Frequency domain nonlinear full-waveform inversion

Yu Geng, Kris Innanen

ABSTRACT

Based on a data-fitting procedure, full waveform inversion (FWI) aims to build high resolution subsurface structures using full waveform information. Although FWI is a highly nonlinear inverse problem, it is usually solved as a local optimization problem under a linear approximation. The gradient calculation of FWI during each iteration is usually studied with the aid of sensitivity kernel, or Fréchet derivative. Recently applications of FWI show that FWI can successfully build high resolution models in shallow regions, where long-to-intermediate wavelength structures can be reconstructed from diving waves and post-critical reflections. When first order scattering is considered during the construction of sensitivity kernel, recently developed reflection waveform inversion (RWI) provides the possibility to retrieve long-to-intermediate wavelengths in deeper regions from pre-critical reflections. In this study, we first present the construction of nonlinear sensitivities under the scattering theory. Extending the sensitivity kernel to higher order can help reduce the nonlinearity and improve the convergence of FWI. To construct higher order sensitivities, the model perturbation from the forthcoming iteration is needed. We then present a two-iteration approach to perform nonlinear FWI in the frequency domain. Finally, we apply this nonlinear FWI on the Marmousi model. The inverted models with different frequency ranges and different initial models show that this nonlinear FWI can build a reliable high resolution model in both shallow and deeper regions.

INTRODUCTION

Full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984) uses full wavefield information to estimate the subsurface models by minimizing the misfit between recorded data and modeled data. Although the full-wave equation is highly nonlinear for the real earth, the nonlinear waveform inverse problem is usually treated as a local optimization problem, which can be solved by a linearized version of the problem iteratively through the repeated calculation of a local gradient. The misfit function in the local optimization problem is not convex, so FWI suffers from the local minima problem, which makes FWI heavily relies on the availability of low frequency or long offset seismic data and a good initial model.

Theoretically, FWI can recover both long and short wavelengths structures correctly with wide-aperture acquisition systems and broadband sources, while successfully application shows that FWI can build high resolution velocity models in shallow regions (Virieux and Operto, 2009), where long-to-intermediate wavelengths structures are recovered by the information carried by diving waves and post-critical reflections, and a migration-like model in the deeper regions sampled by the pre-critical reflections. The spatial resolution of FWI is governed by the relationship established in diffraction tomography, which is the relationship between the scattering angle θ and local model wavenumber component $\mathbf{k}$ (Sirgue and Pratt, 2004; Alkhalifah, 2015; Brossier et al., 2015)

$$\mathbf{k} = \frac{2\omega}{v} \cos \frac{\theta}{2} \mathbf{n},$$

where ω is the angular frequency, v is the local velocity and $\mathbf{n}$ is the normalization of the vector $\mathbf{k}$. Based on this relationship, during updating the subsurface model, FWI uses large-angle scattering data, e.g., diving waves and post-critical reflections, to get the low-to-intermediate wavenumber components, and small-angle scattering data to get the high wavenumber components. Therefore, without long offsets and low frequency information, FWI cannot reconstruct the low wavenumber components, and behaves more like a least-squares migration, which updates mainly the high wavenumber component of the structures. Even with long offset data, the penetration depths of large-angle scattering data are often insufficient to reach the deeper part of the subsurface structures.

To better retrieve the long wavelengths structures and help FWI converge to a global minima, it is straight forward to consider building a good starting model first, especially when very low frequency information can still not be obtained from the recorded data. One way to do so is to retrieve the low frequency information from the seismic data using complex-valued frequencies in frequency domain, or to say time damping in the time domain, such as Laplace-domain and Laplace-Fourier-domain inversion (Shin and Cha, 2008, 2009), or using envelop information obtained form Hilbert transform, such as envelop inversion (Wu et al., 2014). The other way is to retrieve the low wavenumber components of the model directly from reflection data, and several methods have been proposed in both data and image domain, such as reflection waveform inversion (RWI) (Xu et al., 2012), migration based traveltime tomography (MBTT) (Chavent et al., 1994), differential semblance optimization (Symes and Carazzone, 1991) and migration velocity analysis (MVA) (Sava and Biondi, 2004) and so on. In the data domain methods, the model is decomposed into a low wavenumber background part, which is to be updated, and a high wavenumber perturbation part, which is assumed to be known, and a migration at each iteration is required to provide related virtual sources in depth, which introduce the transmission wavepaths from the reflectors to both sources and receivers, where wide-scattering angles can be obtained to help updating the low wavenumber components of the model. While refractions can be used together with reflections to perform a better reconstruction of the background model (Wang et al., 2015; Zhou et al., 2015), simultaneously inversion of the background and perturbation model are also studied in the data domain and mixed data/image domain (Sun and Symes, 2012; Biondi and Almomin, 2014; Wu and Alkhalifah, 2015; Alkhalifah and Wu, 2016) to mitigate nonlinearities of the FWI formulation.

On the other hand, as pointed out in the FWI review (Virieux and Operto, 2009), the gradient is usually studied in the framework of Fréchet derivative, or sensitivity kernel, which indicates the changes in the seismic waveform caused by the model parameter perturbations, and the sensitivity kernel is generally computed with the Born approximation, which indicates a linear relationship between the model perturbations and the change of the waveform and valid in the case of weak and small perturbations. Although different linearizion approach, such as Rytov approximation, can be used in FWI, several work have already been done to study nonlinear sensitivity as in resistivity inversion (Mcgillivray and Oldenburg, 1990), optical diffuse imaging (Kwon and Yazici, 2010) and seismic inversion (Wu and Zheng, 2014; Innanen, 2014, 2015).

Although FWI can be implemented in both time domain and frequency domain, there are a lot of numerical advantages in frequency domain. Only a few frequencies are needed

for the inversion in the frequency domain method (Sirgue and Pratt, 2004), and frequency domain is more natural to perform a multiscale approach, where for each iteration the impedance matrix is only needed to be constructed once. Therefore, in this study, following Innanen's work (2014; 2015), we study the nonlinear FWI in frequency domain. We start from the construction of the nonlinear sensitivities. By adding a perturbation to the nth updated model, the wavefield can be divided into a background wavefield that is obtained in the nth updated model and a perturbation wavefield that is obtained from both nth updated model and the perturbation. Using both background wavefield and perturbation wavefield, we can obtain the nonlinear sensitivities which contain both the zero-order and higher-order sensitivities. Based on this nonlinear sensitivities, we discuss the construction of a nonlinear two-iteration FWI scheme in frequency domain. For each iteration, we first invert the perturbation using linear inversion scheme in the inner iteration, then substitute the obtained perturbation to generate the nonlinear sensitivities for the outer iteration, and the gradient of the outer iteration nonlinear FWI can be constructed using the nonlinear sensitivities and data residual at the nth updated model or data residual from both nth updated model and the perturbation model. It has to be point out that without calculating the gradient of the outer iteration, the perturbation obtained from the inner iteration can provide a direct update of the model, similar as discussed in Kwak et al's work (2014). From the numerical example, we see that the usage of higher order sensitivities provide the transmission wave paths from the scatter points to both sources and receivers at the surface, along which, the wide-scattering angles provide the ability to update the low wavenumber components between the reflectors and the surface. Inversion using only the higher order sensitivities can provide a good updated background model, which can be further used as initial model in FWI. The applications on the Marmousi model show that this nonlinear FWI converges faster than the conventional FWI, and the inversion results shows that this approach can provide a good inversion result even without the low frequency data.

THEORY

In this paper, we will use the isotropic acoustic wave equation with constant density to describe the wave motion. In the space-frequency domain, the wave equation can be written as

$$\left[\omega^2 s(\mathbf{r}) + \nabla^2 \right] P(\mathbf{r}, \mathbf{r}_s, \omega) = -W(\omega)\delta(\mathbf{r} - \mathbf{r}_s), \tag{1}$$

where $P(\mathbf{r}, \mathbf{r}_s, \omega)$ is the wavefield generated by a source located at $\mathbf{r}_s$, ω is the frequency, and $s(\mathbf{r})$ is the squared slowness parameter $s(\mathbf{r}) = v^{-2}(\mathbf{r})$. In this study, we are using point source with spectrum $W(\omega) = 1$ as source term, so the wavefield $P(\mathbf{r}, \mathbf{r}_s, \omega)$ is equivalent to the related Green's function $G(\mathbf{r}, \mathbf{r}_s, \omega)$. The observed data is collected at receivers $\mathbf{r}_g$ as $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$.

Conventional FWI

FWI seeks to estimate subsurface properties through an iterative process by minimize the difference between the synthetic data and the observed data. The misfit function is usually given in a least-square norm as

$$\phi(s_n) = \frac{1}{2} \sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_{\omega} \left\| \delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) \right\|^2, \tag{2}$$

where $\delta P\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n\right) = P\left(\mathbf{r}_g, \mathbf{r}_s, \omega\right) - G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n\right)$ is the data residual between the observed data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$ and the synthetic data $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$ calculated using the updated model s_n at the nth FWI iteration.

Gradient-based method is usually used to find the optimization solution of the problem. Suppose that at the nth iteration, the model is updated from $s_n(\mathbf{r})$ to $s_{n+1}(\mathbf{r})$ through a perturbation $\delta s_n(\mathbf{r})$

$$s_{n+1}(\mathbf{r}) = s_n(\mathbf{r}) + \delta s_n(\mathbf{r}). \tag{3}$$

This perturbation $\delta s_n(\mathbf{r})$ can be determined by

$$\delta s_n(\mathbf{r}) = -\sum_{\mathbf{r}'} H_n^{-1}\left(\mathbf{r}, \mathbf{r}'\right) g_n\left(\mathbf{r}'\right), \tag{4}$$

where $g_n(\mathbf{r})$ and $H_n\left(\mathbf{r}, \mathbf{r}'\right)$ are the gradient and the Hessian matrix calculated at the nth iteration, which are defined as the first and second order derivatives of the misfit function $\phi(s_n)$ with respect to the model s_n, respectively,

$$g_n(\mathbf{r}) = \frac{\partial \phi(s_n)}{\partial s(\mathbf{r})}, H_n\left(\mathbf{r}, \mathbf{r}'\right) = \frac{\partial^2 \phi(s_n)}{\partial s(\mathbf{r}) \partial s(\mathbf{r}')}. \tag{5}$$

The gradient can be calculated through

$$g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_\omega \mathrm{Re}\left(\frac{\partial G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n\right)}{\partial s(\mathbf{r})} \delta P^*\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n\right)\right), \tag{6}$$

where $*$ stands for the complex conjugate, and $\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) / \partial s(\mathbf{r})$ is the Fréchet derivative or the sensitivity. Since it is extremely expensive to explicitly calculate the sensitivity, in FWI, the adjoint-state method (Plessix, 2006) is used to directly calculate the gradient. While the gradient indicates the descent direction to solve the optimization problem, the inverse Hessian matrix helps to alter the direction and length of the gradient vector. Calculation of full form of the Hessian matrix is usually very expensive, so approximation of Hessian is used to form a Gaussian-Newton update (Virieux and Operto, 2009),

$$H\left(\mathbf{r}, \mathbf{r}'\right) \approx H_{GN}\left(\mathbf{r}, \mathbf{r}'\right) = \sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_\omega \mathrm{Re}\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)}{\partial s(\mathbf{r})} \frac{\partial G^*(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)}{\partial s(\mathbf{r}')}\right). \tag{7}$$

When assuming the Hessian matrix as an identity matrix, the model can then be updated through

$$s_{n+1}(\mathbf{r}) = s_n(\mathbf{r}) - \mu_n g_n(\mathbf{r}), \tag{8}$$

where μ_n is the step length which can be determined by a line search method.

From linear to nonlinear sensitivities

The optimization problem of minimizing the misfit function (2) is highly nonlinear, since the relationship between the wavefield and model parameter is nonlinear, as described by the wave equation (1). One direct way to help mitigating this nonlinearity is to have broadband seismic data with wide-offset range and very good initial model before starting

the FWI procedure, however, even with different types of misfit definition proposed (van Leeuwen and Mulder, 2008; Brossier et al., 2015), the gradient used to update the model is still obtained with the sensitivity $\partial G/\partial s$ derived under the Born approximation (as shown in APPENDIX (A-4)) as

$$\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)}{\partial s(\mathbf{r})} \approx \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n), \tag{9}$$

where the Green's functions are calculated in the background model s_n, so it depends only on s_n but not δs_n. Insert the sensitivity equation (9) back to the gradient equation (6), the gradient can be constructed as the multiplication between the incident wavefield from the source point and backpropagated complex conjugate of the data residual from the receiver positions in the frequency domain, which can also be seen as a zero-lag correlation of the incident wavefield and the backpropagated data residuals in the time domain,

$$g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_{\omega} \mathrm{Re}\left(\omega^2 G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)\right). \tag{10}$$

The other way to mitigate the nonlinearity is taking multiple forward scattering into account, which can be performed by introducing nonlinear or to say higher-order sensitivities. Based on the scattering theory as discussed in former work of Innanen (2015), the synthetic data calculated using model $s_{n+1}(\mathbf{r})$ can be written as an expansion of the synthetic data calculated using model $s_n(\mathbf{r})$ using Lippmann-Schwinger equation as shown in APPENDIX (A-2),

$$\begin{aligned}
G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1}\right) \\
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s_n(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_{n+1}) \\
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s_n(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) \\
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s_n(\mathbf{r}') \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega | s_n) \delta s_n(\mathbf{r}'') G(\mathbf{r}'', \mathbf{r}_s, \omega | s_n) \\
+ \dots
\end{aligned} \tag{11}$$

When adding a small variation δs which is localized at the position $\mathbf{r}$ to s_{n+1}, it can be seen as changing the perturbation from δs_n to $\delta s_n + \delta s(\mathbf{r})$ while keep the background model s_n unchanged, then the calculated field is

$$\begin{aligned}
G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})) = G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) \\
+ \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \left[\delta s_n(\mathbf{r}') + \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')\right] G(\mathbf{r}', \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})) \\
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \left[\delta s_n(\mathbf{r}') + \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')\right] G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) \\
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \left[\delta s_n(\mathbf{r}') + \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')\right] \\
\times \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega | s_n) \left[\delta s_n(\mathbf{r}'') + \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'')\right] G(\mathbf{r}'', \mathbf{r}_s, \omega | s_n) + \dots
\end{aligned} \tag{12}$$

On the other hand, if we add the variation $\delta s(\mathbf{r})$ to the background model s_n and keep the perturbation δs_n unchanged, the calculated field is

$$
G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})) = G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n + \delta s(\mathbf{r}))
$$
$$
+ \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n + \delta s(\mathbf{r})) \delta s_n(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})). \quad (13)
$$

By keeping the first order term related to $\delta s(\mathbf{r})$, $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n + \delta s(\mathbf{r}))$ in equation (13) is

$$
G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n + \delta s(\mathbf{r}))
$$
$$
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n + \delta s(\mathbf{r}))
$$
$$
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n)
$$
$$
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}')
$$
$$
\times \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega | s_n) \delta s(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}'') G(\mathbf{r}'', \mathbf{r}_s, \omega | s_n)
$$
$$
+ \dots
$$
$$
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) + \dots . \quad (14)
$$

Insert (14) back to (13), we can get

$$
G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})) = G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n)
$$
$$
+ \omega^2 \int \mathrm{d}\mathbf{r}' \left(G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) + \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}', \omega | s_n) \right) \delta s_n(\mathbf{r}')
$$
$$
\times \left(G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) + \omega^2 G(\mathbf{r}', \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) \right) + \dots
$$
$$
= G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) + \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n)
$$
$$
+ \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \delta s_n(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n)
$$
$$
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) \delta s_n(\mathbf{r}') \delta s(\mathbf{r})
$$
$$
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) \delta s_n(\mathbf{r}') \delta s(\mathbf{r}) + \dots . \quad (15)
$$

which is equivalent to the result obtained by keeping the background model unchanged as in equation(12). Comparing the synthetic data $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r}))$ in the perturbed model $s_{n+1} + \delta s(\mathbf{r})$ in equation (15) and synthetic data $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})$ at the $n + 1th$ iteration in equation (11), the perturbation $\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1}, \delta s(\mathbf{r}))$ caused by this variation $\delta s(\mathbf{r})$ is the difference between these two series:

$$
\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1}, \delta s(\mathbf{r})) = G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1} + \delta s(\mathbf{r})) - G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})
$$
$$
= \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) \delta s(\mathbf{r}) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n)
$$
$$
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) \delta s_n(\mathbf{r}') \delta s(\mathbf{r})
$$
$$
+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) \delta s_n(\mathbf{r}') \delta s(\mathbf{r})
$$
$$
+ \dots . \quad (16)
$$

The sensitivities at $n + 1th$ iteration can then be written as

$$\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})}{\partial s(\mathbf{r})} = \lim_{\delta s \to 0} \frac{\delta G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1}, \delta s(\mathbf{r}))}{\delta s(\mathbf{r})}$$

$$= \left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})}{\partial s(\mathbf{r})}\right)_0 + \left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})}{\partial s(\mathbf{r})}\right)_1 + \dots, \quad (17)$$

where the zero order term is the conventional FWI sensitivities (9)

$$\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})}{\partial s(\mathbf{r})}\right)_0 = \frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)}{\partial s(\mathbf{r})} = \omega^2 G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n), \quad (18)$$

and the first order term is

$$\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_{n+1})}{\partial s(\mathbf{r})}\right)_1 = \omega^4 \int d\mathbf{r}' \delta s_n(\mathbf{r}') \left[G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}_s, \omega | s_n) \right.$$

$$\left. + G(\mathbf{r}_g, \mathbf{r}, \omega | s_n) G(\mathbf{r}, \mathbf{r}', \omega | s_n) G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) \right]. \quad (19)$$

Same as in the conventional FWI sensitivities, the zero order term depends only on the model s_n, while starting from the first order term, the sensitivities depend not only on the model s_n, but also on the perturbation δs_n, which introduce the nonlinear effect to the sensitivities.

Calculation of perturbation δs_n before $n + 1th$ FWI iteration

Note that all the Green's functions used in this new sensitivities (17) for $n + 1th$ model s_{n+1} are still calculated in the nth model s_n. However, to calculate the higher order terms in the new sensitivities (17), we need to know δs_n first. In the former research work, Innanen (2015) showed the perturbation δs_n can be exchanged in to a series related to the nth data residual $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$, according to inverse scattering theory, therefore, the perturbation δs_n can be inverted from the data residual through direct nonlinear inverse scattering. In this study, instead of using direct nonlinear inverse scattering, we are using a linearized inversion to get a perturbation for each model s_n during the FWI iteration.

The data residual $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$, which is the difference between the observed data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$ and the simulated data $G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$ in the nth model s_n, can be seen as the scattered data produced by the difference between the true model s and the nth model s_n, and is related with the perturbation $\Delta s = s - s_n$ through the relation

$$\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) = P(\mathbf{r}_g, \mathbf{r}_s, \omega) - G(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$$

$$= \omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \Delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n) + \dots. \quad (20)$$

By taking only the first order term related to Δs, equation (20) becomes the Born forward modeling, which describe the relationship between the perturbation Δs and the data residual as

$$\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n) = \omega^2 \int d\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega | s_n) \Delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega | s_n). \quad (21)$$

Therefore, instead of calculating the exact direct inverse of the forward modeling (21), the perturbation Δs can be approached by the solution $\delta\tilde{s}$ of a data fitting scheme in a least squares sense, which iteratively minimizes the misfit function

$$\phi(\delta\tilde{s}) = \frac{1}{2}\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\|\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n) - \delta P_{cal}(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)\|^2$$

$$= \frac{1}{2}\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\|\delta S(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)\|^2, \tag{22}$$

where $\delta P_{cal}(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$ is the scattered data calculated using the Born forward modeling (21), and $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$ is the observed scattered data obtained from the difference of the observed data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$ and the simulated data $G(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$ in the nth model as in (20). The perturbation is then obtained by

$$\delta\tilde{s}_{m+1}(\mathbf{r}) = \delta\tilde{s}_m(\mathbf{r}) - \sum_{\mathbf{r}'}\tilde{H}_m^{-1}(\mathbf{r}, \mathbf{r}')\,\tilde{g}_m(\mathbf{r}'), \tag{23}$$

where $\tilde{H}^{-1}$ and $\tilde{g}$ are the inverse of the Hessian and the gradient of the misfit function (22), respectively. Suppose that at the mth iteration, the scattered data residual is $\delta S_m(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$. Similar to FWI, in this inversion, the gradient of misfit function (22) is

$$\tilde{g}_m(\mathbf{r}) = -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\omega^2\text{Re}\left(G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)G(\mathbf{r}, \mathbf{r}_s, \omega|s_n)\delta S_m^*(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)\right), \tag{24}$$

and approximation of the Hessian can also be used instead of using the full Hessian

$$\tilde{H}_{GN}(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\omega^4\text{Re}(G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)G(\mathbf{r}, \mathbf{r}_s, \omega|s_n)$$

$$\times G^*(\mathbf{r}_g, \mathbf{r}', \omega|s_n)G^*(\mathbf{r}', \mathbf{r}_s, \omega|s_n)), \tag{25}$$

so that a Gaussian-Newton updating can be performed. Since the Hessian is diagonally dominant in most case, the Hessian can be approximated by the diagonal term in the approximated Hessian (25) as

$$\tilde{H}_{GN}(\mathbf{r}, \mathbf{r}) = \sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\omega^4\left|G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)\right|^2\left|G(\mathbf{r}, \mathbf{r}_s, \omega|s_n)\right|^2. \tag{26}$$

Line search can also be used to determine the optimized scaling factor during the updating as

$$\delta\tilde{s}_{m+1}(\mathbf{r}) = \delta\tilde{s}_m(\mathbf{r}) - \alpha_m\tilde{g}_m(\mathbf{r}). \tag{27}$$

Nonlinear FWI with nonlinear sensitivities

Once obtain the perturbation as the solution $\delta\tilde{s}$ from (23) or (27) for nth iteration during FWI, we can write a direct update using this perturbation without line searching as

$$\tilde{s}_{n+1}(\mathbf{r}) = s_n(\mathbf{r}) + \delta\tilde{s}(\mathbf{r}), \tag{28}$$

similar as discussed in Kwak et al 's work (2014). Here we will continue to use this perturbation to construct the higher order sensitivities. To avoid the calculation of Green's function $G(\mathbf{r}', \mathbf{r}, \omega|s_n)$ at each point $\mathbf{r}$ in the higher order terms of the sensitivities (17) explicitly, we can calculate these higher order terms by solving the related wave equation. The scattered wavefield from source can be expended as

$$
\begin{aligned}
&\delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}) \\
&= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|\tilde{s}_{n+1}) \\
&= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_n) \\
&\quad + \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega|s_n) \delta\tilde{s}(\mathbf{r}'') G(\mathbf{r}'', \mathbf{r}_s, \omega|s_n) \\
&\quad + \dots
\end{aligned}
\tag{29}
$$

and scattered wavefield from receiver as

$$
\begin{aligned}
&\delta G(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s}) \\
&= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') G(\mathbf{r}', \mathbf{r}, \omega|\tilde{s}_{n+1}) \\
&= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') G(\mathbf{r}', \mathbf{r}, \omega|s_n) \\
&\quad + \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}_g, \mathbf{r}', \omega|s_n) \delta\tilde{s}(\mathbf{r}') \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega|s_n) \delta\tilde{s}(\mathbf{r}'') G(\mathbf{r}'', \mathbf{r}, \omega|s_n) \\
&\quad + \dots .
\end{aligned}
\tag{30}
$$

By substituting scattering wavefield (29) and (30) back into the nonlinear sensitivities (17), the higher order terms of the sensitivities can be reduced to

$$
\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1})}{\partial s(\mathbf{r})} \right)_1 + \dots = \omega^2 \left[\delta G(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s}) G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) \right.
$$
$$
\left. + G(\mathbf{r}_g, \mathbf{r}, \omega|s_n) \delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}) \right] .
\tag{31}
$$

It is obvious that unlike the conventional FWI sensitivity or to say the zero order sensitivities, the higher order sensitivities are the combination of two parts, which are the products between the scattered wavefield and the background wavefield for both source and receiver side. If only the first order scatterings in the scattered wavefield (29) and (30) are considered, then the first order sensitivity can be written as

$$
\left(\frac{\partial G(\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1})}{\partial s(\mathbf{r})} \right)_1 = \omega^2 \left[\delta G_{Born}(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s}) G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) \right.
$$
$$
\left. + G(\mathbf{r}_g, \mathbf{r}, \omega|s_n) \delta G_{Born}(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}) \right] .
\tag{32}
$$

Substituting the higher order sensitivities (31) or first order sensitivities (32) and the zero order sensitivity (18) back to the gradient (6), we can get the gradient of the nonlinear FWI with nonlinear sensitivities as

$$
\begin{aligned}
g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s} \sum_{\mathbf{r}_g} \sum_{\omega} \mathrm{Re} \Big(&\omega^2 \delta P^*(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n) \big(G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) G(\mathbf{r}_g, \mathbf{r}, \omega|s_n) \\
&+ \delta G(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s}) G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) + G(\mathbf{r}_g, \mathbf{r}, \omega|s_n) \delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}) \big) \Big),
\end{aligned}
\tag{33}
$$

and update the model from s_n to s_{n+1} through equation (8).

Note that the data residual in the gradient (33) is still the data residual from the nth iteration, which is used to generate the backpropagated receiver side wavefield and scattered wavefield. Therefore, for each shot, the source side wavefield and scattered wavefield in the higher order sensitivites (31) can be obtained by solving

$$\left(\omega^2 s_n(\mathbf{r}) + \nabla^2\right) G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) = -\delta(\mathbf{r} - \mathbf{r}_s), \tag{34}$$

$$\left(\omega^2 \tilde{s}_{n+1}(\mathbf{r}) + \nabla^2\right) \delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}) = -\omega^2 \delta\tilde{s}(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_n), \tag{35}$$

and with the reciprocity principle $G(\mathbf{r}, \mathbf{r}_g, \omega|s_n) = G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)$, the receiver side wavefield $U_r(\mathbf{r}, \mathbf{r}_g, \omega|s_n) = \delta P^* (\mathbf{r}_g, \mathbf{r}_s, \omega|s_n) G(\mathbf{r}, \mathbf{r}_g, \omega|s_n)$ and the related scattered wavefield $\delta U_r(\mathbf{r}, \mathbf{r}_g, \omega|s_n) = \delta P^* (\mathbf{r}_g, \mathbf{r}_s, \omega|s_n) \delta G(\mathbf{r}, \mathbf{r}_g, \omega|s_n, \delta\tilde{s})$ can be obtained by solving

$$\left(\omega^2 s_n(\mathbf{r}) + \nabla^2\right) U_r(\mathbf{r}, \mathbf{r}_g, \omega|s_n) = -\delta(\mathbf{r} - \mathbf{r}_g)\delta P^* (\mathbf{r}_g, \mathbf{r}_s, \omega|s_n), \tag{36}$$

$$\left(\omega^2 \tilde{s}_{n+1}(\mathbf{r}) + \nabla^2\right) \delta U_r(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s}) = -\omega^2 \delta\tilde{s}(\mathbf{r}') U_r(\mathbf{r}_g, \mathbf{r}', \omega|s_n). \tag{37}$$

When sensitivities up to only first order is used to construct the gradient as in equation (32), the scattered wavefield can be obtained in model s_n instead of model $\tilde{s}_{n+1}$ in equation (35) and (37).

From another point of view, since $\tilde{s}_{n+1}$ in equation (28) itself is already a direct update of the model, we can modify the data residual into

$$\delta P (\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1}) = P (\mathbf{r}_g, \mathbf{r}_s, \omega) - G (\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1})$$
$$\approx P (\mathbf{r}_g, \mathbf{r}_s, \omega) - G (\mathbf{r}_g, \mathbf{r}_s, \omega|s_n) - \delta G(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n, \delta\tilde{s}), \tag{38}$$

which contains not only the data residual in model s_n, but also the data residual of the synthesized scattered data in perturbation $\delta\tilde{s}$, to get the gradient as

$$g_n(\mathbf{r}) = -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\text{Re}\Big(\omega^2 \delta P^* (\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1}) \big(G(\mathbf{r}, \mathbf{r}_s, \omega|s_n)G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)$$

$$+ \delta G(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s})G(\mathbf{r}, \mathbf{r}_s, \omega|s_n) + G(\mathbf{r}_g, \mathbf{r}, \omega|s_n)\delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s})\big)\Big)$$

$$= -\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g}\sum_{\omega}\text{Re}\Big(\omega^2 \delta P^* (\mathbf{r}_g, \mathbf{r}_s, \omega|\tilde{s}_{n+1}) \big(G(\mathbf{r}, \mathbf{r}_s, \omega|\tilde{s}_{n+1})G(\mathbf{r}_g, \mathbf{r}, \omega|\tilde{s}_{n+1})$$

$$- \delta G(\mathbf{r}_g, \mathbf{r}, \omega|s_n, \delta\tilde{s})\delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_n, \delta\tilde{s})\big)\Big). \tag{39}$$

and perform an update from $\tilde{s}_{n+1}$ to s_{n+1} by replacing s_n to $\tilde{s}_{n+1}$ in equation (8).

Implementation of nonlinear FWI

So far we have discussed the calculation of nonlinear sensitivities by dividing the whole scheme into two steps: the first step is calculating the perturbation using a linear waveform inversion, the second step is using the perturbation to generate the higher order terms in the sensitivities. Using this nonlinear sensitivity, we can then get the nonlinear gradient for FWI. The whole scheme for this nonlinear FWI is shown as in Table 1.

Table 1: Nonlinear FWI

Algorithm 1 Algorithm for nonlinear FWI

Input: Recorded seismic data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model s_0
Output: Inverted model s_n

  1: **for** $n = 0, \ldots, n_{max}$ **do**   ▷ *Outer loop for FWI*
  2:       Get data residual $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$ for all frequency
  3:       Initial perturbation $\delta \tilde{s}_0 \leftarrow 0$
  4:       **for** $m = 0, \ldots, m_{max}$ **do**   ▷ *Inner loop for perturbation*
  5:             Calculate $\delta P_{cal}(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$
  6:             Get data residual $\delta S(\mathbf{r}_g, \mathbf{r}_s, \omega|s_n)$
  7:             Calculate the gradient $\tilde{g}_m(\mathbf{r})$ (equation (24))
  8:             Calculate the step length α_m using the line search method
  9:             Update $\tilde{s}_m$ (equation (23))
10:       **end for**
11:       Calculate nonlinear gradient (equation (33) or (39))
12:       Calculate the step length μ_n using the line search method
13:       Update to s_{n+1} (equation (8))
14: **end for**

In the above algorithm, all the frequency components are inverted simultaneously, which is equivalent to the time domain approach. In frequency domain, as Sirgue and Pratt discussed (2004), it is more straightforward to perform the inversion sequentially from the low to high frequencies, and the computational cost is far cheaper than full time-domain approach. Moreover, the frequency interval can be determined according to the half offset-to depth ratio $R_{max} = h_{max}/z$ as

$$\omega_{n+1} = \frac{\omega_n}{\alpha_{min}}, \tag{40}$$

with

$$\alpha_{min} = \frac{1}{\sqrt{1 + R_{max}^2}} \tag{41}$$

Although the frequency interval determined by equation (40) is not constant as the normally used frequency domain interval, which is determined by the maximum recorded time to prevent aliasing, it can provide enough coverage for the vertical wavenumber. A more efficient nonlinear FWI algorithm can be perform with certain frequencies as shown in Table 2.

NUMERICAL EXAMPLES

In this section, we will investigate this nonlinear FWI scheme numerically. We will use a nine point frequency domain finite difference with constant-density (Jo et al., 1996) to model the observed data and synthetic data. DWI will be used to refer the direct inversion as in equation (28) with the perturbation calculated under the linear inversion. FOFWI and NFWI will be used to distinguish the gradient calculated using equation (33) up to first order with data residual in model s_n and (39) with data residual in model $\tilde{s}_{n+1}$, respectively. First, we will discuss the physical meaning of the higher order sensitivities with a

Table 2: Frequency domain nonlinear FWI

Algorithm 2 Algorithm for frequency domain nonlinear FWI

Input: Recorded seismic data $P(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model s_0
Output: Inverted model s_n

  1: **for** $\omega = \omega_{min}, \ldots, \omega_{max}$ **do**   ▷ *Frequency loop*
  2:     **for** $n = 0, \ldots, n_{max}$ **do**   ▷ *Outer loop for FWI*
  3:         Get data residual $\delta P(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
  4:         Initial perturbation $\delta \tilde{s}_0 \leftarrow 0$
  5:         **for** $m = 0, \ldots, m_{max}$ **do**   ▷ *Inner loop for perturbation*
  6:             Calculate $\delta P_{cal}(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
  7:             Get data residual $\delta S(\mathbf{r}_g, \mathbf{r}_s, \omega | s_n)$
  8:             Calculate the gradient $\tilde{g}_m(\mathbf{r})$ (equation (24))
  9:             Calculate the step length α_m using the line search method
10:             Update $\tilde{s}_m$ (equation (23))
11:         **end for**
12:         Calculate nonlinear gradient (equation (33) or (39))
13:         Calculate the step length μ_n using the line search method
14:         Update to s_{n+1} (equation (8))
15:     **end for**
16: **end for**

---

simple model. Then by showing the inversion results in complex model, we will discuss the efficiency and accuracy of this method.

Higher order sensitivities in an one-interface model

As we have discussed in the former section, the zero order term of the new sensitivities (17) is the sensitivity kernel of the conventional FWI, which depends only on the background model s_n and contains the iso-phase or iso-time surfaces where data residual can be back-projected. For example, Figure 1 shows a single frequency zero-order sensitivity kernel built in a homogeneous velocity model, where velocity is $2km/s$, with one source and one receiver. We can see that this kernel can provide the complete resolution ability carried out by all of the arrivals in the seismogram, and it represents the first Fresnel zone at the shallow depth, which associated with the forward scattering, while the secondary Fresnel zones, which are the outer fringes of the sensitivity kernel, are associated with the back scattering. As the width of this zone is relatively limited and decreases with depth as the scattering angle decreases for a fixed source-receiver offset, it is hard to update the long wavelengths component of deep part of the model.

Meanwhile, the higher order sensitivities depend not only on the background model s_n, but also on the perturbation δs_n. For example, a single frequency first-order sensitivity kernel built with same frequency, velocity and acquisition system but with one reflector as perturbation is shown in Figure 2, and for reference, the nonlinear sensitivity as used in gradient (33) calculated in the model with reflector is shown in Figure 3, where in both figures, the yellow line shows the position of the reflector. The two parts of this first order
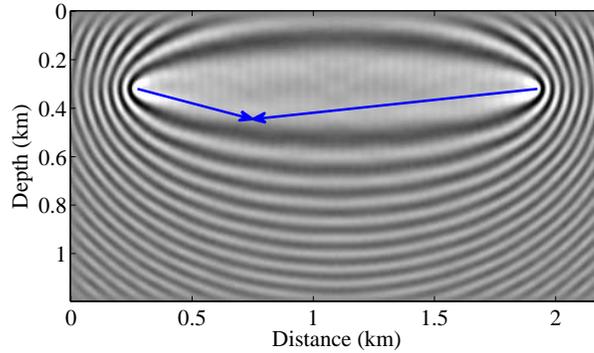
FIG. 1. Zero-order sensitivity kernel in a homogeneous model.

kernel in Figure 2 are built from the product of the scattered wavefield generated from the receiver side and the incident wavefield from the source side and the product of the scattered wavefield generated from the source side and the incident wavefield from the receiver side, which generate two first Fresnel zones connecting both the source position and the reflector (as indicated by cyan line) and the reflector and receiver position (as indicated by red line), respectively. Therefore, during the nonlinear FWI updating, these first order sensitivities provide the ability to update the long wavelengths component of deep part of the model.
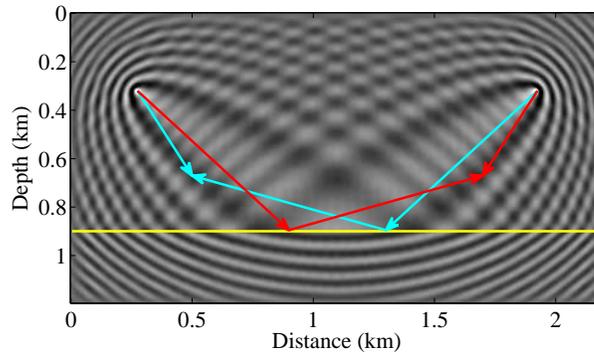
FIG. 2. First-order sensitivity kernel in a homogeneous model with one reflector.
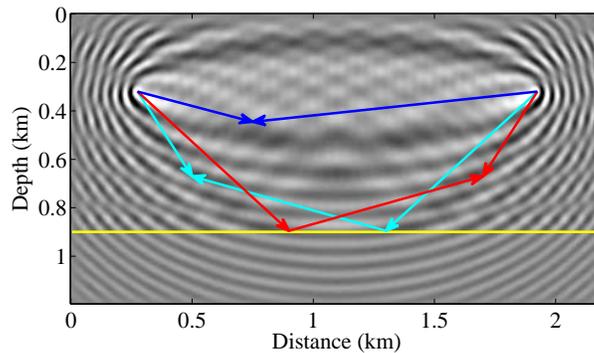
FIG. 3. Nonlinear sensitivity kernel in a homogeneous model with one reflector.

Inversion using first order sensitivities with prior known model perturbation

Here we only present one simple example of frequency domain RWI with prior known perturbation to demonstrate the contribution of the first order sensitivities for updating the

long wavelength structures, and in a companion paper we discuss and review the detail of the usage of this first order sensitivities and reflection data residual to perform a RWI in time domain. With prior known model perturbation, the higher order sensitivities can be seen as the connection between the perturbations of the background velocity and the change of the wavefield. As shown in Figure 4, we add a Gaussian ball and a reflector to the homogeneous background. The model is $2.2km$ in the horizontal direction and $1.2km$ in the depth direction, with spatial interval $10m$ in both directions. The background velocity is $2.5km/s$, the center of the Gaussian ball is $2.8km/s$ and the reflector locates at $z = 1km$. We put both sources and receivers along the surface, with source interval $dsx = 50m$ and receiver interval $dgx = 10m$. Suppose that the reflector information is already known, we are trying to invert the Gaussian ball using the homogeneous background as initial model. Figure 5 shows the inversion result using only the first order sensitivities with one iteration for each frequency in the range of $1Hz$ to $20Hz$ and interval $1Hz$. As expected, the result gives information of the Gaussian ball, which has good resolution along the horizontal direction, and does not have good resolution along the vertical direction. Using this inversion result as the initial model for conventional FWI, we can get a good reconstruction of the Gaussian ball in both horizontal and vertical direction as shown in Figure 6.
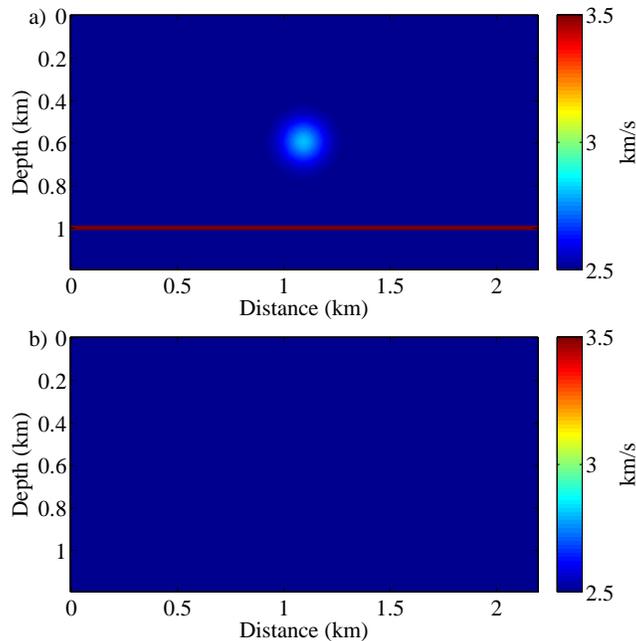
FIG. 4. Velocity model with Gaussian ball and reflector (a) and related background velocity model (b).

Nonlinear FWI test of Marmousi data

The Marmousi model is used to perform the synthetic test of the nonlinear FWI. The true model is shown in Figure 7, we add a $500m$ water layer on the top of the model to reduce the refractions in the data. First we will test with a initial model by smoothing the true velocity using a Gaussian smoother as shown in Figure 8. There are 461 fixed receivers
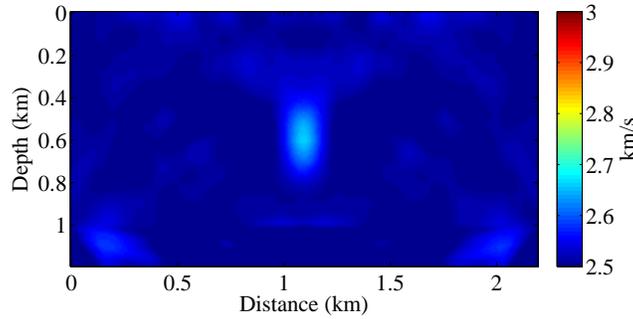
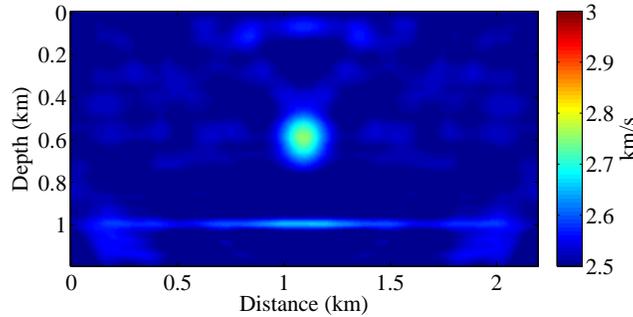FIG. 5. Inversion result using first order sensitivities.

FIG. 6. FWI result using inversion result in Figure 5 as initial model.

with spacing of $20m$ and 46 sources with spacing of $200m$ along the surface, where the first receiver locates at $x = 0m$ and the first source locates at $x = 100m$. With 3 frequency data ($4Hz$, $6.6Hz$, $14.9Hz$) starting from $4Hz$ to $15Hz$, 10 iterations for each frequency are used for FWI, DWI and nonlinear FWI inversion by updating only the velocity under the added water layer, and 5 inner iterations are used for calculating perturbation in FOFWI, DWI and NFWI. Figure 9a, Figure 9b, Figure 9c and Figure 9d shows the inversion results from conventional FWI, FOFWI, DWI and NFWI, respectively. Velocity profiles along $x = 2km, 4km, 6km, 8km$ are shown in Figure 10, where black lines show the true velocity, green lines show the initial velocity, red lines show the conventional FWI inverted velocity, dashed blue lines show the FOFWI inverted velocity, magenta lines show the DWI inverted velocity and solid blue lines show the NFWI inverted velocity. Compared to conventional FWI result, both of the nonlinear FWI results and DWI result give better reconstruction of the velocity, especially in the deeper part, while the NFWI gives the best inversion result in all the four methods. However, it has to be pointed out that in this example, although DWI result is better than conventional FWI result, quality of inversion result obtained by DWI depends on the inner iteration. Therefore, we only compare the conventional FWI results and NFWI results in the following test. Figure 11 shows the norm of the data residual vector of the three inversion methods as a function of iteration numbers for all three frequencies, and DWI data residual is not shown because the velocity is updated directly with the perturbation obtained from linear inversion, where data residual plotted here is not used as a misfit function during iteration. Since in NFWI, both data synthesized in model s_n and $\delta\tilde{s}$ is used to obtain the data residual, it shows a lower data residual at the first iteration, where the data residuals of conventional FWI and FOFWI are almost the same.
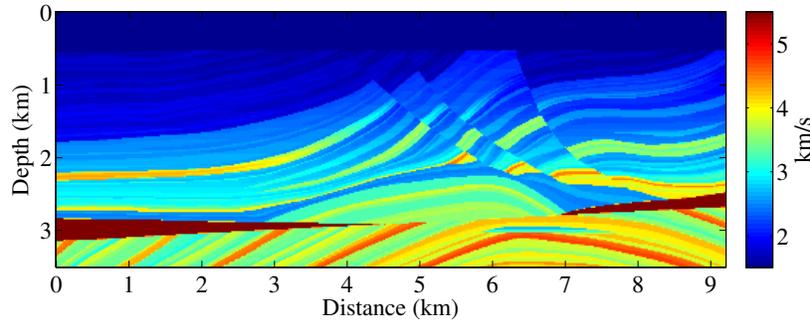
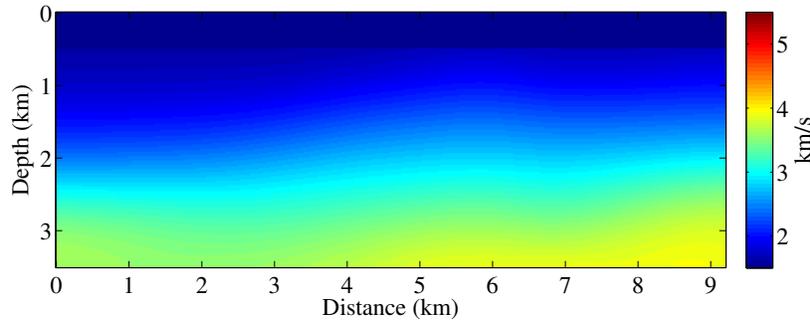FIG. 7. True velocity model of the Marmousi example.

FIG. 8. The initial velocity model of the Marmousi example.

Then we use a linearly increasing model as the initial model, which is exact at the added water layer, but changes linearly along depth direction from $1.5km/s$ to $4km/s$, as shown in Figure 12. The acquisition system is the same as the former test, but 5 frequency data ($2Hz, 3.3Hz, 5.5Hz, 9Hz, 14.9Hz$) starting from $2Hz$ to $15Hz$ are used for the inversion. Figure 13a shows the inversion result from conventional FWI with 10 iterations for each frequency, and Figure 13b shows the inversion result from NFWI with same iterations. Velocity profiles along $x = 2km, 4km, 6km, 8km$ are shown in Figure 14, where black lines show the true velocity, green lines show the initial velocity, blue lines show the NFWI inverted velocity and red lines show the conventional FWI inverted velocity. Since the initial model is far away from the true velocity model, and we update the velocity at the water layer in this test, even with frequency as low as $2Hz$, conventional FWI can only reconstruct rough structures of the model, but NFWI gives accurate high resolution of the model even at the deeper part.

Next we change the starting frequency from $2Hz$ to $4Hz$, and 3 frequencies ($4Hz$, $6.6Hz$, $14.9Hz$, and the last frequency is changed to match the last frequency as in the $2 - 15Hz$ case) are used to get the data for the inversion. During this test, we keep the water layer velocity unchanged, and only update the velocity under the water layer. Figure 15a shows the inversion result from conventional FWI with 30 iterations for each frequency, and Figure 15b shows the inversion result from NFWI with same iterations. Velocity profiles along $x = 2km, 4km, 6km, 8km$ are shown in Figure 16. Without lower frequency information, the conventional FWI gives wrong velocity updates even at shallow part of the velocity, where NFWI can provide reasonable velocity information. However, compared with the inversion results with a better initial model as in Figure 10, the deeper area of the inverted velocity using both methods are not good, which may be caused by the insuffi-
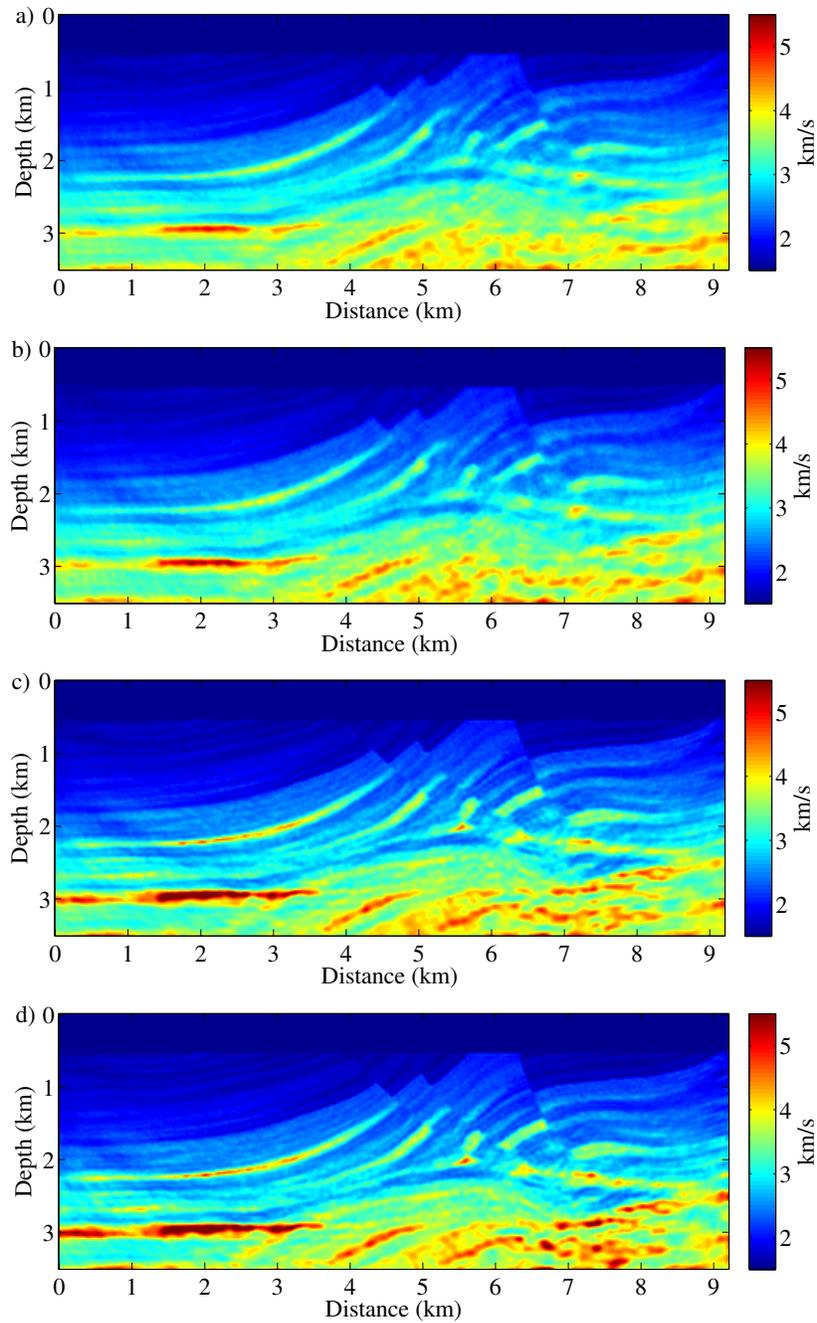
FIG. 9. Inversion using a) FWI b) FOFWI c) DWI and d) NFWI with 10 iterations for 3 frequencies starting from $4Hz$ to $15Hz$.
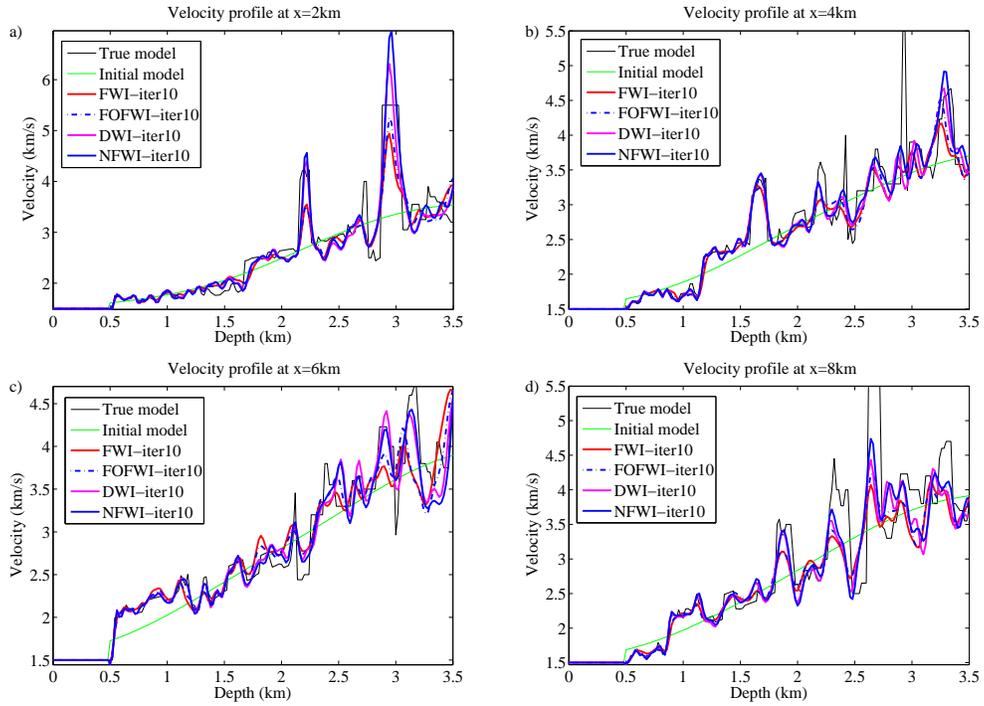
FIG. 10. Velocity profiles along a) $x = 2km$, b) $x = 4km$, c) $x = 6km$ and d) $x = 8km$.
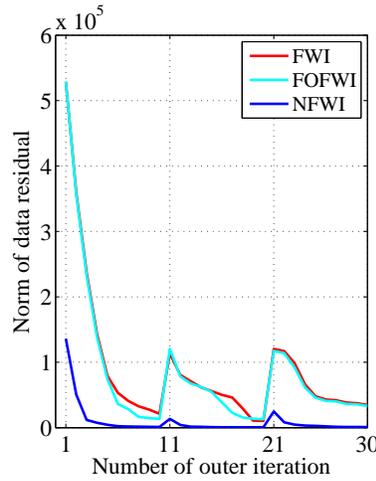

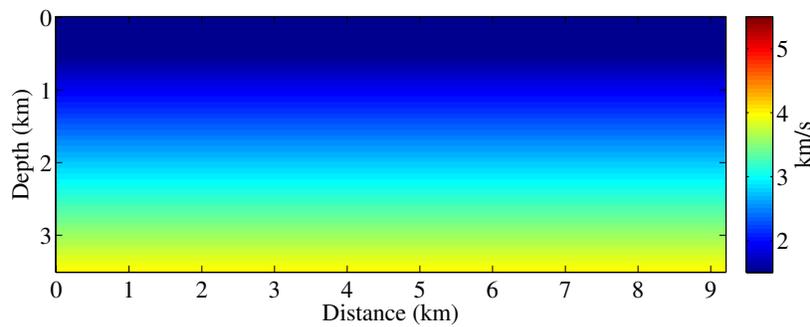FIG. 11. Norm of data residual vector as a function of the number of iterations.

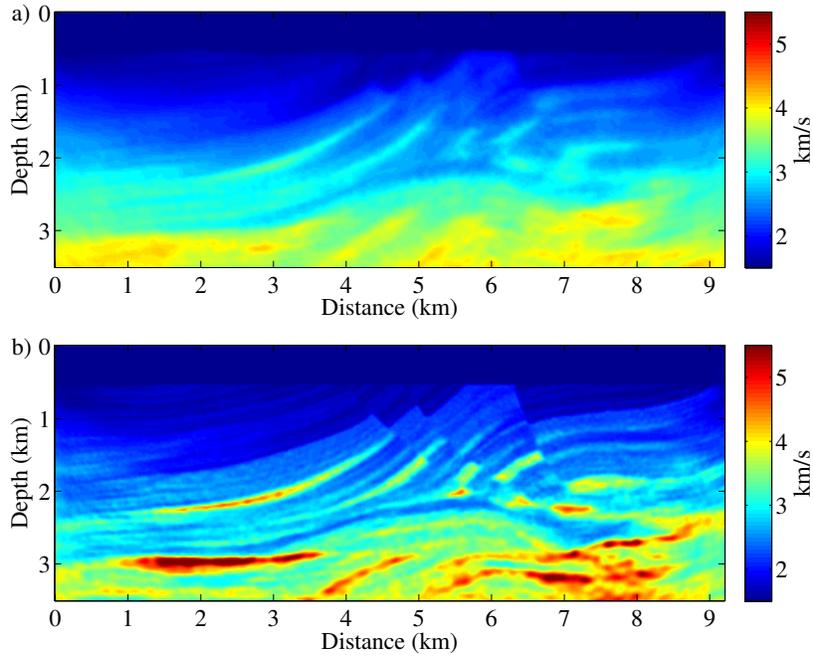FIG. 12. The initial velocity model of the Marmousi example.

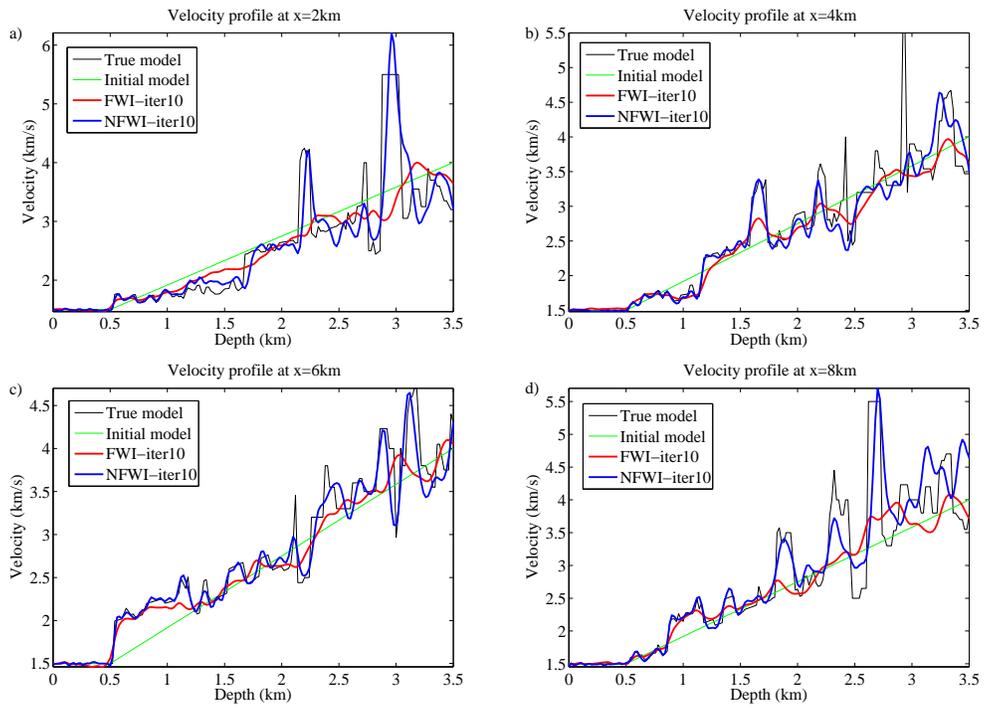FIG. 13. Inversion using a) FWI and b) NFWI with 10 iterations for 5 frequencies starting from $2Hz$ to $15Hz$.

FIG. 14. Velocity profiles along a) $x = 2km$, b) $x = 4km$, c) $x = 4km$ and d) $x = 8km$.

cient wavenumber coverage with only 3 frequencies for the whole inversions. To verify this, we use uniformly sampled 12 frequencies between $4Hz$ to $15Hz$ to perform both conventional FWI and NFWI inversions with the same initial model. Figure 17a shows the inversion result from conventional FWI with 10 iterations for each frequency, and Figure 17b shows the inversion result from NFWI with same iterations. Velocity profiles along $x = 2km, 4km, 6km, 8km$ are shown in Figure 18. Compared to the inversion results in Figure 15 with only 3 frequencies, more frequency information provide better wavenumber coverage, and the inversion results are improved significantly. With the conventional FWI method, velocity is updated incorrectly, especially at the shallow area around $z = 1km$ to $1.5km$ along vertical profile $x = 4km$ and deeper area along the vertical profile $x = 6km$, where the NFWI method provides a fairly good result.
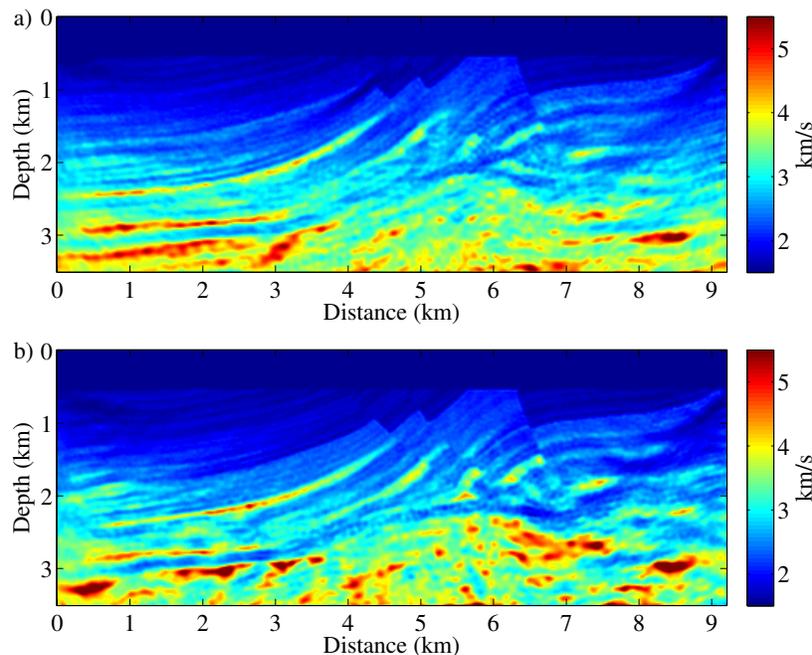

FIG. 15. Inversion using a) FWI and b) NFWI with 30 iterations for 3 frequencies starting from $4Hz$ to $15Hz$.

DISCUSSION AND CONCLUSION

It is shown that the resolution of FWI is determined by the scattering angle and frequency information, which is similar as in diffraction tomography. Successful applications of FWI rely on good initial model and seismic dataset with long offset and low frequency information. Due to the restriction of the availability of scattering angle in the deep regions, reconstructing the long-to-intermediate wavelength structures can still be challenging.

FWI can be performed in both time and frequency domain. Compared to time domain, which is more appropriate to choose certain time window to distinguish certain type of the data needed to different type of inversion, e.g., reflections only in RWI, or diving waves and reflections separately in joint FWI (Wang et al., 2015; Zhou et al., 2015), frequency domain FWI is more natural to perform a multiscale inversion and theoretically only a few discrete frequencies are sufficient to cover the wavenumber spectrum. When initial model
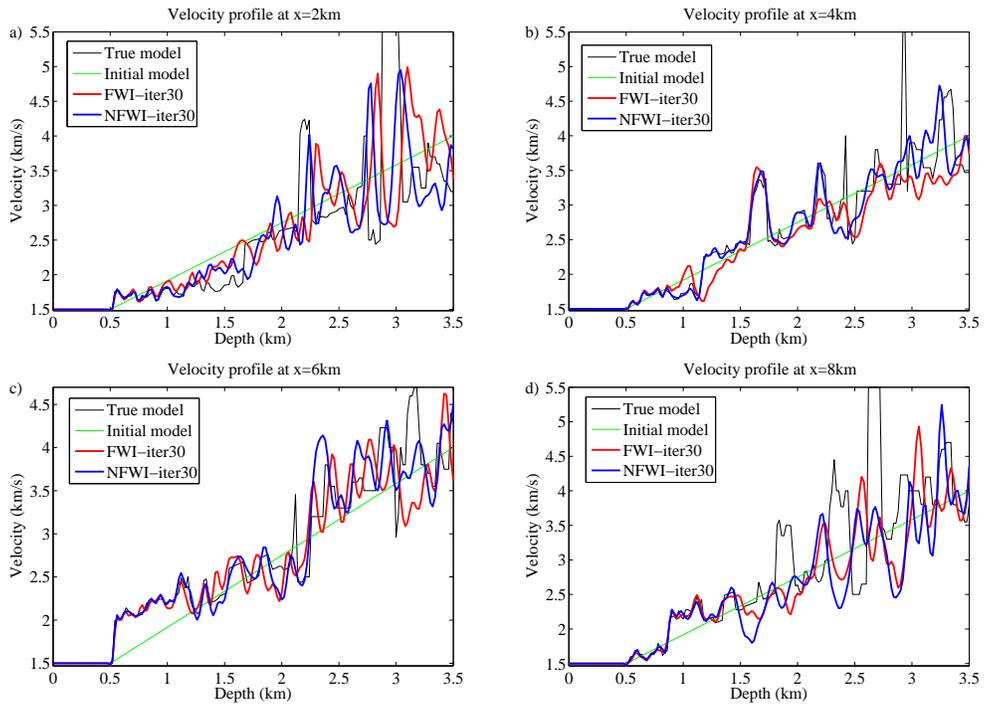
FIG. 16. Velocity profiles along a) $x = 2km$, b) $x = 4km$, c) $x = 4km$ and d) $x = 8km$.
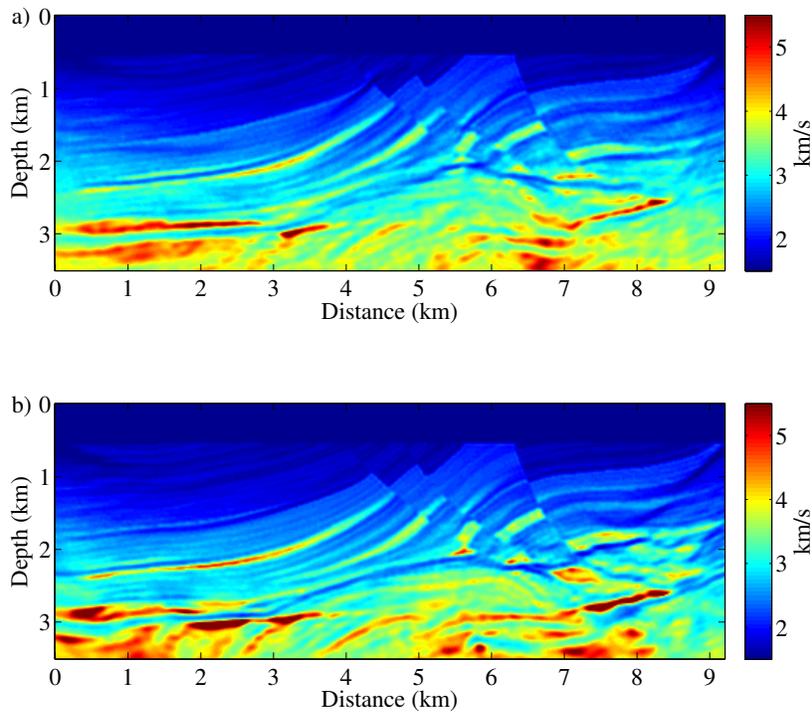

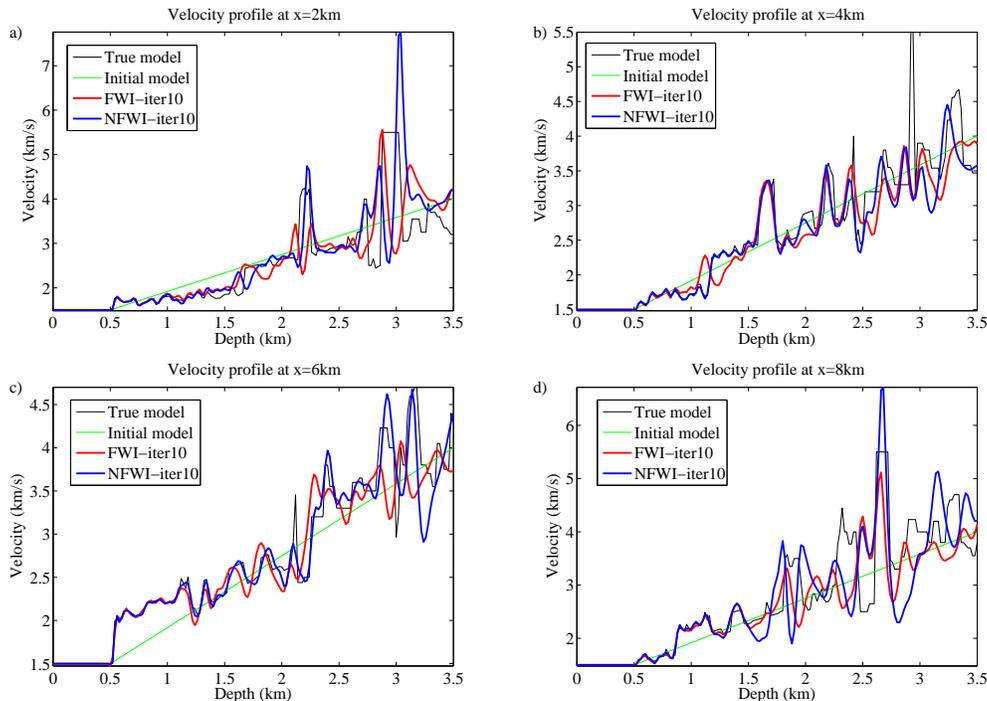FIG. 17. Inversion using a) FWI and b) NFWI with 10 iterations for 12 frequencies starting from $4Hz$ to $15Hz$.

FIG. 18. Velocity profiles along a) $x = 2km$, b) $x = 4km$, c) $x = 4km$ and d) $x = 8km$.

is not close to the complex true model, it is still necessary to use sufficient frequencies to get a reasonable inversion result.

The recent research on nonlinear sensitivities provide the possibility to better handle the nonlinearity in FWI. The sensitivities used in conventional FWI is equivalent to the zero order term in the nonlinear sensitivities, and back-project diving wave and post-critical reflections to the low wavenumber first Fresnel zone, while pre-critical reflections are usually back-projected to the corresponding high-wavenumber migration isochrones. Study on the higher order sensitivities based on the scattering theory shows that these higher order sensitivities can provide transmission wave paths from scatters to both sources and receivers, where wide scattering angle can be obtained to provide the ability to update the long wavelength structures between the reflectors and the surface. Application of these first order sensitivities is studied as in the recent proposed method RWI, and it shows that with prior known perturbation model, first order sensitivities can help update the long wavelength components in the region between reflectors and sources/receivers, and the inverted results can be further used as initial model in conventional FWI.

In this study, by including both zero order and higher order terms in the calculation of sensitivities, we have constructed a two-iteration nonlinear FWI approach in frequency domain, which can provide a better update to the model and converges faster than the conventional FWI. In this nonlinear FWI approach, a linear inversion is used to obtain the perturbation model before calculating the nonlinear FWI gradient. In the linear inversion, the data residual from the current iteration is used as the observed data, and in the following nonlinear FWI, although same data residual can be used in the related misfit function, data residual includes the scattering data synthesized from the linear inversion as well can

provide a better inversion result. The results of a Marmousi model to with different initial model and frequency bands illustrate the convergence characteristics of the nonlinear FWI.

ACKNOWLEDGMENT

APPENDIX - THE BORN APPROXIMATION

Suppose that the model can be split into two parts

$$s(\mathbf{r}) = s_0(\mathbf{r}) + \delta s(\mathbf{r}), \tag{A-1}$$

where $s_0(\mathbf{r})$ contains the low-wavenumber component of the model or the background part, and $\delta s(\mathbf{r})$ contains the high-wavenumber component or the perturbation part. Accordingly the Green's function satisfying equation (1) can be split into the Green's function in the background model and perturbation as

$$G(\mathbf{r}, \mathbf{r}_s, \omega|s) = G(\mathbf{r}, \mathbf{r}_s, \omega|s_0) + \delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_0, \delta s), \tag{A-2}$$

where the perturbation of the Green's function or scattered wavefield can be expressed exactly as

$$
\begin{aligned}
\delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_0, \delta s) &= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_0) \delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s) \\
&= \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_0) \delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_0) \\
&+ \omega^4 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_0) \delta s(\mathbf{r}') \int \mathrm{d}\mathbf{r}'' G(\mathbf{r}', \mathbf{r}'', \omega|s_0) \delta s(\mathbf{r}'') G(\mathbf{r}'', \mathbf{r}_s, \omega|s_0) \\
&+ \dots.
\end{aligned} \tag{A-3}
$$

Under the single-scattering assumption, only first-order term is considered in equation (A-3), which gives the Born approximation

$$\delta G_{Born}(\mathbf{r}, \mathbf{r}_s, \omega|s_0, \delta s) = \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_0) \delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_0), \tag{A-4}$$

$$G_{Born}(\mathbf{r}, \mathbf{r}_s, \omega|s) = G(\mathbf{r}, \mathbf{r}_s, \omega|s_0) + \omega^2 \int \mathrm{d}\mathbf{r}' G(\mathbf{r}, \mathbf{r}', \omega|s_0) \delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_0). \tag{A-5}$$

The Lippmann-Schwinger equation (A-2) describes a nonlinear relationship between the model perturbation δs and the scattered wavefield $\delta G(\mathbf{r}, \mathbf{r}_s, \omega|s_0, \delta s)$, while under the Born approximation (A-4), this nonlinear relationship reduces to a linear one, and the background Green's function and the perturbation of the Green's function can be calculated with the following equations

$$\left[\omega^2 s_0(\mathbf{r}) + \nabla^2\right] G(\mathbf{r}, \mathbf{r}_s, \omega|s_0) = -\delta(\mathbf{r} - \mathbf{r}_s), \tag{A-6}$$

$$\left[\omega^2 s_0(\mathbf{r}) + \nabla^2\right] \delta G_{Born}(\mathbf{r}, \mathbf{r}_s, \omega|s_0, \delta s) = -\omega^2 \delta s(\mathbf{r}') G(\mathbf{r}', \mathbf{r}_s, \omega|s_0) \tag{A-7}$$

by finite-difference method with point source and source term as $\omega^2 \delta s(\mathbf{r}')G(\mathbf{r}', \mathbf{r}_s, \omega|s_0)$, respectively.

REFERENCES

Alkhalifah, T., 2015, Scattering-angle based filtering of the waveform inversion gradients: Geophysical Journal International, **200**, No. 1, 363–373.

Alkhalifah, T., and Wu, Z. D., 2016, The natural combination of full and image-based waveform inversion: Geophysical Prospecting, **64**, No. 1, 19–30.

Biondi, B., and Almomin, A., 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, No. 3, Wa129–Wa140.

Brossier, R., Operto, S., and Virieux, J., 2015, Velocity model building from seismic reflection data by full-waveform inversion: Geophysical Prospecting, **63**, No. 2, 354–367.

Chavent, G., Cément, F., and Gómez, S., 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example, *in* SEG Technical Program Expanded Abstracts, vol. 328, 1179–1182.

Innanen, K. A., 2014, Seismic full waveform inversion with nonlinear sensitivities: CREWES Annual Report, **26**.

Innanen, K. A., 2015, Residual dependent fwi sensitivities based on direct nonlinear inverse scattering: CREWES Annual Report, **27**.

Jo, C. H., Shin, C. S., and Suh, J. H., 1996, An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator: Geophysics, **61**, No. 2, 529–537.

Kwak, S., Kim, Y., and Shin, C., 2014, Frequency-domain direct waveform inversion based on perturbation theory: Geophysical Journal International, **197**, No. 2, 987–1001.

Kwon, K., and Yazici, B., 2010, Born expansion and fréchet derivatives in nonlinear diffuse optical tomography: Computers and Mathematics with Applications, **59**, No. 11, 3377–3397.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, *in* Conference on Inverse Scattering, Theory and Application, Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Mcgillivray, P. R., and Oldenburg, D. W., 1990, Methods for calculating fréchet derivatives and sensitivities for the nonlinear inverse problem: a comparative study: Geophysical Prospecting, **38**, No. 5, 499–524.

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**, No. 2, 495–503.

Sava, P., and Biondi, B., 2004, Wave-equation migration velocity analysis. ii. subsalt imaging examples: Geophysical Prospecting, **52**, No. 6, 607–623.

Shin, C., and Cha, Y. H., 2008, Waveform inversion in the laplace domain: Geophysical Journal International, **173**, No. 3, 922–931.

Shin, C., and Cha, Y. H., 2009, Waveform inversion in the laplace-fourier domain: Geophysical Journal International, **177**, No. 3, 1067–1079.

Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, No. 1, 231–248.

Sun, D., and Symes, W. W., 2012, Waveform inversion via non-linear differential semblance optimization, *in* SEG Technical Program Expanded Abstracts, 1–7.

Symes, W. W., and Carazzone, J. J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, No. 5, 654–663.

Tarantola, A., 1984, Inversion of seismic-reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266.

van Leeuwen, T., and Mulder, W. A., 2008, Velocity analysis based on data correlation: Geophysical Prospecting, **56**, No. 6, 791–803.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, Wcc1–Wcc26.

Wang, H. Y., Singh, S. C., Audebert, F., and Calandra, H., 2015, Inversion of seismic refraction and reflection data for building long-wavelength velocity models: Geophysics, **80**, No. 2, R81–R93.

Wu, R. S., Luo, J. R., and Wu, B. Y., 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, No. 3, Wa13–Wa24.

Wu, R. S., and Zheng, Y. C., 2014, Non-linear partial derivative and its de wolf approximation for non-linear seismic inversion: Geophysical Journal International, **196**, No. 3, 1827–1843.

Wu, Z. D., and Alkhalifah, T., 2015, Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion: Geophysics, **80**, No. 6, R317–R329.

Xu, S., Wang, D., Chen, F., Lambaré, G., and Zhang, Y., 2012, Inversion on reflected seismic wave, *in* SEG Technical Program Expanded Abstracts, vol. 509, 1–7.

Zhou, W., Brossier, R., Operto, S., and Virieux, J., 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, No. 3, 1535–1554.