

OPEN

Seismic waveform tomography with shot-encoding using a restarted L-BFGS algorithm

Received: 9 June 2017
Accepted: 24 July 2017
Published online: 17 August 2017

Ying Rao[1,2] & Yanghua Wang[2]

In seismic waveform tomography, or full-waveform inversion (FWI), one effective strategy used to reduce the computational cost is shot-encoding, which encodes all shots randomly and sums them into one super shot to significantly reduce the number of wavefield simulations in the inversion. However, this process will induce instability in the iterative inversion regardless of whether it uses a robust limited-memory BFGS (L-BFGS) algorithm. The restarted L-BFGS algorithm proposed here is both stable and efficient. This breakthrough ensures, for the first time, the applicability of advanced FWI methods to three-dimensional seismic field data. In a standard L-BFGS algorithm, if the shot-encoding remains unchanged, it will generate a crosstalk effect between different shots. This crosstalk effect can only be suppressed by employing sufficient randomness in the shot-encoding. Therefore, the implementation of the L-BFGS algorithm is restarted at every segment. Each segment consists of a number of iterations; the first few iterations use an invariant encoding, while the remainder use random re-coding. This restarted L-BFGS algorithm balances the computational efficiency of shot-encoding, the convergence stability of the L-BFGS algorithm, and the inversion quality characteristic of random encoding in FWI.

Full-waveform inversion (FWI) is used to extract information from seismic waveforms to reconstruct a subsurface model of the Earth with complex velocities. The inverted velocity images often demonstrate high resolution and high accuracy[1,2]. However, their computational cost obstructs the wide application of FWI, especially for the inversion of seismic reflection data, which are routinely acquired during exploration geophysics[3]. In FWI methods, the subsurface model is updated iteratively following the update direction, which is defined in terms of the gradient. This gradient vector is the first-order derivative of the FWI objective function with respect to the unknown model parameters, and is computed for each shot separately and is formed by cross-correlation between the incident wavefield and the adjoint wavefield. These two wavefields are simulated for each seismic shot individually[4]. Thus, this process requires a very high operational speed, particularly for practical three-dimensional (3D) seismic data with thousands of shots.

The shot-encoding technique assigns different weights, either $+1$ or -1, to all the shots randomly and sums them into a super shot, greatly reducing the number of wavefield simulations needed for FWI[5–9]. While this shot-encoding strategy improves the efficiency of FWI, it undesirably enhances crosstalk noise, which is caused by the interference among the individual shots within the super shot.

For FWI, the most commonly used inversion methods include the gradient-based method and the quasi-Newton method. The latter involves the Hessian matrix, which is the second-order derivative of the FWI objective function. The limited-memory BFGS (L-BFGS) algorithm is one example of a quasi-Newton method[10,11], where BFGS refers to the Broyden-Fletcher-Goldfarb-Shanno algorithm for updating the Hessian matrix or its inverse, and limited-memory means this algorithm does not store these matrices explicitly. The L-BFGS algorithm has proven to be more effective than the gradient-based method[12–14]. However, if the shot-encoding strategy is adopted for FWI, the standard L-BFGS algorithm will cause an unstable convergence[15,16]. This effect occurs because the calculation in each iteration requires the difference in the gradient between consecutive iterations. These gradients are inconsistent since each iteration uses different shot-encoding values and in turn different objective functions. The restarted L-BFGS algorithm developed in this paper ensures the stability of the inversion

[1]State Key Laboratory of Petroleum Resources & Prospecting, China University of Petroleum (Beijing), Beijing, 102249, China. [2]Centre for Reservoir Geophysics, Department of Earth Science and Engineering, Imperial College London, London, SW7 2BP, UK. Correspondence and requests for materials should be addressed to Y.W. (email: yanghua.wang@imperial.ac.uk)

and the imaging quality of the shot-encoded FWI. Furthermore, a higher efficiency and fewer memory requirements ensure the applicability of the shot-encoded FWI method to 3D seismic data.

Shot-encoded FWI

During the application of FWI, the model is updated along the negative direction of the gradient vector[17]. Using the adjoint-state method[4,18], the gradient is obtained through a cross-correlation between the wavefield from forward modelling and the wavefield from residual back-propagation. Back-propagation uses exactly the same computing engine as employed by forward modelling. Therefore, if there are N_S shots in the seismic field data, then there are $2 \times N_S$ wavefield simulations that must be executed at each iteration.

In seismology, a shot gather refers to a group of seismic traces recorded by a series of receivers for which the recorded wavefield has originated from a single source. To reduce the number of wavefield simulations in the inversion process, the shot gathers are multiplied by either $+1$ or -1 randomly and are then stacked into a super-shot gather[5]. Equivalently, the source signatures are multiplied by random codes and are subsequently excited simultaneously to simulate forward wave propagation. For the calculation of the gradient, the residual of the waveform between the shot-encoded synthetic wavefield and the shot-encoded observed wavefield is backward propagated and cross-correlated with the forward wavefield. This cross-correlation will exhibit a crosstalk effect between different shots, which can directly deteriorate the quality of FWI imagery[7].

For typical shot-encoded FWI, seismic shots are assigned into different super shots[16,19,20]. The shots contained within a super shot should be randomly encoded rather than simply superimposed so that the crosstalk effect can be offset by the randomness of encoding. The objective function for the shot-encoded FWI is defined in terms of the weighted wavefields according to the weights in the encoding procedure. Then, the gradient of the objective function consists of both the ordinary gradient (i.e., without shot-encoding) and an extra term representing the crosstalk between different shots. Fortunately, this crosstalk term also clearly indicates that an increase in the randomness of the shot-encoding can effectively suppress the crosstalk effect. This quality improvement occurs because the weights w_i used in the encoding procedure are randomly generated values of $+1$ and -1, and the sum of a large number of $w_i w_j$ (i.e., the multiplication between different encoded shots) in the crosstalk term has an expectation of zero[8,9].

The inversion algorithm used in shot-encoded FWI must be stochastic. Because the model update procedure is stochastic, both gradient-based and Newton-type methods used in conventional deterministic inversions must be properly adjusted to fit within a shot-encoded framework. The restarted L-BFGS algorithm proposed in this paper exploits the features of shot-encoding and the L-BFGS algorithm and uses these features to enhance the computational efficiency and image quality of FWI by mitigating the crosstalk effect.

Restarted L-BFGS algorithm

In FWI, the subsurface model is updated iteratively, and thus, the theoretical wavefield that is calculated based on the subsurface model gradually matches the field observations. Hence, the objective function is commonly described as

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}(\mathbf{m})\|_2^2, \tag{1}$$

where $\mathbf{d}_{\text{obs}}$ is the vector of the observed seismic data, and $\mathbf{d}_{\text{cal}}(\mathbf{m})$ is the vector of the synthetic data calculated based on the model, $\mathbf{m}$. This objective function is minimised gradually through iterative model updates. The model update is defined by the gradient according to a precondition of the inverse Hessian matrix in the form of[21]

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{B}_k \mathbf{g}_k, \tag{2}$$

where k is the iteration number, α_k is the step length, $\mathbf{B}_k$ is the inverse Hessian matrix, and $\mathbf{g}_k = \nabla\phi(\mathbf{m}_k)$ is the gradient vector. Since it exploits the inverse Hessian matrix, this is a quasi-Newton method.

In a BFGS algorithm, the inverse Hessian matrix $\mathbf{B}_k$ is derived recursively from $\mathbf{B}_{k-1}$ using the following relation[22]

$$\mathbf{B}_k = \left(\mathbf{I} - \frac{\mathbf{z}_{k-1}\mathbf{y}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}\right)\mathbf{B}_{k-1}\left(\mathbf{I} - \frac{\mathbf{y}_{k-1}\mathbf{z}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}\right) + \frac{\mathbf{z}_{k-1}\mathbf{z}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}, \tag{3}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{y}_{k-1}$ is the difference in the gradient between two consecutive iterations, and $\mathbf{z}_{k-1}$ is the model difference. Because of its recursive nature, $\mathbf{B}_{k-1}$ does not need to be stored, and $\mathbf{B}_k$ can be calculated quickly and adaptively based on the gradient differences $\{\mathbf{y}_{k-1}, \mathbf{y}_{k-2}, \mathbf{y}_{k-3}, \ldots\}$ and the model differences $\{\mathbf{z}_{k-1}, \mathbf{z}_{k-2}, \mathbf{z}_{k-3}, \ldots\}$ from the previous iterations. Therefore, Equation (3) represents the limited-memory version of the BFGS algorithm (i.e., L-BFGS).

If $\mathbf{y}_{k-1}^T\mathbf{z}_{k-1} > 0$, the inverse Hessian matrix $\mathbf{B}_k$ is positive definite. This is because in the following formula

$$\mathbf{v}^T\mathbf{B}_k\mathbf{v} = \left(\mathbf{v} - \frac{\mathbf{z}_{k-1}^T\mathbf{v}}{\mathbf{z}_{k-1}^T\mathbf{y}_{k-1}}\mathbf{y}_{k-1}\right)^T\mathbf{B}_{k-1}\left(\mathbf{v} - \frac{\mathbf{z}_{k-1}^T\mathbf{v}}{\mathbf{z}_{k-1}^T\mathbf{y}_{k-1}}\mathbf{y}_{k-1}\right) + \frac{(\mathbf{z}_{k-1}^T\mathbf{v})^2}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}, \tag{4}$$

if $\mathbf{B}_{k-1} > 0$, both terms on the right-hand side are nonnegative, while the first term is zero only if $\mathbf{v} = 0$, and the second term is zero only if $\mathbf{z}_{k-1}^T\mathbf{v} = 0$. This positive definite property ensures that the update $-\mathbf{B}_k\mathbf{g}_k$ (Equation 2) is in a descent direction, and consequently the model update during the inversion will converge. If shots are

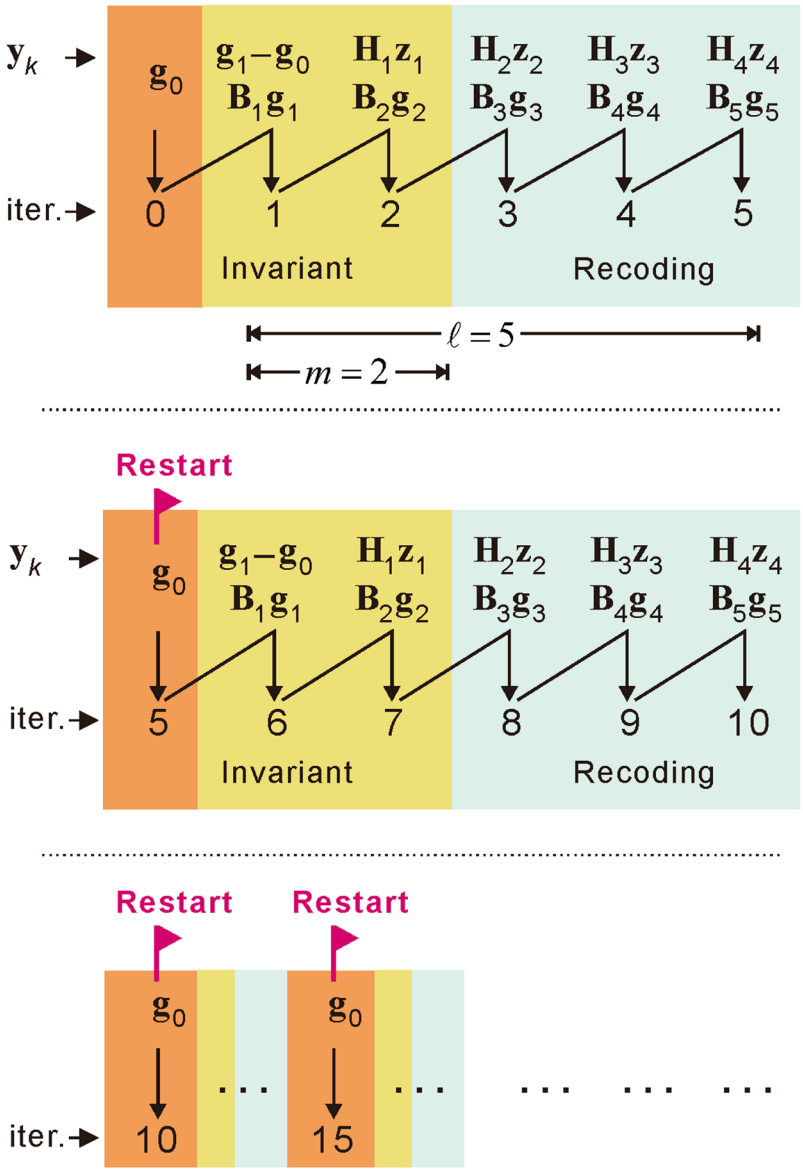


Figure 1. The restarted L-BFGS algorithm. The implementation of the L-BFGS algorithm is restarted after a number of iterations (ℓ). Within the segment between two starts, only the coding of the first few iterations (m) is kept unchanged (yellow) and is the same as the previous one (orange). The shots for the rest of the iterations (grey) are re-coded with a sufficient randomness to suppress crosstalk effects. Within a restarted segment, the gradient differential is given by the secant equation $\mathbf{y}_{k-1} = \mathbf{H}_{k-1}\mathbf{z}_{k-1}$ except for the first iteration, where $\mathbf{y}_0 = \mathbf{g}_1 - \mathbf{g}_0$. Both $\mathbf{H}_{k-1}\mathbf{z}_{k-1}$ and $\mathbf{B}_k\mathbf{g}_k$ are calculated using the limited-memory formulas.

re-coded at each iteration, two gradients $\mathbf{g}_k(\mathbf{m}_k, \mathbf{w}_k)$ and $\mathbf{g}_{k-1}(\mathbf{m}_{k-1}, \mathbf{w}_{k-1})$ are calculated from the objective functions with different sets of encoding, which are $\mathbf{w}_{k-1}$ and $\mathbf{w}_k$. For the gradient difference

$$\mathbf{y}_{k-1} = \mathbf{g}_k - \mathbf{g}_{k-1}, \tag{5}$$

even if the step length α_{k-1} is sufficiently small, the condition $\mathbf{y}_{k-1}^T\mathbf{z}_{k-1} > 0$ is difficult to satisfy. Therefore, shot-encoding will induce instability and potentially non-convergence during FWI.

Here, we propose the restarted L-BFGS algorithm (Fig. 1) containing the following three distinguishing features.

(1) Structurally, the L-BFGS calculation is restarted after every ℓ iterations. The total number of L-BFGS iterations is divided into a series of segments, and each segment consists of ℓ iterations. The recursive L-BFGS calculation is only based on the gradient differences $\mathbf{y}_k$ and model differences $\mathbf{z}_k$ within the individual segment. Therefore, the segment interval ℓ is also referred to as the storage length in the L-BFGS algorithm (i.e., for storing the vectors $\mathbf{y}_k$ and $\mathbf{z}_k$).

(2) Within a segment, the shot-encoding of the first few iterations is kept invariant (yellow) and is the same as the previous iteration (orange). A new reassignment of the weights (either $+1$ or -1) is triggered for the remainder of the iterations (grey). We denote m as the number of iterations that employ invariant coding. Generally, $2 \leq m < \ell$, which maximises the overall randomness of shot-encoding to effectively suppress the crosstalk effect. Meanwhile, maintaining the weights of coding as invariant for a few of iterations, rather than recoding at every iteration, will significantly improve the convergence speed and stability during the inversion.

(3) For the first iteration, the gradient difference is calculated by $\mathbf{y}_0 = \mathbf{g}_1 - \mathbf{g}_0$. For the rest of the iterations, the gradient difference is calculated based on the following secant equation:

$$\mathbf{y}_{k-1} = \mathbf{H}_{k-1}\mathbf{z}_{k-1}, \tag{6}$$

where $\mathbf{H}_{k-1} = \nabla^2\phi(\mathbf{m}_{k-1})$ is the Hessian matrix in the $k-1$-th iteration. Equation (6) represents the gradient differential rather than the gradient difference. Using the gradient differential will improve the accuracy and, in turn, the stability of convergence of the FWI.

Note that the secant equation in Equation (6) differs slightly from the conventional secant equation used in the BFGS algorithm. The latter imposes a condition[23] wherein $\mathbf{H}_k$ is close to the current matrix $\mathbf{H}_{k-1}$ and presents the following approximation as the secant equation:

$$\mathbf{y}_{k-1} = \mathbf{H}_k\mathbf{z}_{k-1}. \tag{7}$$

This approximation can be justified because

$$\mathbf{y}_{k-1} \equiv \mathbf{g}_k - \mathbf{g}_{k-1} = \frac{\mathbf{g}_k - \mathbf{g}_{k-1}}{\mathbf{m}_k - \mathbf{m}_{k-1}}(\mathbf{m}_k - \mathbf{m}_{k-1}), \tag{8}$$

wherein the fraction can be treated as either a forward difference or a backward difference, i.e.,

$$\frac{\mathbf{g}_k - \mathbf{g}_{k-1}}{\mathbf{m}_k - \mathbf{m}_{k-1}} \approx \mathbf{H}_k \approx \mathbf{H}_{k-1}. \tag{9}$$

However, the approximated secant equation in Equation (7) cannot be used to directly calculate $\mathbf{y}_{k-1}$, as it depends on $\mathbf{H}_k$ and not $\mathbf{H}_{k-1}$. In this paper, we use secant Equation (6) to calculate $\mathbf{y}_{k-1}$.

To calculate the Hessian matrix, we use the DFP (Davidon-Fletcher-Powell) method[23], which represents the Hessian matrix based on the gradient difference and the model difference as follows:

$$\mathbf{H}_k = \left(\mathbf{I} - \frac{\mathbf{y}_{k-1}\mathbf{z}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}\right)\mathbf{H}_{k-1}\left(\mathbf{I} - \frac{\mathbf{z}_{k-1}\mathbf{y}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}\right) + \frac{\mathbf{y}_{k-1}\mathbf{y}_{k-1}^T}{\mathbf{y}_{k-1}^T\mathbf{z}_{k-1}}. \tag{10}$$

Exploiting its recursive characteristic, we derive the following formula:

$$\begin{aligned}\mathbf{H}_k \;=\; & \mathbf{V}_{k-1}^T\mathbf{V}_{k-2}^T\cdots\mathbf{V}_1^T\mathbf{H}_0\mathbf{V}_1\cdots\mathbf{V}_{k-2}\mathbf{V}_{k-1} \\ & +\rho_1\mathbf{V}_{k-1}^T\mathbf{V}_{k-2}^T\cdots\mathbf{V}_2^T\mathbf{y}_1\mathbf{y}_1^T\mathbf{V}_2\cdots\mathbf{V}_{k-2}\mathbf{V}_{k-1} \\ & \vdots \\ & +\rho_{k-2}\mathbf{V}_{k-1}^T\mathbf{y}_{k-2}\mathbf{y}_{k-2}^T\mathbf{V}_{k-1} \\ & +\rho_{k-1}\mathbf{y}_{k-1}\mathbf{y}_{k-1}^T,\end{aligned} \tag{11}$$

where $\rho_k = 1/\mathbf{y}_k^T\mathbf{z}_k$ and $\mathbf{V}_k = (\mathbf{I} - \rho_k\mathbf{z}_k\mathbf{y}_k^T)$. According to the secant equation, where $\mathbf{H}_k\mathbf{z}_k = \mathbf{y}_k$, we define the initial Hessian matrix $\mathbf{H}_0$ as

$$\mathbf{H}_0 = \mathbf{y}_0\mathbf{y}_0^T[\mathbf{z}_0\mathbf{y}_0^T]^{-1} \approx \frac{\mathbf{y}_0^T\mathbf{y}_0}{\mathbf{z}_0^T\mathbf{y}_0}\mathbf{I}. \tag{12}$$

Equation (11) does not need to store matrices and is a limited-memory version of the DFP formula. This equation is similar to the L-BFGS formula, which uses the same recursive form as Equation (11) for the inverse Hessian matrices $\mathbf{B}_k$, for $k = 0, 1, \ldots$, and does not need to store these matrices as well. Therefore, in shot-encoded FWI, we use the L-BFGS and DFP algorithms and calculate $\mathbf{B}_k\mathbf{g}_k$ and $\mathbf{H}_k\mathbf{z}_k$ based only on the vectors $\mathbf{y}_k$ and $\mathbf{z}_k$, which are stored within the segment of ℓ iterations.

Results

Accuracy of the gradient differential. The difference in the gradients is an approximation to the gradient differential. The numerical errors contained within $\mathbf{y}_{k-1}$ include any strong perturbations in gradients $\mathbf{g}_k$ and $\mathbf{g}_{k-1}$. However, the secant equation is an accurate solution of the differential of the gradient functional. The gradient differential based on $\mathbf{H}_{k-1}\mathbf{z}_{k-1}$ is more stable than the subtraction of any two gradients. The accuracy of $\mathbf{H}_{k-1}\mathbf{z}_{k-1}$, which replaces $\mathbf{g}_k - \mathbf{g}_{k-1}$, ensures the steady convergence of the iterative inversion.

Gradient differentials are used as constraints for the calculation of the Hessian matrix $\mathbf{H}_k$ and its inverse $\mathbf{B}_k$, which both have a similar limited-memory recursive formula. Because of their recursive characteristics, the

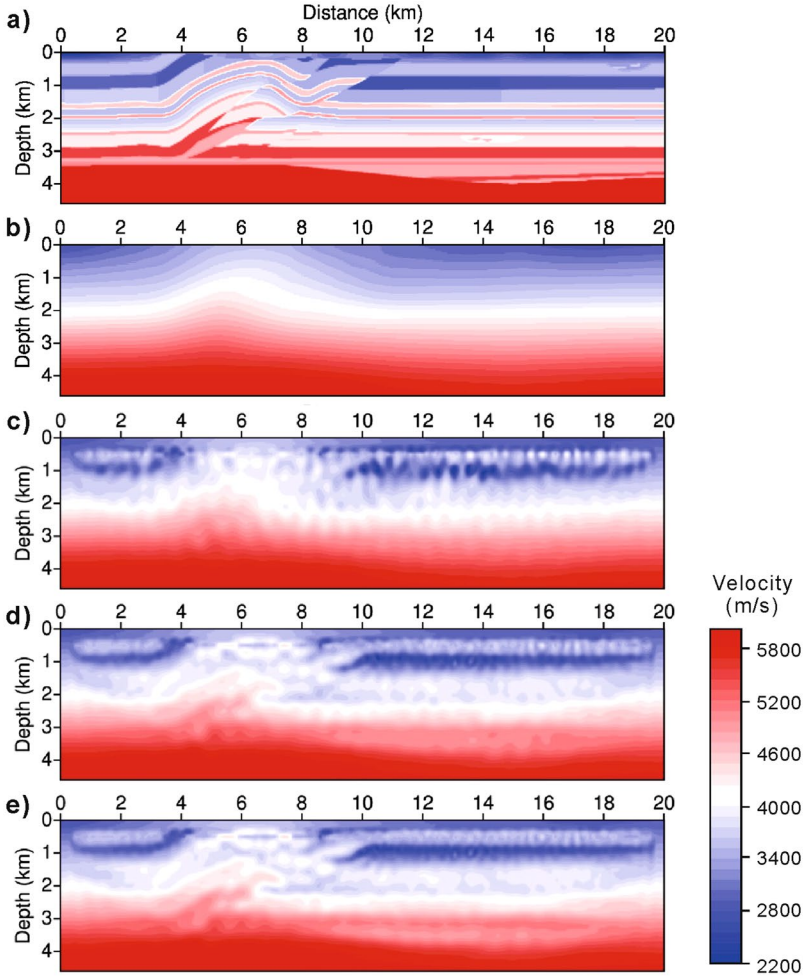


Figure 2. Randomness vs. crosstalk. (**a**) An overthrust model. (**b**) The initial model for FWI. (**c**,**d**,**e**) The inversion results at the 10th, 50th and 100th iterations, using the data set of the lowest frequency band (2−4 Hz). Crosstalk noise is gradually suppressed along with the increased randomness in the restarted L-BFGS algorithm.

influence of the initial $\mathbf{y}_{k-\ell}$ within the invariant encoding procedure gradually decreases in the subsequent iterations wherein the shots are encoded randomly. Therefore, the L-BFGS calculation must be restarted after a given number of iterations.

Randomness vs. crosstalk. Only randomness can suppress the crosstalk effect in shot-encoded FWI. A 2D line of an SEG/EAGE overthrust model[24] (Fig. 2a) is used to test the effectiveness of the restarted L-BFGS algorithm. The model is partitioned into 401 × 93 cells in the horizontal and vertical directions with a 50-m cell size. There are 191 shot points laid atop the model. The spatial interval between the shots is 100 m. The source signature is a Ricker wavelet[25] with a dominant frequency of 7 Hz. Each shot gather consists of 401 receivers. For shot-encoded FWI, all 191 shots are grouped into a single super shot.

The multi-scale inversion method is adopted[26,27], for which the data have been divided into four frequency bands: 2−4, 4−6, 6−8 and 8−10 Hz. Therefore, the FWI procedure consists of four stages. The initial model for the first inversion of the data set within the lowest frequency band is the smoothed version of the true model (Fig. 2b). The initial model for the inversion of any data set within a higher frequency band is the inversion result from the lower frequency band. The inversion of each frequency-band data set is executed over 200 iterations.

Figure 2 also displays the results at the 10th, 50th and 100th iterations, using the data set from the lowest frequency band. These results clearly demonstrate the presence of the crosstalk effect within shot-encoded FWI. The shallow horizontally layered structure appears to be truncated by noise into multiple short segments. The deeper horizontal interface appears curved with numerous fluctuations. However, this crosstalk effect decreases with an increase in the number of iterations as a consequence of an increase in the accumulated randomness of encoding.

The crosstalk effect will be effectively suppressed in the final inversion results because the randomness of shot-encoding is further increased stage by stage along with an increased number of L-BFGS iterations during the multi-scale inversion. In the multi-scale inversion shown in Fig. 3, a total of 800 iterations have been executed. In the final result of the four-stage inversion, the fault structure and shallow layers have been successfully reconstructed using shot-encoded FWI.

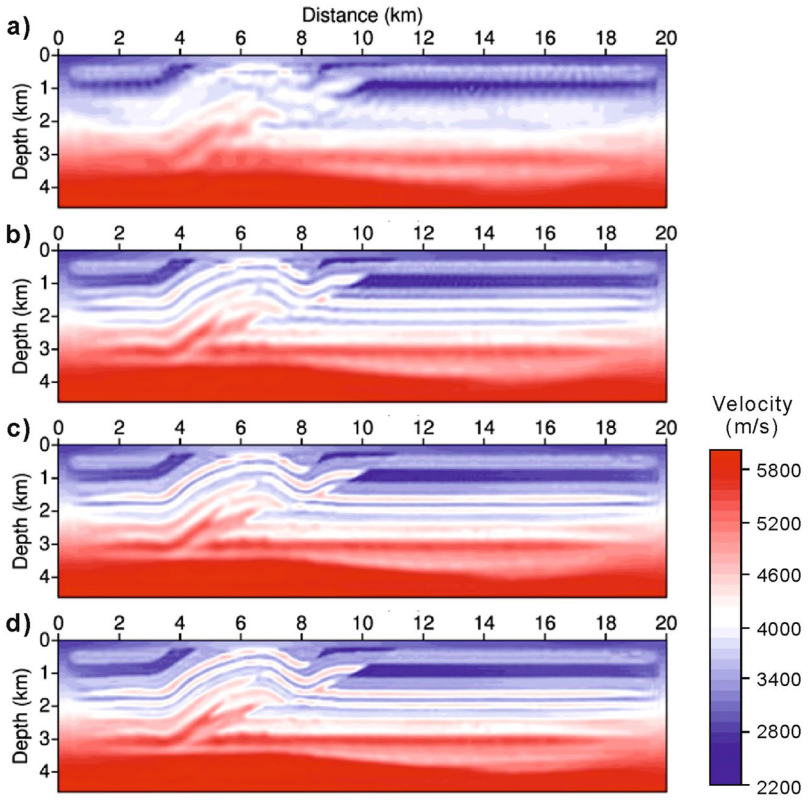


Figure 3. Shot-encoded FWI with the restarted L-BFGS algorithm. (**a**,**b**,**c**,**d**) Reconstructed velocity profiles corresponding to the respective inversion stages of four frequency-band data sets (2−4, 4−6, 6−8, 8−10 Hz). The crosstalk effect is eventually suppressed because of the cumulative randomness in the shot-encoding.

For comparison, Fig. 4 displays the results of standard FWI, without shot-encoding, using the standard L-BFGS algorithm. For each data set, the inversion is executed over 30 iterations. Therefore, a total of 120 iterations are executed. The main structures are clearly recovered in the FWI model. These results are used as a benchmark for verifying the effectiveness of shot-encoded FWI.

Shot-encoded FWI utilising the restarted L-BFGS algorithm (Fig. 3) can produce results similar to those of FWI without shot-encoding utilising the standard L-BFGS algorithm (Fig. 4). The linear features stretching from the top to the bottom in the final results confirm that both the restarted L-BFGS algorithm and the standard L-BFGS algorithm possess the same reconstruction ability during waveform inversion. However, the efficiency of these two inversions differs significantly. The computational efficiency of FWI is determined by the actual number of wavefield simulations required. These wavefield simulations include wave back-propagation and step-length calculations for every iteration. Figure 5 shows the number of wavefield simulations required to achieve convergence. The total number of simulations in the restarted L-BFGS algorithm ($<0.25 \times 10^4$, shaded region in Fig. 5) is substantially less than that in the standard L-BFGS algorithm (7×10^4). Therefore, the restarted L-BFGS algorithm fully reflects an accelerated shot-encoding inversion procedure. Compared with FWI without shot-encoding, the restarted L-BFGS algorithm significantly reduces computational costs.

Applicability to 3D FWI. As the restarted L-BFGS algorithm can improve image quality by suppressing the crosstalk effect and can simultaneously reduce computational costs and storage requirements, shot-encoded FWI is shown to be applicable to cases involving 3D data sets.

The size of the actual 3D overthrust SEG/EAGE model is $4.65 \times 20 \times 20\,\text{km}^3$. The model is partitioned into $93 \times 401 \times 401$ grids with a cell size of $50 \times 50 \times 50\,\text{m}^3$. The source signature is a Ricker wavelet with a domain frequency of 7 Hz. There are 400 shots spread out over the surface at an interval of $1 \times 1\,\text{km}^2$. Each shot gather consists of 32,761 traces. In the inversion, every 100 shots are assembled into a super shot.

The multi-scale inversion method is adopted for the shot-encoded FWI of the four frequency bands similar to that in the 2D example. The initial model is a Gaussian smoothed version of the true velocity model. For the inversion of each frequency band, 100 iterations are executed, and the restarted L-BFGS algorithm is used with $\ell = 5$ and $m = 2$. Figure 6 displays the resulting velocity slices at different depths to compare the true velocity model and the reconstructed velocity model. At depths of 1.0 and 1.5 km, the morphology of an ancient channel is clearly reconstructed. At a depth of 2 km, the main body of the channel can also be inverted, although there is limited wave coverage.

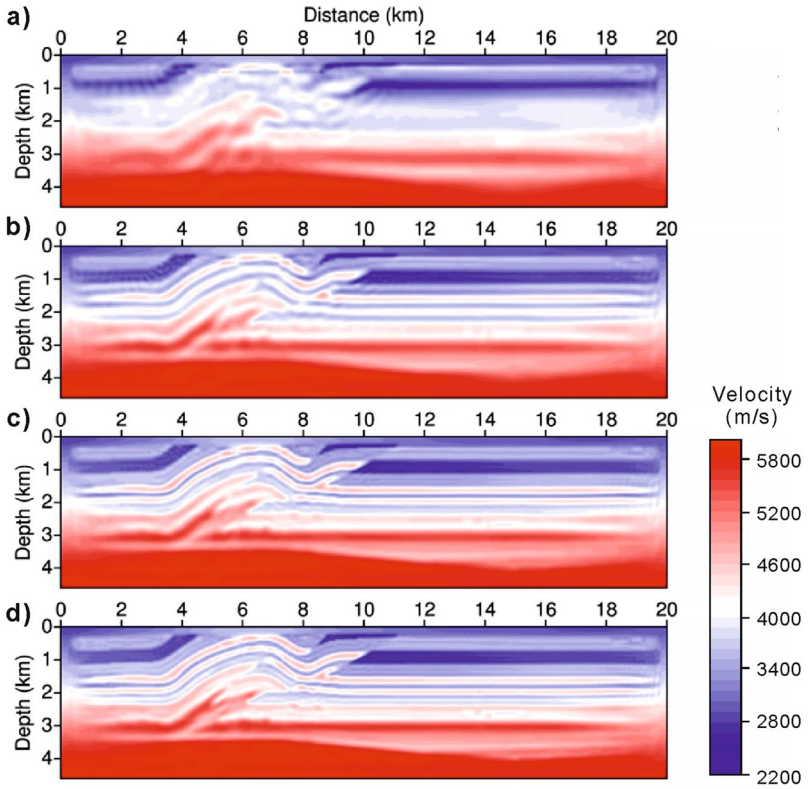


Figure 4. Velocity models reconstructed by FWI without shot-encoding. **(a,b,c,d)** Results corresponding to the respective data sets of four frequency bands. These results are the benchmarks used for the shot-encoded FWI with the restarted L-BFGS algorithm proposed in this paper.

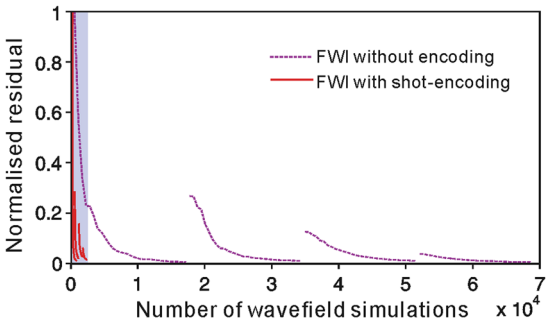


Figure 5. Number of wavefield simulations. Convergence rates of FWI without shot-encoding (purple dashed lines) and of shot-encoded FWI using the restarted L-BFGS algorithm (red solid lines within the shadowed zone). The horizontal axis denotes the number of wavefield simulations, and the vertical axis is the normalised residual. Each line segment corresponds to an inversion of a data set for a particular frequency band.

Conclusions

A restarted L-BFGS algorithm for shot-encoded FWI is proposed to improve the efficiency, enhance the stability and reduce the memory requirement for seismic waveform tomography.

In a standard L-BFGS algorithm, the recursive calculations require the difference of the gradients between consecutive iterations. If shot-encoding is conducted differently in different iterations, the gradients will be inconsistent, and the inversion will not converge steadily. The restarted L-BFGS algorithm replaces the difference in the gradients with an accurate gradient differential using the secant equation. Its accuracy leads to steady convergence. The algorithm also enhances the flexibility of the restart mode of the L-BFGS algorithm with shot-encoding. Because of sufficient randomness in the shot-encoding in the restarted L-BFGS algorithm, crosstalk noise is eventually suppressed through the iteration of the implemented L-BFGS algorithm.

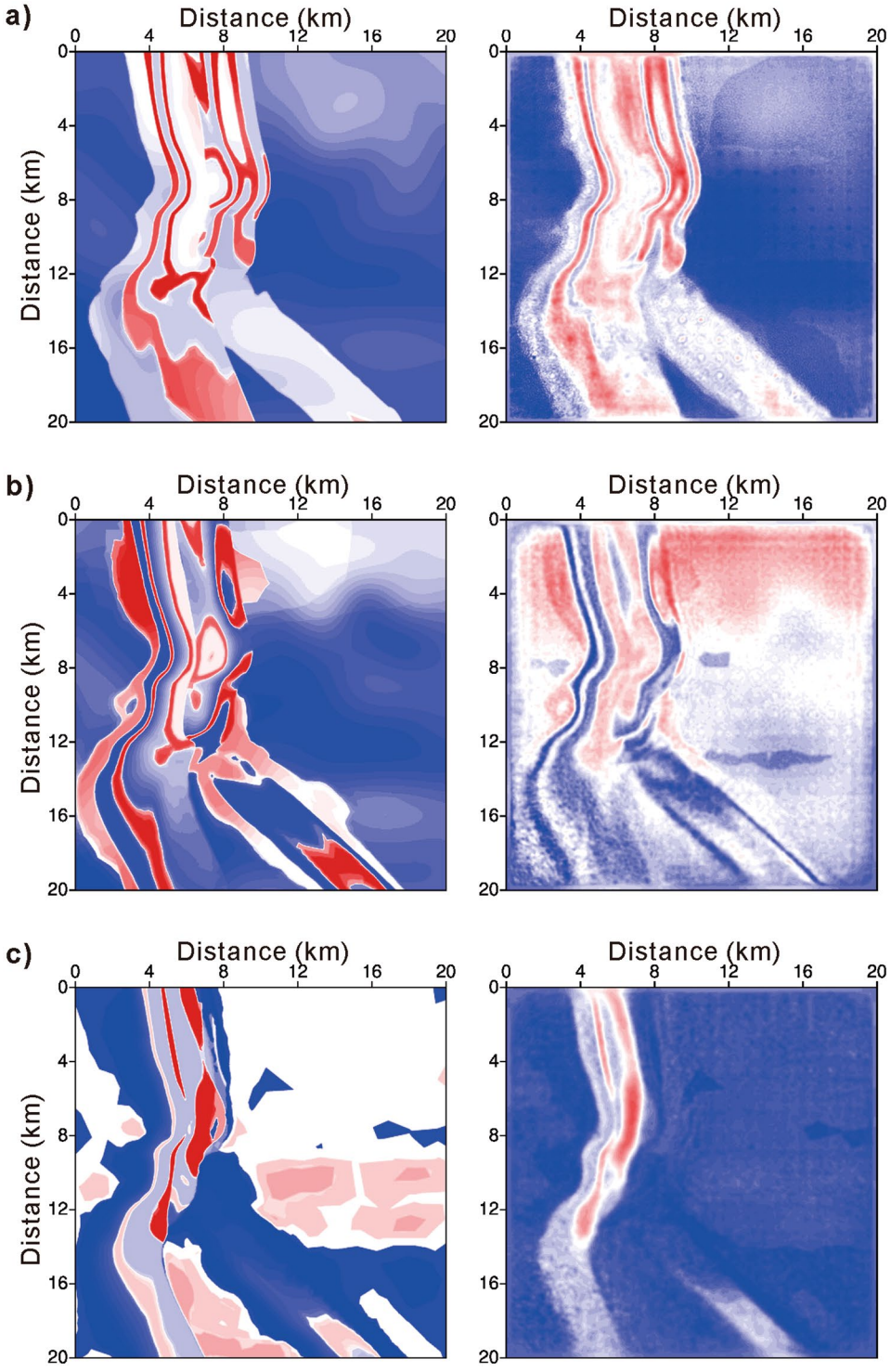


Figure 6. Shot-encoded FWI applicability to a 3D model. (**a**) Velocity slices at a depth of 1 km. (**b**) Velocity slices at a depth of 1.5 km. (**c**) Velocity slices at a depth of 2 km. The left column is the true velocity model, and the right column is the velocity model reconstructed using shot-encoded FWI.

This breakthrough ensures that shot-encoded FWI will be applicable to the vast quantity of existing seismic reflection data that have been acquired by exploration geophysicists. It also holds enormous potential in teleseismic research by preceding teleseismic tomography based on travel time residuals, as there is a better chance to generate high-resolution crustal images using this method and local earthquakes, where the sources are found inside the model, though their exact location and mechanism must also be determined.

References

1. Operto, S., Virieux, J., Dessa, J. X. & Pascal, G. Crustal seismic imaging from multi-fold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern Nankai trough. *Journal of Geophysical Research* **111**(B9), B09306, doi:10.1029/2005JB003835 (2006).
2. Brossier, R., Operto, S. & Virieux, J. Seismic imaging of complex structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics* **74**(6), WCC63–WCC76, doi:10.1190/1.3215771 (2009).
3. Wang, Y. & Rao, Y. Reflection seismic waveform tomography. *Journal of Geophysical Research* **114**, B03304, doi:10.1029/2008JB005916 (2009).
4. Plessix, R. E. 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 (2006).
5. Capdeville, Y., Gung, Y. & Romanowicz, B. Towards global earth tomography using the spectral element method: a technique based on source stacking. *Geophysical Journal International* **162**, 541–554, doi:10.1111/j.1365-246X.2005.02689.x (2005).
6. Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R. & Lee, S. Fast full-wavefield seismic inversion using encoded sources. *Geophysics* **74**(6), WCC177–WCC188, doi:10.1190/1.3230502 (2009).
7. Ben-Hadj-Ali, H., Operto, S. & Virieux, J. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics* **76**(4), R109–R124, doi:10.1190/1.3581357 (2011).
8. Jeong, W., Pyun, S., Son, W. & Min, D.-J. A numerical study of simultaneous-source full waveform inversion with L_1-norm. *Geophysical Journal International* **194**, 1727–1737, doi:10.1093/gji/ggt182 (2013).
9. Jeong, W., Kang, M., Kim, S., Min, D. J. & Kim, W. K. Full waveform inversion using student's *t*-distribution: a numerical study for elastic waveform inversion and simultaneous-source method. *Pure and Applied Geophysics* **172**, 1491–1509, doi:10.1007/s00024-014-1020-7 (2015).
10. Nocedal, J. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation* **35**, 773–782, doi:10.1090/S0025-5718-1980-0572855-7 (1980).
11. Liu, C., Gao, F., Feng, X., Liu, Y. & Ren, Q. Memoryless quasi-Newton (MLQN) method for 2D acoustic full waveform inversion. *Exploration Geophysics* **46**, 168–177, doi:10.1071/EG13090 (2014).
12. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **45**, 503–528, doi:10.1007/BF01589116 (1989).
13. Guitton, A. & Symes, W. W. Robust inversion of seismic data using the Huber norm. *Geophysics* **68**, 1310–1319, doi:10.1190/1.1598124 (2003).
14. Ma, Y. & Hale, D. Quasi-Newton full-waveform inversion with a projected Hessian matrix. *Geophysics* **77**(5), R207–R216, doi:10.1190/GEO2011-0519.1 (2012).
15. Guitton, A. & Díaz, E. Attenuating crosstalk noise with simultaneous source full waveform inversion. *Geophysical Prospecting* **60**, 759–768, doi:10.1111/j.1365-2478.2011.01023.x (2012).
16. Castellanos, C., Metivier, L., Operto, S., Brossier, R. & Virieux, J. Fast full waveform inversion with source encoding and second-order optimization methods. *Geophysical Journal International* **200**, 718–742, doi:10.1093/gji/ggu427 (2015).
17. Wang, Y. Seismic Inversion: Theory and Applications (Wiley Blackwell, 2016).
18. Tarantola, A. Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation (Elsevier, 1987).
19. van Leeuwen, T. & Herrmann, F. Fast waveform inversion without source-encoding. *Geophysical Prospecting* **61**, 10–19, doi:10.1111/j.1365-2478.2012.01096.x (2012).
20. Schiemenz, A. & Igel, H. Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: Application to Valhall ocean-bottom cable data. *Geophysical Journal International* **195**, 1970–1988, doi:10.1093/gji/ggt362 (2013).
21. Byrd, R. H., Hansen, S. L., Nocedal, J. & Singer, Y. A stochastic quasi-Newton method for large-scale optimization. *SIAM Journal on Optimization* **26**, 1008–1031, doi:10.1137/140954362 (2016).
22. Nocedal, J. & Wright, S. J. Numerical Optimization (Springer, 2000).
23. Davidon, W. C. Variable metric method for minimization. *SIAM Journal on Optimization* **1**, 1–17, doi:10.1137/0801001 (1991).
24. Aminzadeh, F., Brac, J., & Kunz, T. 3D salt and overthrust models. SEG/EAGE Modelling Series, No. 1: Distribution CD of salt and overthrust models, SEG Book Series. Tulsa, Oklahoma, ISBN: 9781560800774 (1997).
25. Wang, Y. Frequencies of the Ricker wavelet. *Geophysics* **80**(2), A31–A37, doi:10.1190/geo2014-0441.1 (2015).
26. Ravaut, C. *et al.* Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International* **159**, 1032–1056, doi:10.1111/j.1365-246X.2004.02442.x (2004).
27. Wang, Y. Seismic waveform modelling and tomography, in Gupta, H. K. (ed.), Encyclopaedia of Solid Earth Geophysics, pp. 1290–1301 (Springer Verlag, 2011).

Acknowledgements

The authors are grateful to the Natural Science Foundation of China (grant nos 41474111 and 41622405) and the sponsors of the Centre for Reservoir Geophysics, Imperial College, London, for supporting this research.

Author Contributions

Y.R. implemented the algorithm and conducted the FWI inversion. Y.W. focused on the innovative concepts and finalised the paper. All authors contributed to the preparation of the figures and the revision of the paper.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017