Full waveform inversion by model extension

Guillaume Barnier, Ettore Biondi, and Biondo Biondi

ABSTRACT

The main challenge inherent to full waveform inversion (FWI) is its inability to correctly recover the Earth's subsurface seismic parameters from inaccurate starting models. This behavior is due to the presence of local minima in the FWI objective function. To overcome this problem, we propose a new objective function in which we modify the nonlinear modeling operator of the FWI problem by adding a correcting term that ensures phase matching between predicted and observed data. This additional term is computed by demigrating an extended model variable, and its contribution is gradually removed during the optimization process while ensuring convergence to the true solution. Since the proposed objective function is quadratic with respect to the extended model variable, we make use of the variable projection method. We refer to this technique as full waveform inversion by model extension (FWIME). We provide a theoretical description of our method and we illustrate its potential on two synthetic examples for which FWI fails to retrieve the correct solution. First, by inverting data generated in a borehole setup. Then, by inverting diving waves recorded with a standard surface acquisition geometry. In both cases, we purposely choose a very inaccurate initial model and we show that FWIME manages to recover the true solution.

INTRODUCTION

FWI has the potential of inverting all model scales and providing high resolution subsurface images but it is greatly hampered by its sensitivity to the quality of the initial model, commonly known as cycle-skipping (Virieux and Operto, 2009). To overcome this issue, multiple methods have been developed based on adding optimization parameters or constraints to relax the condition that predicted and observed data should lie within a quarter of a wavelength from each other (Van Leeuwen and Herrmann, 2013; Biondi and Almomin, 2014; Warner and Guasch, 2014; Huang and Symes, 2015). Biondi and Almomin (2014) use an extended model approach to avoid the cycle-skipping issue, but the originally proposed method, based on a nested-scheme algorithm, heavily relies on the user to tune the many inversion parameters employed in the optimization process.

Following their work, we propose a new formulation also relying on an extended Born modeling operator to correct for substantial mismatches between observed and

predicted data. However, there are four main differences. First, we do not separate our model parameter into a background (low-wavenumber) and a strictly high-wavenumber perturbation. Secondly, during our inversion process, we aim at completely removing all the energy present in the extended model perturbation and we attach no physical meaning to it. Therefore, we do not need to apply any scale mixing or wavenumber filtering of any sort during our inversion. Thirdly, the use of the variable projection method to compute the adequate extended model perturbation allows us to better control the phase alignment of the data residuals (Golub and Pereyra, 1973; Rickett, 2013; Huang and Symes, 2015). Finally, our formulation reduces the number of optimization parameters to only one, the tradeoff between the data-fitting and the model regularization terms.

We illustrate the potential of this new technique on two synthetic examples where transmission data are inverted and where FWI fails to converge towards the true model. In the first example, we invert data recorded in a borehole setup. In the second one, we invert direct and diving waves recorded in a surface acquisition geometry. For the FWIME scheme, we do not follow a multiscale approach (Bunks et al., 1995) and all frequencies in the data are used simultaneously.

FWIME THEORY

Objective function formulation

In our formulation, we propose to minimize the following objective function

$$\Phi_\epsilon(\mathbf{m}, \tilde{\mathbf{p}}) = \frac{1}{2} \left\| \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \mathbf{d}^{obs} \right\|_2^2 + \frac{\epsilon^2}{2} \left\| \mathbf{D}\tilde{\mathbf{p}} \right\|_2^2 , \tag{1}$$

where $\mathbf{f}$ is the wave-equation operator, $\mathbf{m}$ is the velocity model, $\mathbf{d}^{obs}$ represents the observed data, $\tilde{\mathbf{B}}$ denotes the extended Born modeling operator, and $\tilde{\mathbf{p}}$ is an extended perturbation. Possible extensions include time-lags, subsurface offsets, or shot records (Biondi and Almomin, 2012, 2014; Huang and Symes, 2015). $\mathbf{D}$ is an invertible version of the differential semblance operator (DSO) that enhances the non-physical extended energy of $\tilde{\mathbf{p}}$ (Symes and Kern, 1994), and is linear with respect to $\tilde{\mathbf{p}}$. ϵ is the tradeoff parameter between the data-fitting and the regularization terms. The subscript in Φ_ϵ indicates that ϵ is a fixed parameter throughout the inversion. Since Φ_ϵ is quadratic with respect to $\tilde{\mathbf{p}}$ (for a fixed $\mathbf{m}$), we decide to use the variable projection method to minimize equation 1, which corresponds to minimizing the following objective function

$$\Phi_\epsilon(\mathbf{m}) \;=\; \frac{1}{2} \left\| \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs} \right\|_2^2 + \frac{\epsilon^2}{2} \left\| \mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) \right\|_2^2 , \tag{2}$$

where $\tilde{\mathbf{p}}_\epsilon^{opt}$ is an extended perturbation, defined as the minimizer of the objective function $\Phi_{\epsilon,\mathbf{m}}$ defined by

$$\Phi_{\epsilon,\mathbf{m}}(\tilde{\mathbf{p}}) = \frac{1}{2}\left\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}} - \left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right)\right\|_2^2 + \frac{\epsilon^2}{2}\left\|\mathbf{D}\tilde{\mathbf{p}}\right\|_2^2. \tag{3}$$

Assuming that the Hessian matrix of $\Phi_{\epsilon,\mathbf{m}}$ is invertible, $\tilde{\mathbf{p}}_\epsilon^{opt}$ is given by

$$\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) = \left[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right), \tag{4}$$

where $*$ denotes adjoint operators. Note that $\tilde{\mathbf{p}}_\epsilon^{opt}$ also depends nonlinearly on $\mathbf{m}$. The data residual component (first term) on the right side of equation 2 is a modified FWI objective function where an additional term $\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})$ is used to ensure the phase alignment between modeled and observed data. During the optimization process, we gradually reduce the contribution of this additional term by adding a regularization term on the right side of equation 2 (which pushes the L_2-norm of $\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})$ to zero). Therefore, finding the global minimizer of this equation is equivalent to finding the global minimizer of the conventional FWI objective function defined by

$$\Phi_{\mathrm{FWI}}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}}\right\|_2^2. \tag{5}$$

FWIME gradient

Equation 2 is minimized using a gradient-based descent method and we show in Appendix A that its gradient is given by

$$\nabla\Phi_\epsilon(\mathbf{m}) = \left[\mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m})\right]\left(\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{obs}\right), \tag{6}$$

where $\mathbf{B}^*$ is the adjoint of the non-extended Born modeling operator and $\mathbf{T}^*$ is the adjoint of the tomographic operator that depends on both $\mathbf{m}$ and the optimal perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$ (Sava and Biondi, 2004; Biondi and Almomin, 2014). The first component of the FWIME gradient in equation 6 (referred to as the Born gradient) is similar to the conventional FWI gradient but with a modified data residual. In addition, the second component (referred to as the tomographic gradient) may update regions of the model wavenumber spectrum missed by the first one (Barnier et al., 2018).

Effect of ϵ on the FWMIE objective function

We analyze the theoretical behavior of the FWIME objective function for two extreme cases where $\epsilon = 0$ and $\epsilon \rightarrow +\infty$. We support this analysis by numerical examples proposed in the next section.

Case when $\epsilon = 0$

When $\epsilon = 0$, the data objective function should (by construction) be approximately constant and equal to zero. This behavior is due to the fact that using an extended perturbation allows us to find $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ that accurately predicts the data that the nonlinear modeling $\mathbf{f}(\mathbf{m})$ has not been able to predict (Symes, 2008). In the specific case where we do not apply any regularization ($\epsilon = 0$), the data matching/misfit is almost identical (and equal to zero) for all models. However, the computation of $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ is achieved by an iterative method (linear conjugate gradient) and is truncated after a finite number of iterations, which may result in a non-perfect matching of the data residual in equation 4.

Case when $\epsilon \rightarrow +\infty$ and connection to the FWI objective function

So far, we have not been able to mathematically prove that FWIME converges to the true solution regardless of the accuracy of the starting model (while using gradient-based descent methods). This is still under investigation. However, if we claim that our method is more robust than conventional FWI, we should ensure that convergence of FWI to the true solution also implies convergence of FWIME. We prove in Appendix B that the FWIME objective function converges (pointwise in $\mathbf{m}$) towards the FWI objective function when ϵ tends to infinity. Therefore, whenever FWI converges to the true solution, FWIME should also converge to the true solution (for a high enough ϵ-value).

Choice of ϵ

For all the numerical examples shown in this report, our choice of ϵ-values are based on a trial and error approach, which lacks efficiency for field data applications. Therefore, we are currently researching an automatic and robust method to select the optimal value for the trade-off parameter ϵ (the unique user-adjustable parameter of our FWIME workflow). As we show throughout this report, this parameter greatly affects the shape of the FWIME objective function, making its estimation a crucial step to ensure the success of our method.

Given an initial model $\mathbf{m}_{init}$, one sufficient condition for either FWI or FWIME problems to converge to the true solution (using a gradient-based descent method)

is that the linearization of the modeling operator (about the initial model estimate $\mathbf{m}_{init}$) is accurate. For FWI, that would imply

$$\mathbf{d}^{obs} = \mathbf{f}(\mathbf{m}_{true}) \approx \mathbf{f}(\mathbf{m}_{init}) + \mathbf{F}(\mathbf{m}_{init})(\mathbf{m}_{true} - \mathbf{m}_{init}), \tag{7}$$

where $\mathbf{F}$ is the Jacobian operator of the modeling operator $\mathbf{f}$. The $\approx$ symbol indicates that the higher-order terms in equation 7 do not necessarily vanish, but do not introduce any local minima in the objective function. Our goal is not to quantify the accuracy to which equation 7 should be satisfied, because we have no control on it. For conventional FWI (and for a given dataset), the shape of the objective function is fixed. If the condition in equation 7 is not satisfied, the only way to recover the true model is by either changing the data (e.g., adding lower-frequency energy), improving the acquisition geometry, or improving the accuracy of our initial guess.

In our proposed method, the trade-off parameter ϵ provides an additional degree of freedom that may allow us to satisfy the FWIME-equivalent of equation 7 regardless of the initial model. We first recast equation 2 into a single-term objective function,

$$\Phi_\epsilon(\mathbf{m}) = \frac{1}{2} \left\| \mathbf{g}_\epsilon(\mathbf{m}) - \mathbf{D} \right\|_2^2, \tag{8}$$

where

$$\mathbf{g}_\epsilon(\mathbf{m}) = \begin{pmatrix} \mathbf{g}_\epsilon^1(\mathbf{m}) \\ \\ \mathbf{g}_\epsilon^2(\mathbf{m}) \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) \\ \\ \epsilon\, \mathbf{D}\, \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) \end{pmatrix}, \tag{9}$$

and

$$\mathbf{D} = \begin{pmatrix} \mathbf{d}^{obs} \\ \mathbf{0} \end{pmatrix}. \tag{10}$$

In the case of FWIME, equation 7 is expressed by

$$\mathbf{D} = \mathbf{g}_\epsilon(\mathbf{m}_{true}) \approx \mathbf{g}_\epsilon(\mathbf{m}_{init}) + \mathbf{G}_\epsilon(\mathbf{m}_{init})(\mathbf{m}_{true} - \mathbf{m}_{init}), \tag{11}$$

where $\mathbf{G}_\epsilon$ (the Jacobian operator of $\mathbf{g}_\epsilon$) is defined by

$$\mathbf{G}_\epsilon(\mathbf{m}) = \begin{pmatrix} \dfrac{\partial \mathbf{g}_\epsilon^1(\mathbf{m})}{\partial \mathbf{m}} \\ \\ \epsilon\, \mathbf{D}\, \dfrac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}} \end{pmatrix}. \tag{12}$$

The derivations of $\dfrac{\partial \mathbf{g}_\epsilon^1(\mathbf{m})}{\partial \mathbf{m}}$ and $\dfrac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}$ are shown in Appendix C.

Finally, given a true model $\mathbf{m}_{true}$ and an initial model $\mathbf{m}_{init}$, we are currently investigating a proof of existence of an ϵ-value and a set of models (that include $\mathbf{m}_{init}$) for which equation 11 holds. Our approach is to study the full FWIME Hessian operator and show that it can be made positive with an adequate choice of ϵ.

Computational cost of FWIME

The main computational cost (and bottleneck) of FWIME remains the variable projection step when the FWIME objective function (equation 2) is evaluated. For each evaluation of equation 2, $\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m})$ must be computed by minimizing $\Phi_{\epsilon,\mathbf{m}}$ in equation 3. This is equivalent to performing a least-squares reverse time migration (LSRTM) with an extended model (i.e., iteratively inverting the Hessian matrix of $\Phi_{\epsilon,\mathbf{m}}$). We are currently investigating ways to accelerate the convergence of the variable projection step. More specifically, we are implementing the approach proposed in Hou and Symes (2015) where the authors provide a computationally efficient method of approximating the inverse Hessian matrix of of $\Phi_{\epsilon,\mathbf{m}}$ for subsurface offsets extension. We are also working on developing this approximation for time-lag extension.

NUMERICAL EXAMPLES

We apply FWIME and conventional FWI on two synthetic problems in which noise-free data are generated using a two-way acoustic isotropic finite-difference scheme. When applying the FWIME workflow, we inject the full data bandwidth without taking a multiscale approach and we use nonlinear conjugate gradient to solve equations 2 and 5. In both examples, we use a time-lag extension for $\tilde{\mathbf{p}}$. The size of the extended axis is case dependent and determined by trial and error.

Inversion of borehole data

This simple transmission experiment is designed to illustrate the basic mechanisms of our method. We place sources in a vertical borehole every 50 m and receivers in the other vertical borehole every 10 m. The distance between the two boreholes is 1 km. The true velocity model is assumed to be uniform and equal to 2.5 km/s. We use a source wavelet containing frequencies ranging from 2.5 Hz to 35 Hz.

We sample the FWIME and FWI objective functions (equations 2 and 5) for uniform velocity models ranging from 2.0 km/s to 3.0 km/s by increments of 0.05 km/s. For each model, we compute the FWIME objective functions for five ϵ-values ranging from 0 to 1.0×10^{-4} (Figure 1). As expected, the FWI objective function presents local minima (dashed blue curve in Figure 1(a)). However, for certain ϵ-values, the FWIME objective function is monotonically decreasing towards the true solution. For these ϵ-values, FWIME managed to remove all local minima and guarantees global

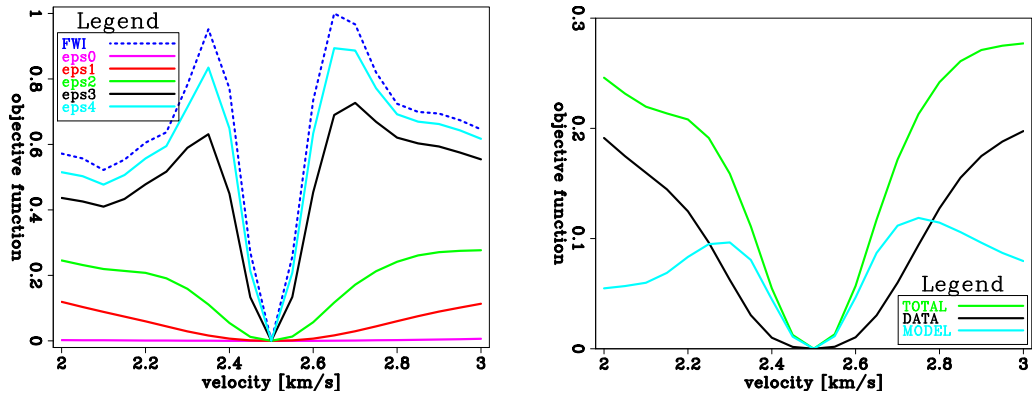
convergence for gradient-based methods when inverting for a model parametrized by a single scalar value (only allowing spatially uniform models). Figure 1(b) shows the breakdown of each component of the total FWIME objective function for ϵ_2 (solid green curve in Figure 1(a)). We observe that the data component is monotonically decreasing towards the true solution (solid black curve in Figure 1(b)) whereas the model displays a non-monotonically decreasing behavior (solid blue curve in Figure 1(b)).

Additionally, we observe that for $\epsilon = 0$, the FWIME objective function is approximately flat and equal to zero (magenta curve in Figure 1). This observation confirms that even for very inaccurate velocity models $\mathbf{m}$, the optimal extended perturbation $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ (equation 4) manages to capture the components present in the observed data $\mathbf{d}^{obs}$ that were not modeled by the nonlinear operator $\mathbf{f}(\mathbf{m})$, thus setting the data-fitting term in equation 2 to zero (up to numerical precision and number of iterations of linear conjugate gradient to minimize Equation 3). Conversely, we can observe in Figure 1 that when ϵ increases, the FWIME objective function converges towards the FWI one, which illustrates the pointwise convergence property mentioned in the previous section.

Figure 2(a) shows the convergence curves (solid lines) for a FWIME workflow applied using a uniform starting velocity model equal to 2.0 km/s. Throughout this inversion, we do not impose the inverted models to be uniform in space. We observe that all components of the FWIME objective function converge to zero. On the same plot, we superimpose the FWI objective function evaluated at each iteration of FWIME (dashed blue line). This curve is not the result of an inversion process, but simply the values that the FWI objective function would have taken for this sequence of inverted models. These observations show that the FWIME optimization path is insensitive to the local minima present in the conventional FWI objective function. In fact, conventional FWI starting from the same initial model fails to converge to the true solution (Figure 2(b)). This analysis is also confirmed by the average velocity of the inverted models from the two optimization schemes (Figure 2(c)). We choose this model metric because of the inherent uncertainty in the conventional traveltime tomography problem (Squires et al., 1994).

Inversion of direct arrivals and refracted waves

We invert synthetic data generated by a laterally invariant velocity model whose profile is shown in Figure 4(a) (dashed pink curve). The model is 12 km wide, 2.4 km deep, and does not contain any reflector. The generated data are therefore solely composed of direct arrivals and refracted energy (i.e., diving waves) (Figure 3(a)). We use a grid spacing of 30 m x 30 m for our finite-difference scheme, we place 400 receivers at the surface every 30 m and 198 sources every 60 m. Our initial model $\mathbf{m}_{init}$ is spatially uniform and set to 2.0 km/s (red curve in Figure 4(a)). In this example, we assume that the only coherent energy in the data lies within the 3-12 Hz frequency range.



(a) (b)

Figure 1: (a) FWIME (solid lines) and FWI (dashed blue line) objective functions for uniform velocity models and for increasing ϵ values ($\epsilon_0 = 0$, $\epsilon_1 = 3.0 \times 10^{-6}$, $\epsilon_2 = 1.0 \times 10^{-5}$, $\epsilon_3 = 5.0 \times 10^{-5}$, $\epsilon_4 = 1.0 \times 10^{-4}$). (b) FWIME objective function components for $\epsilon = \epsilon_2$. The total FWIME objective function (green curve) is identical to the one in Figure 1(a) (green curve). It is the sum of the data component (black curve) and the model component (light blue curve). [**CR**]

Conventional FWI approach

We first conduct a conventional time-domain FWI workflow for two different sets of frequency bands. For each band, we perform 100 iterations of nonlinear conjugate gradient. In the first band, we use energy from 3-5 Hz, which corresponds to what would be the first step of a multiscale approach. In the second band, we simultaneously use all the available frequencies in the data. Figures 4(a) and (b) show three inverted vertical profiles for each FWI workflow. The profiles are extracted at $x = 3$ km, $x = 6$ km, and $x = 9$ km, respectively. In both cases, conventional FWI fails to recover the correct velocity model for any depth greater than 0.4 km.

FWIME approach

We now apply our FWIME algorithm to the same problem by using the full 3-12 Hz frequency bandwidth, and we follow the workflow described in the previous section. We guide the reader through the first iteration of our algorithm. First, we set ϵ to 5.0×10^{-5} on the basis of trial-and-error, and we keep its value constant during the entire optimization scheme. We then compute the optimal extended perturbation $\tilde{\mathbf{p}}_\epsilon^{opt}$ by minimizing the objective function shown in equation 4 (variable projection step) using 35 iterations of linear conjugate gradient. Figure 5(a) show the zero time-lag cross-section of $\tilde{\mathbf{p}}_\epsilon^{opt}$. We observe that the linear optimization process has spread energy into the deeper parts of the section in the form of horizontal layering. Figure 5(b) displays a time-lag common image gather (CIG) extracted in the middle of the section, at $x = 6$ km. As expected, the energy is well focused at zero-lag for the

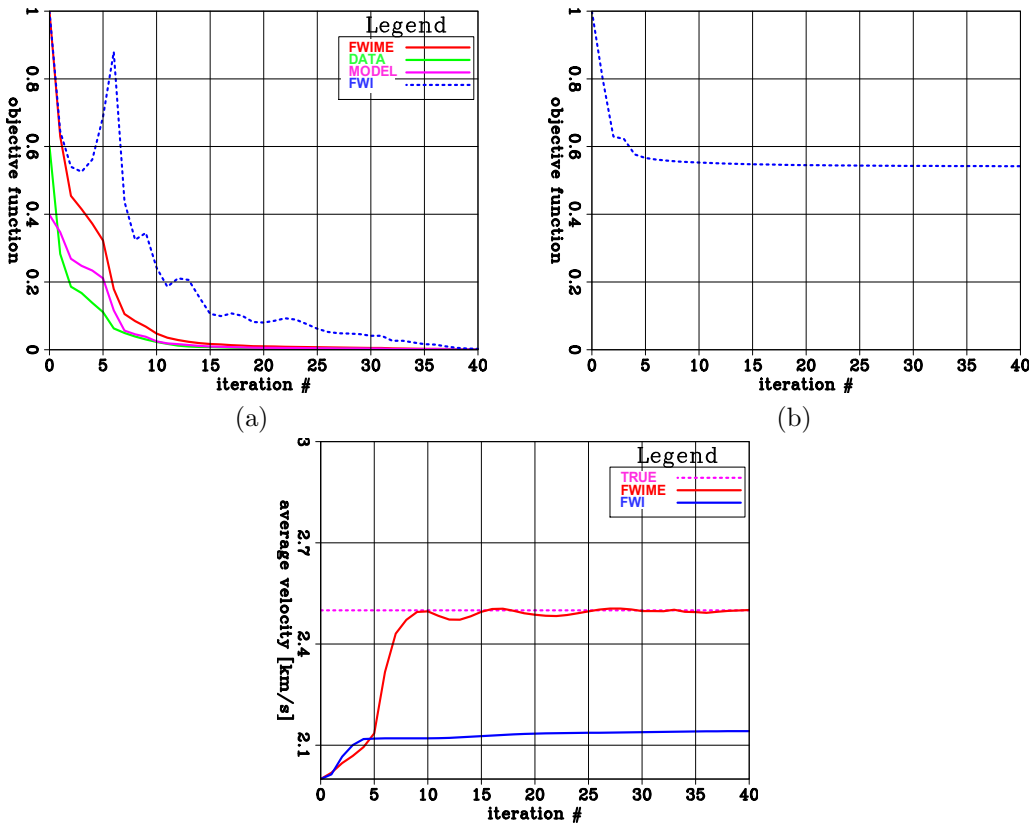


Figure 2: (a) Convergence curves for the FWIME workflow using $\epsilon = \epsilon_1$. The solid curves show the total, data, and model objective functions plotted in red, green and magenta, respectively. The dashed blue line corresponds to the values of the conventional FWI objective function evaluated at each iteration. (b) Convergence curve for conventional FWI. (c) Inverted average velocity as a function of iterations for FWIME (solid red curve) and conventional FWI (dashed blue line). [**CR**]

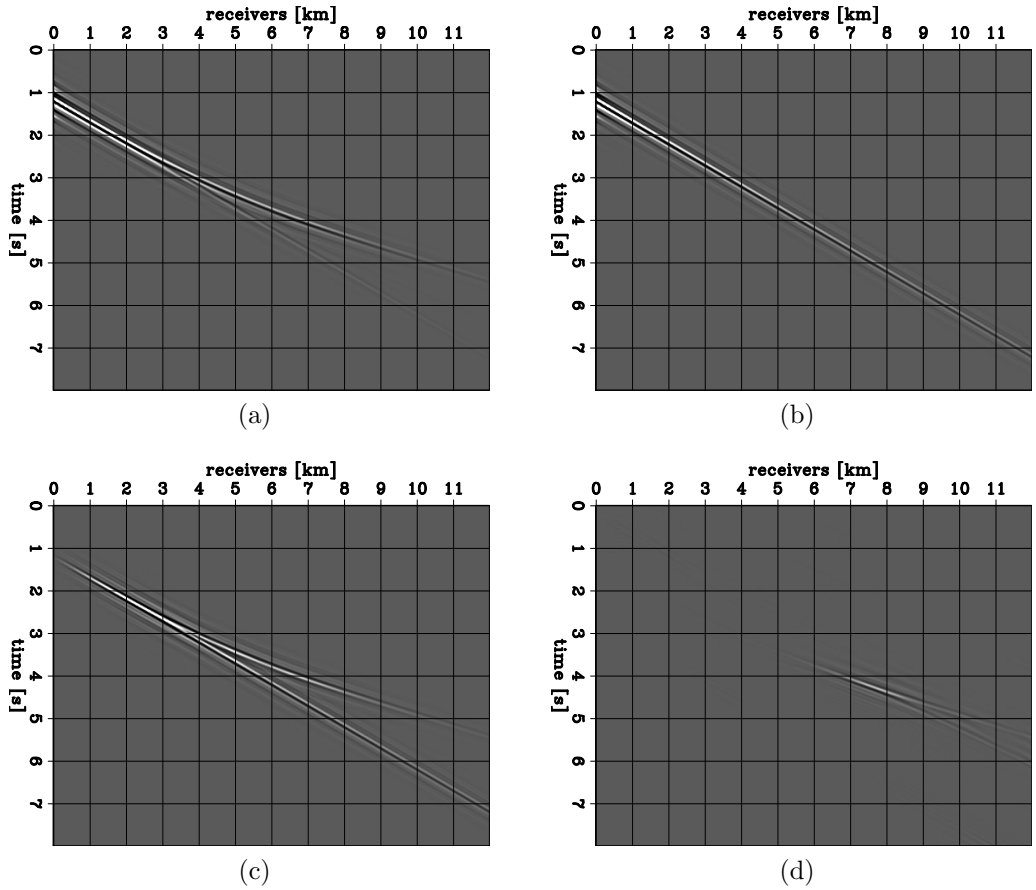
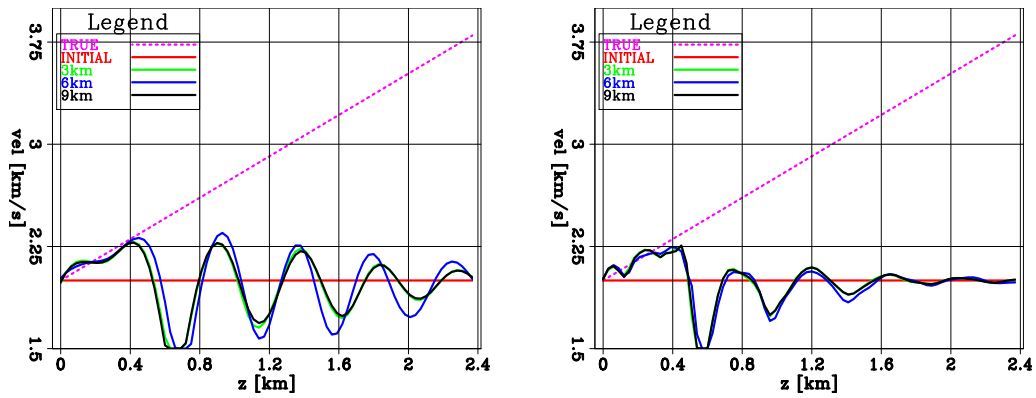


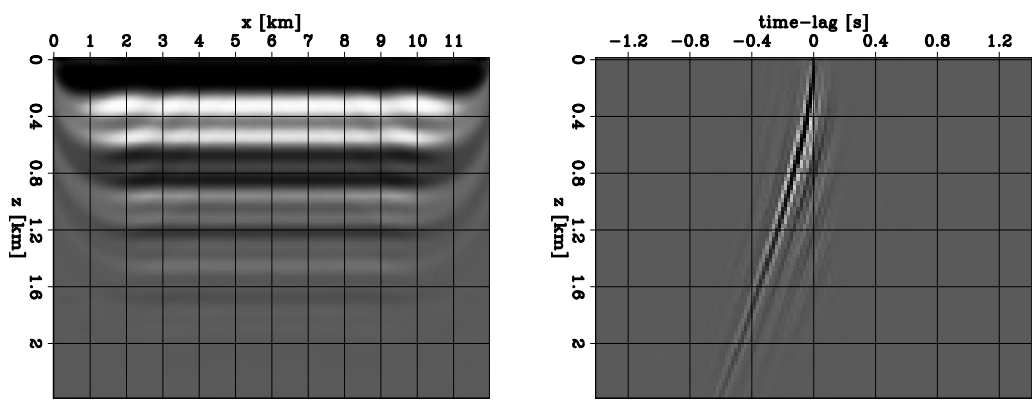
Figure 3: Shot gathers gerenated by a source located at $x = 0$ km. (a) Observed data, $\mathbf{d}^{obs}$. (b) Predicted data computed with the initial velocity model, $\mathbf{f}(\mathbf{m}_{init})$. (c) Initial FWI data residual $\mathbf{r}_{\text{FWI}} = \mathbf{f}(\mathbf{m}_{init}) - \mathbf{d}^{obs}$. (d) Initial FWIME data residuals (after computation of $\tilde{\mathbf{p}}_{\epsilon}^{opt}$), $\mathbf{r}_{\text{FWIME}} = \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\epsilon}^{opt}(\mathbf{m}) - \mathbf{d}^{obs}$. [**CR**]

shallow part of the model where the initial velocity is close to the true one. However, as we move deeper, the initial velocity model becomes increasingly inaccurate, and the coherent energy gradually moves away from the zero-lag axis. This illustrates the importance of using an extended model perturbation in order to capture events in the data that would have been lost otherwise (in this case, the refracted energy).



(a) (b)

Figure 4: Vertical velocity profiles extracted at three horizontal positions from FWI inverted models after 100 iterations. (a) Inverted model profiles using a frequency band of 3-5 Hz. (b) Inverted model profiles using a frequency band of 3-12 Hz. [**CR**]



(a) (b)

Figure 5: (a) Zero time-lag cross-section of the optimal extended perturbation at the first iteration of FWIME. (b) Time-lag CIG of the of the optimal extended perturbation at the first iteration of FWIME, extracted at $x = 6$ km. [**CR**]

The optimization of $\tilde{\mathbf{p}}_\epsilon^{opt}$ is the most computationally intensive step of our workflow, and it is crucial as it plays a double role in the gradient computation. On one hand, it is demigrated in order to generate the correcting term in the data residual, $\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})$. Figure 3(d) shows the FWIME data residual obtained at the first iteration after computation of the optimal model perturbation. For this specific shot position, most of the direct arrivals have been matched but a portion of the diving waves is still present. This is due to the presence of the regularization term since

$\epsilon \neq 0$. Note that this data residual is then injected in the Born and tomographic adjoint operators to compute the first gradient (equation 6). On the other hand, $\tilde{\mathbf{p}}_{\epsilon}^{opt}$ is also used by the tomographic operator itself. So far, we have observed that the quality of the tomographic gradient is quite sensitive to the accuracy with which we solve for the optimal extended perturbation (i.e., the number of iterations in the variable projection step).

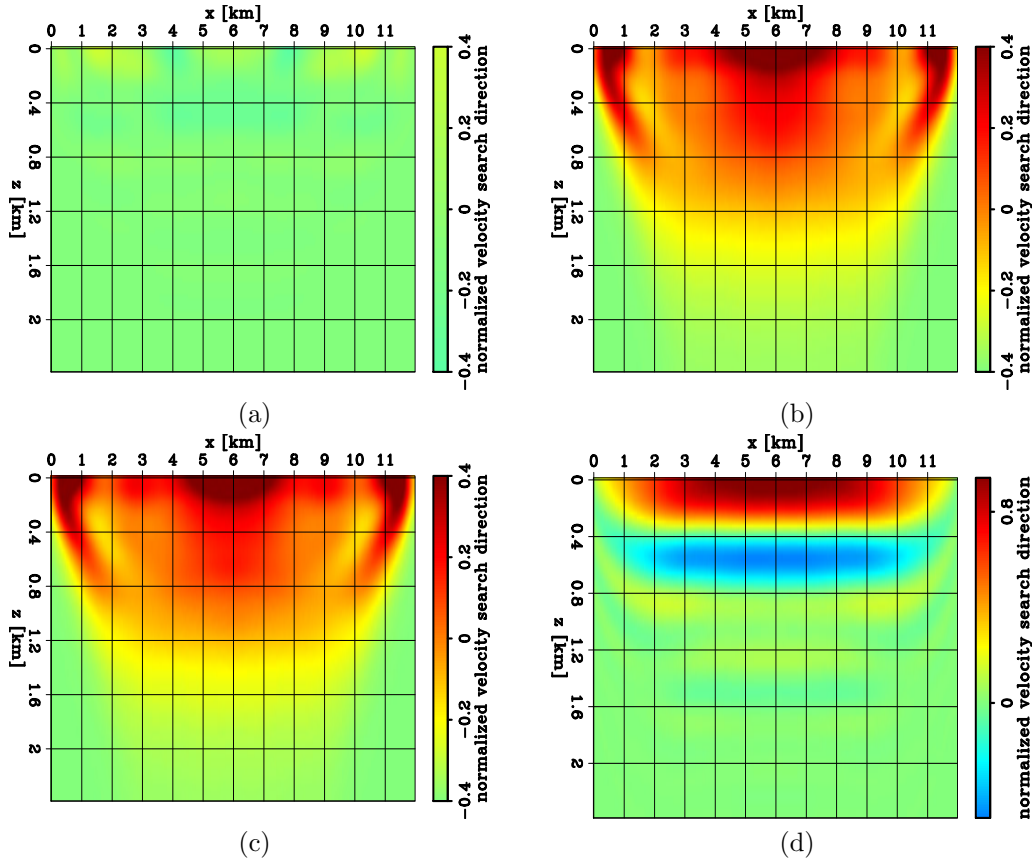


Figure 6: First velocity search directions. (a) Born component of the first search direction. (b) Tomographic component of the first search direction. (c) Total search direction. (d) First FWI search direction for the 3-5 Hz frequency band. [**CR**]

Figures 6(a)-(d) show the Born, tomographic, and total first search directions of our FWIME workflow. The total search search direction seems promising and it is mostly dominated by the tomographic component. Figure 6(d) shows the first FWI search directions for the 3-5 Hz frequency band. We can clearly see that the FWIME search direction has a more balanced amplitude and its phase is accurate up to a much deeper part of the model (approximately 1.2 km). By performing a truncated Gauss-Newton step of the FWI workflow, we would probably improve the amplitude balancing of the search direction in Figure 6(d), but the incorrect phase shift occurring in the very shallow parts would remain.

We conduct 10 iterations of our FWIME workflow followed by 100 iterations of conventional FWI. Figure 7(a) shows three inverted profiles (extracted at $x = 3$ km,

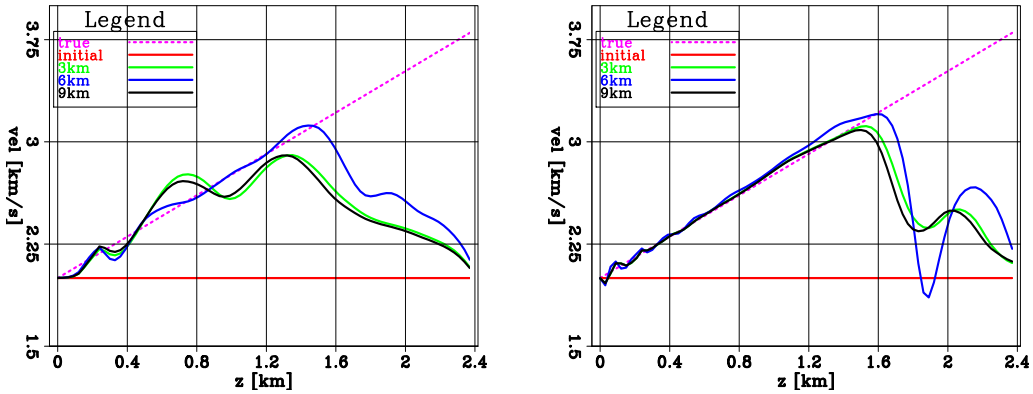


Figure 7: (a) Vertical inverted profiles after 10 iterations of FWIME extracted at $x = 3$ km, $x = 6$ km, and $x = 9$ km. (b) Vertical inverted profiles after 10 iterations of FWIME followed by 100 iterations of conventional FWI, extracted at $x = 3$ km, $x = 6$ km, and $x = 9$ km. [**CR**]

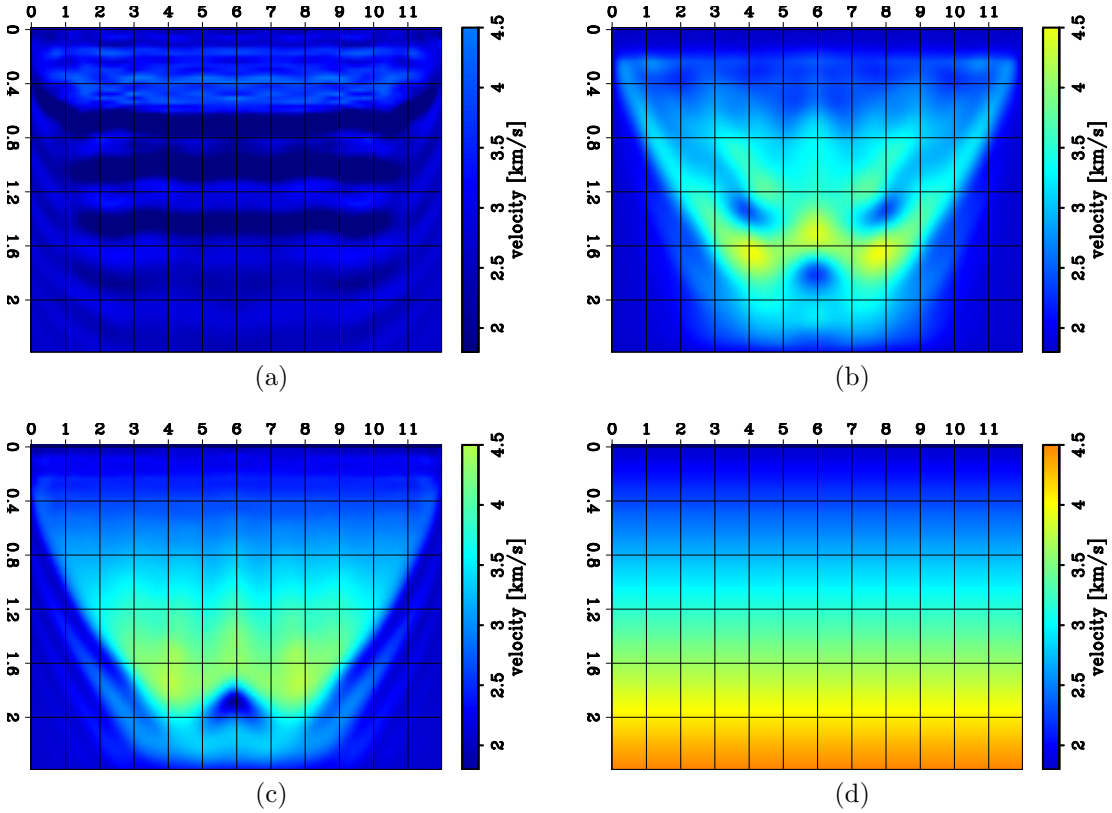


Figure 8: Inverted models using a 3-12 Hz frequency bandwidth. (a) Inverted model after 100 iterations of conventional FWI. (b) Inverted model after 10 iterations of FWIME. (c) Inverted model after 10 iterations of FWIME followed by 100 iterations of FWI. (d) True model. [**CR**]

$x = 6$ km, and $x = 9$ km) after 10 iterations of FWIME. The FWIME workflow managed to capture the long wavenumber components of the model up to a depth of approximately 1.2 km. Furthermore, our inability to recover deeper parts of the model is likely stemming from a lack of offset range in our acquisition geometry, and not from a cycle-skipping issue. We then use the FWIME inverted model as the initial model for a conventional FWI workflow using the full frequency bandwidth. Figure 7(b) shows the inverted profiles after 100 iterations of conventional FWI. Figures 8(a)-(d) show the 2D-panels of the inverted models for FWI, FWIME, and FWIME followed by FWI. We can observe that once FWIME captured the long wavelength component of the model, FWI was able to accurately recover the true model without the need to adopt a multi-scale approach.

CONCLUSIONS

We develop a new method that addresses the cycle-skipping issues inherent to conventional FWI. We modify the nonlinear modeling operator of the FWI problem by adding a correcting term that ensures phase matching between predicted and observed data, and we gradually reduce its contribution over iterations by adding a regularization term controlled by a single parameter. This formulation mitigates the need for the user to tune many parameters during the inversion process or to follow a multiscale approach as in FWI. We show its potential and success on two synthetic transmission problems where FWI fails to recover the true model.

ACKNOWLEDGEMENTS

We would like to thank the Stanford Exploration Project affiliate companies for financial support. Guillaume Barnier would also like to thank Joseph Dellinger, Esteban Diaz, Simon Luo, and the BP America Subsurface Technical Centre (STC) for their useful advice and constructive comments.

REFERENCES

Barnier, G., E. Biondi, and B. Biondi, 2018, Tomographic full waveform inversion with variable projection: Presented at the 80th EAGE Conference and Exhibition 2018.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (tfwi) by combining full waveform inversion with wave-equation migration velocity anaylisis: Presented at the 2012 SEG Annual Meeting.

———, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B., P. Sava, et al., 1999, Wave-equation migration velocity analysis: 69th Ann. Internat. Mtg Soc. of Expl. Geophys, 1723–1726.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Golub, G. H. and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on numerical analysis, **10**, 413–432.

Hou, J. and W. W. Symes, 2015, An approximate inverse to the extended born modeling operatoran approximate inverse operator: Geophysics, **80**, R331–R349.

Huang, Y. and W. W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: SEG Technical Program Expanded Abstracts 2015, 1326–1331, Society of Exploration Geophysicists.

Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, 874–881.

Sava, P. and B. Biondi, 2004, Wave-equation migration velocity analysis. i. theory: Geophysical Prospecting, **52**, 593–606.

Squires, L. J., P. L. Stoffa, and G. Cambois, 1994, Borehole transmission tomography for velocity plus statics: Geophysics, **59**, 1028–1036.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical prospecting, **56**, 765–790.

Symes, W. W. and M. Kern, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **42**, 565–614.

Van Leeuwen, T. and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Warner, M. and L. Guasch, 2014, Adaptive waveform inversion-fwi without cycle skipping-theory: Presented at the 76th EAGE Conference and Exhibition 2014.

APPENDIX A

We derive the gradient expression in equation 6. We define

$$\mathbf{r}_d(\mathbf{m}) = \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}} \tag{13}$$

$$\mathbf{r}_m(\mathbf{m}) = \mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}). \tag{14}$$

The gradient of Φ_ϵ is given by

$$\nabla\Phi_\epsilon(\mathbf{m}) = \left(\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{r}_d(\mathbf{m}) + \epsilon^2 \left(\frac{\partial \mathbf{r}_m(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{r}_m(\mathbf{m}), \tag{15}$$

and we have

$$\begin{aligned}
\left(\frac{\partial \mathbf{r}_d(\mathbf{m})}{\partial \mathbf{m}}\right)^* &= \left(\frac{\partial \mathbf{f}(\mathbf{m})}{\partial \mathbf{m}}\right)^* + \left(\frac{\partial(\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}))}{\partial \mathbf{m}}\right)^* \\
&= \mathbf{B}^*(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{B}}(\mathbf{m})}{\partial \mathbf{m}}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \\
&= \mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})^*, \qquad (16)
\end{aligned}$$

where $\mathbf{T}^*(\mathbf{m}) = \left(\frac{\partial \tilde{\mathbf{B}}(\mathbf{m})}{\partial \mathbf{m}}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\right)^*$. Similarly,

$$\left(\frac{\partial \mathbf{r}_m(\mathbf{m})}{\partial \mathbf{m}}\right)^* = \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{D}^*. \qquad (17)$$

Equation 15 becomes

$$\begin{aligned}
\nabla\Phi_\epsilon(\mathbf{m}) &= \left[\mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})^*\right]\mathbf{r}_d(\mathbf{m}) + \epsilon^2\left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \mathbf{D}^*\mathbf{r}_m(\mathbf{m}) \\
&= \left[\mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m})\right]\mathbf{r}_d(\mathbf{m}) + \left(\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}\right)^* \left[\tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{r}_m(\mathbf{m})\right]. \qquad (18)
\end{aligned}$$

Since $\tilde{\mathbf{p}}_\epsilon^{opt}$ satisfies equation 4, we have

$$\left[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\right]\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) = \tilde{\mathbf{B}}^*(\mathbf{m})(\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m})). \qquad (19)$$

Therefore,

$$\begin{aligned}
\tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{r}_d(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{r}_m(\mathbf{m}) &= \tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}}\right) + \epsilon^2\mathbf{D}^*\mathbf{D}\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) \\
&= \left[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\right]\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \tilde{\mathbf{B}}^*(\mathbf{m})(\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m})) \\
&= \mathbf{0}. \qquad (20)
\end{aligned}$$

Finally, equation 15 reduces to

$$\nabla\Phi(\mathbf{m}) = \left[\mathbf{B}^*(\mathbf{m}) + \mathbf{T}^*(\mathbf{m})\right]\left(\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) - \mathbf{d}^{\mathbf{obs}}\right). \qquad (21)$$

Note that the fact that equation 19 is satisfied allows us to avoid the expensive computation of $\frac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}$.

APPENDIX B

We show that the FWIME objective function converges pointwise in $\mathbf{m}$ towards the FWI objective function (equation 5) when $\epsilon \to +\infty$.

1. We first prove that $\forall\, \mathbf{m},\ \lim_{\epsilon\to\infty} \frac{\epsilon^2}{2}\|\mathbf{D}\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 = 0$, and
2. we conclude that $\forall\, \mathbf{m},\ \lim_{\epsilon\to\infty} \Phi(\mathbf{m},\epsilon) = \Phi_{\mathrm{FWI}}(\mathbf{m})$,

where Φ is the FWIME objective function. The optimal extended perturbation $\tilde{\mathbf{p}}^{opt}$ is given by

$$\begin{aligned} \tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon) &= \left[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right) \\ &= \frac{1}{\epsilon^2}\left[\frac{1}{\epsilon^2}\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right). \end{aligned} \tag{22}$$

When $\epsilon \to +\infty$, we can make the following approximation

$$\frac{1}{\epsilon^2}\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \mathbf{D}^*\mathbf{D} \approx \mathbf{D}^*\mathbf{D}. \tag{23}$$

Therefore,

$$\begin{aligned} \tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon) &\approx \frac{1}{\epsilon^2}\left[\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right) \\ &\approx \frac{1}{\epsilon^2}\tilde{\mathbf{q}}(\mathbf{m}), \end{aligned} \tag{24}$$

where $\tilde{\mathbf{q}}(\mathbf{m}) = \left[\mathbf{D}^*\mathbf{D}\right]^{-1}\tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{d}^{obs} - \mathbf{f}(\mathbf{m})\right)$. Assuming that $\tilde{\mathbf{q}}(\mathbf{m})$ is bounded, we can deduce that for any $\mathbf{m}$,

- $\|\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$, and that
- $\frac{\epsilon^2}{2}\|\mathbf{D}\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$.

From equations 2 and 5,

$$\Phi(\mathbf{m},\epsilon) - \Phi_{\text{FWI}}(\mathbf{m}) = \frac{1}{2}\|\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2 + \frac{\epsilon^2}{2}\|\mathbf{D}\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m})\|_2^2 - \frac{1}{2}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2. \tag{25}$$

Moreover,

$$\|\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2 = \|\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2 + \|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m})\|_2^2 + 2\,\left(\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m}). \tag{26}$$

$\|\tilde{\mathbf{p}}^{opt}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$ implies that $\|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{\text{opt}}(\mathbf{m},\epsilon)\|_2^2 \underset{\epsilon\to\infty}{\to} 0$, and by the Cauchy-Schwarz inequality,

$$\left|\left(\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{\text{opt}}(\mathbf{m},\epsilon)\right| \leq \|\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2\, \|\tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{\text{opt}}(\mathbf{m},\epsilon)\|_2. \tag{27}$$

Therefore,

- $\left(\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\right)^* \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}^{\text{opt}}(\mathbf{m},\epsilon) \underset{\epsilon\to\infty}{\to} 0$, and
- $\|\mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{p}}_{\text{opt}}(\mathbf{m},\epsilon) - \mathbf{d}^{\text{obs}}\|_2^2 - \|\mathbf{f}(\mathbf{m}) - \mathbf{d}^{\text{obs}}\|_2^2 \underset{\epsilon\to\infty}{\to} 0.$

Finally, we can deduce that

$$\forall\, \mathbf{m},\ \lim_{\epsilon\to\infty} \Phi(\mathbf{m},\epsilon) = \Phi_{\text{FWI}}(\mathbf{m}). \tag{28}$$

APPENDIX C

Here, we derive the expression of the FWIME linearized forward modeling operator (Jacobian matrix of $\mathbf{g}_\epsilon^1$) that we refer to as $\mathbf{G}_\epsilon^1$. It is given by

$$\mathbf{G}_\epsilon^1(\mathbf{m}) = \mathbf{B}(\mathbf{m}) + \mathbf{T}_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m}) + \tilde{\mathbf{B}}\tilde{\mathbf{P}}_\epsilon(\mathbf{m}), \tag{29}$$

where $\tilde{\mathbf{P}}_\epsilon(\mathbf{m}) = \dfrac{\partial \tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})}{\partial \mathbf{m}}$. The main difficulty is to compute $\tilde{\mathbf{P}}_\epsilon(\mathbf{m})$ (or its effect when applied to a model perturbation). We differentiate each side of equation 19 with respect to $\mathbf{m}$, and we define

- $\mathbf{E} = \dfrac{\partial}{\partial \mathbf{m}}\Big(\big[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\big]\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\Big)$, and
- $\mathbf{F} = \dfrac{\partial}{\partial \mathbf{m}}\Big(\tilde{\mathbf{B}}^*(\mathbf{m})(\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m}))\Big)$.

$\mathbf{E}$ can be expressed by

$$\begin{aligned}\mathbf{E} &= \frac{\partial}{\partial \mathbf{m}}\Big(\big[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\big]\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m})\Big) \\ &= \frac{\partial}{\partial \mathbf{m}}\big[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\big]\tilde{\mathbf{p}}_\epsilon^{opt}(\mathbf{m}) + \mathbf{H}_{\Phi_{\epsilon,\mathbf{m}}}\tilde{\mathbf{P}}_\epsilon(\mathbf{m}),\end{aligned} \tag{30}$$

where $\mathbf{H}_{\Phi_{\epsilon,\mathbf{m}}} = \big[\tilde{\mathbf{B}}^*(\mathbf{m})\tilde{\mathbf{B}}(\mathbf{m}) + \epsilon^2\mathbf{D}^*\mathbf{D}\big]$ is the Hessian matrix of $\Phi_{\epsilon,\mathbf{m}}$ (equation 3). Therefore,

$$\mathbf{E} = \tilde{\mathbf{W}}_{\tilde{\mathbf{B}}(\mathbf{m})\mathbf{p}_\epsilon^{opt}} + \tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{T}_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m}) + \mathbf{H}_{\Phi_{\epsilon,\mathbf{m}}}\tilde{\mathbf{P}}_\epsilon(\mathbf{m}), \tag{31}$$

where $\tilde{\mathbf{W}}$ is the wave equation migration velocity analysis (WEMVA) operator (Biondi et al., 1999) extended in time-lags. Similarly,

$$\mathbf{F} = \tilde{\mathbf{W}}_{\mathbf{d}^{\mathbf{obs}} - \mathbf{f}(\mathbf{m})} - \tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{B}(\mathbf{m}). \tag{32}$$

Combining equations 31 and 32, we can write

$$\mathbf{H}_{\Phi_{\epsilon,\mathbf{m}}}\tilde{\mathbf{P}}_\epsilon(\mathbf{m}) = \tilde{\mathbf{W}}_{-\mathbf{r}_d} - \tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{T}_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m}) - \tilde{\mathbf{B}}^*(\mathbf{m})\mathbf{B}(\mathbf{m}), \tag{33}$$

where $\mathbf{r}_d = \mathbf{f}(\mathbf{m}) + \tilde{\mathbf{B}}(\mathbf{m})\mathbf{p}_\epsilon^{opt} - \mathbf{d}^{\mathbf{obs}}$ are the FWIME objective function data residuals. Therefore, the application of $\mathbf{G}_\epsilon$ to a model perturbation $\Delta\mathbf{m}$ is given by

$$\mathbf{G}_\epsilon^1(\mathbf{m})\Delta\mathbf{m} = \Big(\mathbf{B}(\mathbf{m}) + \mathbf{T}_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m})\Big)\Delta\mathbf{m} + \tilde{\mathbf{B}}(\mathbf{m})\tilde{\mathbf{P}}_\epsilon(\mathbf{m})\Delta\mathbf{m}, \tag{34}$$

$\tilde{\mathbf{P}}_\epsilon(\mathbf{m})\Delta\mathbf{m}$ is the solution of the following linear system,

$$\mathbf{H}_{\Phi_{\epsilon,\mathbf{m}}}\mathbf{x} = \mathbf{y}, \tag{35}$$

where $\mathbf{y}$ is expressed by

$$\mathbf{y} = \tilde{\mathbf{W}}_{-\mathbf{r}_d}\Delta\mathbf{m} - \tilde{\mathbf{B}}^*(\mathbf{m})\left(\mathbf{T}_{\tilde{\mathbf{p}}_\epsilon^{opt}}(\mathbf{m}) + \mathbf{B}(\mathbf{m})\right)\Delta\mathbf{m}. \tag{36}$$

Equation 35 can be solved using a linear conjugate gradient scheme.