Reformulating TFWI

Ali Almomin

ABSTRACT

Tomographic full waveform inversion (TFWI) provides a robust but expensive method to invert the seismic data. Scale separation of the model greatly reduces the cost but adds complexity to theory and the implementation of the inversion. In this paper, I provide two approaches that reduce the complexity of TFWI. First, I rederive TFWI with one model only in an abstract formulation that is applicable to any form of wave-equation. Then, I provide a new approximation to the inversion that can potentially provide more accurate results.

INTRODUCTION

Previously, we reduced the cost of TFWI by separating the extended model into two components: a non-extended smooth background and an extended rough perturbation (Biondi and Almomin, 2012; Almomin and Biondi, 2013). This might have caused some confusion on the resulting relationship and balance between these two parameters and their relationship to the original model. Furthermore, the interpretation of these two parameters limited the way we could separate them and increased the difficulty of moving to different wave-equations, such as the elastic. To overcome these limitations, I first derive the "original" TFWI in an abstract way that makes it applicable to different forms of the wave-equation. Then, I derive TFWI using the two-parameter approach. Next, I rederive TFWI while keeping one abstract model. Finally, I provide an alternative method to derive TFWI that is potentially more accurate than the previous TFWI and closer to the original TFWI.

THEORY

To solve any nonlinear inversion problem in a gradient-based method, we only need to evaluate two quantities: the objective function and the gradient. Evaluating the objective function requires the forward modeling operator whereas evaluating the gradient requires the linearized adjoint operator, i.e., the adjoint of the linearized forward operator. The linearized forward operator can be computed simply by taking the derivative of the forward modeling operator with respect to the model space. This linearized operator relates a (preferably small) perturbation of the model space to a perturbation in the data space. In the wave-equation, the effect of linearized operator is an explicit scattering of the waves by the model perturbations or an

imaging these perturbations by the adjoint operator. I will refer to this type of scattering as linearization scattering. For the purposes of this report, I will ignore the regularization term of TFWI and only focus on the data fitting term. Moreover, a tilde ($\sim$) above a parameter or operator indicates that it is extended.

TFWI

First, we look at TFWI without any approximations or shortcuts, which I refer to as the original TFWI. The original TFWI objective function is:

$$J_{\text{TFWI}}(\widetilde{\mathbf{m}}) = \frac{1}{2}\|\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) - \mathbf{d}_{\text{obs}}\|_2^2 = \frac{1}{2}\|\mathbf{r}(\widetilde{\mathbf{m}})\|_2^2, \tag{1}$$

where $\widetilde{\mathbf{m}}$ is the extended model, $\widetilde{\mathbf{f}}$ is the extended forward modeling operator, $\mathbf{d}_{\text{obs}}$ is the observed surface data, and $\mathbf{r}$ is the residual. In the extended modeling operator, all parts of the extended model interact with the wavefield in every propagation step, which is the main source of cost increase in computation. The gradient of the objective function is:

$$\frac{\partial J_{\text{TFWI}}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \left(\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}}\right)^* \mathbf{r}(\widetilde{\mathbf{m}}). \tag{2}$$

The derivative of the residual is computed as:

$$\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \frac{\partial \widetilde{\mathbf{f}}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \widetilde{\mathbf{L}}(\widetilde{\mathbf{m}}), \tag{3}$$

where $\widetilde{\mathbf{L}}$ is the extended linearized modeling operator. I can now compute the gradient as follows:

$$\mathbf{g}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{L}}^*(\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}). \tag{4}$$

Although the previous equations are simple and direct, they can be very expensive and difficult to compute, depending on the extension axes of choice. Therefore, we need to find a cheaper approach to approximating these computations.

Two-parameter approximation

The previously published approximation to TFWI utilizes the two-parameter approximation that separates the extended model into two components:

$$\widetilde{\mathbf{m}} \approx \mathbf{b} + \widetilde{\mathbf{p}}, \tag{5}$$

where $\mathbf{b}$ is the non-extended smooth background component and $\widetilde{\mathbf{p}}$ an extended rough perturbation component. The underlying assumption is that the extended portion will not have a smooth component. It is also easy to understand conceptually: we are separating the parts of the model that affect the transmission (or forward scattering)

from the parts that affect the reflection (or back scattering). Now, I can approximate the modeling operator using Taylor's expansion as follows:

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{f}}(\mathbf{b}+\widetilde{\mathbf{p}}) \approx \mathbf{f}(\mathbf{b}) + \widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}}. \tag{6}$$

This approximate modeling operator simply adds the data resulting from propagating the wavefield with the smooth component of the model (the $\mathbf{f}(\mathbf{b})$ term) to the data resulting from scattering the wavefield with the rough component of the model (the $\widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}}$ term). It is important to notice that this scattering is a result of the approximation we used by model separation, which is similar to but independent from the linearization scattering due to gradient calculation. I will refer to this second type of scattering as approximation scattering. Both the linearization scattering and approximation scattering use Taylor's expansion, which might make things a bit confusing. Moreover, the scattering with the rough component happens only once, so multiples cannot be modeled with this operator unless the rough component was added to the smooth background. Using the approximate operator, the residuals can be calculated as:

$$\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}) = \mathbf{f}(\mathbf{b}) + \widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}} - \mathbf{d}_{\text{obs}}. \tag{7}$$

Next, I calculate the residual derivative with respect to the perturbation as:

$$\frac{\partial \mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})}{\partial \widetilde{\mathbf{p}}} = \widetilde{\mathbf{L}}(\mathbf{b}), \tag{8}$$

and the residual derivative derivative with respect to the background as:

$$\frac{\partial \mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})}{\partial \mathbf{b}} = \frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}} + \frac{\partial \widetilde{\mathbf{L}}(\mathbf{b})}{\partial \mathbf{b}}\widetilde{\mathbf{p}} = \mathbf{L}(\mathbf{b}) + \mathbf{T}(\mathbf{b},\widetilde{\mathbf{p}}), \tag{9}$$

where $\mathbf{T}$ is the tomographic operator that is a function of two variables: a propagation variable and a scattering variable. This tomographic operator has two scattering terms because we first applied the approximation linearization, and then perturbed that model using the linearization scattering, which is effectively a scattering of a scatterer. In other words, it is the second derivative of the modeling operator with respect to the model. Finally, I can calculate the gradient of $\widetilde{\mathbf{p}}$ as:

$$\mathbf{g}_{\widetilde{\mathbf{p}}}(\mathbf{b},\widetilde{\mathbf{p}}) = \widetilde{\mathbf{L}}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}), \tag{10}$$

and the gradient of $\mathbf{b}$ as:

$$\mathbf{g}_{\mathbf{b}}(\mathbf{b},\widetilde{\mathbf{p}}) = \mathbf{L}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}) + \mathbf{T}^*(\mathbf{b},\widetilde{\mathbf{p}})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}). \tag{11}$$

One issue with the gradient of $\mathbf{b}$ is the term $\mathbf{L}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})$ which can potentially cycle-skip since the operator is not extended. Another issue is that multiples will not be modeled (neither by the non-extended part nor by the extended part) unless we feed the extended perturbation back into the non-extended background. To circumvent these problems, Almomin and Biondi (2013) proposed the nested inversion scheme where the objective function was separated into two components and the observed

data was redefined in order to avoid this term in the gradient of $\mathbf{b}$. This nested scheme requires a balance between how many iterations we perform in each inner loop as well as a good balance on how to mix and separate the smooth and rough components. Having these components in different dimensions only adds complexity to the understanding of the problem.

One-parameter approximation

The previous approximation is based on the fact that the extended part of the model only contains rough components. Furthermore, it requires a nested inversion scheme with several steps that ensure different scales of the model are simultaneously inverting. Also, implementing the previous scheme for other wave-equations, such as the elastic wave-equation, might not be trivial in terms of which variable to extend and how to separate and mix them. Therefore, I rederive the previous approximation in an equivalent but generalized way using one parameter and an abstract separation operator as follows:

$$\widetilde{\mathbf{m}} = \mathbf{C}\widetilde{\mathbf{m}} + [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}, \tag{12}$$

where $\mathbf{C}$ is the separation operator. For the operator approximation to be accurate, I need the following inequality to hold:

$$\|\mathbf{C}\widetilde{\mathbf{m}}\| \gg \|[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}\|. \tag{13}$$

Now, I can approximate the modeling operator using Taylor's expansion as follows

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}} + [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}. \tag{14}$$

For the previous equation to be efficient, the operator $\mathbf{C}$ needs to restrict the extended part of the model. From the two requirements of the operator $\mathbf{C}$, we can see that separating the model into smooth and rough components is not the only possibility. In fact, it is more optimal to separate it into a non-extended part and an exteded part, both of which can contain rough or smooth components. This latter separation will make the approximation scattering even smaller compared to the propagation term. The residual can now be written as:

$$\mathbf{r}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}} - \mathbf{d}_{\text{obs}}. \tag{15}$$

Next, I calculate the residual derivative as follows:

$$\begin{aligned}\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} &= \frac{\partial \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} + \frac{\partial \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}}[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}} + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}] \\ &= \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{C} + \widetilde{\mathbf{T}}(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}})\mathbf{C} + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}].\end{aligned} \tag{16}$$

Since I used one model parameter only, the terms can be cancelled resulting in the following residual derivative:

$$\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{T}}(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}})\mathbf{C}. \tag{17}$$

Finally, I can calculate the gradient as follows:

$$\widetilde{\mathbf{g}}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{L}}^*(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}) + \mathbf{C}^*\widetilde{\mathbf{T}}^*(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I}-\mathbf{C}]\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}). \tag{18}$$

Alternative approximation

Now I present an alternative formulation to approximate the expensive TFWI. Instead of first approximating the modeling operator and then taking the derivative of the residual, I first take the "expensive" derivative followed approximating the calculation. This way, I can focus on approximating the propagation operator only since it is the most expensive step, instead of approximating the modeling operator. To do so, I need to rewrite our modeling operator and gradient as functions of the propagation operator. The modeling operator can be written as:

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) = \mathbf{K}^*\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{K}\mathbf{s}, \tag{19}$$

where $\mathbf{K}$ is a spatial padding operator, $\widetilde{\mathbf{P}}$ is the extended propagation operator, and $\mathbf{s}$ is the source. Next, I write the linearized modeling operator as:

$$\widetilde{\mathbf{L}}(\widetilde{\mathbf{m}}) = \mathbf{K}^*\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{D}\widetilde{\mathbf{U}}_\mathbf{s}(\widetilde{\mathbf{m}}), \tag{20}$$

where $\mathbf{D}$ is a second time derivative, and $\widetilde{\mathbf{U}}_\mathbf{s}$ is the extended scattering operator that convolve the input with a source wavefield to calculate a scattered wavefield. The source wavefield that is used in the scattering operator is calculated as:

$$\mathbf{u}_\mathbf{s}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{K}\mathbf{s}. \tag{21}$$

I can write the adjoint modeling operator as follows:

$$\mathbf{L}^*(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{U}}^*_\mathbf{s}(\widetilde{\mathbf{m}})\mathbf{D}^*\widetilde{\mathbf{P}}^*(\widetilde{\mathbf{m}})\mathbf{K}. \tag{22}$$

By examining the previous equations, we can see that the only expensive step is the propagation step. Therefore, I will approximate the propagation by a cheaper alternative, such as:

$$\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{D}\widetilde{\mathbf{S}}([\mathbf{I}-\mathbf{C}]\widetilde{\mathbf{m}})\widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}}), \tag{23}$$

where $\widetilde{\mathbf{S}}$ is a scattering operator. Finally, we use this approximate propagation operator in equations (19) to (22). Another possiblity is using a multi-scattering (recursive) propagator that can take more orders of scatterings to better approximate the extended propagation operator. Notice that the adjoint operator using the approximate propagation will not give an exact gradient of the objective function, it can be very close, depending on how well we approximated the propagation operator.

CONCLUSIONS

Scale separation of the model can add complexity to the theory and implementation of TFWI. I rederive TFWI with one model parameter in an abstract way that is applicable to any form of wave-equation. This new derivation reduces the complexity of TFWI and any ambiguity related to how different model components interact and update in the inversion. I also provide an alternative approximation to TFWI by re-ordering the gradient calculation and the Taylor's expansion of the propagation operator. The alternative approximation is easier to implement and can potentially provide more accurate results.

REFERENCES

Almomin, A. and B. Biondi, 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations: SEP-Report, **149**, 51–58.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEP-Report, **147**, 1–12.