Full Waveform Inversion without Low Frequency Data – Retrieving Buried Low Wavenumber Information

Wenyi Hu, Jiefu Chen, Jianguo Liu, and Aria Abubakar

Abstract

Full waveform inversion (FWI) is an advanced seismic processing method for high resolution subsurface geophysical property reconstruction through a data-fitting procedure. When low frequency seismic data are not available, conventional FWI often suffers from the cycle-skipping issue caused by the severe nonlinearity nature of the standard L2 norm objective function, inducing strong artifacts in the reconstructed geophysical property models. After a brief introduction of the FWI theory, we provide an overview of the cycle-skipping phenomenon and its cause, and the recent research progress on this topic. Various algorithms and strategies that attempt to tackle the cycle-skipping problem are compared and discussed. These strategies include the scattering angle filtering, model extension, combination of FWI and traveltime inversion, nonlinear frequency spectrum extension, phase unwrapping, and the Beat Tone FWI. Instead of compiling an exhaustive list of the existing approaches, which is impractical, we introduce a classification of these methods from a signal processing perspective. This will enable a geophysicist or a signal processing engineer to properly apply and continuously improve these approaches for challenging subsurface exploration projects.

Introduction and Motivation

In a typical seismic data acquisition survey for hydrocarbon exploration, sources and the associated receivers are usually deployed on the earth surface. Seismic waves excited by the sources propagate toward the subsurface, penetrate the formations, and eventually arrive at the receivers, carrying a great amount of the subsurface geophysical property information [1]. For decades, the community of exploration geophysics has been using the tomography-migration-

Wenyi Hu is with Advanced Geophysical Technology, Houston, TX, Email: wenyi.hu@agtgeo.com
Jiefu Chen is with the Department of Electrical and Computer Engineering, University of Houston, Email: jchen84@uh.edu
Jianguo Liu is with Advanced Geophysical Technology, Houston, TX, Email: jianguo.liu@agtgeo.com
Aria Abubakar is with Schlumberger Sugar Land Technology Centers, Sugar Land, TX, Email: aabubakar@slb.com

AVA (amplitude variation with angle) workflow to analyze and process seismic data to determine the characteristics of formations [1]-[3]. The tomography-migration approach provides a low resolution velocity model with the correct kinematic information overlaid with a high resolution earth image containing the detailed subsurface structural information. This approach requires extensive user interactions, and the products of this approach, tomography velocity model and migration earth image, are not immediately ready for the next workflow such as rock property analysis and reservoir characterization. Furthermore, in complex geological environments, building a kinematically accurate velocity model is very challenging. These limitations of the tomography-migration approach, along with the rapid development of computation power, precipitated the significant progress of FWI research in the last decade. Although FWI has not matured to become a standard seismic processing procedure in the current industry, we have witnessed many successful FWI applications, where shallow velocity anomalies were resolved reliably and reservoir structures were reconstructed accurately. These capabilities, which the conventional seismic processing tools do not possess, are valuable for geohazard avoidance and better reservoir characterization.

FWI takes into account the full waveform information of seismic data to reconstruct an earth model including not only the geologic structures but also the geophysical properties. Properly taking care of various frequencies, offsets, azimuths, and other factors of seismic data, FWI is expected to generate the maps of geophysical properties continuously covering a broadband wavenumber spectrum [4]. Due to the extremely large number of unknowns, most FWI methods employ gradient-based local optimization algorithms [5]-[9] that may converge to a local minimum when the inversion starts at high frequencies or the starting velocity model is not sufficiently close to the true model. In current seismic data acquisition environments, reliable low frequency components below 5 Hz do not practically exist in typical seismic datasets [10]. Due to this fact, the local minimum problem, commonly referred to as the cycle-skipping issue, remains as one of the biggest challenges today to fully utilize the FWI capabilities. In recent years, many

research efforts have been devoted aiming to overcome this challenge and some of the achievements have greatly benefited from the advances in signal processing. From a signal processing perspective, one of the effective approaches to the cycle-skipping mitigation without acquiring low frequency data is to retrieve the low wavenumber information buried in the high frequency data. In this context, the key component of FWI research on cycle-skipping suppression is how to make low wavenumber information "visible" through various signal processing techniques.

There are some surveys in the literature that give excellent reviews of the FWI theory, algorithms, applications, and the cycle-skipping phenomenon [11], [12]. After a brief review of the conventional seismic processing technologies, we provide an introduction to the FWI theory, the fundamental mechanism behind the cycle-skipping phenomenon, the theoretical and practical aspects of various algorithms, workflows, and strategies developed to solve this issue, the essence of the underlying challenges and difficulties we are facing, and the implications for future directions of this research topic. This overview will serve as an accessible guide to both geophysicists and signal processing engineers to explore the potential opportunities in the FWI research for subsurface exploration.

Conventional Approach and FWI

In a conventional tomography-migration-AVA workflow, the first step is to build a smooth velocity model honoring the kinematic information only, as shown in Figure 1a (we need to mention that we present an exceptionally good model in Figure 1a for the purpose of delivering the concept). According to the wavenumber spectra, the tomography velocity model substantially deviates from the true velocity model (Figure 1c) in terms of spatial resolution because the tomography resolution is approximately determined by the Fresnel zone radius [13] [14], which is dependent not only on the wavelength, but also on the propagation length. In the second step, with the kinematically correct velocity model generated in the first step, the recorded seismic data are tracked backward from the receivers to the subsurface domain through a migration engine to

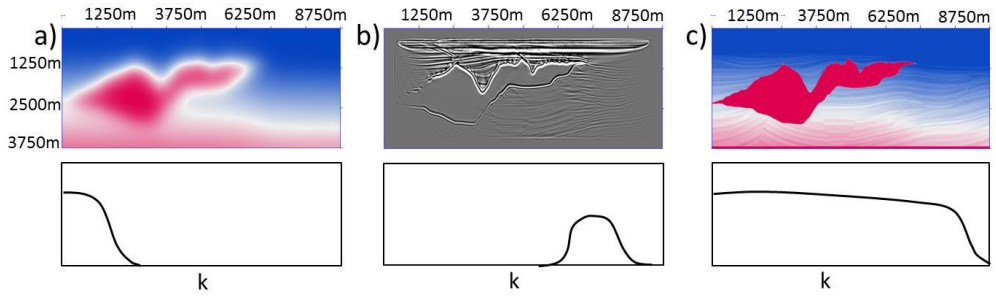

Figure 1. a) Tomography velocity model and its wavenumber spectrum; b) migration image and its wavenumber spectrum; c) true velocity model and its wavenumber spectrum.

accurately position the reflectors and the scattering points in the spatial domain, providing valuable subsurface structural information at high resolution. Combining the tomography velocity and the migration image, a broad wavenumber spectrum may be recovered but a significant spectrum gap always exists as shown in Figure 1a and Figure 1b. On the other hand, FWI attempts to fill this gap by delivering a velocity model with a continuous wavenumber spectrum similar to the true full-band velocity model in Figure 1c. The spectra shown in Figure 1 are schematic diagrams for ease of explanation.

Without losing generality, the cost function of a standard FWI can be defined in the frequency domain as follows:

$$C(\mathbf{v}) = \sum_{f=1}^{N_f} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \left\| S_{r,s,f}(\mathbf{v}) - M_{r,s,f} \right\|^2 , \tag{1}$$

where S corresponds to the simulated data, M corresponds to the measurement data, $\mathbf{v}$ contains the unknown model parameters such as seismic velocities, mass-density, etc., to be inverted, N_s, N_r, and N_f are the number of sources, receivers, and frequencies, respectively. In practice, an additional regularization cost function such as in [7], [9] is added in (1) to arrive at meaningful inverted parameters. The gradient of the cost function with respect to the model parameters can be calculated by performing a residual back-propagation procedure [5] to update the velocity model to minimize the data discrepancy. The velocity model update is carried out in an iterative manner until the data mismatch is within a predefined error tolerance or the gradient is small [7].

Nonlinearity, Local Minimum, and Cycle-skipping

Minimization of the cost function (1) is a nonlinear optimization problem and its level of nonlinearity increases with frequency [15]. The mathematical details of the nonlinearity analysis are described as follows.

Assuming the current velocity model is $\mathbf{v_0}$, to evaluate the cost function (1), the following governing wave equation is solved to obtain the simulated data,

$$\nabla^2 p_0 + \omega^2 s_0^{\,2} p_0 = -i\omega\rho Q, \tag{2}$$

where p_0 is the wavefield, $s_0 = 1/v_0$ is the slowness, and Q is the source. To construct the sensitivity matrix, the slowness model s_0 is perturbed as

$$\nabla^2(p_0 + \delta p) + \omega^2(s_0^{\,2} + \delta s^2)(p_0 + \delta p) = -i\omega\rho Q \tag{3}$$

where δp is the wavefield response to the perturbation of the slowness model δs^2, i.e., the scattered wavefield. Subtracting equation (2) from (3), we have

$$\nabla^2 \delta p + \omega^2 s_0^{\,2} \delta p = -\omega^2 \delta s^2 \left[p_0(\mathbf{r}', \mathbf{r}_s) + \delta p(\mathbf{r}', \mathbf{r}_s)\right], \tag{4}$$

where $\mathbf{r}$' is the subsurface point where the slowness is perturbed by δs^2 and $p_0(\mathbf{r}', \mathbf{r}_s)$ represents the wavefield response at $\mathbf{r}$' to the source located at $\mathbf{r}_s$. In a standard FWI algorithm, with the weak scattering assumption that $\delta p << p_0$, the linear approximation of the sensitivity kernel can be derived by dropping δp from the right side of equation (4):

$$\frac{\delta p}{\delta s^2} \approx -\omega^2 \int_{\mathbf{r}'} G(\mathbf{r}_r, \mathbf{r}') p_0(\mathbf{r}', \mathbf{r}_s), \tag{5}$$

where G is the Green's function [16].

In many situations, a starting model is not sufficiently close to the true model (not within the basin of attraction of the global minimum) [12] and the weak scattering assumption is violated, hence the above linear approximation is no longer justified and the algorithm may converge to a local minimum v_L instead of the true solution v_T, as shown in Figure 2. An effective strategy for handling this issue is to initiate the FWI under the lower frequency regime, where the non-

convexity nature of the cost function diminishes due to the alleviated nonlinearity as shown in Figure 2. Consequently, the same starting model v_0 may converge toward the global minimum at the low frequency, thus finding an approximate solution v_A. Next, the algorithm switches back to the high frequency regime using v_A as the new starting model, enabling the algorithm to find the true solution v_T.

In FWI research, the local minimum issue is commonly observed as the cycle-skipping phenomenon, which may occur when the simulated data and measured data are out of phase for more than a half period. A cycle-skipping phenomenon might be prevented if the optimization is performed at a lower frequency to reduce the phase mismatch within a half period. Unfortunately, in exploration geophysics, low enough frequency data are practically unavailable. Due to this fact, cycle-skipping has been recognized as one of the main challenges in FWI applications. Since FWI is essentially an operator mapping the surface data domain to the subsurface spatial domain, the importance of low frequency data is manifested by the established link between frequency and wavenumber $k=\omega/v$. From this perspective, if the low wavenumber components in the subsurface model can be reconstructed reliably without using any low frequency data, the cycle-skipping problem can be solved. With this observation and this strategy, our task is redefined as: how to retrieve low wavenumber information buried in high frequency seismic data? To find the answer, first we need to investigate the spatial resolution of FWI.

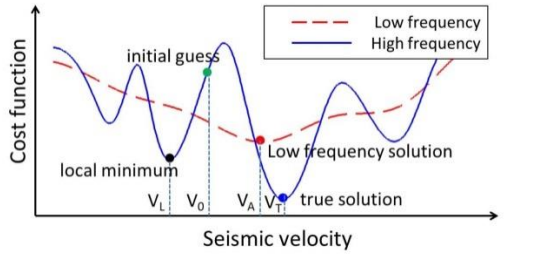

Figure 2. A one variable optimization problem.

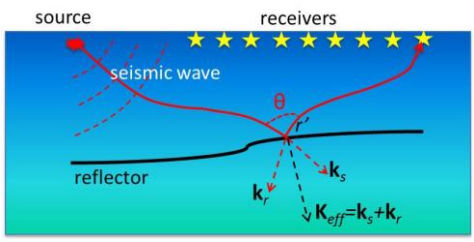

Figure 3. FWI resolution analysis.

Spatial Resolution of FWI

For a specific source *m* and a receiver *n*, the Fréchet derivative of the function **S**(**v**) in the cost function (1) with respect to the velocity *v'* at the subsurface point **r**' is

$$\partial S_{m,n} / \partial v' = \gamma G(\mathbf{r}', \mathbf{r}_m) G(\mathbf{r}', \mathbf{r}_n), \quad (6)$$

where $G(\mathbf{r}', \mathbf{r})$ is the Green's function denoting the response at **r**' to an impulse excited at location **r**, γ is a complex coefficient. Discarding the amplitude and phase information and assuming the far field approximation, the sensitivity kernel in (6) can be represented by

$$G(\mathbf{r}', \mathbf{r}_m) G(\mathbf{r}', \mathbf{r}_n) = A_s e^{i\mathbf{k}_s \cdot \mathbf{r}'} A_r e^{i\mathbf{k}_r \cdot \mathbf{r}'} = A e^{i\mathbf{k}_{eff} \cdot \mathbf{r}'}, \quad (7)$$

where A_s and A_r are the amplitude of the incident wave and the scattered wave, $\mathbf{k}_{eff}=\mathbf{k}_s+\mathbf{k}_r$ is the effective wavenumber dictating the model reconstruction spatial resolution. Namely,

$$k_{eff} = (2\omega / v') \cos(\theta / 2), \quad (8)$$

where θ is the subsurface scattering angle [11], as shown in Figure 3. According to equation (8), frequency is not the only factor that determines the spatial resolution of FWI. The subsurface scattering angle θ plays an important role too. With a wide range of scattering angle (e.g., sources and receivers located around the targets), seismic data at a single frequency should be sufficient to cover a wideband spectrum of wavenumbers. This observation directly led to the technique of scattering angle based filtering for low wavenumber component retrieval [17].

Scattering angle based filtering method

Based on formulation (8), for monochromatic seismic waves, the effective wavenumber can be precisely controlled by selecting the subsurface scattering angle. In other words, by allowing only a specific portion of the seismic energy with large scattering angles to contribute to the velocity update, a subsurface reconstruction free of cycle-skipping might be possible. After that, more data with wider scattering angle range can be fed into the inversion engine to resolve the finer features. To apply this idea to practical scenarios, the scattering angle based filtering method was developed to precondition the gradient, in purpose of retaining only low wavenumber information [17]. The key component of the scattering angle based filtering method is the extension of the

standard FWI gradient by adding a velocity normalized time lag ζ. This extra dimension in the gradient allows one to extract the scattering angle information at any subsurface location. With this technique, the FWI gradient g(**r**') becomes the extended version

$$g^E(\mathbf{r}',\zeta)=\omega^2\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}G(\mathbf{r}',\mathbf{r}_s)u_r(\mathbf{r}')e^{-4i\omega\frac{\zeta}{v(\mathbf{r}')}}, \tag{9}$$

where the normalized time lag $\zeta=v(\mathbf{r}')\tau/4$, τ is a time lag variable, and u_r is the back-propagated residual. We also have

$$\cos^2(\theta/2)=|\mathbf{k}|^2/k_\zeta^2, \tag{10}$$

where **k** is the effective wavenumber vector and k_ζ is the wavenumber corresponding to the normalized time lag ζ. Formulation (10) serves as a scattering angle filter to identify and exclude the information buried in the extended gradient associated with small scattering angles. The extended gradient after the filtering has the form

$$g^{FE}(\mathbf{r}',\zeta)=F^{-1}w(\mathbf{k},v,\omega)Fg^E(\mathbf{r}',\zeta), \tag{11}$$

where F and F^{-1} are the Fourier and inverse Fourier transform, and the window function w is

$$w(\mathbf{k},v,\omega)=\begin{cases}1, & \cos(\theta/2)>v|\mathbf{k}|/\omega\\ 0, & \cos(\theta/2)\le v|\mathbf{k}|/\omega\end{cases}. \tag{12}$$

After the filtering, the summation of the remaining energy in the extended gradient is expected to contain only the low wavenumber information. An attractive feature of this method is the quantitative control over the reconstruction resolution.

Instead of using the extended gradient approach, the scattering angle filtering technique has also been practiced through a local plane wave decomposition method [18]. Namely, after both the extrapolated source-side and the receiver-side wavefields are decomposed into local plane waves characterized by their propagation direction angles θ_s and θ_r, a predefined angle domain selective filter is applied to construct a gradient with controlled wavenumber components:

$$g(\mathbf{r}') = \frac{2}{v^3(\mathbf{r}')} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T \sum_{\theta_s} \sum_{\theta_r} A(\theta_s, \theta_r) \left[\hat{G}(\mathbf{r}', \mathbf{r}_s, t, \theta_s) \hat{u}_r(\mathbf{r}', \mathbf{r}_r, t, \theta_r) \right] dt , \quad (13)$$

where $\wedge$ represents the wavefield plane wave decomposition. A denotes the angle domain selective filter. Although this approach extracts the scattering angle information without introducing an additional dimension in the gradient, thus avoid the 4-D FFT, the plane wave decomposition procedure and the angle domain filtering need to be performed at every time step and every subsurface location, hence the computational expense might be prohibitive.

FWI with extended model space

Similar to the gradient extension, another category of FWI methods extends the velocity model to address the cycle-skipping issue from a different perspective. Without extracting the subsurface scattering angle information, this category of FWI applies the model extension along the time lag axis to linearize the wave equation. The idea of velocity model extension to overcome the nonlinearity issue was originated from the differential semblance optimization [19], a strategy proposed for wave-equation based migration velocity analysis, and later was generalized to FWI to reduce the nonlinearity [20].

With the observation that, under some circumstances in FWI applications, the discrepancy between the total wavefield and the background wavefield (or the incident wavefield) can often be approximately parameterized by a traveltime shift τ, i.e.,

$$p_0(\mathbf{r}', \mathbf{r}_s, 0) + \delta p(\mathbf{r}', \mathbf{r}_s, 0) \approx p_0(\mathbf{r}', \mathbf{r}_s, \tau), \quad (14)$$

thus the validity of the linear relationship between the model perturbation and the wavefield response is retained by extending the model space in an extra dimension τ

$$\nabla^2 \delta p + \omega^2 s_0^{\,2}(\tau = 0) \delta p = -\omega^2 \delta s^2(\tau) \tilde{*} p_0(\mathbf{r}', \mathbf{r}_s), \quad (15)$$

where $\tilde{*}$ denotes the convolution in the time shift dimension τ. Correspondingly, the cost function of the FWI problem with the extended velocity model space can be posed as

$$C(\mathbf{s}) = \sum_{f=1}^{N_f}\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\left\| S_{r,s,f}(\mathbf{s}) - M_{r,s,f} \right\|^2 + \lambda\left\| \tau|\mathbf{s}^2 \right\|^2 \quad . \tag{16}$$

This method is called tomographic full waveform inversion [21].

Comparing (16) with the standard FWI cost function (1), we note two main differences: 1) the velocity model in the tomographic FWI has four dimensions with an extension in the time shift parameter τ; 2) there is a penalty term in the cost function of the tomographic FWI, whose function is to enforce the time shift τ to gradually converge to 0. In other words, this method allows more degrees of freedom within the inversion process to incorporate the kinematic correction into the waveform data fitting procedure, but at an additional cost of the extension.

FWI with time shift minimization

As discussed previously, when simulated data and measured data are out of phase for more than a half period, a cycle-skipping issue is inevitable using a standard FWI algorithm. Based on this fact, another category of solutions to the cycle-skipping problem was developed, i.e., the traveltime inversion combined with a standard FWI algorithm, where the former is capable of consisting the low wavenumber components and the latter is able to resolve the high wavenumber details. One realization of this strategy is called the wave-equation reflection traveltime inversion algorithm (WERTI) [22]. WERTI seeks to reliably estimate the traveltime shift by introducing the dynamic image warping (DIW), an extension to the dynamic warping algorithm for speech recognition. WERTI finds the traveltime shifts by solving a constrained optimization problem

$$\begin{aligned} &\tau(\mathbf{r}_r, t; \mathbf{r}_s) = \arg\min_l \frac{1}{2}\int d\mathbf{r}_s d\mathbf{r}_r dt\left[p^s(\mathbf{r}_r, t; \mathbf{r}_s) - p^m(\mathbf{r}_r, t + l(\mathbf{r}_r, t; \mathbf{r}_s); \mathbf{r}_s)\right]^2 \\ &\left|\partial\tau/\partial t\right| \le \sigma_t, \quad \left|\partial\tau/\partial\mathbf{r}_r\right| \le \sigma_r, \quad \left|\partial\tau/\partial\mathbf{r}_s\right| \le \sigma_s \end{aligned} \tag{17}$$

where l is the shift field. Here, p^s is the simulated data and p^m is the measured data. This optimization problem can be solved through a dynamic programming method [23]. σ_t, σ_r, and σ_s are the parameters of the constraints for controlling the rates of the shift variation in directions t,

$\mathbf{r}_r$, $\mathbf{r}_s$. The estimated time shift τ is then fed into another inverse problem for the traveltime shift minimization with the following cost function:

$$C\left(\mathbf{s}^b,\mathbf{s}^r\right)=\frac{1}{2}\sum_s\sum_r\int dt\tau^2\left(\mathbf{r}_r,t,\mathbf{r}_s\right). \tag{18}$$

The cost function (18) decomposes the slowness model into a smooth background model $\mathbf{s}^b$ honoring the kinematic information and a rough reflection model $\mathbf{s}^r$. A gradient based optimization algorithm is applied to update $\mathbf{s}^b$ and $\mathbf{s}^r$ in an alternating manner. With this strategy, the high wavenumber structures are gradually shifted to the correct positions without being trapped into the cycle-skipping ambiguity because the kinematic information contained in the background velocity becomes more and more accurate while the inversion process proceeds.

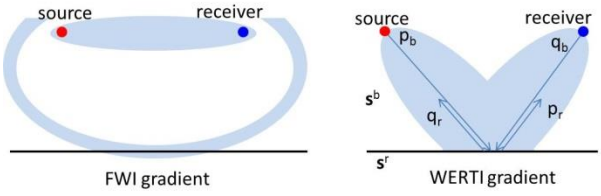

Figure 4. Spatial coverage of gradients of FWI and WERTI.

The explicit form of the gradient of (18) with respect to $\mathbf{s}^b$ is

$$\nabla_{\mathbf{s}^b}C=-\int d\mathbf{r}_s dt\mathbf{q}_r\mathbf{p}_b-\int d\mathbf{r}_s dt\mathbf{q}_b\mathbf{p}_r \tag{19}$$

with $\mathbf{p}_b$ and $\mathbf{p}_r$ denoting the source-side wavefields and $\mathbf{q}_b$ and $\mathbf{q}_r$ denoting the receiver-side back-propagated wavefields. The spatial coverage of the resulting gradient is schematically shown in Figure 4. Within a WERTI $\mathbf{s}^b$ update, the reflection slowness model $\mathbf{s}^r$ serves as a set of secondary sources producing the wavefields $\mathbf{q}_b$ and $\mathbf{q}_r$, thus converting a reflection-like propagation into a transmission-like propagation between the source and the scattering point, and between the receiver and the scattering point. Because reflection seismic energy is generally associated with high wavenumber components while transmission energy is associated with relatively low wavenumber components, with these transmission modes, one can expect to recover the low wavenumber components in the background model to avoid the cycle-skipping problem.

The same idea has been exploited by [24] in their development of the so-called reflection waveform inversion (RWI). Similarly, RWI splits the velocity model into a high wavenumber component model called the perturbation and a low wavenumber component model called the reference model. In its iterative inversion process, the perturbation model is generated through a true-amplitude migration using the fixed reference model for the wave propagation. After that, de-migration is implemented to generate a set of simulated data and this dataset is used for the residual calculation to optimize the reference velocity. This workflow shares the same philosophy of the migration velocity analysis (MVA) except that the ray-based wave propagation engine is replaced by the wave equation based engine and no residual moveout (NMO) picking is required. One of the main limitations of RWI is the lack of low wavenumber information in the shallow regions, which is dominantly contributed from the diving waves excluded by RWI. Natural integration of FWI and RWI within an inversion process may overcome this limitation [25].

Adaptive matching filter (AMF) based FWI is another variant of this methodology [26]. AMF is a filter with the capability of measuring the nonstationary phase differences between two signals. Inserting the AMF $\mathbf{f}(\gamma,t)$ into the cost function of FWI, we have the following least square inverse problem, where γ is the stretching variable:

$$\min\left\|\sum_{\gamma}\mathbf{s}(t)-\mathbf{m}(\gamma t)\mathbf{f}(\gamma,t)\right\|^2. \tag{20}$$

Another similar algorithm is the adaptive waveform inversion (AWI) [27], where a Wiener filter is employed to measure the time shift between the simulated data and the measured data to reconstruct a tomography-like result to prevent the cycle-skipping. AWI consists of two stages. In the first stage, a Wiener filter is designed through the minimization of the following objective function to match the simulated trace p^s to the recorded trace p^m

$$C_1=\left\|p^s*w-p^m\right\|^2. \tag{21}$$

In the second stage, another minimization problem is solved to obtain the model. The corresponding objective function is

$$C_2 = \|Tw\|^2 / \|w\|^2 , \quad (22)$$

where T are the weights of the filter coefficients. AWI was successfully applied to a 3D field dataset without a good starting velocity model [27].

Because the time shift minimization based approaches separate the information contained in the seismic data into kinematic part and waveform part, spectrum gaps might be expected in the reconstructed models.

FWI with synthesized low frequency data

Realizing the failures in FWI practice on seismic data are mainly caused by its high nonlinearity related to the lack of low frequency components, some researchers directed their efforts toward synthesizing artificial low frequency data using high frequency data via mathematical transformations. Laplace domain FWI [10] applies the Laplace transform embedded in its forward modeling engine to the wavefield $u(t)$, i.e., $\tilde{u}(s) = \int_0^\infty u(t)e^{-st}dt$. The resulting wavefield $\tilde{u}(s)$ in the Laplace domain are equivalent to the zero frequency component of the damped wavefield in the time domain while the damping effect is determined by the Laplace damping constant s. The inversion part of the Laplace domain FWI is essentially same as a standard FWI except that the data misfit is evaluated based on the logarithmic ratio of the simulated wavefields and the measured wavefields, both in the Laplace domain. The cost function is defined as

$$C(\mathbf{v}) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \left\| \ln\left[\tilde{S}_{r,s}(\mathbf{v})\right] - \ln\left[\tilde{M}_{r,s}\right] \right\|^2 , \quad (23)$$

where $\tilde{S}$ and $\tilde{M}$ denote the simulated and measured wavefields in the Laplace domain respectively. Because the Laplace domain FWI inverts the synthesized seismic data at zero frequency, the inversion process resembles the inversion of Poisson's equation and a smooth model is expected. This analysis is consistent with the numerical experiment reported by [10]. The same idea naturally leads to the Laplace-Fourier domain FWI [28], where the zero frequency components in

the Laplace domain FWI is generalized to low frequencies by $\tilde{u}(s) = \int_0^\infty u(t)e^{-i\omega t}e^{-\sigma t}dt$, and σ is the Laplace damping constant. Theoretically, any low frequency components can be synthesized by the Laplace-Fourier transformation. With a workflow in the order of Laplace domain FWI, Laplace-Fourier domain FWI, and a conventional frequency domain FWI, the low wavenumber components, the intermediate wavenumber components, and the high wavenumber components can be reconstructed sequentially to cover a continuous broadband wavenumber spectrum in the inverted velocity model. The Laplace FWI has been applied to the field data acquired in the Gulf of Mexico to reconstruct a large scale salt structure, yielding encouraging inversion results nearly independent of the frequency content of the data [29]. A successful Ocean Bottom Seismic (OBS) data inversion case study using the logarithmic phase misfit in the Laplace domain was reported in [30].

Another nonlinear operator based method is the envelope inversion [31]. The envelope of seismic data shows relatively less fluctuation than the original time domain seismic data, implying that it may carry some ultra-low frequency information induced by the ultra-low wavenumber structures in the subsurface. The cost function of the envelope FWI is defined as

$$C(\mathbf{v}) = \sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\int_0^T \left\| E\left[S_{r,s}(\mathbf{v},t)\right]^p - E\left[M_{r,s}(t)\right]^p \right\|^2 dt, \tag{24}$$

where p is a data preconditioning factor to apply the weight as a function of time. The instant envelope extraction operator E applied on signal $f(t)$ is defined by

$$E(f(t)) = \sqrt{f^2(t) + f_H^2(t)}, \tag{25}$$

where $f_H(t)$ is the Hilbert transform of the real function $f(t)$. Again, as other approaches for cycle-skipping suppression, an envelope FWI is usually followed by a standard FWI to build the high wavenumber structures on top of the envelope inversion results with ultra-low spatial resolution.

Relying on the DC and low frequency information augmented by the nonlinear operators applied on the seismic data amplitudes, the Laplace domain FWI and the envelope FWI are

intrinsically sensitive to the signal-to-noise ratio of seismic data, casting doubts on their robustness in general field data applications.

The FWI algorithm based on extrapolated low frequency data, the so-called EFWI, also falls into this category [32]. The EFWI exploits a nonlinear signal processing technique to decompose recorded seismic data into atomic events, each of which is defined by a smooth phase function and a smooth amplitude function. This nonlinear signal processing technique, the so-called phase tracking, can be posed as a nonlinear least-square optimization problem:

$$J(a_n, b_n) = \frac{1}{2}\|u(\omega,x) - d(\omega,x)\|^2 + \lambda_1 \sum_n \left\|\nabla_\omega^2 b_n(\omega,x)\right\|^2 + \lambda_2 \sum_n \left\|\nabla_x b_n(\omega,x)\right\|^2 + \lambda_3 \sum_n \left\|\nabla_{\omega,x} a_n(\omega,x)\right\|^2 \quad (26)$$

$$u(\omega,x) = \sum_n^N w(\omega) a_n(\omega,x) e^{ib_n(\omega,x)} \quad (27)$$

where d is the measured data, u is the predicted data constructed by the atomic events, ∇ denotes the gradient, $w(\omega)$ is the known wavelet, λ_1, λ_2, λ_3 are the empirically determined regularization parameters. The cost function (26) is minimized using a gradient based algorithm to obtain the amplitude function a_n and the phase function b_n. After that, with the nondispersive earth assumption, another optimization is performed to obtain the following approximations:

$$a_n(\omega,x) \approx \alpha_n(x); \qquad b_n(\omega,x) \approx \omega\beta_n(x) + \phi_n(x)\,. \quad (28)$$

With the approximation (28), one is able to synthesize any flat-spectrum atomic events outside of the frequency band of the recorded seismic data by

$$u^e(\omega^e, x) = \sum_n^N w(\omega^e) a_n(x) e^{i\left(\omega^e \beta_n(x) + \phi_n(x)\right)}, \quad (29)$$

where ω^e is the frequency outside of the observed data frequency band and u^e is the extrapolated low frequency data. Although the Marmousi model numerical experiment shows some successful inversion results, the multiple assumptions made in this method need to be further justified through extensive synthetic data and field data testing.

FWI immune to phase ambiguity

In the context of frequency domain data fitting, the root cause of cycle-skipping phenomenon in FWI can be partially removed if both measured and simulated data are preprocessed by a phase unwrapping tool. This group of approaches [33] [34] resolving the ambiguity in the data domain without touching the FWI algorithm can potentially be effective but the development of the phase unwrapping algorithm are extremely challenging, especially for 3D.

Inspired by interference beat tone, an acoustic phenomenon commonly used by musicians for turning check, The Beat Tone FWI was developed to suppress the cycle-skipping phenomenon [35]. The idea of the Beat Tone FWI is based on the phase modulation / demodulation concept. This method utilizes two seismic datasets extracted at two slightly different high frequencies. The interferences between the two datasets are exploited by the cost function

$$C(\mathbf{v}) = \sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\left\|\Phi\left[S_{r,s}(\mathbf{v},\omega_2)/S_{r,s}(\mathbf{v},\omega_1)\right] - \Phi\left[M_{r,s}(\omega_2)/M_{r,s}(\omega_1)\right]\right\|^2, \tag{30}$$

where Φ is the phase extraction operator, ω_1 and ω_2 are two frequencies extracted from the

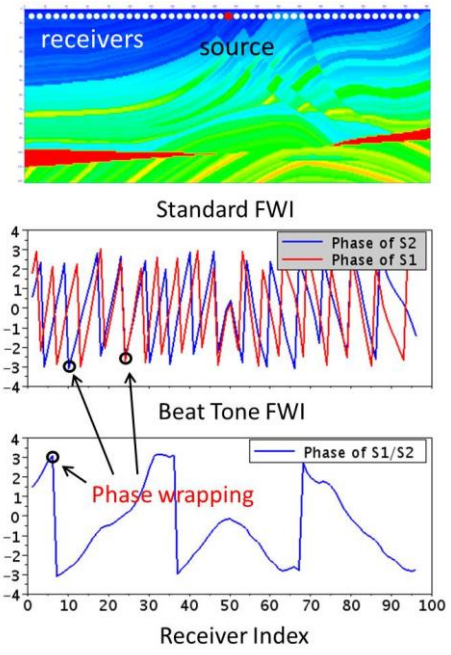

Figure 5. True velocity model; phases of data S_1 measured at 8.5 Hz and S_2 measured at 7 Hz; and phase difference between S_1 and S_2, serving as Beat Tone FWI data.

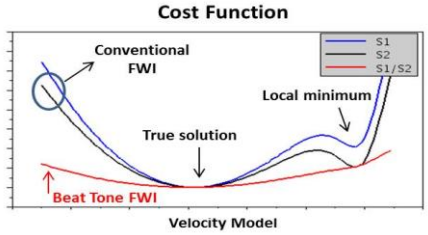

Figure 6. Diagram showing the mechanism of local minimum mitigation using Beat Tone FWI.

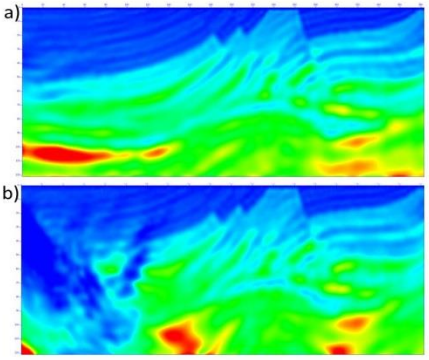

Figure 7. a) Beat Tone FWI; b) conventional FWI.

measured data with $|\omega_2 - \omega_1| \ll \min(\omega_2, \omega_1)$. To understand the mechanism of the Beat Tone FWI, assuming there are two seismic signals with the plane wave approximation: $u_1 = A_1 e^{i\int k_1 dr}$ and $u_2 = A_2 e^{i\int k_2 dr}$, we have $u_1/u_2 = (A_1/A_2)e^{i\int (k_1 - k_2) dr}$. Since we abandon the amplitude information in the cost function (30), the spatial resolution of the gradient of the cost function is mainly determined by the back-propagated data residual and the subsurface scattering angle, replacing the effective wavenumber formulation (8) by

$$k_{eff}^{BT} = 2\big(k_1(\mathbf{r}') - k_2(\mathbf{r}')\big)\cos(\theta/2) = \big(2|\omega_1 - \omega_2|/v'\big)\cos(\theta/2). \quad (31)$$

In other words, the mapping from the low-frequency data to the low-wavenumber model established by the wave equation is now replaced by the mapping from the frequency-difference in the data to the low-wavenumber model. The effect of phase ambiguity resolving of this method can be observed in Figure 5, where S_1 is the data recorded at 8.5 Hz and S_2 is recorded at 7 Hz. The number of phase wrapping occurrences is significantly reduced with the Beat Tone FWI strategy and low wavenumber structures are expected to be reconstructed reliably. Accordingly, the number of local minima may reduce as well, as shown in Figure 6. In Figure 7, the velocity models reconstructed by a conventional FWI method and the Beat Tone FWI are compared, where the strong artifacts induced by the cycle-skipping are observed in the conventional FWI result. On the other hand, with the same seismic datasets, the Beat Tone FWI successfully mitigated the cycle skipping issue. One main disadvantage of the Beat Tone FWI is the amplified scattering effect observed as a noise-like pattern existing in the objective function when the two selected frequencies are too close.

Discussion and Conclusions

In this paper, we reviewed an advanced seismic data processing technology, the full waveform inversion, with focus on one of its main challenges known as the cycle-skipping phenomenon which prevents the FWI technology from being widely applied for subsurface exploration. After analyzing the relationship between cycle-skipping, local minimum, and nonlinearity, and

investigating the spatial resolution formulation of the conventional FWI algorithm, we identify that one of the solutions to the cycle-skipping problem is to retrieve the low wavenumber information buried in the band limited recorded seismic data without resorting to extra low frequency data. Although discussing all the methods of cycle-skipping mitigation is not possible, this article gives a reasonable review of five main categories of these approaches.

The first category is the FWI incorporated with a subsurface scattering angle based filter. This method was inspired by the observation that the effective wavenumber of a subsurface geophysical model (i.e., the wavenumber of model reconstruction), is mainly determined by the frequency of the seismic data, the velocity at the scattering point, and the subsurface scattering angle. While the frequencies of the data and the velocity model are fixed, manipulation of the scattering angles by selecting the wave energy within a specific propagation direction range helps emphasize the low wavenumber structures and prevent the high wavenumber components from contributing in the inversion process at the early stages. Usually large scattering angles are associated with seismic data with long offsets covering many wavelengths. Hence, under these conditions, the linear approximation of the sensitivity kernel may no longer be justified. Therefore, we are confronted with a dilemma of alleviating the level of nonlinearity on one side while aggravating the nonlinearity on the other side. The advantage of this method is that the spatial resolution of the reconstructed model can be quantitatively controlled.

The second category is the FWI with extended velocity model and the third category is the FWI with traveltime shift minimization. These two categories share a similar strategy, combination of a traveltime tomography and an FWI. Both categories recognize that the cycle-skipping issue can be avoided if the kinematic information is corrected in the velocity model before the full amplitude and phase information participate in the inversion for the fine geometry feature reconstruction. The FWI with an extended model realizes this strategy by introducing an additional time shift dimension into the velocity model and gradually enforcing the time shift towards zero as the inversion proceeds. On the other hand, the FWI with traveltime shift

minimization directly works in the data domain to correct the kinematic information by inverting the traveltime difference between the simulated and the measured data estimated by various signal processing techniques. The limitation of these two approaches is that the inversion consists of two parts, the tomographic part and the standard FWI part. The spatial resolution of the former is mainly determined by the first Fresnel zone while the resolution of the latter is dictated by the data frequencies. Consequently, the overall wavenumber spectrum can be uneven or discontinuous, depending on the survey geometry and the seismic data frequency spectra. A major concern of the model extension based methods is the significant additional computational cost.

Instead of digging into high frequency seismic data to recover low wavenumber information, another category of approaches attempts to achieve this goal by synthesizing low frequency data using nonlinear transformations. However, these low frequency data do not physically exist. The Laplace domain FWI and the envelope FWI apply nonlinear operators to shape the seismic waveform for the low frequency data augmentation. Although the limited field data testing examples show some encouraging inversion performance, the sensitivity to noise and other uncertainties becomes a main concern. EFWI decomposes the measurement data to atomic events parameterized by the approximate amplitude functions and the approximate phase functions. These atomic events serve as the Green's functions which are assumed to be expandable across the whole frequency band under the smooth phase and smooth amplitude assumptions. Without rigorous theoretical justifications and thorough numerical validation testing, these assumptions may raise substantial doubts about the validity of this approach in field data applications.

The fifth category of the algorithms reviewed in this paper tries to remove the root cause of the cycle-skipping, the phase wrapping phenomenon observed in the phase pattern shown in seismic data. A direct phase unwrapping procedure applied on seismic data is a natural solution, but the technical difficulty remains to be the bottleneck. The Beat Tone FWI, on the other hand,

implicitly suppresses the phase wrapping by utilizing the interferences between two datasets extracted at two slightly different frequencies. These two datasets are injected into the Beat Tone FWI engine to retrieve the low wavenumber information hidden in the interference between the two relatively high frequencies. With this strategy, the number of phase wrapping occurrences is substantially reduced and many local minima are eliminated. The limitation of this approach is that the phase difference between two datasets separated by two slightly different frequencies shows different pattern from that of a true low frequency dataset. Further study is needed in the future to explore the cause of this discrepancy and the potential solutions to suppress these noise-like phenomena.

Till today, the cycle-skipping issue has not been completely solved although the above mentioned approaches mitigate the issue to some extent. The purpose of this article is to help geophysicists and signal processing engineers to have a clear understanding of the current research status and the major challenges existing in this inter-discipline area, which is apparently critical for identifying the potential opportunities of innovation for this interesting and important task in seismic processing technology development along the journey of subsurface exploration.

References

[1] K. Aki and P. Richards, "Quantitative Seismology," 2nd Ed. University Science Books, ISBN 0-935702-96-2, 704pp, 2002.

[2] M. J. Woodward, D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, "A decade of tomography," *Geophysics*, vol. 73, no. 5, pp.VE5-VE11, 2008.

[3] Öz Yilmaz, "Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data," Society of Exploration Geophysicists, ISBN (print): 978-1-56080-094-1, ISBN (online): 978-1-56080-158-0, 2001.

[4] T. Alkhalifah, "Full model wavenumber inversion (FMWI): an emphasize on the appropriate wavenumber continuation," *Geophysics*, 81, R89-R98, 2016.

[5] G. Pratt and C. Shin, "Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion," *Geophysical Journal International*, vol. 133, no. 2, pp. 341-362, 1998.

[6] A. Abubakar, W. Hu, T. M. Habashy, and P. M. van den Berg, "Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data," *Geophysics*, vol. 74, no. 6, pp. WCC47-WCC58, 2009.

[7] W. Hu, A. Abubakar, and T. M. Habashy, "Simultaneous multifrequency inversion of full-waveform seismic data," *Geophysics*, vol. 74, no. 2, pp. R1-R14, 2009.

[8] A. Abubakar, G. Pan, M. Li, L. Zhang, T. M. Habashy, and P. M. van den Berg, "Three-dimensional seismic full waveform inversion using the finite difference contrast source inversion method," *Geophysical Prospecting*, vol. 59, no. 5, pp. 874-888, 2011.

[9] W. Hu, A. Abubakar, T. M. Habashy, and J. Liu, "Preconditioned non-linear conjugate gradient method for frequency domain full waveform seismic inversion," *Geophysical Prospecting*, vol. 59, no. 3, pp. 477-491, 2011.

[10] C. Shin and Y. H. Cha, "Waveform inversion in the Laplace domain," *Geophysical Journal International*, vol. 173, no. 3, pp. 922-931, 2008.

[11] J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCC1-WCC26, 2009.

[12] T. Alkhalifah, "Full waveform inversion in an anisotropic world: where are the anisotropic parameters hiding?" EAGE Publications, ISBN Number: 9789462821927, 2016.

[13] P. R. Williamson, "A guide to the limits of resolution imposed by scattering in ray tomography," *Geophysics*, vol. 56, no. 2, pp. 202-207, 1991.

[14] T. Alkhalifah, “Insights into the data dependency on anisotropy: an inversion prospective,” *Geophysical Prospecting*, 64, 505–513, 2016.

[15] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, 60, 1457–1473, 1995.

[16] R. Gunther and B. Biondi, “Ignoring density in waveform inversion,” *STANFORD EXPLORATION PROJECT*, p.71, 2007.

[17] T. Alkhalifah, "Scattering-angle based filtering of the waveform inversion gradients," *Geophysical Journal International*, vol. 200, no. 1, pp. 363-373, 2015.

[18] X. Xie, "An angle-domain wavenumber filter for multi-scale full-waveform inversion," In *SEG Technical Program Expanded Abstracts 2015*, pp. 1132-1137, 2015

[19] W. W. Symes and J. J. Carazzone, "Velocity inversion by differential semblance optimization," *Geophysics*, vol. 56, no. 5, pp. 654-663, 1991.

[20] W. W. Symes, "Migration velocity analysis and waveform inversion," *Geophysical prospecting*, vol. 56, no. 6, pp. 765-790, 2008.

[21] B. Biondi and A. Almomin, "Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion," *Geophysics*, vol. 79, no. 3, pp. WA129-WA140, 2014.

[22] Y. Ma and D. Hale, "Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion," *Geophysics*, vol. 78, no. 6, pp. R223-R233, 2013.

[23] D. Hale, "Dynamic warping of seismic images," *Geophysics*, vol. 78, pp. S105-S115, 2013.

[24] S. Xu, D. Wang, F. Chen, G. Lambaré, and Y. Zhang, "Inversion on reflected seismic wave," In *SEG Technical Program Expanded Abstracts 2012*, pp. 1-7, 2012.

[25] T. Alkhalifah and Z. Wu, “The natural combination of full and image-based waveform inversion,” *Geophysical prospecting*, 64, 19–30, DOI: 10.1111/1365-2478.12264, 2016.

[26] H. Zhu and S. Fomel, "Building good starting models for full-waveform inversion using adaptive matching filtering misfit," *Geophysics*, vol. 81, no. 5, pp. U61-U72, 2016.

[27] M. Warner and L. Guasch, “Robust adaptive waveform inversion,” In *SEG Technical Program Expanded Abstracts 2015*, pp. 1059-1063, Society of Exploration Geophysicists, 2015.

[28] C. Shin and Y. H. Cha, "Waveform inversion in the Laplace–Fourier domain," *Geophysical Journal International*, vol. 177, no. 3, 1067-1079, 2009.

[29] W. Ha, and C. Shin, “Laplace-domain full-waveform inversion of seismic data lacking low-frequency information,” *Geophysics*, 77(5), pp.R199-R206, 2012.

[30] R. Kamei, G. R. Pratt, and T. Tsuji, “On acoustic waveform tomography of wide-angle OBS data—strategies for pre-conditioning and inversion,” *Geophysical Journal International*, 194(2), pp.1250-1280, 2013.

[31] R. Wu, J. Luo, and B. Wu, "Seismic envelope inversion and modulation signal model," *Geophysics*, vol. 79, no. 3, pp. WA13-WA24, 2014.

[32] Y. Li and L. Demanet, "Full-waveform inversion with extrapolated low-frequency data," *Geophysics*, vol. 81, no. 6, pp. R339-R348, 2016.

[33] N. K. Shah, M. R. Warner, J. K. Washbourne, L. Guasch, and A. P. Umpleby, "A phase-unwrapped solution for overcoming a poor starting model in full-wavefield inversion," In *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, 2012.

[34] Y. Choi and T. Alkhalifah, “Unwrapped phase inversion with an exponential damping,” *Geophysics*, 80, R251-R264, 2015.

[35] W. Hu, "FWI without low frequency data - beat tone inversion," In *2014 SEG Annual Meeting*. Society of Exploration Geophysicists, 2014.