Linear wavefield optimization using a modified source

Tariq Alkhalifah[1]

[1]Physical Sciences and Engineering
King Abdullah University of Science and Technology
Mail box # 1280
Thuwal 23955-6900
Saudi Arabia

Abstract. Recorded seismic data are sensitive to the Earth's elastic properties, and thus, they carry information of such properties in their waveforms. The sensitivity of such waveforms to the properties is nonlinear causing all kinds of difficulties to the inversion of such properties. Inverting directly for the components forming the wave equation, which includes the wave equation operator (or its perturbation), and the wavefield, as independent parameters enhances the convexity of the inverse problem. The optimization in this case is provided by an objective function that maximizes the data fitting and the wave equation fidelity, simultaneously. To enhance the practicality and efficiency of the optimization, I recast the velocity perturbations as secondary sources in a modified source function, and invert for the wavefield and the modified source function, as independent parameters. The optimization in this case corresponds to a linear problem. The inverted functions can be used directly to extract the velocity perturbation. Unlike gradient methods, this optimization problem is free of the Born approximation limitations in the update, including single scattering and cross talk that may arise for example in the case of multi sources. These specific features are shown for a simple synthetic example, as well as the Marmousi model.

AMS subject classifications: 52B10, 65D18, 68U05, 68U07

Key words: waveforms, inversion, acoustic wavefields, optimization.

1 Introduction

Recording waves that may originate from active, or natural sources, including ambient noise is now prevalent in many applications ranging from medical imaging, reverse engineering, non-destructive testing, and, of course, delineating the Earth physical properties. The resulting signals carry information of the object they originated from and the medium they travelled through. The state of these waves as a function of space and time are referred to as wavefields. These functions depend on the source of the wavefield energy and the medium they reside in (Aki and Richards, 1980). A special kind of

wavefield is the Green's function (Green, 1828), which represents the wavefield response to a specific point source in time and space (or just space in most practical applications, considering our band limited signals). So, wavefields tend to be a superposition or summation of these Green's functions weighted by the actual sources of energy in the wavefield, as well as the sources of scattering (secondary sources) (Hughes, 1995; Innanen, 2006; Wu and Aki, 1988). In real life, wavefields are only known at the sensor (recording device) locations. In our computing devices, we solve for these wavefields using the appropriate wave equation (considering the physical nature of the medium), for a given source of energy (location and structure) and given medium properties. If within the simulation process of waves, the source or the medium properties are not representative of the true source or medium properties under investigation, the wavefield would usually be wrong and its values at the simulated sensors would differ from those measured in the real experiment. In classic waveform inversion, we use such differences, measured in many ways, to update the source information and the medium properties or at least one of them (Tarantola, 1987). An integral part of this process is the accuracy of the wavefield, which connects these unknowns to the measurements, and often satisfies a particular wave equation, or specifically its partial differential equation (PDE) form in time or frequency. For the specific problem of waves propagating within a medium, having the accurate wave equation for a specific medium, implies having the accurate form, the medium information, the wavefield and source function. The classic inversion method suffers from the sinusoidal nature of waves, and thus, faces issues related to cycle skipping and the highly nonlinear relation between the medium properties and the wave behavior. Improvements in the performance of waveform inversion is crucial to many applications as the cost of the process is high (Bunks et al., 1995; Pratt et al., 1996; Virieux and Operto, 2009; Wu and Toksoz, 1987).

An approach to reduce the nonlinearity of waveform inversion is provided loosening the constraint on the wave equation and allowing the wavefield to fit the data regardless of the velocity model (Abubakar et al., 2009; Kleinman and den Berg, 1992; van Leeuwen and Herrmann, 2013; Xiong, 1989). As a result, the optimization problem includes, at least, two terms, or two objectives: reducing the modelled wavefield misfit to the data and increasing its compliance to the wave equation. Using such an optimization, van den Berg and Kleinman (1997) and Abubakar et al. (2009) invert for the medium perturbations and the source contrasts. On the other hand, van Leeuwen and Herrmann (2013) elected to invert for the wavefield and the medium perturbations. The philosophy behind both approaches is supported by the inversion iterative nature. Since the initial velocity model is assumed wrong (it provides wavefields that do not fit the data), then why do we need to constrain the wavefield to the wave equation. The wave equation is as good as its operator, which is driven primarily by the model. However, in both implementations, updating the velocity model is an integral part of the iterative process.

Since the scattering series, and specifically, the Lippmann-Schwinger equation (Wu and Zheng, 2014) suggests that the wavefield can be constructed from the background model and scattering (secondary or contrast sources), we can formulate an optimization

for the wavefield and the secondary sources, and initially bypass inverting for the source of nonlinearity given by the medium perturbations. The inversion for the medium perturbations can happen in a follow up step. Thus, in this case, the wave equation operator remains stationary, corresponding to the background model allowing for faster wavefield solutions. Though this approach seems exposed to the weaknesses of the scattering series (the convergence issue (Innanen, 2006; Jakobsen and Wu, 2016)), the loose implementation of the Lippmann-Schwinger equation (used as a penalty) allows us to avoid such limitations.

In this paper, I outline an algorithm that utilizes the wave equation's linear relation to the wavefield and the force function. Finding the wavefield that fits the data and satisfies an accurate wave equation is the objective of waveform inversion. However, we often do not have the accurate wave equation operator controlled by the model information. If we absorb the model perturbations into the source function, we can formulate a linear optimization problem to invert for the wavefield and a modified source. This optimization has valuable features in efficiency, as well as accuracy. We will analyse such features on a simple model, as well as the Marmousi model. Full inversion implementations and more complicated models are topics of a follow up paper.

2 Theory

The suggested inversion allows us to start with an initial wave equation corresponding to an initial knowledge of the medium, an initial knowledge of the state of the wavefield, and an initial knowledge of the source. These could be guesses, possibly good ones, and possibly based on using other approaches to obtain the initial values. We then update all three components of the wave equation or two of them or any of them, in any order or together. To establish the relation between perturbations in the model and the resulting changes in the wavefield or the data, we utilize perturbation theory admitting the infamous Born scattering series (Born, 1926). For brevity, I use wavefields, and all relative variables, represented in the frequency domain.

2.1 The Lippmann Schwinger equation

The wave equation in the frequency domain can be discretized to form the following linear equation:

$$\mathbf{L}\mathbf{u} = \mathbf{f}, \tag{2.1}$$

where $\mathbf{L}$ is the impedance kernel (or matrix), $\mathbf{u}$ is a function (vector) holding discrete values of the complex-valued wavefield over the range of the model space for a single frequency, and $\mathbf{f}$ is the source vector described also within that space, also possibly complex valued. For point grid sources in the domain of interest represented by the identity matrix, $\mathbf{I}$, the Green's function satisfies a similar equation:

$$\mathbf{L}\mathbf{G} = \mathbf{I}, \tag{2.2}$$

where $\mathbf{G}$ is the Green's matrix with columns made up of the wavefield response to a particular grid point source. Thus, each column of $\mathbf{G}$ spans the model space. As a result, the wavefield can also be evaluated using a simplified version of Green's theorem (ignoring boundary conditions) as follows:

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{f}. = \mathbf{G}\mathbf{f}. \tag{2.3}$$

Considering the impedance matrix for a simpler (or known) model, $\mathbf{L_0}$, the new Green's function satisfies the following equation:

$$\mathbf{L_0}\mathbf{G_0} = \mathbf{I}. \tag{2.4}$$

Assuming the difference $\mathbf{V} = \mathbf{L} - \mathbf{L_0}$ (medium perturbations) causes a Green's function (wavefield) perturbation, $\mathbf{G} - \mathbf{G_0}$, then the total Green's function satisfies the following Lippmann Schwinger equation:

$$\mathbf{G} = \mathbf{G_0} + \mathbf{G_0}\mathbf{V}\mathbf{G}. \tag{2.5}$$

We can then solve for the full Green's function as follows:

$$\mathbf{G} = (\mathbf{I} - \mathbf{G_0}\mathbf{V})^{-1}\mathbf{G_0}, \tag{2.6}$$

which is hard to implement numerically as it includes an inverse of a large matrix, which tends to be unstable when $\mathbf{V}$ is large (the matrix less diagonal dominant). The Born series is extracted by expanding equation 2.6 using the Neumann series (Born-Neumann expansion). The expanded series is not guaranteed to converge, especially if $\mathbf{V}$ is large (i.e. the determinant of $\mathbf{V}\mathbf{G_0}$ is bigger than 1).

If we consider $\mathbf{V}$ to be small, we can replace $\mathbf{G}$ with $\mathbf{G_0}$ in the right hand side of equation 2.5 to obtain:

$$\mathbf{G_{approx}} = \mathbf{G_0} + \mathbf{G_0}\mathbf{V}\mathbf{G_0}, \tag{2.7}$$

which is the Born approximation, and it is the essence of the gradient based update for FWI, as we will see next.

2.2 The Inversion

In classic implementations of waveform inversion, we seek the velocity model information (or what is missing from it, perturbations) from the difference between the modelled and measured data. To do so we formulate an optimization problem that utilizes the wave equation, or any form of it, to obtain data (our wavefields at the measuring points) that are similar to the measured ones. Thus, an optimization problem in seeking the true perturbation $\hat{\mathbf{V}}$ can have the following form:

$$\hat{\mathbf{V}} = min_{\mathbf{V}} J(\mathbf{V}) = min_{\mathbf{V}} \frac{1}{2} |\mathbf{d} - C\mathbf{G}(\mathbf{V})\mathbf{f}|_2^2, \tag{2.8}$$

such that $(\mathbf{L_0}+\mathbf{V})\mathbf{G}=\mathbf{I}$. Here, C projects the wavefield to the receiver locations in which the measured data, $\mathbf{d}$, reside. The operator $\mathbf{L_0}$ corresponds to the wave equation for the initial (background) velocity, and $\mathbf{V}$ is assumed zero at the beginning of the inversion process. In many applications, including classic FWI, $\mathbf{V}$ is assumed to be block diagonal with nonzero elements spanning only the model space. The gradient, computed using the adjoint state method, is, thus, given by (Pratt, 1999)

$$\nabla_{\mathbf{V}} J = \mathbf{u}^*(\mathbf{L_0}+\mathbf{V})^{-1}C^T\Delta\mathbf{d}, \tag{2.9}$$

where for conventional waveform inversion, $\nabla_{\mathbf{V}} J$ is also given by elements that span the model space. Here, $\Delta\mathbf{d}=\mathbf{d}-C\mathbf{G}(\mathbf{V})\mathbf{f}$, and the symbol $*$ stands for the complex conjugation. The gradient here constitutes the adjoint of the Born approximation equation 2.7. In this case, the perturbation can be updated using

$$\mathbf{V}_{\text{new}} = \mathbf{V} - A_{\mathbf{V}}\nabla_{\mathbf{V}} J, \tag{2.10}$$

where $A_{\mathbf{V}}$ can be the Hessian or any approximation of it, or as simple as a predetermined step length.

We can also establish a gradient for $\mathbf{f}$:

$$\nabla_{\mathbf{f}} J = (\mathbf{L_0}+\mathbf{V})^{-1}C^T\Delta\mathbf{d}, \tag{2.11}$$

which is the time reversal of the residual data (Artman et al., 2010). Its update, thus, is given by $\mathbf{f}_{\text{new}} = \mathbf{f} - A_{\mathbf{f}}\nabla_{\mathbf{f}} J$, where $A_{\mathbf{f}}$ is the again a form of the Hessian. Thus, we can iteratively update $\mathbf{V}$ and $\mathbf{f}$ using their gradients. The Green's function here is computed using the wave equation, and it is dependent on the velocity. This dependency is the reason for the nonlinearity of the objective function as the Green's function is nonlinearly dependent on perturbations in velocity, as evidence by the Born series.

If we treat the Green's function as an independent variable of the perturbation by loosening the constraint on the wave equation (Kleinman and den Berg, 1992; Xiong, 1989), we can formulate the following two-term optimization problem:

$$J(\mathbf{V},\mathbf{G},\mathbf{f}) = \frac{1}{2}|C\mathbf{G}\mathbf{f}-\mathbf{d}|_2 + \frac{1}{2}\epsilon|(\mathbf{L_0}+\mathbf{V})\mathbf{G}-\mathbf{I}|_2, \tag{2.12}$$

where $\mathbf{G}$ is the full Green's function, and $\mathbf{I}$, as described earlier, is the identity matrix. In this case, the gradient with respect to perturbations in the wave equation operator is given by

$$\nabla_{\mathbf{V}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \epsilon\mathbf{G}^*\Delta\mathbf{I}, \tag{2.13}$$

where, $\Delta\mathbf{I}=(\mathbf{L_0}+\mathbf{V})\mathbf{G}-\mathbf{I}$. However, unlike wavefields, which extend the model space, Green's functions can be as large as squared the model space. Meanwhile, the gradient with respect to the Green's function as an independent function is given by:

$$\nabla_{\mathbf{G}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \epsilon(\mathbf{L_0}+\mathbf{V})^*\Delta\mathbf{I} + \mathbf{f}^*C^T\Delta\mathbf{d}, \tag{2.14}$$

and the gradient with respect to the source is given by

$$\nabla_{\mathbf{f}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \mathbf{G}^* C^T \Delta \mathbf{d}, \tag{2.15}$$

which again constitutes a time reversal with the full Green's function. Here, I try to constrain three parameters with two terms in the objective function, which can be a source of nonuniqness in the inversion and additional constraints are required.

2.3 Inversion with the Lippmann Schwinger equation

If we replace the wave equation with the Lippmann Schwinger equation, the conventional optimization problem can be formulated as follows:

$$\mathbf{V} = min_{\mathbf{V}} |\mathbf{d} - C\mathbf{G}\mathbf{f}|_2^2, \tag{2.16}$$

such that $\mathbf{G}$ satisfies equation 2.5, in which C again projects the wavefield to the receiver locations. This optimization problem is equivalent to FWI. If we replace $\mathbf{G}$ with $\mathbf{G_{approx}}$, we obtain the linearized inversion (Tarantola, 1987), which can be used to obtain optimized Gauss Newton updates for a full waveform inversion (Wu and Alkhalifah, 2015).

Moving back to the extended form (equation 2.12), the optimization in this case has the following two-term structure

$$J(\mathbf{V},\mathbf{G},\mathbf{f}) = \frac{1}{2}|C\mathbf{G}\mathbf{f} - \mathbf{d}|_2 + \frac{1}{2}\epsilon|\mathbf{G} - \mathbf{G_0} - \mathbf{G_0}\mathbf{V}\mathbf{G}|_2, \tag{2.17}$$

where again $\mathbf{G}$ is the full Green's function satisfying $\mathbf{L}\mathbf{G} = \mathbf{I}$, and $\mathbf{G_0}$ is the background Green's function satisfying $\mathbf{L_0}\mathbf{G_0} = \mathbf{I}$. Thus, the gradient with respect to the perturbation $\mathbf{V}$:

$$\nabla_{\mathbf{V}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \epsilon \mathbf{G}^* \mathbf{G}_0^* \Delta \mathbf{G}, \tag{2.18}$$

which somewhat resembles the classic FWI gradient, but for the residual, which here is the Green's function, $\Delta\mathbf{G} = \mathbf{G} - \mathbf{G_0} - \mathbf{G_0}\mathbf{V}\mathbf{G}$. The gradient with respect to the Green's function is given by:

$$\nabla_{\mathbf{G}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \epsilon(\mathbf{I} - \mathbf{V}^*\mathbf{G}_0^*)\Delta\mathbf{G} + \mathbf{f}^* C^T \Delta\mathbf{d}, \tag{2.19}$$

and the gradient with respect to the source is given by

$$\nabla_{\mathbf{f}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \mathbf{G}^* C^T \Delta\mathbf{d}, \tag{2.20}$$

which again constitutes a time reversal with the full Green's function. Since $\mathbf{f}$ is a function that controls the data linearly, there is a potential tradeoff between it and the Green's function. Thus, if $\mathbf{f}$ is unknown, we will need to constrain it.

As mentioned earlier, Green's functions can have dimensions squared the model space to cover sources at every model point. Inverting for it, however, can be cumbersome though maybe useful as mentioned above to allow for an expanded inversion for

the operator perturbation. However, for practical implementations, I substitute $\mathbf{u} = \mathbf{G}\mathbf{f}$ into equation 2.17 after multiplying the second term by $\mathbf{f}$, to obtain:

$$J(\mathbf{V},\mathbf{u},\mathbf{f}) = \frac{1}{2}|C\mathbf{u}-\mathbf{d}|_2 + \epsilon\frac{1}{2}|\mathbf{u}-\mathbf{G_0}\mathbf{f}-\mathbf{G_0}\mathbf{V}\mathbf{u}|_2. \tag{2.21}$$

This form provides an opportunity to reduce the complexity of the problem as we will see next.

3 The reduction of the problem

I will outline two forms of this proposed reduction, which are based on the following substitution:

$$\mathbf{f_e} = \mathbf{f} - \mathbf{V}\mathbf{u}. \tag{3.1}$$

The idea is to postpone the inversion for $\mathbf{V}$, which is the source of nonlinearity ($\mathbf{V}\mathbf{u}$), to a separate stage. This approach attempts to extract the wavefield from the background medium using the scattering series in an optimization formulation.

3.1 Focusing on the wavefield

Substituting equation 3.1 into equation 2.12 yields the reduced linear optimization given by

$$J(\mathbf{u},\mathbf{f_e}) = \frac{1}{2}|C\mathbf{u}-\mathbf{d}|_2 + \frac{1}{2}\epsilon|\mathbf{u}-\mathbf{G_0}\mathbf{f_e}|_2. \tag{3.2}$$

We now have two unknowns, the wavefield and a modified source function supposedly holding information corresponding to the actual source and secondary sources (perturbations). If we invert for one of these two potential unknowns, this optimization problem is convex, and can be used to solve for $\mathbf{u}$ using the following linear equations:

$$\begin{pmatrix} C \\ \epsilon\mathbf{L_0} \end{pmatrix}\mathbf{u} = \begin{pmatrix} \mathbf{d} \\ \epsilon\mathbf{f_e} \end{pmatrix}. \tag{3.3}$$

with

$$\mathbf{f_e} = \mathbf{f_e} + \mathbf{\Delta f_e}, \tag{3.4}$$

and $\mathbf{\Delta f_e} = \epsilon(\mathbf{L_0}\mathbf{u} - \mathbf{f_e})$.

The least square form of equation 3.3 is given by

$$\left(C^TC + \epsilon^2\mathbf{L_0}^*\mathbf{L_0}\right)\mathbf{u} = C^T\mathbf{d} + \epsilon^2\mathbf{L_0}^*\mathbf{f_e}, \tag{3.5}$$

with

$$\mathbf{f_e} = \mathbf{L_0}\mathbf{u}. \tag{3.6}$$

Equations 3.5 and 3.6 are solved sequentially in an alternate fashion.

3.2 An alternative formulation

Utilizing the Lippmann-Schwinger formulation 2.5 in both terms of the extended optimization, I obtain

$$J(\mathbf{V},\mathbf{u},\mathbf{f}) = \frac{1}{2}|\mathbf{d} - C\mathbf{G_0 f} - C\mathbf{G_0 V u}|_2 + \frac{1}{2}\epsilon|\mathbf{u} - \mathbf{G_0 f} - \mathbf{G_0 V u}|_2. \tag{3.7}$$

Considering that the known (initial) velocity model is used to develop the initial wavefield (stripped of the source function) $\mathbf{G_0}$, then the difference between the true wavefield, part of which is represented in the measured data, and this wavefield is given by the perturbation $\mathbf{V}$. In classic full waveform inversion, the difference in the wavefields at the recording stations is used to update the background model using the Born approximation $\mathbf{V_0}$.

The gradient for this optimization, using the adjoint state method, is given by:

$$\begin{aligned} \nabla_{\mathbf{V}} J &= \mathbf{u}\mathbf{G}_0^* C^T \mathbf{\Delta d} + \mathbf{ffluG}_0^* \mathbf{\Delta u} \\ \nabla_{\mathbf{u}} J &= \epsilon \mathbf{\Delta u} - \mathbf{V}^* \mathbf{u}^* \left(\epsilon \mathbf{\Delta u} + C^T \mathbf{\Delta d} \right) \\ \nabla_{\mathbf{f}} J &= -\mathbf{G}_0^* \left(\epsilon \mathbf{\Delta u} + C^T \mathbf{\Delta d} \right), \end{aligned} \tag{3.8}$$

where $\mathbf{\Delta d} = \mathbf{d} - C\mathbf{G_0 f} - C\mathbf{G_0 V u}$ and $\mathbf{\Delta u} = \mathbf{u} - \mathbf{G_0 f} - \mathbf{G_0 V u}$. Again this optimization offers a large degree of freedom and a high level of non uniqueness in which additional constraints and regularizations are needed.

Using the reduction process based on including of the model perturbation in a new force function 3.1, we obtain the following objective function:

$$J(\mathbf{u},\mathbf{f_e}) = \frac{1}{2}|\mathbf{d} - C\mathbf{G_0 f_e}|_2 + \frac{1}{2}\epsilon|\mathbf{u} - \mathbf{G_0 f_e}|_2. \tag{3.9}$$

In this case, the solution satisfies the following linear form:

$$\begin{pmatrix} \epsilon \mathbf{G_0} \\ C\mathbf{G_0} \end{pmatrix} \mathbf{f_e} = \begin{pmatrix} \epsilon \mathbf{u} \\ \mathbf{d} \end{pmatrix}, \tag{3.10}$$

where

$$\Delta \mathbf{u} = \mathbf{u} - \mathbf{G_0 f_e}, \tag{3.11}$$

and $\mathbf{u} = \mathbf{u} + \Delta\mathbf{u}$, or in another form $\mathbf{u} = \mathbf{G_0 f_e}$. We continue iterating between equations 3.10 and 3.11 until, for example, $|\Delta\mathbf{u}|_2$ is small. This is similar to the set 3.5 and 3.6, with replacing the roles of wavefield and force function.

We, then, can use $\mathbf{f_e}$ and $\mathbf{u}$, with a known $\mathbf{f}$ to invert for $\mathbf{V}$.

3.3 Inverting for V

Now the estimation of medium perturbations, $\mathbf{V}$, can be applied in a separate step. If we know the force function, $\mathbf{f}$, or at least the space component of it the problem is generally trivial. However, if this function is unknown, we might need to inject constraints, like promoting sparsity in $\mathbf{f}$, to the problem. We can formulate an optimization to find the minimum of:

$$J(\mathbf{V},\mathbf{f}) = |\mathbf{f_e} - \mathbf{f} + \mathbf{V}\mathbf{u}|_2^2. \tag{3.12}$$

Such an optimization can be performed efficiently since no modelling is involved and these functions often span the model space. The gradients are given by:

$$\nabla_{\mathbf{V}} J(\mathbf{V},\mathbf{f}) = \mathbf{u}^* \Delta\mathbf{f}, \tag{3.13}$$

and

$$\nabla_{\mathbf{f}} J(\mathbf{V},\mathbf{G},\mathbf{f}) = \Delta\mathbf{f}, \tag{3.14}$$

where $\Delta\mathbf{f} = \mathbf{f_e} - \mathbf{f} + \mathbf{V}\mathbf{u}$. Using the gradients, we can update one of these parameters or both as needed. In solving for both $\mathbf{V}$ and $\mathbf{f}$, we are trying to identify in $\mathbf{f_e}$ the actual sources, $\mathbf{f}$, and the secondary ones, $\mathbf{V}\mathbf{u}$, identified in Kleinman and den Berg (1992) as contrast sources. However, our identification is given directly by the perturbation $\mathbf{V}$. In using the gradient method to obtain $\mathbf{V}$, we are vulnerable to crosstalk when $\mathbf{f}$ contains more than one event.

However, if the true source function, $\mathbf{f}$, is known, or at least its space component is known, we can solve for $\mathbf{V}$ directly using equation 3.1. It is given by a direct division:

$$\mathbf{V} = \frac{\mathbf{f} - \mathbf{f_e}}{\mathbf{u}} \approx \frac{\mathbf{u}^*(\mathbf{f} - \mathbf{f_e})}{\mathbf{u}^*\mathbf{u} + \alpha}, \tag{3.15}$$

where α is a small positive number to guarantee that the denominator is bigger than zero. This division, or deconvolution, slightly mitigates crosstalk. It is, however, dependent on a wavefield in the denominator. Since we are pursuing in the division a velocity or medium update, we should utilize smoothing operators. Such operators can regularize the division to provide us with a smooth version courtesy of shaping operators (Fomel, 2007). I utilize such smooth divisions in the examples below.

4 A potential algorithm

The derived formulas offer the opportunity to suggest a number of implementation strategies. The key feature here is that the estimation of the medium perturbations can be done independently using a separate step. The inversion of the wavefield and the modified source, like any inversion, will depend on how accurate the background model as such an inversion depends on the size of the scatterer. The implementation based on the extended objective function 3.2 in which the wavefield and the modified source are inverted

allows for additional degrees of freedom to help us converge in spite of the potentially inaccurate background (initial) model. However, an opportunity to update the background model is helpful, granted it does not add much to the cost of the inversion. So the right balance between robustness and efficiency is needed.

A frequency domain implementation is favourable here considering that the wavefield and the source function have reduced dimensions in such a domain. In this case, each frequency will require an inversion of the matrix involved in equations 3.3 or 3.10. We can utilize this fact by updating the velocity model per frequency as we move from low to high frequency. The only parameter we might want to change in the left hand side (The matrix) is ϵ. This parameter, as van Leeuwen and Herrmann (2013) suggested, does not have to change frequently. So it can also be updated per frequency.

This implementation that allows for a single LU decomposition per frequency is synonymous to the cost of solving the wave equation in the frequency domain, and thus, offers what might be the most efficient rendition of an inversion. Since the extended equation forces the data fitting in the initial iterations, it is also somewhat immune to cycle skipping. However, like any inversion, there is no guarantee of convergence.

I share an Algorithm of a potential inversion implementation using the least squares form of equations 3.10 and 3.11, in which we place the modified source as the central parameter. This algorithm explicitly clarifies the role of the background Green's function more vividly. Note that we can pull $\mathbf{G_0^*}$ from the least square version of equation 3.10, and thus, the modified force function satisfies:

$$\left(\epsilon^2\mathbf{G_0}+C^TC\mathbf{G_0}\right)\mathbf{f_e}=\left(\epsilon^2\mathbf{u}+C^T\mathbf{d}\right), \tag{4.1}$$

which suggests that the inverted $\mathbf{f_e}$ attempts in average, to fit the current wavefield and data (weighted by ϵ). For large ϵ, $\mathbf{f_e}$ is controlled mainly by the background wavefield, and thus, it is closer to $\mathbf{f}$ and the problem is closer to conventional FWI, with the danger of cycle skipping if the initial velocity is bad. For $\epsilon=0$, $\mathbf{f_e}$ is the time reversal source image. For an intermediate ϵ it contains both components (background wavefield and time reversal) weighted by ϵ. In the algorithm, we assume $\mathbf{f}$ is known, which is often the case with active experiments.

5 Examples

At this stage, let us analyse some of the features of the reduced formulation using a simple two-layer model. The model shown in Figure 1a has a velocity of 2 km/s in the first layer and 3.5 km/s in the second layer. Using a Helmholtz solver for a frequency of 5 Hz and a source function given in Figure 1b, we solve for the wavefield with the real part shown in Figure 2a. Considering a background (initial) homogeneous model with velocity equal to 2.5 km/s (a value between the first and second layers velocity), we obtain a wavefield with the real part shown in Figure 2b for the same force function. Thus, the background wave equation operator provides, as expected, circular wavefronts. Thus,

Algorithm 1 Modified source inversion algorithm

Input: Observed seismic data $\mathbf{d}$; initial velocity providing $\mathbf{L_0}$; Number of iterations *niter*; Source function $\mathbf{f}$; all associated with the ith frequency; The weight ϵ;
Output: The model perturbation $\mathbf{V}$.

Initialize: $\mathbf{u} \leftarrow \mathbf{G_0 f}$;
for $i = 1 \ldots N$ **do**
  $\mathbf{G_0} \leftarrow$ Solve $(\mathbf{L_0} + \mathbf{V})\mathbf{G_0} = \mathbf{I}$;
  for $k = 1 \ldots niter$ **do**
    $\mathbf{f_e} \leftarrow$ Solve $\left(\epsilon^2 \mathbf{G_0^* G_0} + \mathbf{G_0^*} C^T C \mathbf{G_0}\right)\mathbf{f_e} = \mathbf{G_0^*}\left(\epsilon^2 \mathbf{u} + C^T \mathbf{d}\right)$ using for example conjugate gradient methods;
    $\mathbf{u} \leftarrow \mathbf{G_0 f_e}$;
  end for
  $\mathbf{V} \leftarrow \frac{\mathbf{f} - \mathbf{f_e}}{\mathbf{u}}$;
  Increase ϵ if needed;
end for

we can appreciate how the layering effected the wavefield, especially in the region above the layer (acting as a secondary source). If we move the complexity of the model (the pertubation) to the source function, we can use the background wave equation operator to obtain the same complex wavefield. Using the exact wavefield shown in Figure 2a., we compute $\mathbf{f_e}$ from equation 3.1, which is shown in Figure 3a. Applying the Helmholtz solver with the constant background velocity for the force function f_e, we obtain the wavefield shown in Figure 3b. In fact, for the homogeneous model, the Green's function can be described analytically. Thus, the complexity of this new wavefield is the result of the modified source function. The resulting wavefield is very close to the true one shown in Figure 2a. The mild differences are attributed to the boundary condition that is different for $\mathbf{L}$ and $\mathbf{L_0}$, and ignored here.

This resulting wavefield along with $\mathbf{f_e}$ can be used to compute the perturbation using a gradient method given by equation 3.13 or a smooth division (Fomel, 2007), of equation 3.15. The smooth division, applied here, includes a 7-point smoother in both space directions. The smoother is needed since we are dealing with a single frequency and a single source in calculating the perturbation. The purpose of this simple experiment is to demonstrate the validity of the formulations. The need for the smoothing is evident in the estimated perturbation, $\mathbf{V}$, in Figure 4a. It is obtained using the wavefield in Figure 3b and $\mathbf{f_e}$ in Figure 3a, as well as the original source, with the gradient method. It clearly contains an imprint of the wavefield. On the other hand, the smooth division with a smoothing of 7 points in both space axes, results in the perturbation in Figure 4b. Obviously, the smooth division provided a better perturbation estimate than the gradient. Even with a 7-point smoother applied to the resulting velocity perturbation, δm obtained from the gradient approach, (Figure 4c) calculated from $\mathbf{V}$ in Figure 4a is not very stable compared to the velocity perturbation using the smooth division (Figure 4d). In both

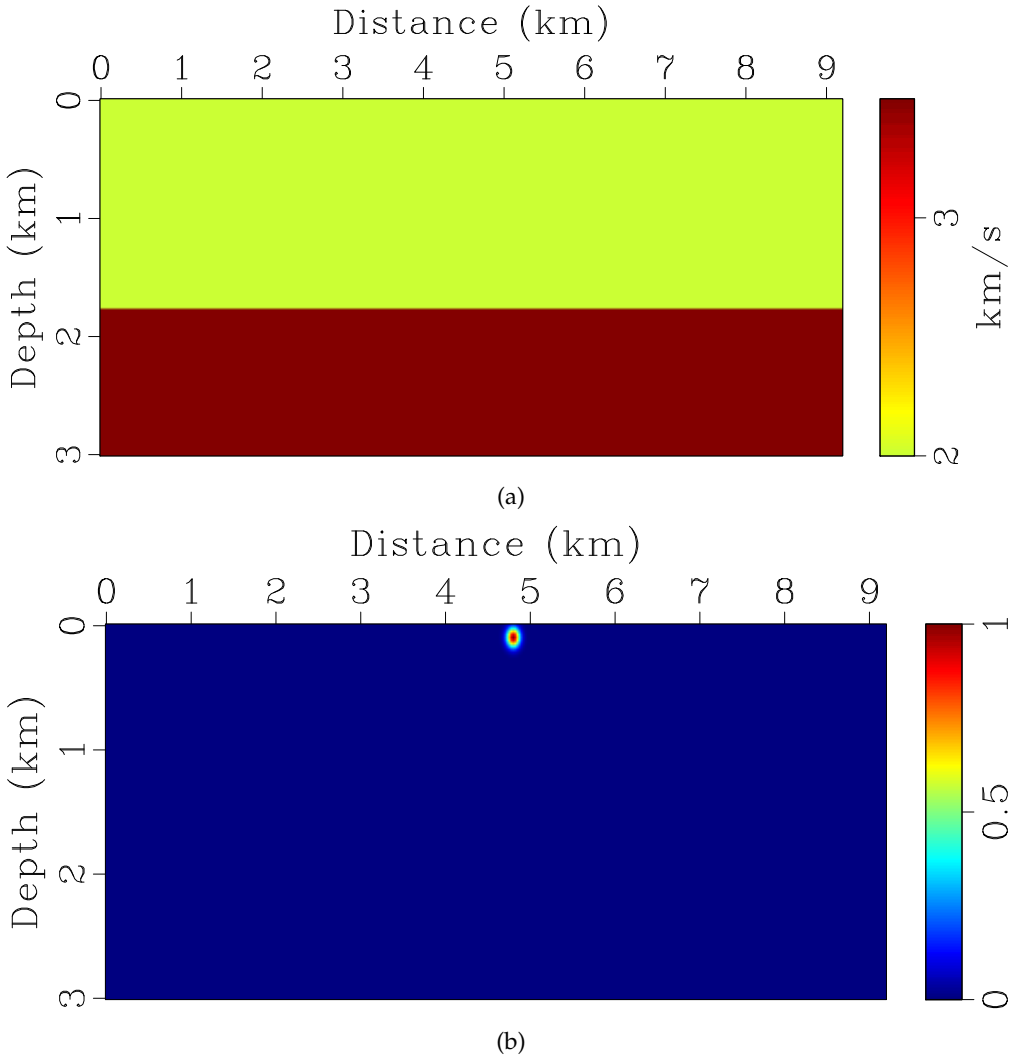


Figure 1: a) A simple two-layer model with the velocity in first layer 2.0 km/s, and the second layer 3.5 km/s. b) The source function used to produce the single frequency wavefield.

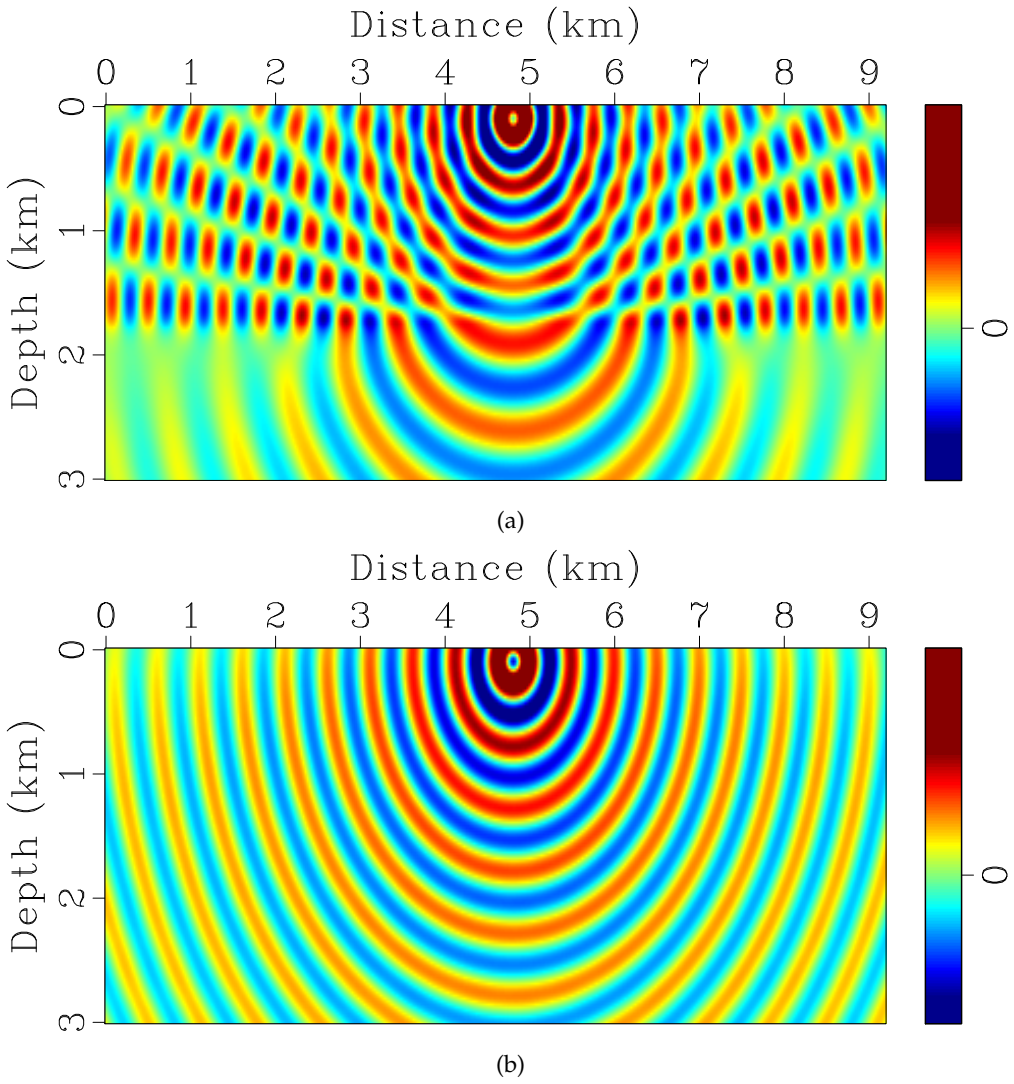


Figure 2: a) The real part of the resulting wavefield for the model in Figure 1a and the corresponding $\mathbf{L}$, as well the source function in Figure 1b. b) The real part of the wavefield for the backgound homogeneous model corresponding to $\mathbf{L_0}$ with a velocity of 2.5 km/s and the same source function.

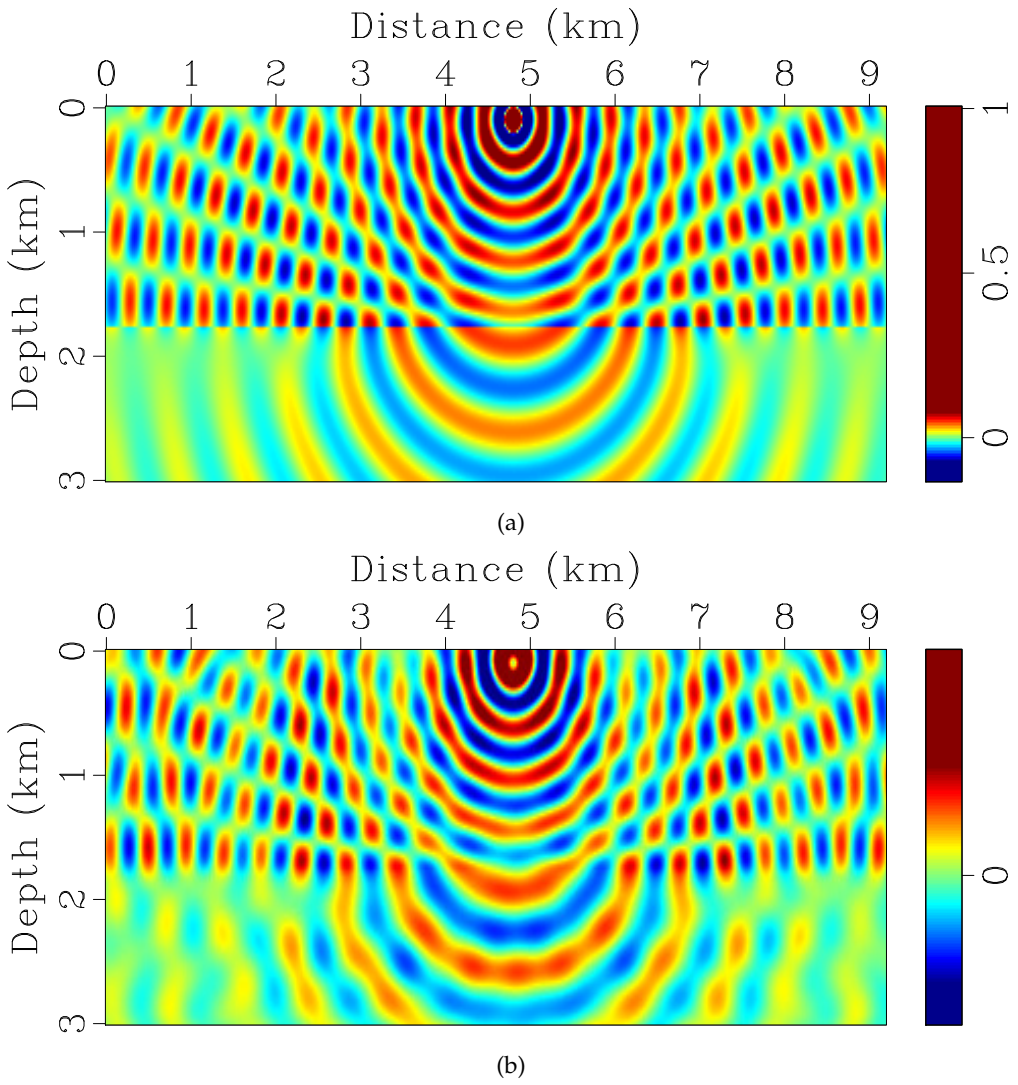


Figure 3: a) The modified source function $\mathbf{f_e}$ obtained using equation 3.1 with the exact wavefield in Figure 2a and the true $\mathbf{V}$ and $\mathbf{f}$. b) The real part of the wavefield solved using $\mathbf{L_0 u} = \mathbf{f_e}$ to compare with Figure 2a.

cases, the required update of -0.5 in the first layer and 1.0 km/s in the second layer is apparent in average in the resulting perturbations. These perturbations are obtained using the exact $\mathbf{f_e}$, but it shows the potential of the approach, and its closed loop nature.

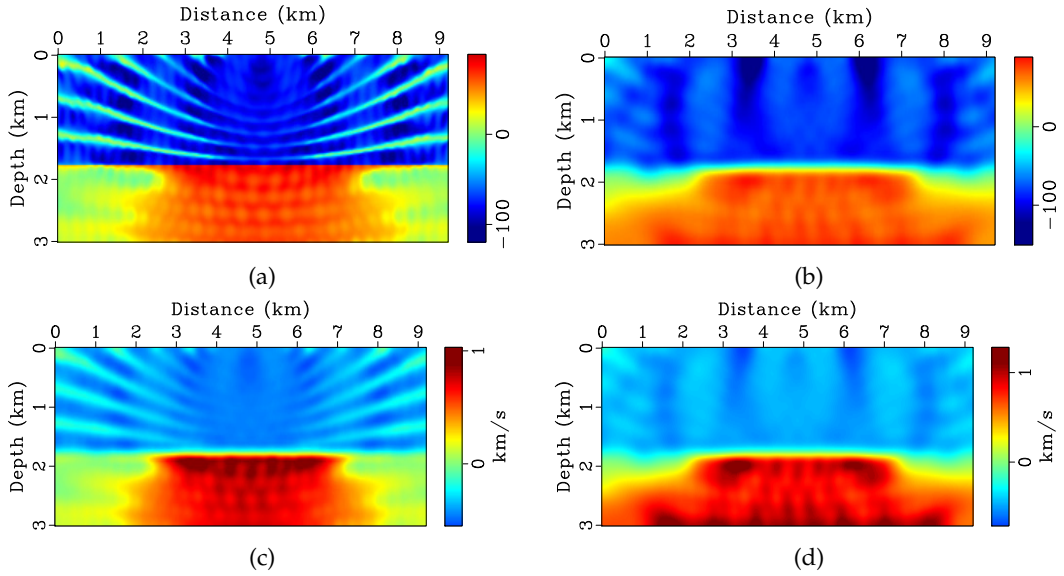


Figure 4: a) The perturbation $\mathbf{V}$ obtained by using the gradient method (using equation 3.13 iteratively), and b) by using a smooth division. c)The corresponding velocity perturbation for the gradient case a) after smoothing it with the same window used in the smooth division. d) The corresponding perturbation in velocity for the smooth division.

One of the features I promote with the direct (smooth) division, to obtain the velocity perturbation, is the mitigation of crosstalk artifacts often associated with simultaneous inversions of multi sources (Krebs et al., 2009). The crosstalks are often associated with gradient-based updates in waveform inversion (correlation replacing deconvolution), as we compare state and adjoint state wavefields assuming the continuity in the wavefield at perturbation points. For multi sources, the comparison includes energy from unrelated events of the state and adjoint state wavefields interacting with each other forming artifacts. Since the inversion, promoted here, evaluates the velocity perturbation using direct inversion, the results are immune from such artifacts. Figure 5b shows the wavefield for the same model in Figure 1a but for a 3-source function, not uniformly spaced (Figure 5a). As expected the wavefield is complicated. We use it again to compute $\mathbf{f_e}$, as shown in Figure 6a. Considering the homogeneous background model we used above, and solving the Helmholtz wave equation in the background model with $\mathbf{f_e}$, we obtain the wavefield in Figure 6b. It is reasonably similar to the wavefield shown in Figure 5b.

As above, we next invert for the velocity perturbation for the three simultaneous sources case. The velocity perturbation obtained using the gradient method and a 7-point smoother is shown in Figure 7a. Though the perturbations are close to what we expect, they include a lot of artifacts. On the other hand, direct division using a smooth

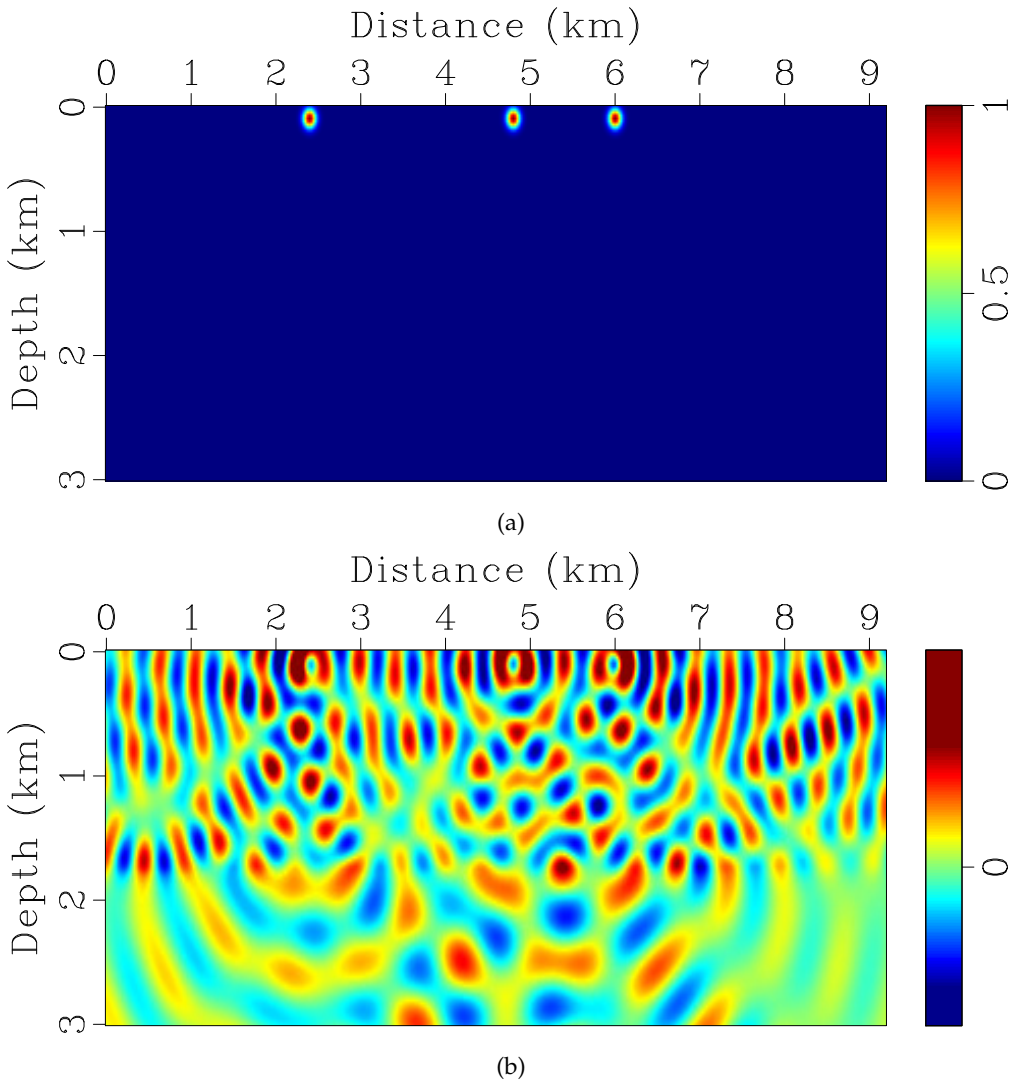


Figure 5: a) The source function with three sources. b) The real part of the resulting wavefield for the model in Figure 1a corresponding $\mathbf{L}$, and the source function in a).

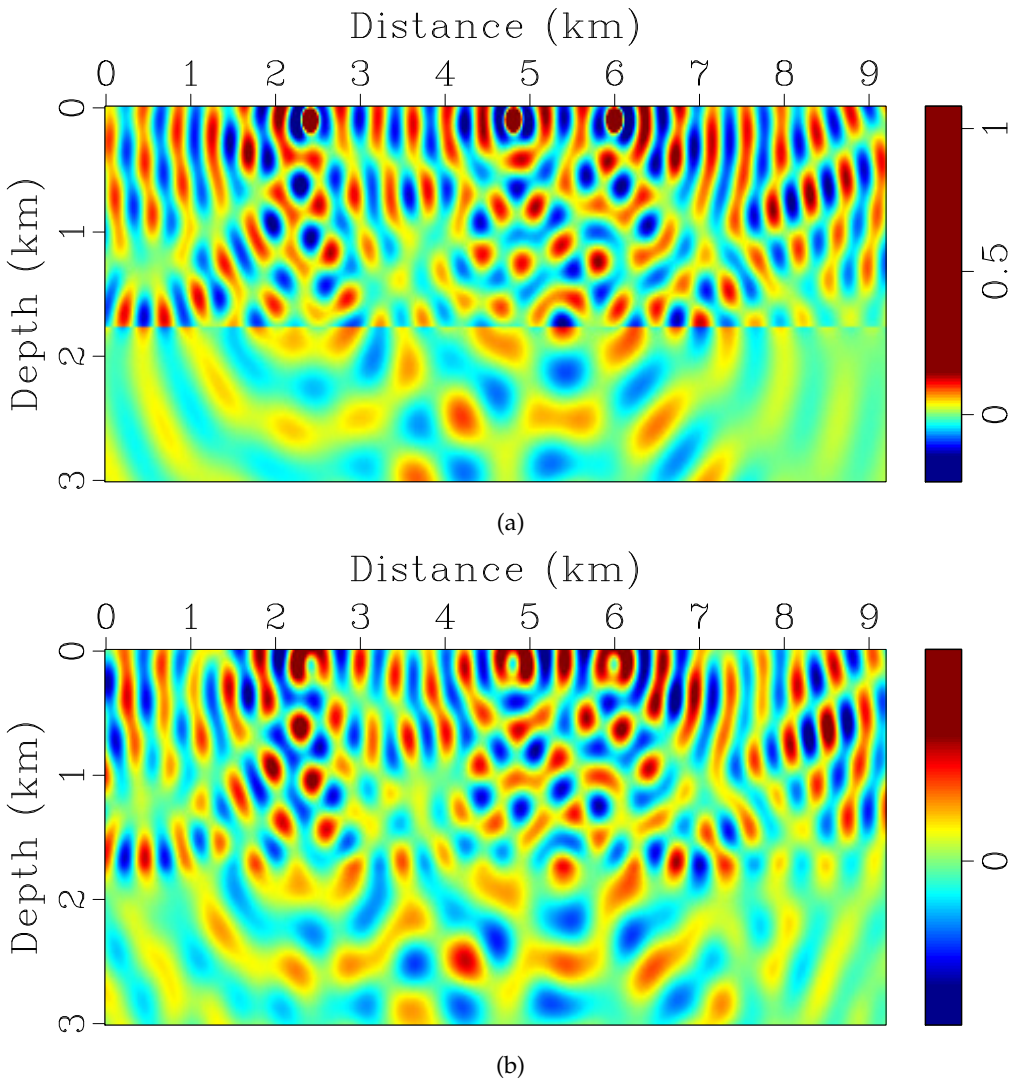


Figure 6: a) The modified source function $\mathbf{f_e}$ obtained using equation 3.1 with the exact wavefield in Figure 5b and the true $\mathbf{V}$ and $\mathbf{f}$. b) The real part of the wavefield solved using $\mathbf{L_0 u} = \mathbf{f_e}$.

division with a 7-point window admits the velocity perturbation shown in Figure 7b). This perturbation is relatively free from artifacts and compared to the single source case (Figure 4d, benefited from more illumination, and thus, the change between layers is more regular laterally.

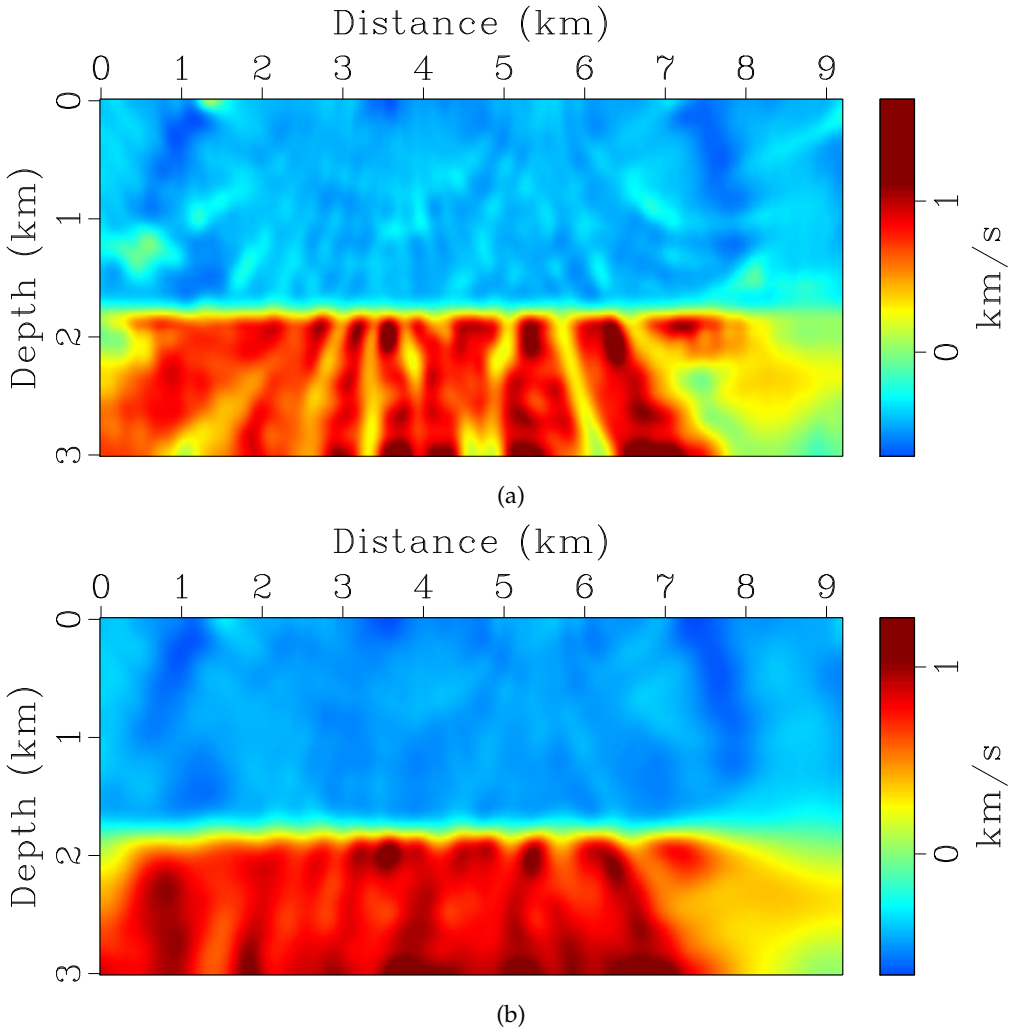


Figure 7: a)The velocity perturbation for the gradient case after smoothing it with the same window used in the smooth division. b) The corresponding perturbation in velocity for the smooth division.

I repeat the above experiment with the Marmousi model shown in Figure 8a. I consider a homogeneous background given by a velocity of 2.5 km/s. Thus, the perturbations we seek are shown in Figure 8b. For a change, I use a higher frequency of 10 Hz for this example. Figure 9a shows the wavefield computed by solving the Helmholtz wave equation for the source function given in Figure 1b. I then compute the modified source

$\mathbf{f_e}$ from the perturbation shown in Figure 8b. Solving the background wave equation for the velocity of 2.5 km/s using the modified source results in the wavefield shown in Figure 9b. It looks similar to the exact wavefield. The difference is caused by not including the boundary condition in the modified source. However, such a difference is small with respect to inversion standards. We use this wavefield and the modified source to compute the velocity perturbation shown in Figure 10a using smooth division with a 7-point smoother. For comparison, the velocity perturbation computed from the wavefield and modified source using a gradient method is shown in Figure 10b. As we saw before, the direct division provides a better inversion as we compare the inverted perturbations with the true ones shown in Figure 8b. The difference is expected to be larger for simultaneous sources. So we repeat the experiment with the simultaneous sources function shown in Figure 5a. The resulting velocity perturbation by division is shown in Figure 11a. Meanwhile, the gradient approach admitted an artifact infested velocity perturbation shown in Figure 11b.

6 Discussions

The main objective of this paper is to introduce the two step inversion in which the model estimation is handled in a separate step. Thus, I focussed on the concept and performing some simple numerical analysis. We will include the process of inversion in a sequel paper.

Of course, the accuracy of the inversion in the first step (for $\mathbf{u}$ and $\mathbf{f_e}$) depends on the background model. Despite that the introduction of the modified source function in equation 3.1 is exact, as expected, finding the wavefield and modified source function that satisfies the background wave equation and the data is slightly more involved when the background model is far from the true model. Thus, in the implementation, I suggest to update the wave equation operator as we scale up the frequency. This proposed methodology tries to balance the need for efficiency with maintaining a semblance of accuracy. However, there are many potential strategies and options to utilize the two-step implementation. The cost of the inversion depends on the implementation strategy. As discussed earlier. we anticipate that one matrix inversion will be needed per frequency. In this case, the cost is similar to the cost of solving the wave equation. The cost of the smooth division to obtain the velocity perturbation is negligible. However, due to the expected variations in illumination depending on the acquisition, a line search will be needed to find the best update along the velocity perturbation vector direction.

Among the most important features of the new formulation is the reduction of crosstalk in the case of simultaneous sources. Since this feature is related to the mitigation of crosstalk, we expect other sources of crosstalk, like multiples, will benefit from this feature. The smoothness used in the division is natural. It will help in the case of simultaneous sources. Such smoothness is often used with FWI gradients to reduce noise. For a fair comparison we used the same smoothness for the division and gradient methods.

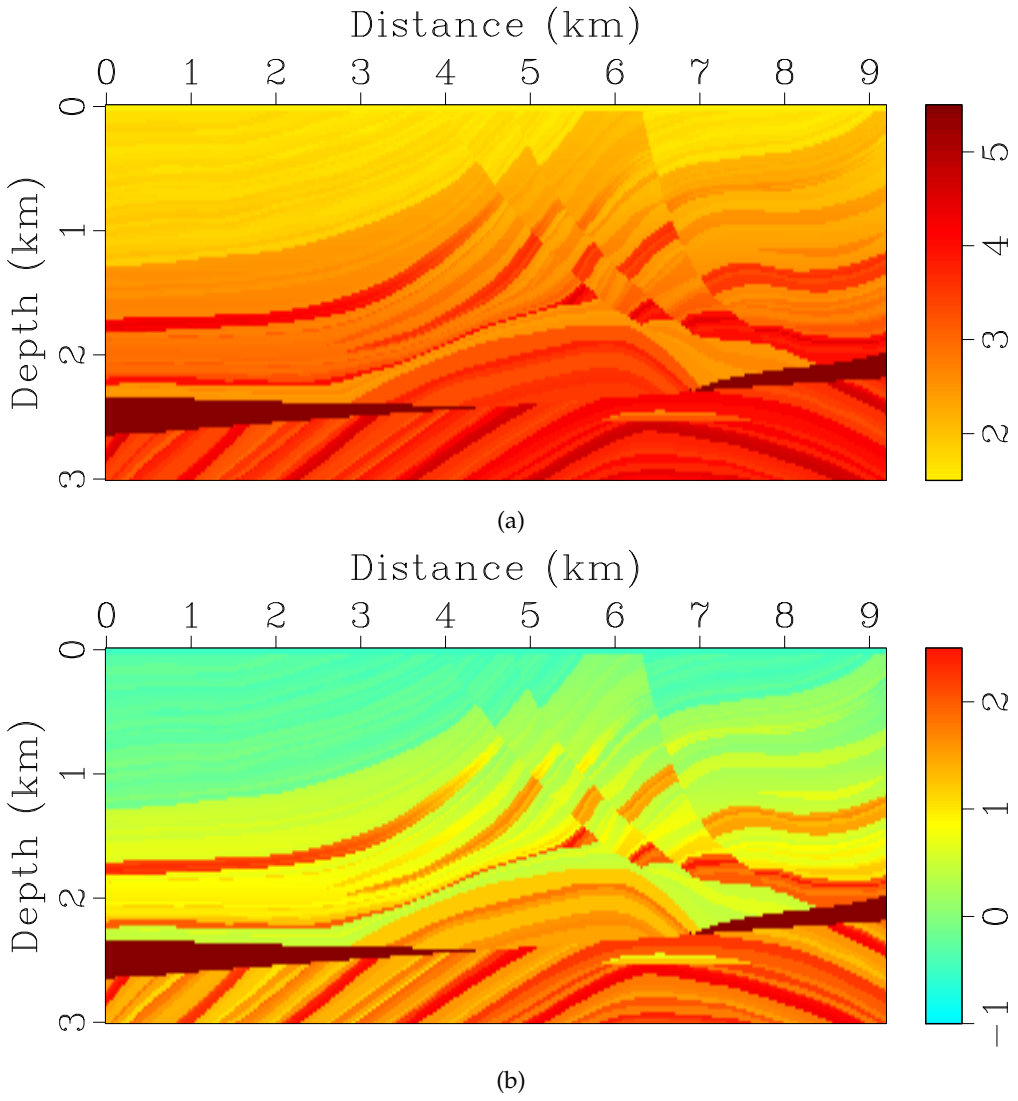


Figure 8: a) The Mamrousi model. b) The velocity perturbation we seek considering a constant background velocity model of 2.5 km/s.

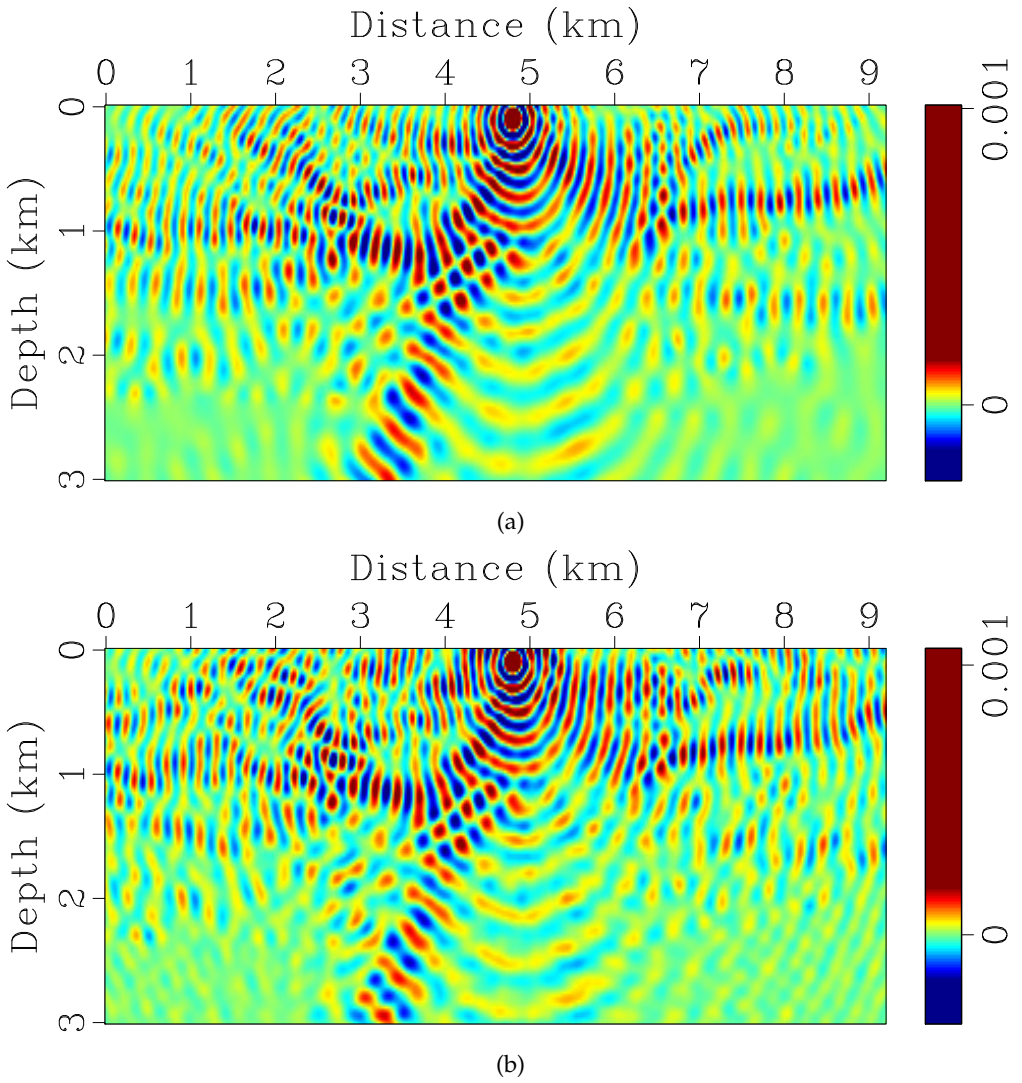
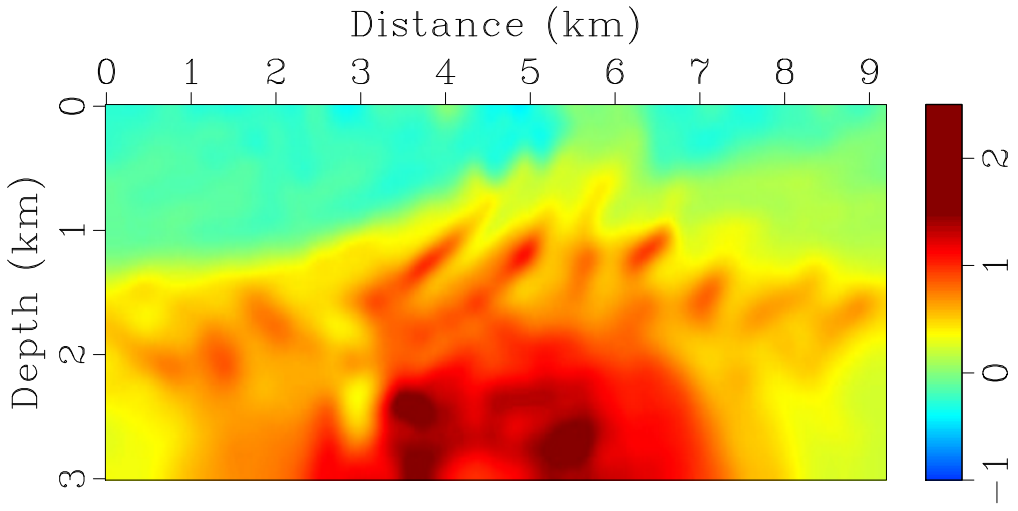
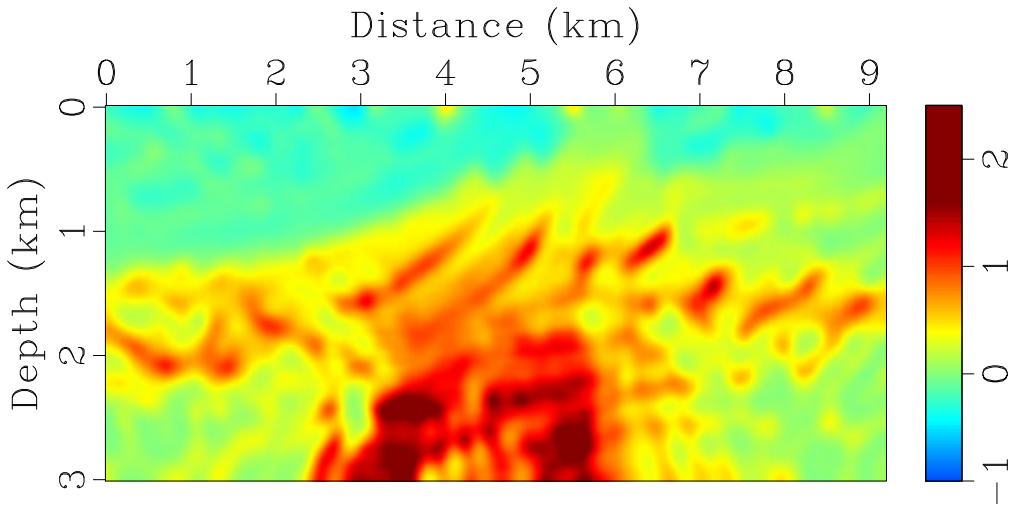


Figure 9: a) The true wavefield obtained by solving the Helmholtz wave equation for the true velocity model shown in Figure 8a. b) The wavefield corresponding to the background 2.5 km/s model with the modified source.



(a)



(b)

Figure 10: a) The inverted model using the smooth division. b) The inverted model using the gradient method. Both inverted models using the true $\mathbf{f_e}$ and computed wavefield in Figure 9b

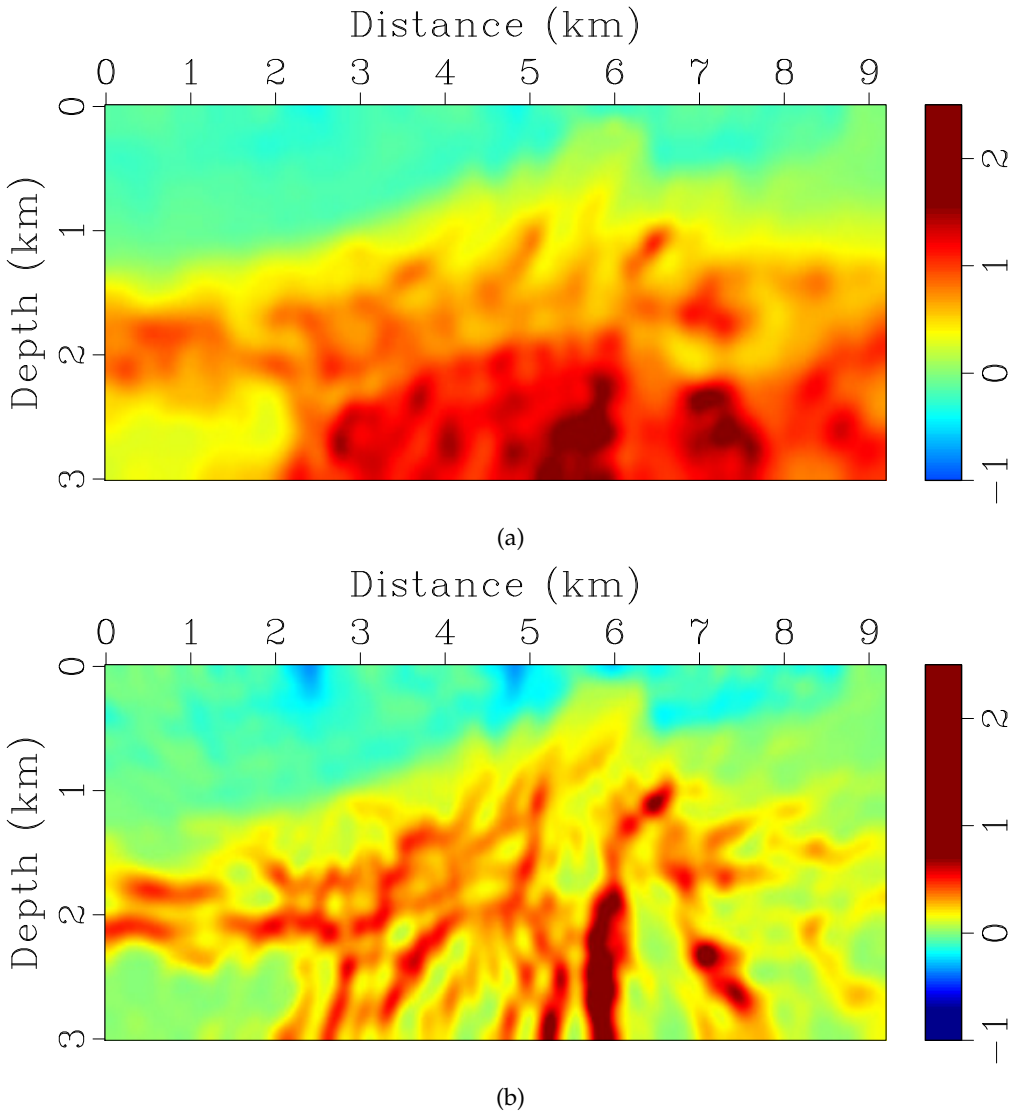


Figure 11: a) The smooth division inverted model for the simultaneous source case. b) The inverted model using the gradient method for the simultaneous source case.

7 Conclusions

I developed an efficient waveform inversion strategy that relies on a convex optimization problem in inverting for the wavefield and a modified source. The wavefield and modified source can be used to directly invert for the velocity perturbation in a separate step. The efficiency of the strategy is provided by relying on the background wavefield for such a convex inversion in which we iteratively invert for the wavefield and the modified source function. In addition, the direct inversion for the perturbation is immune to crosstalk in the case of using multi sources. Some of the features of the approach is demonstrated on a simple two-layer model, as well as the Marmousi model.

8 Acknowledgments

I thank KAUST for its support. I also thank the SWAG group for many useful discussions. I especially thank Chao Song for many fruitful exchanges. I also thank the assistant editor, Mohammad Akbar Zuberi and an anonymous reviewer for their helpful in reviewing the paper.

References

Abubakar, A., W. Hu, T. M. Habashy, and P. M. van den Berg, 2009, Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data: GEOPHYSICS, **74**, WCC47–WCC58.

Aki, K., and P. Richards, 1980, Quantitative Seismology: Theory and Methods: Freeman.

Artman, B., I. Podladtchikov, and B. Witten, 2010, Source location using timereverse imaging: Geophysical Prospecting, **58**, 861–873.

Born, M., 1926, Quantenmechanik der stovorgnge: Zeitschrift fr Physik, **38**, 803827.

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

Fomel, S., 2007, Local seismic attributes: GEOPHYSICS, **72**, A29–A33.

Green, G., 1828, An essay on the application of mathematical analysis to the theories of electricity and magnetism: Nottingham, England: T. Wheelhouse, 1012.

Hughes, T., 1995, Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods: Comput. Methods Appl. Mech. Engrg., **127**, 387–401.

Innanen, K. A., 2006, A multidimensional acoustic forward scattering series model of reflected and transmitted wavefield events: SEG Technical Program Expanded Abstracts 2006, 2961–2965.

Jakobsen, M., and R. Wu, 2016, Renormalized scattering series for frequency-domain waveform modelling of strong velocity contrasts: Geophysical Journal International, **206**, 880–899.

Kleinman, R., and P. den Berg, 1992, A modified gradient method for two- dimensional problems in tomography: Journal of Computational and Applied Mathematics, **42**, 17 – 35.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-waveform seismic inversion using encoded sources: Geophysics, **74**, WCC177–WCC188.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. G., Z.-M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.

Tarantola, A., 1987, Inverse Problem Theory: Elsevier.

van den Berg, P. M., and R. E. Kleinman, 1997, A contrast source inversion method: Inverse Problems, **13**, 1607.

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.

Wu, R., and M. Toksoz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: GEOPHYSICS, **52**, 11–25.

Wu, R. S., and K. Aki, 1988, Introduction: Seismic wave scattering in three-dimensionally heterogeneous earth, *in* Scattering and attenuation of seismic waves: Birkhäuser Verlag, 1–6. (PAGEOPH).

Wu, R.-S., and Y. Zheng, 2014, Non-linear partial derivative and its de wolf approximation for non-linear seismic inversion: Geophysical Journal International, **196**, 1827–1843.

Wu, Z., and T. Alkhalifah, 2015, Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion: GEOPHYSICS, **80**, R317–R329.

Xiong, Z., 1989, Dreidimensionale geoelektromagnetische umkehraufgaben: Dissertation, Univ. Gottingen.