Seismic Waveform Tomography with Simplified Restarting Scheme
Ying Rao, Yanghua Wang, and Dechao Han

Abstract—For shot-encoded seismic waveform tomography, the restarted L-BFGS algorithm is an effective technique to suppress the crosstalk effect among encoded seismic shots. It restarts the L-BFGS calculation at each iteration segment, consisted of a group of iterations, and re-codes the individual shots randomly not only at the beginning but also at the inside of the iteration segment. Here we simplified this scheme using an invariant shot-encoding within each iteration segment and re-coding individual shots only at the beginning of the segment. This simplification did compromise the image quality at the early stage of inversion, as the crosstalk effect appeared on the inversion result of low-frequency data. However, it eventually achieved both the computation efficiency and the good quality for a multi-scale inversion procedure, which inverts seismic data from low-frequency components to high-frequency components in sequence.

Index Terms—full-waveform inversion (FWI), limit-memory BFGS (L-BFGS) algorithm, restarting scheme, seismic tomography, simultaneous sources.

I. INTRODUCTION
Seismic waveform tomography, or full waveform inversion (FWI), exploits information contained in seismic waveforms for reconstructing subsurface velocity images. The inverted velocity models are expected to show high resolution and high accuracy if compared to travel time inversion results [1]–[3]. However, the computational cost limits the application of the FWI method to field seismic data, especially to reflection seismic data. The computational cost is spent mainly on the waveform simulation, which is needed in the inversion not only for waveform fitting but also for velocity-model updating. The inversion iteratively updates the velocity model and defines the updating direction in terms of the gradient vector, which is the first-order derivative of the inversion objective function [4]. This gradient vector is formed by cross-correlation between the incident waveform and the adjoint waveform [5], [6]. Conventionally, the FWI method implements simulation of these two waveforms for each individual shot independently, and thus demands a huge computational time, particularly for 3D cases with practically thousands of shots.

In a shot-encoded FWI, the individual shots are weighted randomly and summed into a super shot. This shot-encoding technique greatly reduces the number of waveform simulations required by the inversion and improves the computational efficiency of FWI [7]–[13]. However, shot-encoding may have a crosstalk effect due to the interference among the individual shots within the super shot. For example, we define the objective function with shot-coding in the L2-norm as

$$\phi(m) = \frac{1}{2} \left[ \sum_{i} (d_{obs,i} - d_{cal,i}(m))^2 \right].$$

where \(d_{obs,i}\) and \(d_{cal,i}\) are the observed and calculated seismic data after coding, \(m\) is the model to be inverted, \(is\) and \(js\) are the indexes of shots, and \(kr\) is the index of receivers. Then, we have the gradient vector corresponding to equation (1) as

$$g = -\sum_{is} \left( \sum_{js} w_{is} w_{js} \frac{\partial d_{cal,js}}{\partial m} \Delta d_{is} + \sum_{js} w_{is} w_{js} \frac{\partial d_{cal,js}}{\partial m} \Delta d_{js} \right).$$

The second term presented above is the crosstalk term between shot \(is\) and shot \(js\).

Numerically, the crosstalk effect arises from the cross-correlation, that forms the gradient, between the source wavefield of one shot and the data-residual backpropagation wavefield of another shot, and vice versa. The crosstalk effect existing in the gradient vector consequently generates undesirable artifacts or noise in the inverted model image. These artefacts in the inverted model cannot be filtered with clear identification in the model image. Jeong et al. in [14] and [15] analyze the crosstalk effect in three common cases in which the inversion objective functions are defined in the L1 norm, the L2 norm and the t-distribution. Rao and Wang in [13] propose a restarting scheme to suppress the crosstalk effect for the shot-encoded FWI method. This algorithm attempts to increase the randomness, by re-coding the individual shots randomly after a group of iterations, so-called the iteration segment.

The restarting scheme in [13] is designed especially in the framework of the limited-memory BFGS (L-BFGS) algorithm, which has proven to be more effective than conventional gradient methods in the inversion. The L-BFGS algorithm is a quasi-Newton method that preconditions the model update direction by exploiting the property of the Hessian matrix, the second-order derivative of the inversion objective function. The L-BFGS algorithm accumulates the gradient and the Hessian information in all previous iterations and defines the updating direction in terms of the gradient vector and the Hessian information in all previous iterations. The L-BFGS algorithm is an effective technique to exploit information contained in seismic waveforms for seismic imaging, simultaneous sources.

Index Terms—full-waveform inversion (FWI), limit-memory BFGS (L-BFGS) algorithm, restarting scheme, seismic tomography, simultaneous sources.

I. INTRODUCTION
Seismic waveform tomography, or full waveform inversion (FWI), exploits information contained in seismic waveforms for reconstructing subsurface velocity images. The inverted velocity models are expected to show high resolution and high accuracy if compared to travel time inversion results [1]–[3]. However, the computational cost limits the application of the FWI method to field seismic data, especially to reflection seismic data. The computational cost is spent mainly on the waveform simulation, which is needed in the inversion not only for waveform fitting but also for velocity-model updating. The inversion iteratively updates the velocity model and defines the updating direction in terms of the gradient vector, which is the first-order derivative of the inversion objective function [4]. This gradient vector is formed by cross-correlation between the incident waveform and the adjoint waveform [5], [6]. Conventionally, the FWI method implements simulation of these two waveforms for each individual shot independently, and thus demands a huge computational time, particularly for 3D cases with practically thousands of shots.

In a shot-encoded FWI, the individual shots are weighted randomly and summed into a super shot. This shot-encoding technique greatly reduces the number of waveform simulations required by the inversion and improves the computational efficiency of FWI [7]–[13]. However, shot-encoding may have a crosstalk effect due to the interference among the individual shots within the super shot. For example, we define the objective function with shot-coding in the L2-norm as

$$\phi(m) = \frac{1}{2} \left[ \sum_{i} (d_{obs,i} - d_{cal,i}(m))^2 \right].$$

where \(d_{obs,i}\) and \(d_{cal,i}\) are the observed and calculated seismic data after coding, \(m\) is the model to be inverted, \(is\) and \(js\) are the indexes of shots, and \(kr\) is the index of receivers. Then, we have the gradient vector corresponding to equation (1) as

$$g = -\sum_{is} \left( \sum_{js} w_{is} w_{js} \frac{\partial d_{cal,js}}{\partial m} \Delta d_{is} + \sum_{js} w_{is} w_{js} \frac{\partial d_{cal,js}}{\partial m} \Delta d_{js} \right).$$

The second term presented above is the crosstalk term between shot \(is\) and shot \(js\).

Numerically, the crosstalk effect arises from the cross-correlation, that forms the gradient, between the source wavefield of one shot and the data-residual backpropagation wavefield of another shot, and vice versa. The crosstalk effect existing in the gradient vector consequently generates undesirable artifacts or noise in the inverted model image. These artefacts in the inverted model cannot be filtered with clear identification in the model image. Jeong et al. in [14] and [15] analyze the crosstalk effect in three common cases in which the inversion objective functions are defined in the L1 norm, the L2 norm and the t-distribution. Rao and Wang in [13] propose a restarting scheme to suppress the crosstalk effect for the shot-encoded FWI method. This algorithm attempts to increase the randomness, by re-coding the individual shots randomly after a group of iterations, so-called the iteration segment.

The restarting scheme in [13] is designed especially in the framework of the limited-memory BFGS (L-BFGS) algorithm, which has proven to be more effective than conventional gradient methods in the inversion. The L-BFGS algorithm is a quasi-Newton method that preconditions the model update direction by exploiting the property of the Hessian matrix, the second-order derivative of the inversion objective function. The L-BFGS algorithm accumulates the gradient and the Hessian information in all previous iterations and defines the updating direction in terms of the gradient vector and the Hessian information in all previous iterations. The L-BFGS algorithm is an effective technique to exploit information contained in seismic waveforms for seismic imaging, simultaneous sources.
gradient difference between two successive iterations, and these two gradients are inconsistent since two iterations use different shot-encodings. The restarted L-BFGS algorithm can improve the stability of convergence by restarting the L-BFGS recursive calculation at each iteration segment. Rao and Wang [13] suggest to re-code the individual shots not only at the beginning but also at the inside of the iteration segment, in order to increase the effective randomness and to suppress the crosstalk effect for the shot-encoded FWI method.

In this paper, the prime objective is to simplify the restarted L-BFGS algorithm, by re-coding the individual shots only at the beginning of each iteration segment but not necessarily at the inside of the segment. This simplification will make the successive gradients within an iteration segment to be consistent, to satisfy the basic requirement of the L-BFGS calculation. However, it will compromise the inversion quality with a crosstalk effect at the early stage of a multi-scale inversion, in which we invert seismic data from low to high frequency components in sequence. We will demonstrate that the simplified scheme is able to mitigate the crosstalk effect existed only at the inversion image of the low-frequency data set, and to produce high-quality inversion images eventually and to achieve computational efficiency at the same time, after we invert seismic data from low-frequency components to high-frequency components in sequence.

II. RECAP OF THE RESTARTED L-BFGS ALGORITHM

In the shot-encoded FWI, all shots are assigned to a super shot. The weights used in shot-encoding \( w = (w_1, w_2, \ldots, w_N) \) satisfy \( E(wW) = \delta_i \), where \( N_S \) is the total number of shots within the super shot, and \( \delta_i \) is the Kronecker delta. The weights are randomly distributed values of \(+1\) and \(-1\) [13]. This random weighting scheme is applied also to travel time tomography, in which a random data subset is used in each iteration of the nonlinear inversion [18].

The gradient vector is obtained by cross-correlation between the waveform of forward simulation and the waveform of residual back-propagation [5], [6]. The back-propagation uses the exact same forward simulation engine. In a conventional FWI method without shot-encoding, it needs to execute \( 2 \times N_S \) times of waveform simulations if there are \( N_S \) shots in the data set. The shot-encoded FWI puts all shots together as a super shot in the waveform simulation for the forward and backward propagations and leads to significant efficiency. However, since the gradient vector is formed by cross-correlation between these two waveforms, the cross-correlation will have the crosstalk effect among the individual shots in the super shot.

Obviously, this crosstalk effect exhibited in the gradient vector will affect the quality of the inversion image. The crosstalk effect can be offset by the randomness in shot-encoding, as the shots are randomly encoded rather than simply superimposed on a super shot. The crosstalk effect can be further decreased if shots within a super shot are widely scattered. To suppress this crosstalk effect, ultimately, the randomness in shot-encoding is essential. This is because the sum of \( wW \) in the cost term has an expectation of zero, if there is a large number of trials in random generation [15]. The restarted L-BFGS algorithm proposed by [13] attempts to mitigate the crosstalk effect through a sufficient randomness in the shot-encoded FWI.

Figure 1a illustrates the workflow of the restarted L-BFGS algorithm, which restarts the L-BFGS recursive calculation in every five iterations. Within these five iterations, the first two iterations keep the shot-encoding unchanged and each of the remaining iterations has a new shot-encoding. We refer to it as Scheme A in the demonstration that follows.

The L-BFGS calculation needs the gradients of all previous iterations and the gradient differences between consecutive iterations. In Fig. 1a, the gradient vector at the \( k \)th iteration is denoted as \( g_k \), and the difference of gradients between two consecutive iterations is denoted as

\[
y_{k+1} = g_{k+1} - g_k.
\]

For any two consecutive iterations, if the shots are re-coded, the inversion objective functions will be different, and two gradients \( g_k \) and \( g_{k+1} \) will be inconsistent. Consequently, the L-BFGS calculation which is based on the gradient difference will be unstable and potentially presents as non-converge.

Treating the gradient difference between consecutive iterations as a discrete approximation to the gradient differential, which is an analytical continuous function, Rao and Wang [13] propose to calculate the gradient difference \( y_{k+1} \) using the secant equation, instead of subtracting two gradients directly. The secant equation is

\[
y_{k+1} = H_{k+1} y_k + z_{k+1},
\]

where \( z_{k+1} \) is the model difference between two iterations, \( z_{k+1} = m_{k+1} - m_k \), \( m_k \) is the solution at the \( k \)th iteration, and \( H_{k+1} \) is the Hessian matrix, the second derivative of the objective function with respect to the current solution \( m_k \). The gradient differential \( y_{k+1} = H_{k+1} y_k + z_{k+1} \) is used whenever the shots are re-coded between two consecutive iterations. It is calculated as

\[
H_{k+1} y_k + z_{k+1} = V_{k+1}^T V_{k+1} y_k + V_{k+1}^T V_{k+1} z_{k+1}.
\]

Recall that two gradients \( g_{k+1} \) and \( g_k \) in eq. (3) are calculated from different objective functions, because of the change to the shot-encoding. In eq. (4), in which \( z_{k+1} \) is just a model update and does not depend on the objective function, the Hessian matrix \( H_{k+1} \) relies on the objective function only at the \( (k-1) \)th iteration. Therefore, when we change the shot encoding and change the objective functions in two consecutive iterations, eq. (4) works and eq. (3) does not.
Using the secant equation to calculate the gradient differential, instead of the gradient difference, will improve the accuracy. Then, if the condition \( y_k^T z_{k+1} > 0 \) is satisfied, it will guarantee the stability of convergence of the inversion [13].

### III. SIMPLIFIED RESTARTING SCHEME

In this paper, we propose to simplify the restarted L-BFGS algorithm, and refer to it as Scheme B. As illustrated in Fig. 1b, we restart shot-encoding in every five iterations, the same as in Scheme A, but keep the shot-encoding invariant within this iteration segment. In this way, we attempt to achieve a good stability and convergence rate of the iterative inversion.

In FWI, the model update is expressed in the form [19]:

\[
m_{k+1} = m_k - \alpha_k B_k g_k,
\]

where \( \alpha_k \) is the step length, and \( B_k \) is the inverse Hessian matrix. Using the L-BFGS form, the vector \( B_k g_k \) is calculated recursively by

\[
B_k g_k = V_k^T V_k B_0 V_1 V_2 \ldots V_k g_k + \rho_k V_k^T z_k z_k^T V_{k+1} g_{k+1} + \cdots
\]

\[
+ \rho_k V_k^T z_k z_k^T V_{k+1} g_{k+1},
\]

Both eq. (5) and eq. (7) are calculated based on vectors \( y_k \) and \( z_k \). The Hessian matrix \( H_k \) and the inverses \( B_k \) are never constructed explicitly and are not stored in any iteration.

Denote the number of iterations within a restart segment by \( \ell \), and the number of iterations with invariant encoding by \( m \), \( 0 \leq m \leq \ell \). In Scheme A, as illustrated in Fig. 1a, \( \ell = 5 \) and \( m = 2 \). Within a restart segment \( \ell \), \( y_k \) for the initial \( m \) iterations with invariant encoding is calculated based on the gradient difference, and \( y_k \) for the rest where shots are re-coded is calculated using the secant equation (eq. 4). In Scheme B, as illustrated in Fig. 1b, \( \ell = 5 \), in which \( y_k \) is calculated based on the gradient difference, as the shot-encoding is kept invariant for \( \ell \) iterations. Only when it restarts a new segment where shots are re-coded is \( y_k \) calculated using the secant equation.

### IV. THE EFFECTIVENESS OF THE SIMPLIFIED SCHEME

To demonstrate the effectiveness of this simplified scheme, we use a 2D line of SEG/EAGE overthrust model in waveform inversion. A layered structure is the most difficult feature to be reconstructed by tomographic inversion using seismic reflection data.

We discretize the true velocity model (Fig. 2a) into 401×93 grids with 50 m cell size, and generate synthetic gathers from 191 shots, which are placed along the top of the model, with a shot interval of 100 m. We use a Ricker wavelet with 7 Hz dominant frequency [20], [21] as the source wavelet. Each shot gather consists of traces from 401 receivers. We use this synthetic data set as the input to the waveform inversion, in which we put all shots into a single super shot.

We separate the synthetic data set into four frequency bands by band-pass filtering. The four sets of frequency bands are 2–4, 4–6, 6–8, 8–10 Hz, respectively. Then, we implement the inversion in a multi-scale manner. We start the inversion with the first data set of the lowest frequency band for building low-wavenumber components of the model and invert the higher frequency-band data sets sequentially for reconstructing higher wavenumber components of the model. In this way, we achieve a steady convergence of the multi-scale inversion.

We use a smoothed version of the true velocity model (Fig. 2b) as the initial velocity model for the first inversion. Then, we use the inversion result of the lower frequency-band data set as the initial model for the inversion of the higher frequency band data set. We execute 200 iterations for the inversion of each frequency band.
Fig. 4. Crosstalk effect of seismic waveform inversion. Inverted data set is the lowest frequency band. Inversion method is restarted L-BFGS algorithm Scheme A with $\ell = 5$ and $m = 2$, and Scheme B with $m = \ell = 5$.

To evaluate the crosstalk effect in these two schemes, we chose a smoother region of the model, between depths 0.8 and 1.1 km and between distances 10.5 and 18.95 km, to quantify the intensity, using the following variance measurement:

$$\varepsilon = \text{var}(\mathbf{m}_k - \overline{\mathbf{m}}_{30}),$$

where $\mathbf{m}_{30}$ is the reconstructed model by a conventional FWI without shot-encoding (the image is not shown here), after 30 iterations of the standard L-BFGS algorithm, and $\mathbf{m}_k$ is the reconstructed model from the shot-encoded FWI, after $k$ iterations, using either Scheme A or Scheme B of the restarted L-BFGS algorithm. We remove the velocity variation, generated by the conventional FWI, within the selected area and treat the remainder as the crosstalk effect in the inversion image. The intensity of the crosstalk effect, shown in Fig. 4, further confirms the observation that Scheme B (Fig. 3b) has a higher amount of crosstalk noise than Scheme A (Fig. 3a). Consequently, for waveform inversion of high-frequency data sets, the number of iterations needed for Scheme B might be more than that required by Scheme A. But the overall convergence rates of these two schemes are very similar.

For the multi-scale inversion in which we invert high-frequency data sets sequentially, we effectively increase the randomness of shot-encoding gradually when we increase the number of L-BFGS iterations cumulatively. Fig. 5 demonstrates that we can suppress the crosstalk effect eventually in the final inversion results, obtained from either Scheme A or Scheme B. The four panels of each subfigure are the reconstructed velocity models corresponding to the four inversion stages, respectively, using data sets from different frequency-bands ($2-4, 4-6, 6-8, 8-10$ Hz).

Figure 3 compares the inversion images, obtained from the inversion of the first data set with the lowest frequency band ($2-4$ Hz), using either Scheme A (Fig. 3a) or Scheme B (Fig. 3b). Each scheme shows the intermediate results at the 10th, 50th and 100th iteration. The inversion results of this low-frequency data set clearly shows that the crosstalk effect, behaving in random distribution, is stronger in Scheme B than that in Scheme A. Whereas Scheme A reveals the basic flat interfaces of layers, Scheme B shows blurred interface contours.

V. CONCLUSION

For shot-encoded seismic waveform inversion, we have simplified the restarted L-BFGS algorithm with an invariant shot-encoding within each iteration segment, rather re-coding the shots within the iteration segment. We have demonstrated that this simplified scheme does have the crosstalk effect in the inversion image of low-frequency seismic data but is able to suppress this crosstalk effect eventually, when we increase the accumulative randomness of shot-encoding.
The advantage of this simplified restarting scheme is its simplicity in terms of the implementation. Because of the invariant shot-encoding within an iteration segment, the definition of the objective functions and in turn the gradients of successive iterations are consistent. We can use the gradient difference in the iterative inversion.

A trade-off of this simplified restarting scheme is that we need more iteration-segments to suppress the crosstalk effect and to achieve the same image quality as the result of the restarted L-BFGS algorithm. More number of iteration-segments means more iterations. However, this extra number of iterations is within a practically acceptable range. In summary, this simplified restarting scheme for the shot-encoded FWI method balances the computation efficiency, the inversion convergence, and the image quality in a controllable manner.

REFERENCES