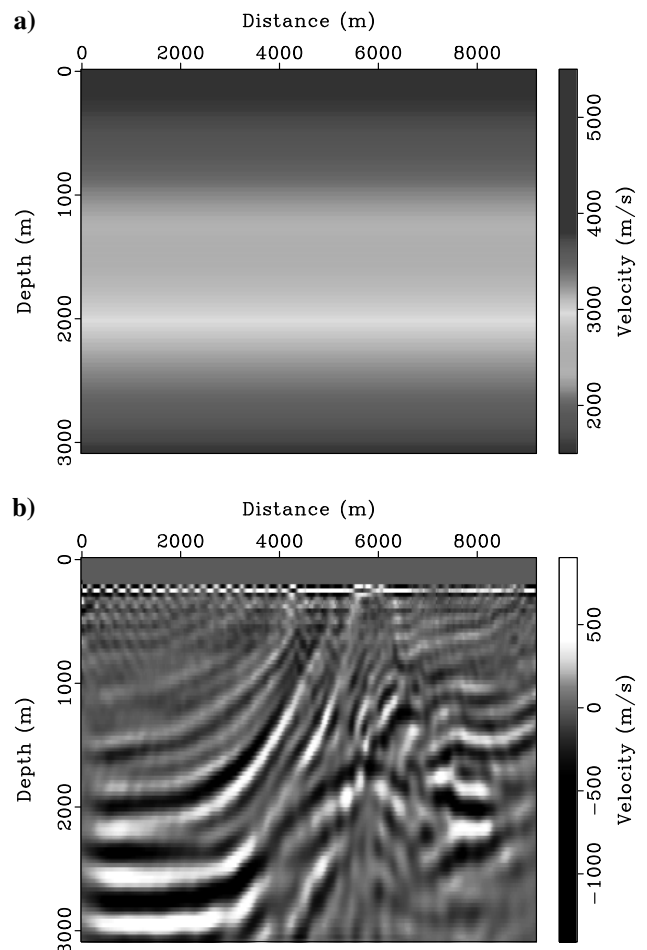


Figure 2. (a) The initial velocity and (b) the inverted  $w$  after 10 iterations with  $v$  fixed by the initial velocity.

$$\mathbf{A}\mathbf{p}_{\text{all}} = \mathbf{B}(v+w)\mathbf{p}_{\text{all}} + \mathbf{f}. \quad (8)$$

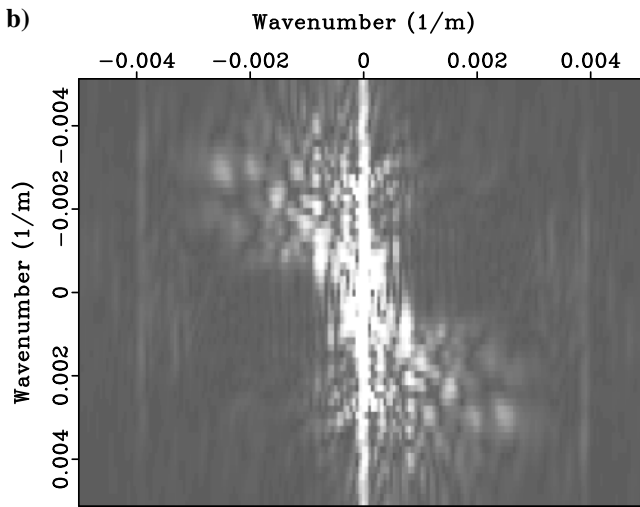
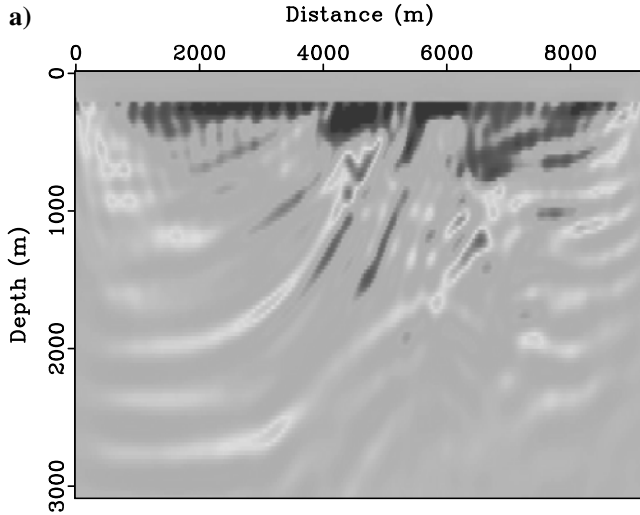
Applying the Born approximation, and specifically, substituting  $\mathbf{p}_{\text{all}} = \mathbf{p} + \mathbf{q}$  yields

$$\mathbf{A}\mathbf{q} - \mathbf{B}(v)\mathbf{q} \approx \mathbf{B}'(v)\mathbf{p} \cdot^* w, \quad (9)$$

where  $\mathbf{B}'(v)$  is defined as

$$\mathbf{B}'(v) = \begin{pmatrix} 0 & b'(v) & 0 & \cdots & 0 & 0 \\ 0 & 0 & b'(v) & \cdots & 0 & 0 \\ \cdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & b'(v) \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \quad (10)$$

and



$$b'(v)p = - \sum_{i=1}^M \left( \sum_{j=1}^N a_{ij} 2 \sin(|\mathbf{k}_j|v\Delta t) |\mathbf{k}_j| \Delta t \right) \times \int \hat{p}(\mathbf{k}, t) 2 \cos(|\mathbf{k}|v_i \Delta t) e^{i\mathbf{k} \cdot \mathbf{x}} d\mathbf{k}. \quad (11)$$

We define the inner product operator here as

$$(h \cdot^* m)(x, t) = h(x, t)m(x), \quad \forall x, t, \quad (12)$$

for the space time function  $h(x, t)$  and space function  $m(x)$ . Instead of solving the optimization problem 6 and 7, we use the objective function introduced in Alkhalifah and Wu (2014), given by

$$\min_{v,w} J(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g}\|_2^2 \quad (13)$$

$$\text{s.t. } \mathbf{A}\mathbf{p} = \mathbf{B}(v)\mathbf{p} + \mathbf{f}, \mathbf{A}\mathbf{q} - \mathbf{B}(v)\mathbf{q} = \mathbf{B}'(v)\mathbf{p} \cdot^* w. \quad (14)$$

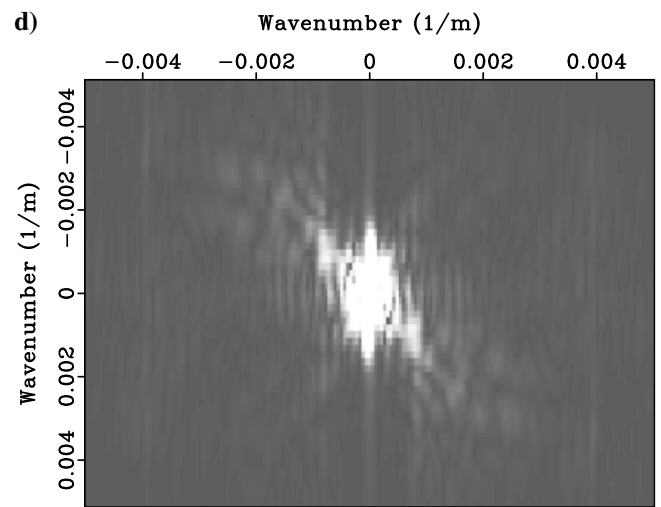
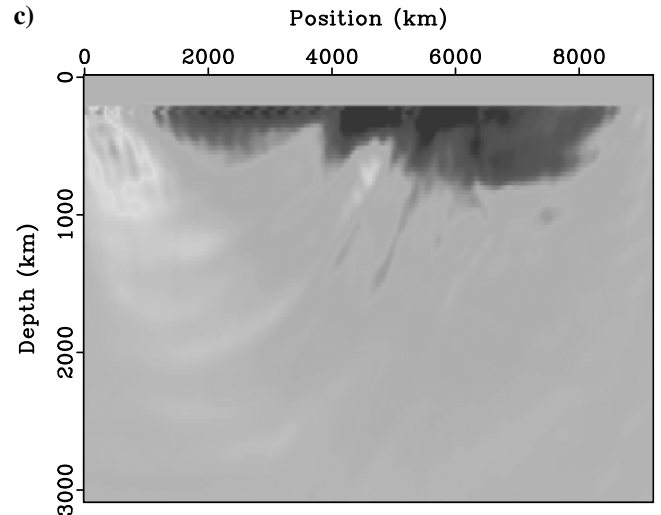


Figure 3. (a) The gradient of standard FWI and (b) its spectral distribution. (c) The gradient of the modified objective function with the improved  $w$  and (d) its spectral distribution.

**Algorithm 1. Reflected-waveform inversion algorithm.**

**Input:** Initial velocity  $v_0$ ,  $\alpha$ ,  $\epsilon$ .

**Output:** Inverted  $v_n$ ,  $w_n$ , and Error.

**function** GRADIENT\_CALCULATION  $v, w$

Calculate the different components I, II, III, IV, and obtain the Error.

Solve the optimization problem 16 to obtain the best  $\beta$ .

Calculate the approximate gradient with respect to  $v$  ( $\nabla_v J$ ) as  $\alpha(\beta\text{III} + \text{I} + \text{II} + \text{IV})$ .

Calculate the gradient with respect to  $w$  ( $\nabla_w J$ ) as I.

Return Error,  $\nabla_v J$ , and  $\nabla_w J$ .

**end function**

**function** OPTIMIZATION  $v_0$

Initialize  $w = 0$ ,  $n = 0$ .

**while** the exit condition is not satisfied **do**

[Error,  $\nabla_v J$ ,  $\nabla_w J$ ] = Gradient Calculation ( $v_n, w_n$ );

update  $v_n, w_n$  as  $v_{n+1}$  and  $w_{n+1}$ .

**end while**

**end function**

In addition, instead of optimizing only  $v$ , which is done in standard FWI and RWI, we optimize the  $v$  and  $w$  simultaneously. Although we use the Born approximation in deriving the above formulation, it is straightforward to prove that the above objective function  $J(\mathbf{p}, \mathbf{q})$  has its minima at  $w = 0, v = v_{\text{exact}}$ , which is the same as standard FWI. If we solve the problem in the subspace  $v = v_0$ , it reduces to a simple least-squares migration at the current velocity. However, if we solve the above problem in the subspace  $w = 0$ , it reduces to the standard FWI method. However, we keep  $v$  and  $w$  in the optimization to reduce the nonlinearity of the objective function. In addition, the objective function is quadratic with respect to the fine-scale part of the velocity  $w$ . Thus, the behavior of this objective function is expected to be better than standard FWI. If we take  $w$  to equal the reverse time migration image, it will reduce to the standard RWI (Plessix et al., 1995; Xu et al., 2012a, 2012b; Zhou et al., 2012; Wang et al., 2013). Our proposed approach has a least-squares inversion of the image component that helps to provide the optimal image we need for RWI. With the modified objective function, unlike classic RWI, our proposed approach uses diving wave information, which is important for inverting for the background velocity in the shallow zone (Alkhalifah and Choi, 2012).

Although the derivation above is based on the spectral wavefield extrapolation approach (Wu and Alkhalifah, 2014b), the proposed method is valid for a conventional finite-difference implementation

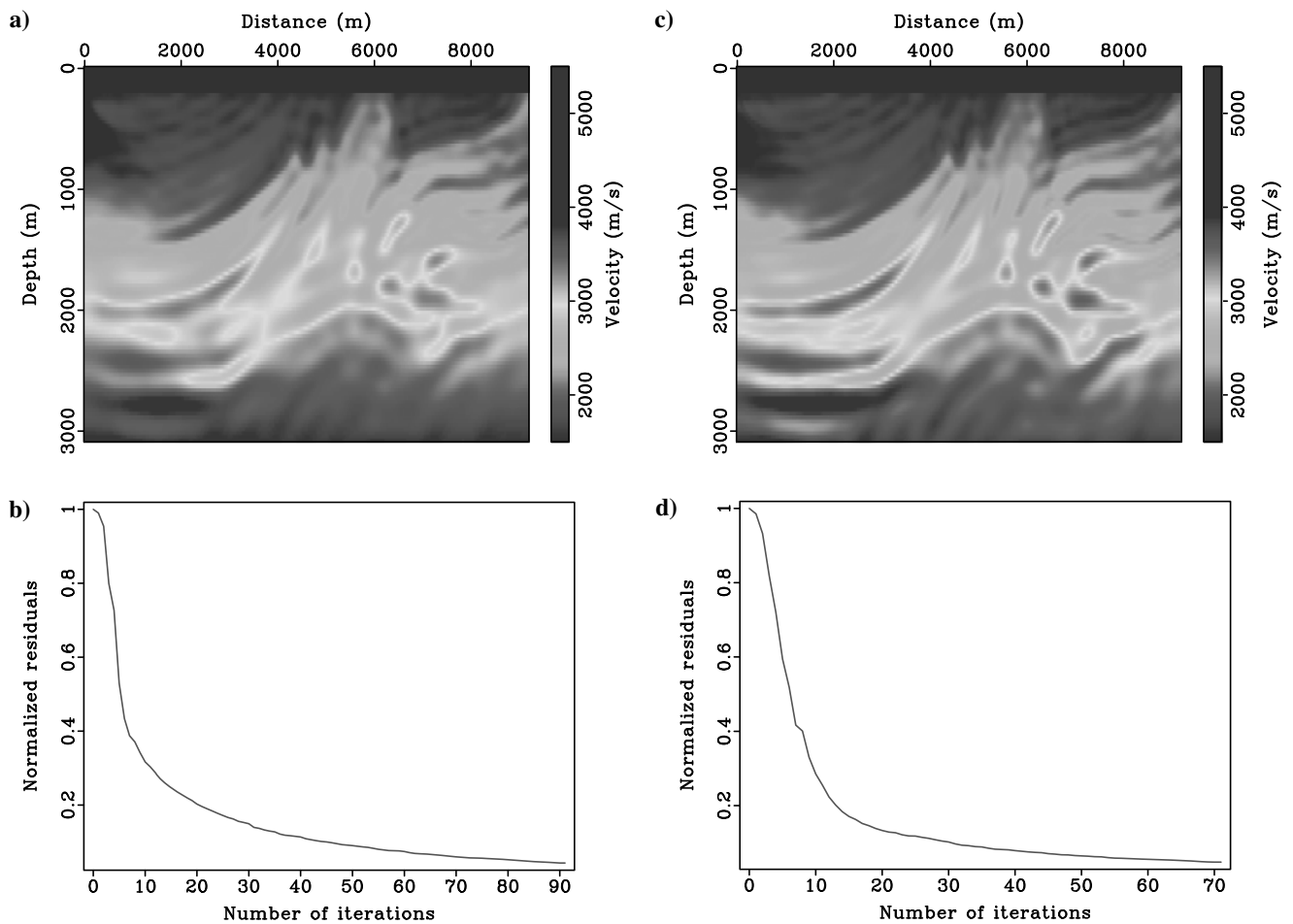


Figure 4. (a) Inverted  $v$  with  $\alpha = 1/20$ , (b) the history of convergence, (c) inverted  $v$  with  $\alpha = 1/10$ , and (d) the history of convergence.

as well. In this situation, the operator  $\mathbf{B}(v)\mathbf{p} = v^2\Delta p$  is the discrete stiffness matrix and  $\mathbf{A}$  is the discrete mass matrix.

### GRADIENT ANALYSIS AND OPTIMIZATION

After obtaining the objective function, we can use the adjoint state method (Plessix, 2006) to derive the gradient with respect to  $v$  and  $w$  as

$$\frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial v} = \text{I} + \text{II} + \text{III} + \text{IV}, \quad \frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial w} = \text{I}, \quad (15)$$

in which

$$\text{I} = \langle \lambda, \mathbf{B}'(v)\mathbf{p} \rangle_t, \quad \text{II} = \langle \lambda, \mathbf{B}'(v)\mathbf{q} \rangle_t, \quad (16)$$

$$\text{III} = \langle \lambda, \mathbf{B}'(v)\mathbf{p} \cdot^* w \rangle_t, \quad \text{IV} = \langle \mu, \mathbf{B}'(v)\mathbf{p} \rangle_t, \quad (17)$$

and  $\lambda = (\mathbf{A} - \mathbf{B}(v))^{-T}(\mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g})$  is the wavefield propagating backward using the residual at the receiver,  $\mu = (\mathbf{A} - \mathbf{B}(v))^{-T}(\mathbf{B}^T(\lambda \cdot w))$  is the wavefield propagating backward using  $\mathbf{B}^T(\lambda \cdot w)$  as the source. The detailed derivation is shown in Appendix B. In the calculation of the above gradient, we need four forward-modeling operations for each gradient calculation. The cost is about double that of standard FWI.

In the above gradient, I is just the gradient of the standard FWI with the residual given by  $\mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g}$  and II + IV is similar to the gradient of the standard RWI (Xu et al., 2012a, 2012b; Zhou et al., 2012; Wang et al., 2013). The III is an additional term. Let us have a look at what this additional term looks like for a simple two-layer problem. We consider a model with a velocity of 2000 m/s in the top layer and 3000 m/s in the bottom layer. The grid interval is 40 m. The interface between the two layers is at 5000-m depth. The source is located at 2500 m, and the receiver is located at 12,500 m on the surface. The term II + IV, shown in Figure 1a, is the smooth update along the reflection raypaths. However, this additional term in the gradient (Figure 1a) has a singular (nonsmooth) component corresponding to the reflection at the image point. Thus, the com-

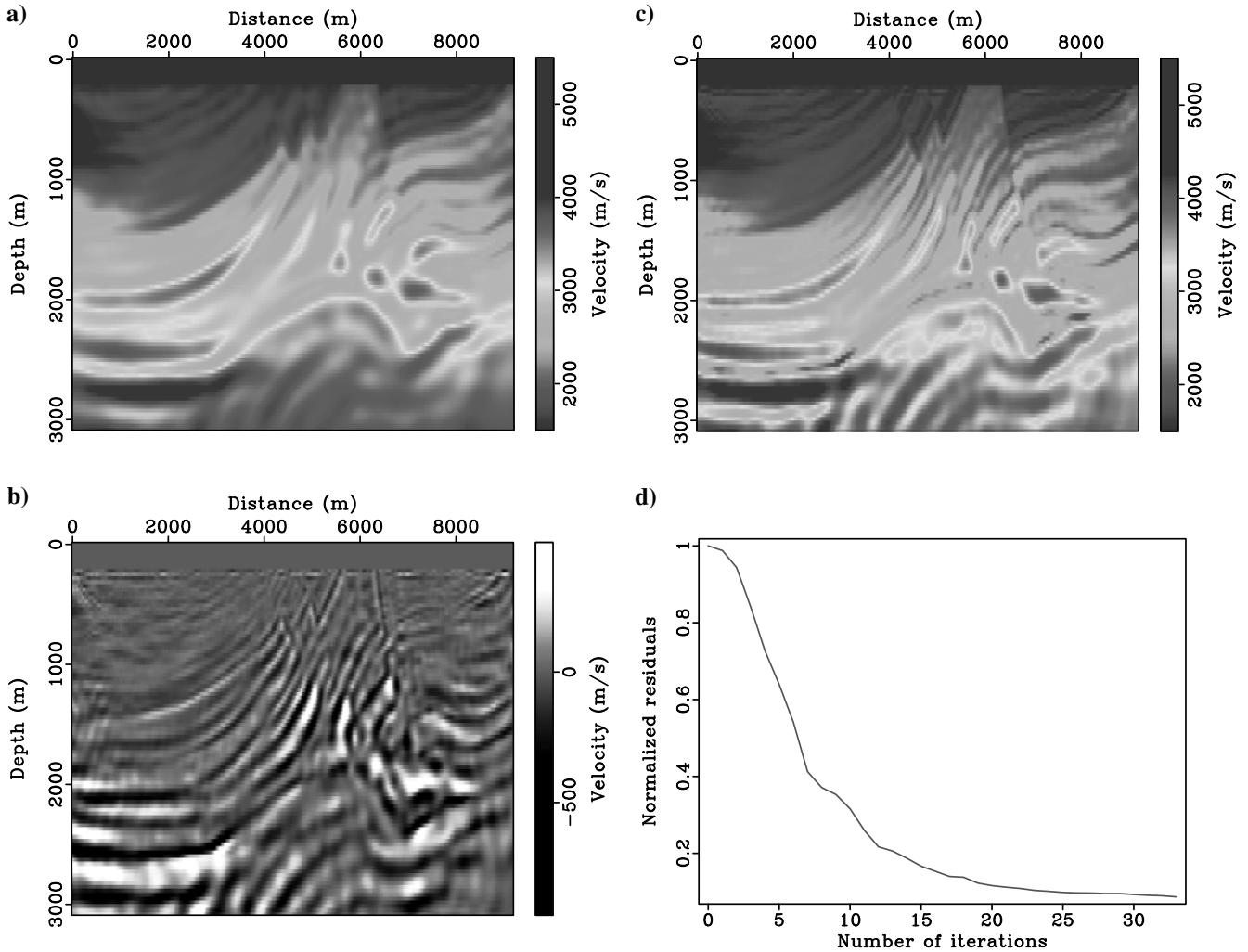


Figure 5. (a) Inverted  $v$ , (b)  $w$ , (c) velocity  $v + w$  with  $\alpha = 1/5$ , and (d) the history of convergence.

bined RWI gradient updates along the wavepath, but also at the reflector, which may cause nonlinearity. Figure 1b shows the third term in our derivation, which looks similar to the singular component in Figure 1a except with an opposite sign. However, the amplitude of term III is smaller than II + IV. If we directly sum the two together, the result is shown in Figure 1c, which still has singularity at the image points. Alternatively, we combine them with a weight,  $\beta$ , applied to the singular component to obtain a smoother gradient. We do that by solving the following optimization problem:

$$\min_{\beta} \|\beta \text{III} + \text{I} + \text{II} + \text{IV}\|_s \quad (18)$$

$$\text{s.t. } \|(1 - \beta)\text{III}\|_{\rho^2} \leq \epsilon \|\text{I} + \text{II} + \text{III} + \text{IV}\|_{\rho^2}, \quad (19)$$

where  $\|\cdot\|_s$  is the norm, which describes the smoothness of the function. We use the following expression:

$$\|u\|_s = \|\Delta u\|_{\rho^2} = \sqrt{\left\| \frac{\partial^2 u}{\partial x^2} \right\|_{\rho^2}^2 + \left\| \frac{\partial^2 u}{\partial z^2} \right\|_{\rho^2}^2} \quad (20)$$

as the description of smoothness. After optimizing the coefficients, we obtain the new approximate gradient as shown in Figure 1d. We see that this gradient has less singularity than Figure 1a and 1c. In our implementation, we always choose  $\epsilon = 1000$  to make sure the approximate gradient is smooth enough.

Let us investigate the gradient behavior for different perturbations  $w$  for the Marmousi model, the detail of which will be described in the next section. The initial velocity shown in Figure 2a is linearly increasing with depth from 1500 to 4000 m/s. We first fix  $v$  and optimize  $w$ . After 10 iterations, the resulting  $w$  is shown in Figure 2b. Using this improved  $w$ , we can get the gradient with respect to  $v$ , which is shown in Figure 3c. Figure 3a shows the gradient for the standard FWI. We can see that Figure 3c has more low-wavenumber information than the standard FWI gradient (Figure 3a). To emphasize this aspect, we Fourier transform both gradients to the wavenumber domain as shown in Figure 3b and 3d. It can be seen that the energy in Figure 3d focuses more on the low-wavenumber components than that in Figure 3b.

About solving the optimization problem, we can use the nested optimization method (Biondi and Almomin, 2014), which optimizes (an extended version of)  $w$  in an inner loop at each iteration, but it is relatively expensive because we need to solve a least-squares migration at every FWI iteration. Note that if  $w = 0$ , the gradient is just the standard FWI gradient, and the method becomes standard FWI. However, to fully use the features of this method and update the background velocity smoothly, we can update  $w$  with larger weights and background  $v$  with smaller weights, through adding a weight function between the gradient with respect to  $v$  and  $w$ . We suggest that after calculating the gradient  $\nabla J = ((\partial J/\partial v), (\partial J/\partial w))$ , we add a scale factor  $\alpha$  and use

$$\nabla_{\alpha} J = \left( \alpha \frac{\partial J}{\partial v}, \frac{\partial J}{\partial w} \right), \quad (21)$$

as our gradient. We first choose  $\alpha$  as a small value and focus on updating the background along the wavepath of the diving and reflected wave, and then increase it to gradually add the reflector to the background. If we choose  $\alpha = 0$ , it is just the least-squares

migration at the current velocity. Thus, convergence is obtained when  $v$  approaches its true value and  $w$  approaches zero. This translates to the fact that the velocity model managed to provide the reflections that fit the data accurately, and thus,  $w$  is not needed to

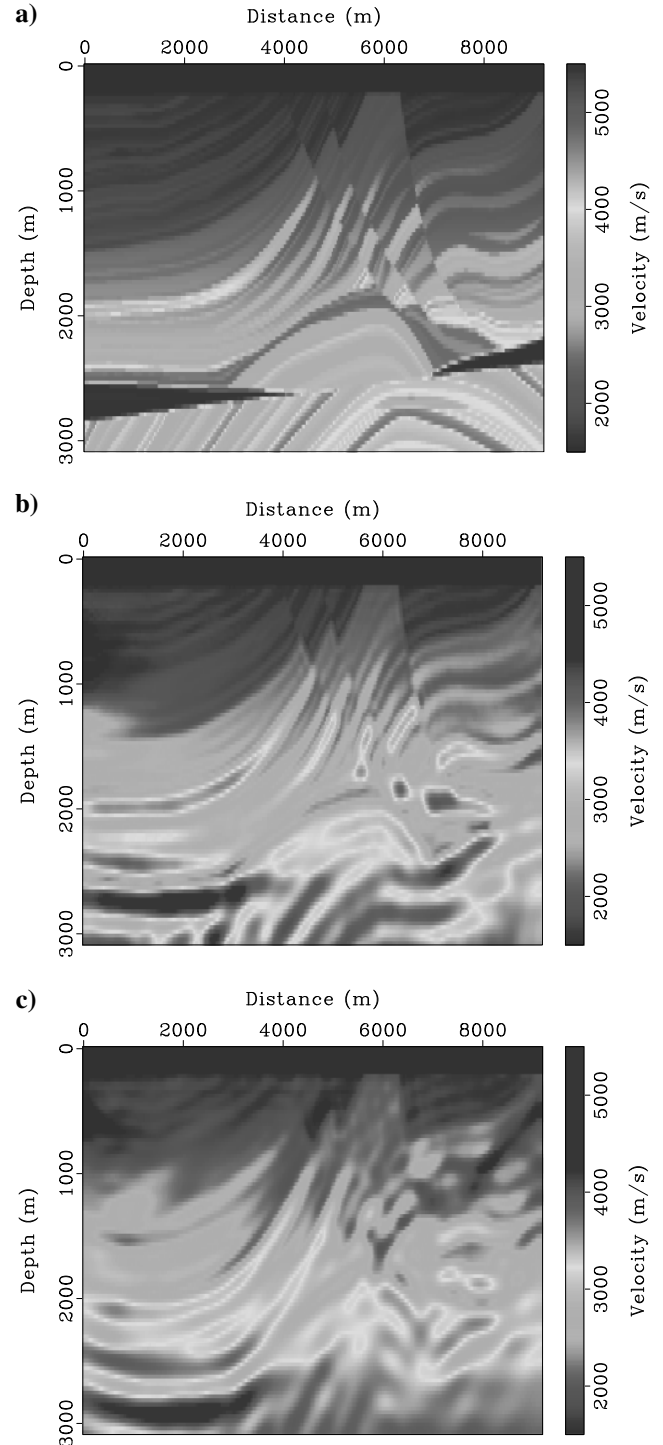


Figure 6. (a) The Marmousi velocity model, (b) the inverted velocity using standard FWI starting with the velocity in Figure 5a, and (c) the inverted velocity using standard FWI with the linear initial velocity model.

produce these reflections. Based on the above, we summarize in Algorithm 1.

### NUMERICAL EXAMPLES

Let us apply this method to the Marmousi model. We add a thin water layer near the surface. The wavefields are extrapolated on a mesh size of  $32 \times 32$  m using a Ricker wavelet with the main frequency of 5 Hz after filtering out frequencies below 4 Hz. The shot interval is 256 m. During the inversion procedure, we do not update  $v$ ,  $w$  in the water layer and we set  $v$  to the exact water velocity and  $w = 0$ . We first invert using  $\alpha = 1/20$ . The inverted  $v$  is shown in Figure 4a, and the misfit progression is shown in Figure 4b. After this inversion using  $\alpha = 1/20$ , we use the result in Figure 4a as the initial model and reset  $w = 0$  for the next stage in which we set  $\alpha = 1/10$ . The result of this stage is shown in Figure 4c with the convergence measure shown in Figure 4d. Finally, we implement the optimization using  $\alpha = 1/5$  with the result of  $\alpha = 1/10$  as the initial model. The inverted  $v$ ,  $w$ ,  $v + w$  with  $\alpha = 1/5$  are shown

in Figure 5a–5c, respectively. The convergence history is shown in Figure 5d. Because the result of  $\alpha = 1$  is similar to  $\alpha = 1/5$ , we do not show it here. For a comparison, we use the inverted  $v$  shown in Figure 5a as our initial velocity to do a standard FWI to get the inverted velocity as shown in Figure 6b. Because the background velocity is accurate enough, the difference between Figure 6b and 5c is small. For comparison, the exact velocity model is shown in Figure 6a and the inverted result of standard FWI with the same linear increasing initial velocity is shown in Figure 6c. Because of the nonlinearity (or cycle-skipping problem), the standard FWI fell into a local minimum. In addition, we show the observed data (Figure 7a), the predicted data (Figure 7b) with initial velocity, residual with initial (Figure 7c) and inverted velocity (Figure 7d) for a shot at 4608 m at the same scale. We can see that the residual of the inverted velocity is much smaller than that of the initial one. From Figures 4a, 4c, and 5a, we can see that as we increase  $\alpha$ , the procedure is adding more and more details to the background velocity model. We still have residuals ( $w \neq 0$ ), especially in the deeper part, and those correspond to our inability to attain the

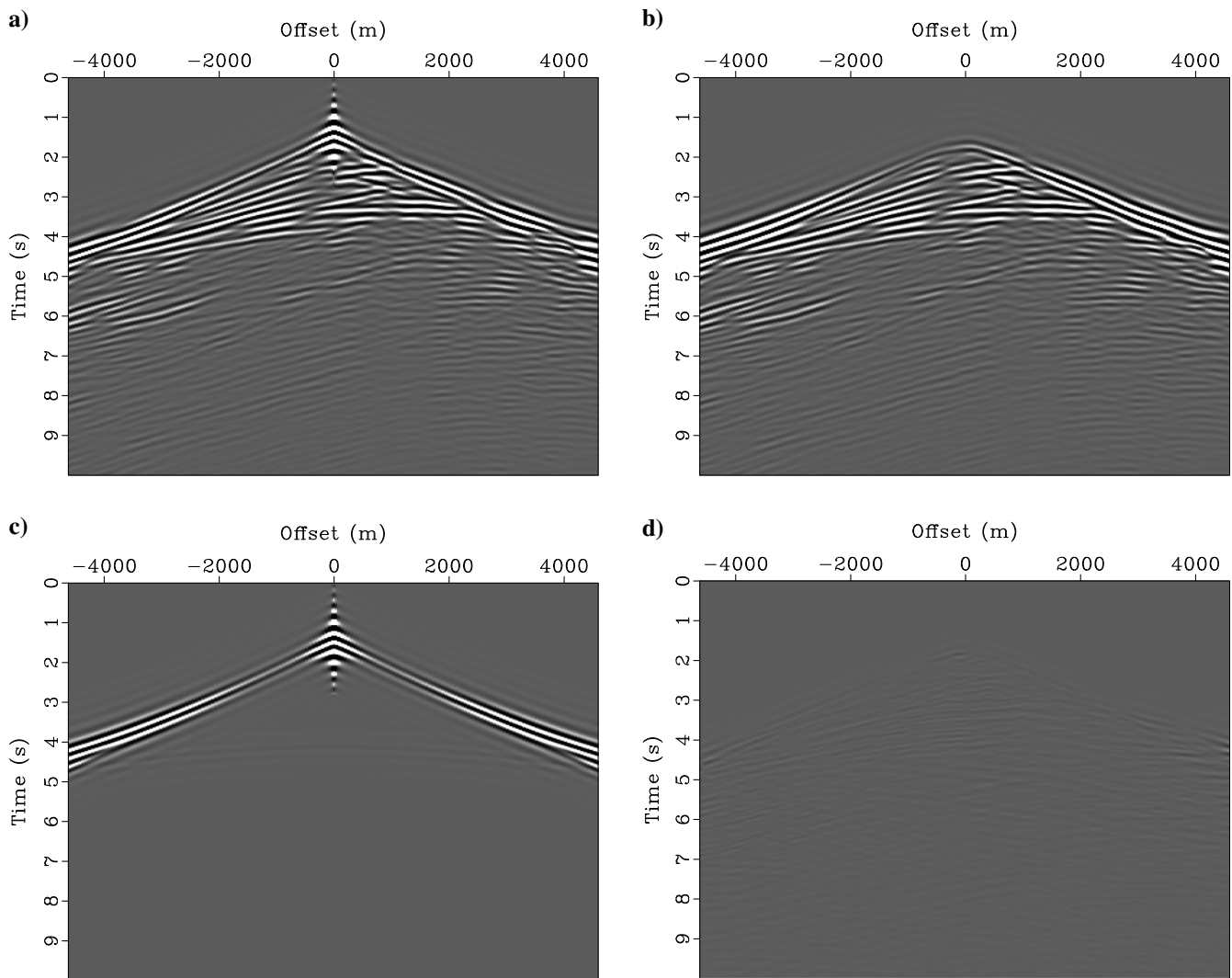


Figure 7. (a) The exact recorded data for the shot at 4608 m, (b) the predicted data with the initial velocity, (c) the residual with the initial velocity, and (d) the residual with the inverted velocity.



necessary model wavenumbers from the data to fully recover the velocity at that depth. Still,  $w$  is the perturbations, which can be added directly to the model. The perturbations, like images, will have a low-resolution representation of the model. However, because we are

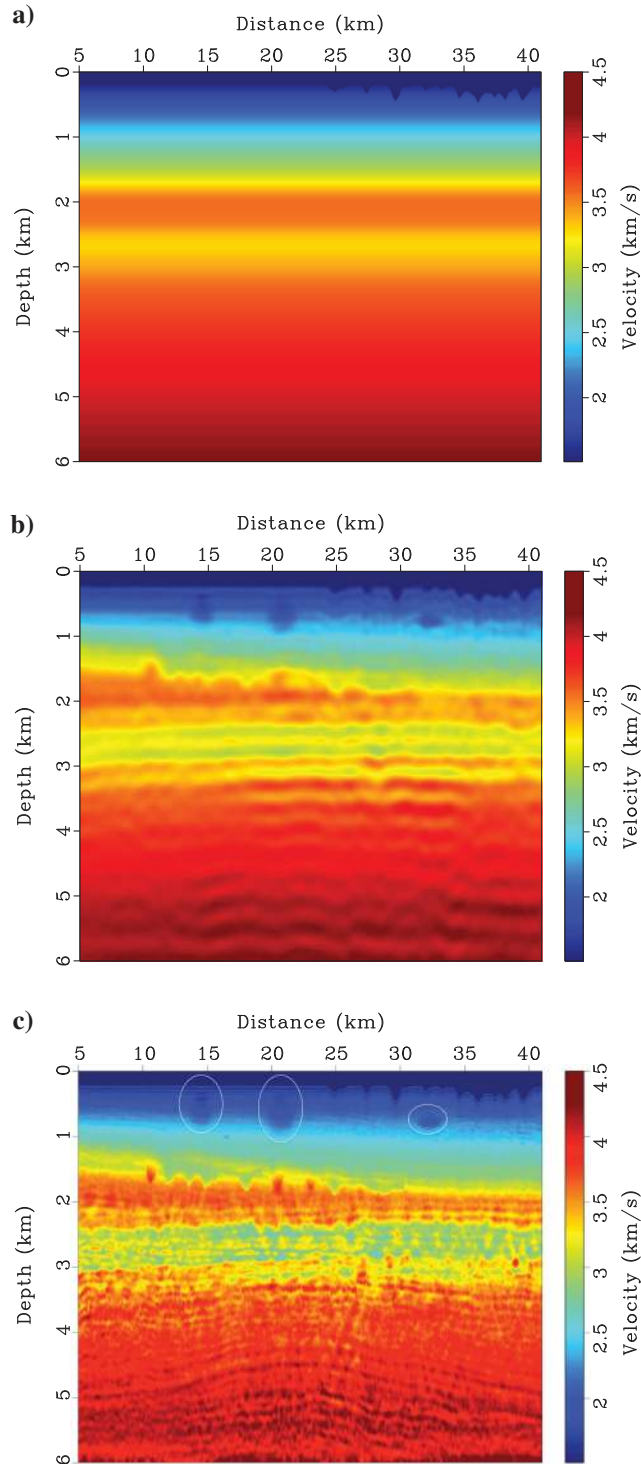


Figure 8. (a) Initial velocity, (b) the inverted  $v$  with the low-frequency data using our proposed method, and (c) the inverted  $v$  after sequential standard FWI with the initial velocity as the inverted  $v$  from the inverted velocity shown in panel (b).

using the RWI and FWI in the implementation, this limitation is physically caused by the acquisition configuration of our data.

Our next example is the 2014 Chevron Gulf of Mexico imaging challenge synthetic data set. It was distributed as a blind test to apply FWI. The data set includes 1600 shots records with 25-m shot sampling at the depth of 15 m. Each shot has 321 receivers with 25-m sampling at a depth of 15 m. The data represent a marine acquisition, and they are generated with an isotropic elastic wave equation. The maximum recording time is 8 s. The data have a low signal-to-noise ratio (S/N) below 3 Hz and strong noise even in the 3–5 Hz range. To increase the S/N and reduce the boundary effect near the source, we convert the data into 641 receivers located on both sides of each shot, relying on the reciprocity principle. To reduce the computational cost associated with the inversion, we use an approach suggested by Díaz and Guitton (2011) and Reker et al. (2014), and we divide the data set into eight groups. We then invert one group after another. We continue until we have covered all the groups, and then we move to a new  $\alpha$ . An initial velocity was provided and is shown in Figure 8a. We apply the proposed inversion method with  $30 \times 30$ -m space sampling and 4-ms time sampling. We use the method proposed by Hicks (2002) to address the fact that shots and receivers are not on the grid points. To avoid the large noise at low frequencies, we apply the proposed approach on the frequency bands of 3–3.2, 3–3.6, 3–4.2, and 3–5.2 Hz, sequentially, with the same  $\alpha$  strategy mentioned in the Marmousi data inversion. The frequency strategy is based on the approach suggested by Sirgue and Pratt (2004). For each frequency band, standard FWI is applied following the proposed method. In addition, we apply a Gaussian smoothing to the gradient with respect to  $v$  to reduce the artifact caused by noise and multiple and elastic effects. The inverted  $v$  is shown in Figure 8b. To speed up the inversion at the later stage, we apply standard FWI (without preconditioning and with preconditioning) on the frequency bands of 3–6.7, 3–8.7, 3–11.3, 3–14.6, 3–18.9, and 3–21 Hz, sequentially, starting from the inverted  $v$  from our proposed method. The preconditioned gradient of the standard FWI is taken to be the multiplication of the original gradient by  $v^2$ . The preconditioned FWI tries to enhance the deeper part of the velocity model. The final inverted velocity

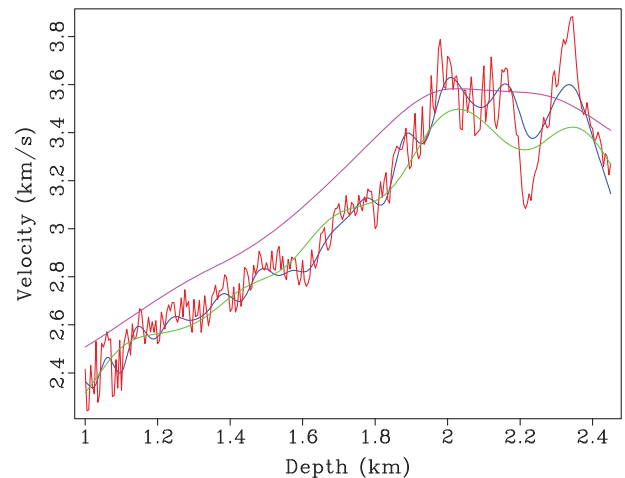


Figure 9. Well-log comparison; pink curve: initial velocity, red curve: exact velocity, green curve: inverted  $v$  with our proposed method, blue curve: inverted  $v$  after sequential standard FWI started from the inverted  $v$  as shown in Figure 8b.

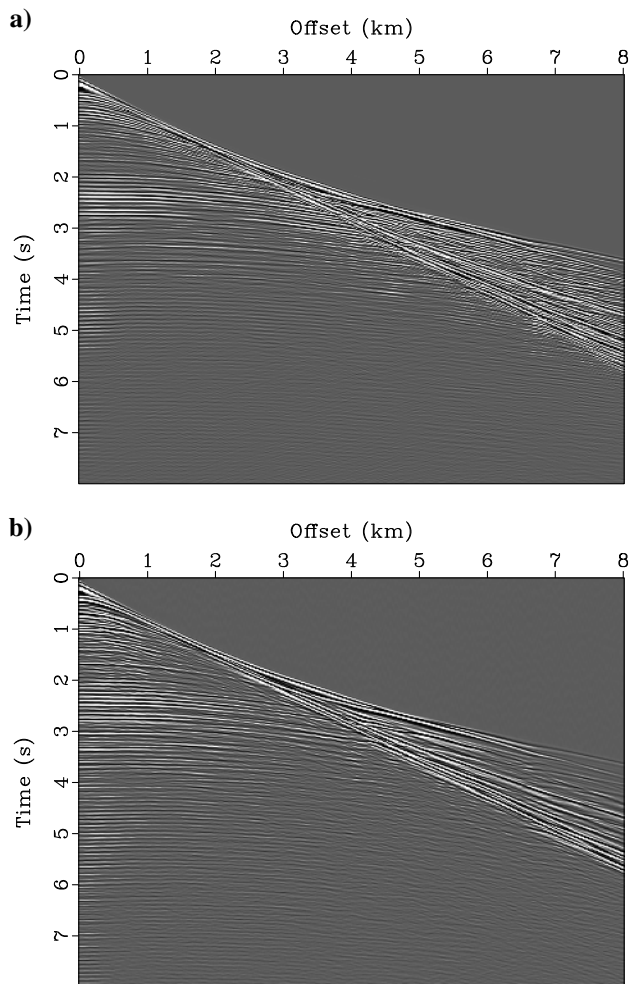


Figure 10. Data comparison: (a) the modeled data and (b) the observed data.

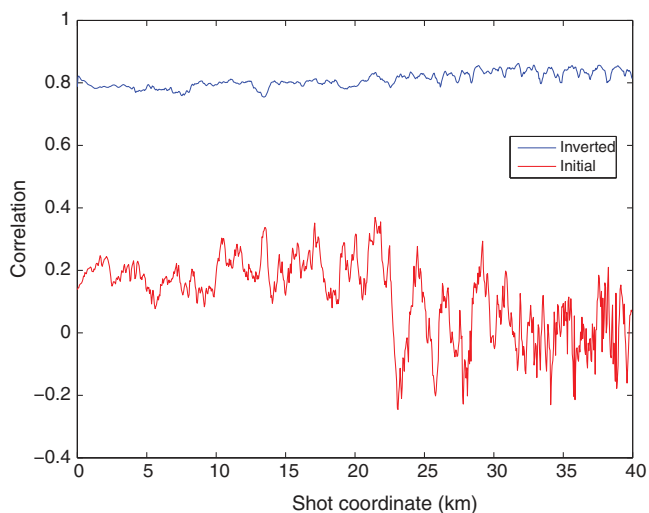


Figure 11. The average correlation for each shot.

is shown in Figure 8c. Figure 9 shows the well logs at 39,375 m from a 1000- to a 2500-m depth for comparison. We can see from Figure 9 that the initial velocity (pink line) is far away from the exact velocity (red line). Figure 10a and 10b shows data comparison between the modeled (Figure 10a) using the inverted velocity and observed data (Figure 10b). We can see that the predicted data are similar to the observed one. To show the accuracy of the inverted model, we define the quality factor as

$$\text{Correlation}_s = \sum_r \frac{\sum_t p_s(x_r, t) g_s(x_r, t)}{n_r \sqrt{\sum_t p_s(x_r, t) p_s(x_r, t) \sum_t g_s(x_r, t) g_s(x_r, t)}} \quad (22)$$

in which  $g_s(x_r, t)$ ,  $p_s(x_r, t)$  are the observed and modeled wavefields from the shot  $s$  and receiver  $r$ , and  $x_s, x_r$  are the relative locations. If the inverted model can produce exactly the same data as the observed one,  $\text{correlation}_s = 1$ . We show the correlation for different shots between the data generated by our inverted model and the observed data. Thus, Figure 11 indicates that the inverted model can predict reasonably accurate the observed data. Figure 12a

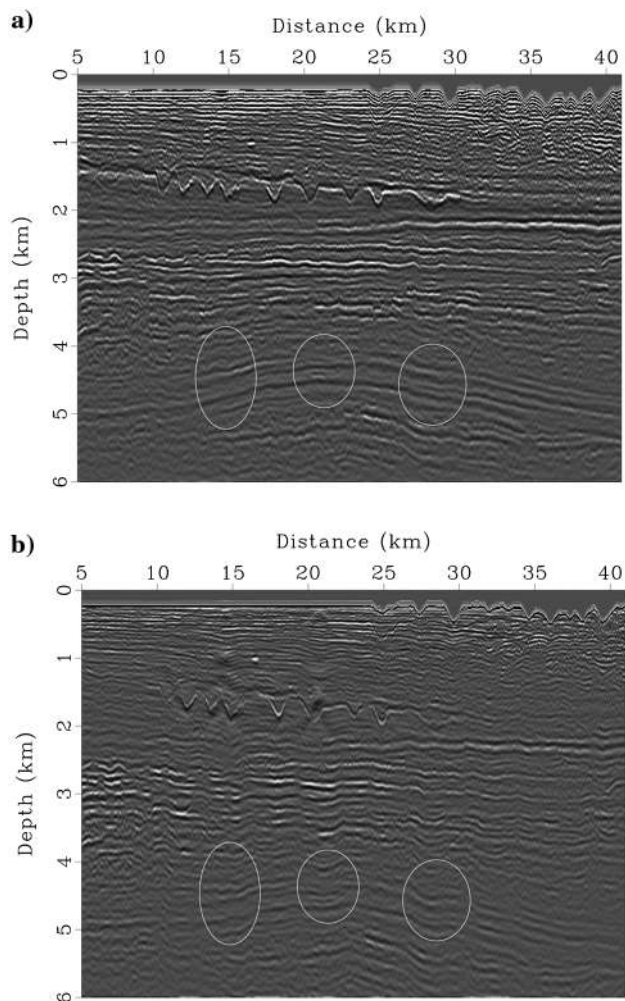


Figure 12. The reverse time migration result with (a) the inverted velocity and (b) initial velocity.

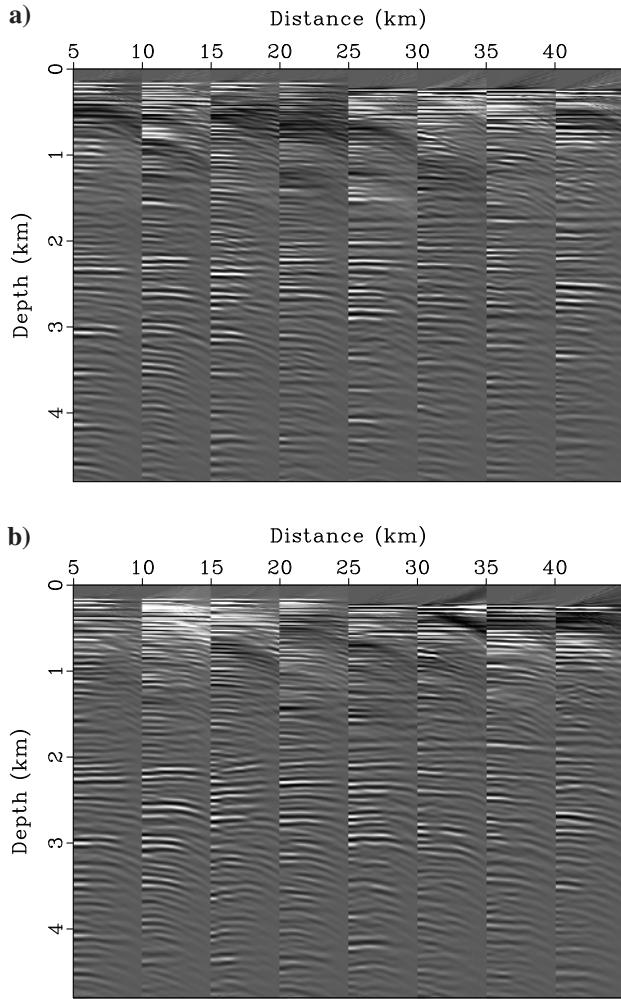


Figure 13. Angle gathers from 0° to 45° using (a) the inverted velocity and (b) initial velocity.

and 12b shows the reverse time migration result with the inverted and initial velocities. The image in Figure 12a is more continuous and better focused. Significant improvements are shown in the three oval areas due to the inverted low-velocity anomalies marked with the oval in Figure 8c. To compare more, we compute angle gathers using the space shift imaging condition (Sava et al., 2005). The resulting angle gathers for every 5 km from angles 0° to 45° using the initial velocity and the inverted one are shown in Figure 13a and 13b. The angle gathers produced by the inverted velocity are flatter.

### CONCLUSIONS

We use a new objective function that combines the data generated by the background velocity (conventional forward modeling) and those generated from the perturbed velocity (Born modeling), to fit the observed data. We invert for the background velocity and the perturbation simultaneously. It is equivalent to mixing FWI and RWI without extended images, and thus it is not prohibitively expensive. To fully use the method to update the smooth component of the velocity at early stage, we add a small weight to the gradient with respect to the background velocity. Application to data generated

using the Marmousi model with frequencies larger than 4 Hz yields accurate results, even when starting with a linearly increasing velocity model. The application to the 2014 Chevron Gulf of Mexico imaging challenge synthetic data set shows the potential of the proposed method.

### ACKNOWLEDGMENTS

We thank KAUST for its support and the SWAG group for the collaborative environment. We thank Chevron for providing the SEG2014 data set. We also thank the associate editor A. Guitton, H. Wang, and two anonymous reviewers for their fruitful suggestions and comments.

### APPENDIX A

#### OPTIMIZED-EXPANSION-BASED LOW-RANK APPROXIMATION

For constant velocity, after applying the spatial Fourier transform  $\hat{p}(\mathbf{k}, t) = F\{p(\mathbf{x}, t)\}$ , the acoustic wave equation 1 is given by

$$\frac{\partial^2 \hat{p}}{\partial t^2} + v^2 |\mathbf{k}|^2 \hat{p} = 0, \quad (\text{A-1})$$

where  $\mathbf{k}$  is the wavenumber vector.

A second-order time-marching scheme with a multidimensional inverse Fourier transform leads to the following familiar expression (Etgen and Brandsberg-Dahl, 2009; Zhang and Zhang, 2009):

$$p(\mathbf{x}, t + \Delta t) + p(\mathbf{x}, t - \Delta t) = \int \hat{p}(\mathbf{k}, t) (2 \cos(|\mathbf{k}|v\Delta t)) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k}. \quad (\text{A-2})$$

The low-rank method (Fomel et al., 2010) tries to approximate the two variable functions  $2 \cos(vk\Delta t)$  with

$$2 \cos(vk\Delta t) \approx \sum_{i=1, j=1}^{i=M, j=N} a_{i,j} 2 \cos(v_i k \Delta t) 2 \cos(v k_j \Delta t), \quad (\text{A-3})$$

through a matrix decomposition. The optimized-expansion-based low-rank method (Wu and Alkhalifah, 2014b) attempts to predict the best  $a_{ij}, v_i, k_j$  by solving the following minimization problem:

$$\min_{v_i, k_j} \min_{a_{ij}} \max_{(v, k) \in \Omega} \left| W(v, k) - \sum_{i=1, j=1}^{i=M, j=N} a_{i,j} W(v_i, k) W(v, k_j) \right|, \quad (\text{A-4})$$

where  $W(v, k) = 2 \cos(vk\Delta t)$  and  $\Omega = [v_{\min}, v_{\max}] \times [0, k_{\max}]$ . In FWI, the minimum and maximum velocities in the model are usually assumed to be known, and  $k_{\max}$  is decided by the mesh parameters. This feature will let us evaluate the coefficients needed for the extrapolation only once at the start, and we will use it throughout the whole full-waveform procedure.

## APPENDIX B

## GRADIENT CALCULATION

The most important ingredient in the optimization problem is the gradient calculation. Similar to [Ma and Hale \(2013\)](#), we can calculate the gradient of the objective function 13 with respect to  $v$  and  $w$  as follows:

$$\delta J(\mathbf{p}, \mathbf{q}) = \frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial v} \delta v + \frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial w} \delta w, \quad (\text{B-1})$$

$$= \langle \mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g}, \mathbf{C}(\delta \mathbf{p} + \delta \mathbf{q}) \rangle, \quad (\text{B-2})$$

and

$$= \langle \mathbf{C}^T(\mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g}), \delta \mathbf{p} + \delta \mathbf{q} \rangle, \quad (\text{B-3})$$

where  $\langle \cdot, \cdot \rangle$  is the summation over the time and space variables for the multiplication of two functions. Similar to the derivation in equation 8 and assuming that  $\mathbf{f}$  is a known variable, we can get the perturbed wavefields  $\delta \mathbf{p}$ ,  $\delta \mathbf{q}$  satisfy

$$\mathbf{A} \delta \mathbf{p} - \mathbf{B}(v) \delta \mathbf{p} = \mathbf{B}'(v) \mathbf{p} \cdot^* \delta v \quad (\text{B-4})$$

and

$$\begin{aligned} \mathbf{A} \delta \mathbf{q} - \mathbf{B}(v) \delta \mathbf{q} &= \mathbf{B}'(v) \mathbf{q} \cdot^* \delta v + \mathbf{B}'(v) \delta \mathbf{p} \cdot^* w \\ &+ \mathbf{B}'(v) \mathbf{p} \cdot^* \delta w + \mathbf{B}'(v) \mathbf{p} \cdot^* w \cdot^* \delta v. \end{aligned} \quad (\text{B-5})$$

By solving the above equation, substituting  $\delta \mathbf{p}$  and  $\delta \mathbf{q}$  in equation B-1 and applying the transpose operator, we then have

$$\delta J(\mathbf{p}, \mathbf{q}) = \langle \langle \lambda, \mathbf{B}'(v) \mathbf{p} \rangle_t, \delta v \rangle_s, \quad (\text{B-6})$$

$$+ \langle \langle \lambda, \mathbf{B}'(v) \mathbf{q} + \mathbf{B}'(v) \mathbf{p} \cdot^* w \rangle_t, \delta v \rangle_s, \quad (\text{B-7})$$

$$+ \langle \lambda, \mathbf{B}'(v) \delta \mathbf{p} \cdot^* w \rangle + \langle \langle \lambda, \mathbf{B}'(v) \mathbf{p} \rangle_t, \delta w \rangle_s, \quad (\text{B-8})$$

and

$$= \langle \mathbf{I} + \mathbf{II} + \mathbf{III} + \mathbf{IV}, \delta v \rangle_s + \langle \mathbf{I}, \delta w \rangle_s, \quad (\text{B-9})$$

where  $\lambda = (\mathbf{A} - \mathbf{B}(v))^{-T}(\mathbf{C}(\mathbf{p} + \mathbf{q}) - \mathbf{g})$  is the wavefield propagating backward using the residual at the receiver  $\langle \cdot, \cdot \rangle_t$  (or  $\langle \cdot, \cdot \rangle_s$ ) is the summation over time (or space) variable and the four terms are defined as

$$\mathbf{I} = \langle \lambda, \mathbf{B}'(v) \mathbf{p} \rangle_t, \quad (\text{B-10})$$

$$\mathbf{II} = \langle \lambda, \mathbf{B}'(v) \mathbf{q} \rangle_t, \quad (\text{B-11})$$

and

$$\mathbf{III} = \langle \lambda, \mathbf{B}''(v) \mathbf{p} \cdot^* w \rangle_t, \quad (\text{B-12})$$

And the fourth term IV can be calculated as

$$\langle \mathbf{IV}, \delta v \rangle_s = \langle \lambda, \mathbf{B}'(v) \delta \mathbf{p} \cdot^* w \rangle, \quad (\text{B-13})$$

$$= \langle \mathbf{B}'(v)^T (\lambda \cdot^* w), \delta \mathbf{p} \rangle, \quad (\text{B-14})$$

and

$$= \langle \langle \mu, \mathbf{B}'(v) \mathbf{p} \rangle_t, \delta v \rangle_s, \quad (\text{B-15})$$

where  $\mu = (\mathbf{A} - \mathbf{B}(v))^{-T}(\mathbf{B}^T(\lambda \cdot^* w))$  is the wavefield propagating backward using  $\mathbf{B}^T(\lambda \cdot^* w)$  as source. Thus, the term IV satisfies that

$$\mathbf{IV} = \langle \mu, \mathbf{B}'(v) \mathbf{p} \rangle_t. \quad (\text{B-16})$$

Hence, the gradients with respect to  $v$  and  $w$  are

$$\frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial v} = \mathbf{I} + \mathbf{II} + \mathbf{III} + \mathbf{IV}, \quad \frac{\partial J(\mathbf{p}, \mathbf{q})}{\partial w} = \mathbf{I}. \quad (\text{B-17})$$

## REFERENCES

- Albertin, U., G. Shan, and J. Washbourne, 2013, Gradient orthogonalization in adjoint scattering-series inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1058–1062.
- Alkhalifah, T., and Y. Choi, 2012, Taming waveform inversion non-linearity through phase unwrapping of the model and objective functions: *Geophysical Journal International*, **191**, 1171–1178, doi: [10.1111/j.1365-246X.2012.05699.x](#).
- Alkhalifah, T., and Z. Wu, 2014, FWI and MVA the natural way: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, WeE10612.
- Almomin, A., and B. Biondi, 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1048–1052.
- Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: *Geophysics*, **79**, no. 3, WA129–WA140, doi: [10.1190/geo2013-0340.1](#).
- Biondi, B., and W. Symes, 2004, Angle domain common image gathers for migration velocity analysis by wavefield continuation imaging: *Geophysics*, **69**, 1283–1298, doi: [10.1190/1.1801945](#).
- Bozda, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: *Geophysical Journal International*, **185**, 845–870, doi: [10.1111/j.1365-246X.2011.04970.x](#).
- Choi, Y., and T. Alkhalifah, 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: *Geophysical Prospecting*, **60**, 748–758, doi: [10.1111/j.1365-2478.2012.01079.x](#).
- Clément, F., G. Chavent, and S. Gómez, 2001, Migration-based traveltimes waveform inversion of 2-D simple structures: A synthetic example: *Geophysics*, **66**, 845–860, doi: [10.1190/1.1444974](#).
- Díaz, E., and A. Guitton, 2011, Fast full waveform inversion with random shot decimation: 81st Annual International Meeting, SEG, Expanded Abstracts, 2804–2808.
- Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: Application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79th Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.
- Fomel, S., L. Ying, and X. Song, 2010, Seismic wave extrapolation using a lowrank symbol approximation: 80th Annual International Meeting, SEG, Expanded Abstracts, 3092–3096.
- Guitton, A., and W. Symes, 2003, Robust inversion of seismic data using the Huber norm: *Geophysics*, **68**, 1310–1319, doi: [10.1190/1.1598124](#).
- Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite difference schemes using kaiser windowed sinc functions: *Geophysics*, **67**, 156–165, doi: [10.1190/1.1451454](#).

- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Presented at Conference on inverse scattering: Theory and application, 206–220.
- Ma, Y., and D. Hale, 2013, Wave-equation reflection travelttime inversion with dynamic warping and full-waveform inversion: *Geophysics*, **78**, no. 6, R223–R233, doi: [10.1190/geo2013-0004.1](https://doi.org/10.1190/geo2013-0004.1).
- Plessix, R., Y. De Roeck, and G. Chavent, 1995, Automatic and simultaneous migration velocity analysis and waveform inversion of real data using a MBTT/WKBJ formulation: 75th Annual International Meeting, SEG, Expanded Abstracts, 1099–1102.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503, doi: [10.1111/j.1365-246X.2006.02978.x](https://doi.org/10.1111/j.1365-246X.2006.02978.x).
- Reker, S., M. Huiskes, and R. Plessix, 2014, Applications of random dynamic shot decimation in full waveform inversion: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, WeELI109.
- Sava, P., B. Biondi, and J. Etgen, 2005, Wave-equation migration velocity analysis by focusing diffractions and reflections: *Geophysics*, **70**, no. 3, U19–U27, doi: [10.1190/1.1925749](https://doi.org/10.1190/1.1925749).
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248, doi: [10.1190/1.1649391](https://doi.org/10.1190/1.1649391).
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790, doi: [10.1111/j.1365-2478.2008.00698.x](https://doi.org/10.1111/j.1365-2478.2008.00698.x).
- Tang, Y., S. Lee, A. Baumstein, and D. Hinkley, 2013, Tomographically enhanced full wavefield inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1037–1041.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Wang, H., S. C. Singh, H. Jian, and H. Calandra, 2012, Integrated inversion of subsurface velocity structures using wave equation tomography and full waveform inversion: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-0472.1](https://doi.org/10.1190/segam2012-0472.1).
- Wang, S., F. Chen, H. Zhang, and Y. Shen, 2013, Reflection-based full waveform inversion (RFWI) in the frequency domain: 83rd Annual International Meeting, SEG, Expanded Abstracts, 877–881.
- Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: *Geophysics*, **79**, no. 3, WA13–WA24, doi: [10.1190/geo2013-0294.1](https://doi.org/10.1190/geo2013-0294.1).
- Wu, Z., and T. Alkhalifah, 2014a, Full waveform inversion based on the optimized gradient and its spectral implementation: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, TuE10607.
- Wu, Z., and T. Alkhalifah, 2014b, The optimized expansion based low-rank method for wavefield extrapolation: *Geophysics*, **79**, no. 2, T51–T60, doi: [10.1190/geo2013-0174.1](https://doi.org/10.1190/geo2013-0174.1).
- Wu, Z., and T. Alkhalifah, 2014c, Spectral implementation of full waveform inversion based on reflections: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, TuP0111.
- Xu, S., D. Wang, F. Chen, G. Lambare, and Y. Zhang, 2012a, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-1473.1](https://doi.org/10.1190/segam2012-1473.1).
- Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare, 2012b, Full waveform inversion for reflected seismic data: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, W024.
- Zhang, Y., and G. Zhang, 2009, One-step extrapolation method for reverse time migration: *Geophysics*, **74**, no. 4, A29–A33, doi: [10.1190/1.3123476](https://doi.org/10.1190/1.3123476).
- Zhou, H., L. Amundsen, and G. Zhang, 2012, Fundamental issues in full waveform inversion: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-0878.1](https://doi.org/10.1190/segam2012-0878.1).
- Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving and reflected waves for velocity model building with impedance inversion based on scale separation: *Geophysical Journal International*, **202**, 1535–1554, doi: [10.1093/gji/ggv228](https://doi.org/10.1093/gji/ggv228).