Full waveform inversion using the excitation representation of the source wavefield

Mahesh Kalita*, KAUST and Tariq Alkhalifah, KAUST

SUMMARY

Full waveform inversion (FWI) is an iterative method of data-fitting, aiming at high resolution recovery of the unknown model parameters. However, it is a cumbersome process, requiring a long computational time and large memory space/disc storage. One of the reasons for this computational limitation is the gradient calculation step. Based on the adjoint state method, it involves the temporal cross-correlation of the forward propagated source wavefield with the backward propagated residuals, in which we usually need to store the source wavefield, or include an extra extrapolation step to propagate the source wavefield from its storage at the boundary. We propose, alternatively, an amplitude excitation gradient calculation based on the excitation imaging condition concept that represents the source wavefield history by a single, specifically the most energetic arrival. An excitation based Born modeling allows us to derive the adjoint operation. In this case, the source wavelet is injected by a cross-correlation step applied to the data residual directly. Representing the source wavefield through the excitation amplitude and time, we reduce the large requirements for both storage and the computational time. We demonstrate the application of this approach on a 2-layer model with an anomaly and the Marmousi II model.

INTRODUCTION

Full waveform inversion (FWI) is a powerful tool to estimate the unknown subsurface parameters quantitatively with high resolution by minimizing the misfit between the observed and the simulated seismic data (Tarantola, 1984). It is an iterative process of updating the model parameters with a quantity that depends on the gradient of the misfit function. No extra computation of the huge Jacobian matrix is required since the gradient is equivalent to the image in reverse time migration process (Lailly, 1983; Plessix, 2006). Despite the huge improvements in the computational power, it can be a challenge to make FWI efficient, especially in large-scale 3D problems (Virieux and Operto, 2009). One of the main reasons for the cost in FWI is the gradient calculation step which involves expensive temporal cross-correlation. It requires an order of nt (time samples in a trace) floating point operations (FLOPs) at the time of the correlation for every model point. Additionally, depending on the size of the wavefield, defined by the model and data, a large memory space/disk storage is required to store the history of the source wavefield. The boundary saving scheme is a solution (Berkhout, 1988; Clapp, 2008), but at the cost of an additional wavefield extrapolation.

In this abstract, we employ the excitation imaging condition in conjunction with FWI, which helps mitigate the memory issue in an efficient manner. It considers the most energetic arrival of the source wavefield (Chang and McMechan, 1986; Nguyen

and McMechan, 2013), resulting in the reduction of the FLOPs during the imaging process. Additionally, we focus on investigating the forward problem using the excitation method. We demonstrate its versatility on a 2 layer model with an anomaly and the Marmousi II model. We will also show field data examples at the meeting.

THEORY

The excitation approach represents the source wavefield by its peak amplitude and arrival time, referred to as excitation amplitude and excitation time (t_{ex}), respectively. Therefore, it doesn't entertain multipathing (Nguyen and McMechan, 2013). Mathematically, it is given by

$$G_{0_{ex}}(\mathbf{x}, t; \mathbf{x}_s) = G_0(\mathbf{x}, t; \mathbf{x}_s) \delta(t - t_{ex}(\mathbf{x}; \mathbf{x}_s)) = \begin{cases} G_0(\mathbf{x}, t_{ex}; \mathbf{x}_s) & \text{for } t = t_{ex}; \\ 0 & \text{otherwise;} \end{cases} \quad (1)$$

where G_0 is the Green's function, an impulse response at the image point $\mathbf{x}$, due to a source positioned at $\mathbf{x} = \mathbf{x}_s$. The subscript ex stands for the excitation representation.

Figure 1: (a) Scattered data using excitation method due to a point perturbation (δv) in background homogeneous velocity (v_0), (b) data modeled with actual velocity ($v_0 + \delta v$) – data modeled with velocity v_0 .

The forward problem: Born modeling with the excitation approach

FWI with Excitation approach

The Born approximation establishes a linear relationship between scattered wavefield and its cause (model perturbation). Mathematically, it is a temporal convolution process of two Green's functions, one being the 2nd order temporal differentiation of the source wavefield and another one the receiver wavefield to the point of perturbation. The scattered wavefield using excitation approximation (equation (1)) is

$$\delta G_{ex}(\mathbf{x}_r, t, \mathbf{x}_s) = \int d\mathbf{x} \frac{\delta v(\mathbf{x})}{v_0^2(\mathbf{x})} \int d\tau \frac{\partial^2}{\partial t^2} \{G_0(\mathbf{x}, t - \tau; \mathbf{x}_s) \delta(t - t_{ex} - \tau)\} G_0(\mathbf{x}_r, \tau; \mathbf{x})$$

where $\delta G_{ex}(\mathbf{x}_r, t; \mathbf{x}_s)$ is the scattered impulse response due to a perturbation δv in the slowness square $1/v_0^2$, recorded at the receiver position $\mathbf{x} = \mathbf{x}_r$. For a band-limited source $f(t)$, however, the above equation accommodates another convolution of $G_{0_{ex}}$ with $f(t)$. Therefore, the scattered data using Born and excitation approximation satisfies the following wave equation:

$$\begin{aligned} (\nabla^2 - \frac{1}{v_0^2} \frac{\partial^2}{\partial t^2}) \delta G_{ex}(\mathbf{x}_r, t; \mathbf{x}_s) &= \frac{\delta v(\mathbf{x})}{v_0^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \{G_{0_{ex}}(\mathbf{x}, t; \mathbf{x}_s) \delta(t - t_{ex} - \tau) \star f(t)\} \\ &= \frac{\delta v(\mathbf{x})}{v_0^2(\mathbf{x})} G_0(\mathbf{x}, t_{ex}; \mathbf{x}_s) \frac{\partial^2}{\partial t^2} \{f(t - t_{ex})\} \quad (2) \end{aligned}$$

where $\star$ represents a convolution process in time. Similar to the conventional Born modeling, excitation method requires two *for* loops: the first one computes the excitation time and amplitude to the point of perturbation and the second one solves equation (2). We display the scattered data induced by a single point perturbation in a homogeneous velocity $v_0 = 2.0$ km/s in Figure 1(a). It is similar to the difference of the data modeled with the actual velocity $v = v_0 + \delta v$ and the background velocity v_0 , respectively as shown in Figure 1(b). Thus, the excitation Born modeling can accurately reproduce the single scattered wavefield. The FWI gradient is its adjoint.

The inverse problem: FWI with the excitation approach

In FWI, the excitation approximation replaces the source wavefield by a modified source function wavelet where its amplitude is scaled by the excitation amplitude and time of origin is shifted by the excitation time. Therefore, the gradient of FWI, based on L^2 norm misfit function, and using the excitation approximation for a source at $\mathbf{x} = \mathbf{x}_s$ is given by:

$$\begin{aligned} g_{ex}(\mathbf{x}) &= \frac{-2}{v_0^3(\mathbf{x})} [\{G_0(\mathbf{x}, t; \mathbf{x}_s) \delta(t - t_{ex}) \star \frac{\partial^2 f(t)}{\partial t^2}\} \bullet \overleftarrow{\Delta d}(\mathbf{x}, t; \mathbf{x}_s)] \\ &= \frac{-2}{v_0^3(\mathbf{x})} G_0(\mathbf{x}, t_{ex}; \mathbf{x}_s) [\frac{\partial^2 f(t - t_{ex})}{\partial t^2} \bullet \overleftarrow{\Delta d}(\mathbf{x}, t; \mathbf{x}_s)] \quad (3) \end{aligned}$$

where Δd is the data residual between the predicted and observed data, with $\overleftarrow{(\cdot)}$ and $\bullet$ represent the backward propagation of wavefield and the zero-lag cross-correlation process in time, respectively. Therefore, the sufficient information from the source wavefield history is only the maximum amplitude and its arrival time in order to compute the gradient. Thus,

this results in significant saving of memory space/disk-storage. Also, because of the source wavelet, the gradient now involves a very small window for cross-correlation, ensuring a very few FLOPs at the time of imaging. Consequently, this method leads in reduction of computational cost compared to its conventional counterpart.

We can formulate the problem in matrix form as follows,

$$g_{ex}(\mathbf{x}) = \langle \mathbf{F} \mathbf{u}_\delta, \mathbf{L}^* \Delta d \rangle, \quad (4)$$

where $\langle, \rangle$ represents the dot product between the two vectors. $\mathbf{F}$ is a lower triangular Toeplitz matrix with j^{th} row given by $(\frac{\partial^2 f}{\partial t^2}|_{t_j}, \dots, \frac{\partial^2 f}{\partial t^2}|_{t_2}, \frac{\partial^2 f}{\partial t^2}|_{t_1}, \dots, 0)$ that acts on the spike source wavelet u_δ with non-zero element at the excitation time t_{ex} , representing the causal convolution process. $\mathbf{L}^*$ is the adjoint of the wave equation operator, signifying the reverse extrapolation of the adjoint source in time. Using the properties of dot product, we can rewrite equation (4) as $g_{ex}(\mathbf{x}) = \langle \mathbf{u}_\delta, \mathbf{F}^* \mathbf{L}^* \Delta d \rangle$ where $\mathbf{F}^*$ is the cross-correlation operator. Again, the differentiation property of cross-correlation operation allows switching the position of the operator $\mathbf{F}^*$ and $\mathbf{L}^*$ as follows,

$$\begin{aligned} \mathbf{L}^* \mathbf{F}^* \Delta d &= (\frac{\partial^2}{\partial t^2} - v_0^2 \nabla^2) \{ \frac{\partial^2 f(t)}{\partial t^2} \star \overleftarrow{\Delta d} \} \\ &= \frac{\partial^2 f(t)}{\partial t^2} \star (\frac{\partial^2}{\partial t^2} - v_0^2 \nabla^2) \overleftarrow{\Delta d} = \mathbf{F}^* \mathbf{L}^* \Delta d. \end{aligned}$$

It suggests that we need to cross-correlate the second derivative of the source function f with Δd before the reverse extrapolation. Since u_δ contains only one non-zero element given by $\frac{G_0(\mathbf{x}, t_{ex}; \mathbf{x}_s)}{v_0^3(\mathbf{x})}$, the imaging process requires the adjoint wavefield only at the time of origin (t_{ex}), and thus, a single FLOP. Therefore, the computation of gradient using the excitation approach is given by:

$$g_{ex}(\mathbf{x}) = \langle \mathbf{u}_\delta, \mathbf{L}^* \mathbf{F}^* \Delta d \rangle. \quad (5)$$

Pseudo-Hessian with excitation approach

Shin et al. (2001) demonstrated that the diagonal of the pseudo-Hessian plays a vital role in compensation the geometrical spreading loss by scaling the gradient with a proper weight. In the excitation approach, the scaling is simply an arithmetic division of the gradient by the square of the excitation amplitude. In order to strictly obey the excitation approximation, we prefer the divisor to be excitation amplitude square, rather than the source normalized term $(\sum_{it=1}^{nt} G(\mathbf{x}, it; \mathbf{x}_s)^2)$ despite both share almost the same computational cost. Finally, the model update $\tilde{g}_{ex}$ at an iteration is given by:

$$\tilde{g}_{ex}(\mathbf{x}) = \frac{\mathbf{L}^* \Delta \hat{d} \text{ at } t = t_{ex}(\mathbf{x})}{\mathbf{u}_\delta \text{ at } t = t_{ex}(\mathbf{x})}, \quad \text{with } \Delta \hat{d} = \mathbf{F}^* \Delta d; \quad (6)$$

which is very similar to the excitation amplitude imaging condition (Nguyen and McMechan, 2013), with a subtle difference in the respective adjoint sources. Nguyen and McMechan (2013) back propagate the data, whereas we back propagate $\Delta \hat{d}$ (the data-residual cross-correlated with the source wavelet). Comparing with conventional FWI, excitation approach scales

FWI with Excitation approach

a different wavelet. The former scales down the cross-correlated Δd with source wavefield by the total energy of the source wavefield. On the other hand, the excitation approach scales down $\Delta \hat{d}$ by the maximum amplitude of the source wavefield.

Comments on adjoint source of excitation approach

Equation (5) reveals that the source side term $\mathbf{u}_s$ in the gradient does not have the source signature in it. Therefore, the adjoint source also needs to be deconvolved from the source signature, which is partially accomplished by the temporal cross-correlation process (equivalent to deconvolution in terms of phase) of the source wavelet to the conventional adjoint source. Also, the cross-correlation process will allow $\Delta \hat{d}$ to have better signal to noise ratio than Δd .

Analogy of Kirchhoff algorithm with excitation approach

The virtual source for the scattered wavefield with excitation approximation (equations (2) and (3)) resembles a Kirchhoff implementation for the source side, which is represented by $A(\mathbf{x}; \mathbf{x}_s) \frac{\partial^2}{\partial t^2} f(t - \tau(\mathbf{x}; \mathbf{x}_s))$, where $A(\mathbf{x}; \mathbf{x}_s)$ and $\tau(\mathbf{x}; \mathbf{x}_s)$ are the amplitude and travel time respectively. The time, τ is analogous to t_{ex} and $A(\mathbf{x})$ to $G_0(\mathbf{x}, t_{ex})$, with a difference owing to the band-limited nature of the source function. However, the excitation representation of the source wavefield is immune to the well-known shortcomings of the Kirchhoff methods such as low illumination in the shadow zone or the infinite amplitude at the caustics, as it is calculated using wavefields.

Multipathing: a shortcoming of excitation approximation

The excitation method is more attractive to the large data sets, especially the 3D problem, because of its computational advantages over the conventional FWI. However, its main weakness is handling multipathing in the source wavefield. It is intrinsically handled in conventional FWI by utilizing the full source wavefield with all its complexity. However, an intermediate solution of storing more than one maximum of the source wavefield also can ameliorate the excitation method in the presence of multipath (Jin et al., 2015; Nguyen and McMechan, 2015).

EXAMPLES

We demonstrate the versatility of excitation method in FWI on a 2-layer velocity model with an anomaly and the Marmousi II model. The observed data for inversion is simulated using the finite difference method, with a 2nd and 4th order accuracy in time and space, respectively.

A 2-layer model with anomaly: The high velocity anomaly is centered at (1 km, 0.25 km) as shown in Figure 2(a). We use a Ricker source wavelet with a peak frequency of 2 Hz. Figure 2(b) displays the initial velocity model. To make the inversion process much easier, we put constraints on both the upper and lower limits of the velocity in the update process. Also, update is done only in the first layer of the model. Figures 2(c) and 2(d) display the update after 50 iterations with the excitation method and the conventional approach, respectively. They suggest a good reconstruction of the anomaly in the inversion process. The update over iterations using the excitation method is similar to its conventional counterpart. However, the

former method has the upper hand by completing the process in 30 minutes whereas the latter in 45 minutes, which is about 50% additional cost, not to mention the memory features.

Marmousi II: The velocity model (Figure 3(a)) contains 679 and 139 grid points in x and z direction, respectively. We place 160 sources and 679 receivers on the surface to record the data set. Also, we follow a hierarchical multi-scale strategy (Bunks et al., 1995) to invert for two sub datasets of 2 and 4 Hz dominant frequency, successively. Figure 3(b) displays the initial velocity model (linearly increasing with depth after the water layer). The velocity model after 150 iterations (Figure 3(c)) is used as the initial model for 2nd stage. The final update is shown in Figure 3(d). To compare, we display the associated results with that from the conventional approach in Figures 3(e) and 3(f). A closer look on the depth profiles at $x = 8$ km (Figure 3(g)) and $x = 12$ km (Figure 3(h)) reveals some subtle differences in the update between the excitation method and its conventional counterpart. It is due to, most probably, the lost of some multipath energy from the source side, a weakness of the excitation approach. However, the excitation method saves 25% computational time compared to the conventional method to compute a gradient.

DISCUSSION

The excitation approach utilizes the most energetic part of the source wavefield in order to compute the gradient. In the Marmousi example, where multipathing is prevalent, such limitation had minor effect on the quality of the inversion. Most of the energy to the gradient computation is usually extracted from a single scattering optimal (energy wise) path of the waves. In some cases, multipathing must be addressed accurately to have the successful inversion.

Alternatively, we can have even a more efficient implementation by extracting the source wavefield excitation information from solving the eikonal equation or ray tracing (Chang and McMechan, 1986; Nichols, 1996), but at the cost of accuracy. Despite the high-frequency asymptotic nature of such approximations, the often simple nature of the source wavefield, compared to the receiver wavefield, allow us to use such approximations, especially at the FWI early iterations where the model is still generally smooth.

CONCLUSION

We develop Born modeling and FWI based on the excitation representation of the source wavefield. It, specifically, simplifies the source wavefield by the maximum amplitude and its arrival time, and thus, reduces the enormous storage requirements for the source wavefield in the conventional implementation of FWI. Gradient evaluation is done only at the time of excitation, no dot product is needed, resulting in a reduction of the number of FLOPs for such computation. Although, multipathing is a shortcoming of this method, such limitation had minor effects in the inversion process, especially in the early iterations, as observed in the inversion of Marmousi II model.

FWI with Excitation approach

Figure 2: The panel depicts (a) a 2-layer velocity model with a high velocity anomaly, (b) the starting velocity model for inversion; the update after 50 iterations with (c) the excitation approach, and (d) the conventional approach.

Figure 3: (a) Marmousi II model, (b) the starting model; the velocity update after (c) 150 iterations, (d) 300 iterations with excitation method. Conventional FWI updates after (e) 150 iterations, and (f) 300 iterations. Comparisons of the depth profiles at (g) $x = 8.0$ km and (h) $x = 12.0$ km for the two methods.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

- Berkhout, A. J., 1988, A holographic approach to acoustic control: *Journal of the Audio Engineering Society*, **36**, 977–995.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, <http://dx.doi.org/10.1190/1.1443880>.
- Chang, W.-F., and G. A. McMechan, 1986, Reverse-time migration of offset vertical seismic profiling data using the excitation-time imaging condition: *Geophysics*, **51**, 67–84, <http://dx.doi.org/10.1190/1.1442041>.
- Clapp, R. G., 2008, Reverse time migration: Saving the boundaries: Stanford Exploration Project, 136.
- Jin, H., G. A. McMechan, and B. Nguyen, 2015, Improving input/output performance in 2D and 3D angle-domain common-image gathers from reverse time migration: *Geophysics*, **80**, no. 2, S65–S77, <http://dx.doi.org/10.1190/geo2014-0209.1>.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: Theory and application: Society for Industrial and Applied Mathematics, 206–220.
- Nguyen, B. D., and G. A. McMechan, 2013, Excitation amplitude imaging condition for prestack reverse-time migration: *Geophysics*, **78**, no. 1, S37–S46, <http://dx.doi.org/10.1190/geo2012-0079.1>.
- Nguyen, B. D., and G. A. McMechan, 2015, Five ways to avoid storing source wavefield snapshots in 2D elastic prestack reverse time migration: *Geophysics*, **80**, no. 1, S1–S18, <http://dx.doi.org/10.1190/geo2014-0014.1>.
- Nichols, D. E., 1996, Maximum energy traveltimes calculated in the seismic frequency band: *Geophysics*, **61**, 253–263, <http://dx.doi.org/10.1190/1.1443946>.
- 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, <http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x>.
- Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: *Geophysical Prospecting*, **49**, 592–606, <http://dx.doi.org/10.1046/j.1365-2478.2001.00279.x>.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, <http://dx.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, <http://dx.doi.org/10.1190/1.3238367>.