A unified formulation of nonlinear, linearized and reflection waveform inversion

Pengliang Yang

School of Mathematics, Harbin Institute of Technology, 150001, China

E-mail: ypl.2100@gmail.com

October 20, 2023

Key points

- Nonlinear, linearized, reflection-based and joint waveform inversion are framed in a unified formulation.

- Gradients and multiparameter preconditioner are developed using chain rule and logarithmic scaling.

- Reflection-based waveform inversion requires more computation to retrieve background model and perturbation.

Abstract

This article formulates nonlinear full waveform inversion (FWI), linearized wave-form inversion (also referred to as least-squares migration, LSM), reflection waveform inversion (RWI) and joint full waveform inversion (JFWI) within a unified mathematical framework. The consistency of this formulation relies on the logarithmic transformation of the physical parameters and the first order Taylor expansion of the perturbed wave operator. It sheds some physical insight to relate model perturbation with reflectivity in a quantitative manner, which is particularly important to prescribe Born scattering equation. The log scaling also simplifies the computation of the gradient of the misfit functional, while yielding a concise expression of the multiparameter pseudo-Hessian preconditioner under zero-offset source-receiver geometry. The gradient of the misfit functional of linearized and nonlinear inverse problems under various parameterizations are derived thanks to the chain rule. The extension of conventional FWI to RWI and JFWI requires more sophisticated computational procedure, in order to decipher a background model from reflection data.

Introduction

Wave propagation is a fundamental tool to probe the complex structures of the Earth. Seismic tomography based on wave information (both amplitude and phase) is a very powerful technique to understand the properties of the subsurface. Among various of seismic tomography methods, full waveform inversion (FWI) is one of the most promising approaches to fully explore the variety of the signals for high resolution seismic imaging (Virieux and Operto, 2009). The quest for the information of the Earth is formulated as an inverse problem in least-squares sense (Tarantola and Valette, 1982), solved iteratively by minimizing the data misfit between observed seismogram and predicted wavefield according to given subsurface parameters. Nowadays, the FWI methodology has been extensively used in both global seismology and exploration geophysics, for hydrocarbon bearing reservoir prospecting (Sirgue

et al., 2010; Operto et al., 2015), and imaging the Earth structure at crustal and regional scale (Fichtner et al., 2009, 2010).

Since the early 1D inversion of seismic impedance (Bamberger et al., 1982), finding the subsurface parameters has been considered as a typical PDE-constrained optimization, while the Fréchet derivatives of the optimization objective can be computed efficiently based on the optimal control theory (i.e. the adjoint state method). The seminar works of Lailly (1983) and Tarantola (1984a) under acoustic approximation reveal that the calculation of the gradient of the misfit functional with respect to different model parameters strongly resembles the seismic migration process: It is nothing else than cross-correlation between the forward incident field and the receiver field propagating backwards in time, despite a scaling of the scattering matrix related to the parameterization of the wave equation. The extension of FWI to elastic regime (Mora, 1987) significantly eliminates the inaccuracy of acoustic approximation, while the importance of low-wavenumber initial model and heavy computation expense are highlighted. In the absence of a precise macromodel, the use of a successive multiscale strategy from low to high frequencies is particularly important for successful inversion (Bunks et al., 1995). The development of frequency domain FWI (Pratt, 1999) leads to the continuation of sparsely selected frequencies for efficient implementation (Sirgue and Pratt, 2004).

The importance of the quality of an initial model has been well recognized for FWI. The predicted data from a very crude model results in the mismatch of the traveltime more than half the period, leading to the so-called the cycle skipping phenomenon so that FWI converges to a local minimum instead of global minimum (Virieux and Operto, 2009). Besides the multiscale inversion, a number of attempts have been made on different misfit functions, using cross-correlation time-shifts (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010), instaneous phase and envelope (Bozdağ et al., 2011), deconvolution (Luo and Sava, 2011; Warner and Guasch, 2016), a combination of time-frequency analysis (Fichtner et al., 2008), and the optimal transport distance (Engquist and Froese, 2014; Métivier et al., 2016), just

to name a few.

It has been discovered that the model update in classical FWI is mainly driven by early arrivals, which corresponds to direct and diving waves coming from transmitted energy at intermediate and far offset. This suggest the use of wide aperture, long offset data (Virieux and Operto, 2009), while the potential of weak reflections concentrated at near offset is not fully exploited. This is indeed deviated from the original motivation of (Tarantola, 1984a), to make use of seismic reflections. Migration-based traveltime tomography proposed by Chavent et al. (1994) is an early attempt to use reflected data (Plessix et al., 1995; Clément et al., 2001). The low-frequency update along the wavepath in a rabbit-ear shape highlighted by Xu et al. (2012) arose a wide range of interest of reflection waveform inversion (RWI) in exploration geophysics. This idea has been further explored by Brossier et al. (2015). The joint full waveform inversion (JFWI) approach by Zhou et al. (2015, 2018) extends RWI by taking into account both long-offset refractions and short-offset reflections.

The nonlinear FWI requires the computation of the gradient of the misfit to construct a descent direction and the estimation of an appropriate step length via line search so that the inversion can update the model parameters iteratively. When a good background model is available, the model perturbations can also be estimated. This also motivates the innovation of linearized waveform inversion by Tarantola (1984b), which is often coined as least-squares migration to estimate a seismic image from reflection data (Nemeth et al., 1999).

This paper is not intended to do an exhaustive review of all aspects of FWI in the literature, but focusing on developing a consistent formulation of nonlinear FWI, linearized waveform inversion and recent reflection waveform inversion. Using the adjoint state method, we derive the gradient of the misfit functional among all these approaches in a succinct fashion, under various kinds of model parameterization. Insightful connections and distinctions among these methods will be drawn towards a unified mathematical framework and a consistent implementation. The log scaling of the model parameters also simplifies the computation of the gradient of the misfit functional, while yielding a concise expression of

the multiparameter pseudo-Hessian preconditioner. The logarithmic impedance is found to uniquely define the reflectivity image in least-squares migration.

Forward problem: A system of linear PDE

We start our development by considering isotropic acoustic wave equation of the first order

$$
\begin{cases}
\rho(\mathbf{x})\partial_t \mathbf{v}(\mathbf{x},t) + \nabla p(\mathbf{x},t) = f_{\mathbf{v}}(\mathbf{x},t), & \mathbf{x} \in X, t \in [0,T] \\
\partial_t p(\mathbf{x},t) + \kappa(\mathbf{x})\nabla \cdot \mathbf{v}(\mathbf{x},t) = f_p(\mathbf{x},t), & \mathbf{x} \in X, t \in [0,T]
\end{cases}
\tag{1}
$$

where $\mathbf{v} = [v_x, v_y, v_z]^{\mathrm{H}}$ is the particle velocity including three directional components v_x, v_y and v_z; p denote the acoustic pressure; $f_{\mathbf{v}} = [f_{v_x}, f_{v_y}, f_{v_z}]^{\mathrm{H}}$ collects the sources of horizontal and vertical forcing while f_p stand for an explosive source term; ρ is the density, κ is the bulk modulus which can be expressed via density ρ and P-wave velocity V_p, i.e., $\kappa = \rho V_p^2$. Clearly, the pressure and particle velocity fields are functions of both time t and space $\mathbf{x}$, while the medium parameters ρ and κ vary only in space $\mathbf{x}$. The forward equation in (1) may be written compactly in a linear matrix system

$$
\underbrace{\begin{bmatrix} \rho\partial_t & \nabla \\ \kappa\nabla\cdot & \partial_t \end{bmatrix}}_{A(m)} \underbrace{\begin{bmatrix} \mathbf{v} \\ p \end{bmatrix}}_{u} = \underbrace{\begin{bmatrix} f_{\mathbf{v}} \\ f_p \end{bmatrix}}_{f}
\tag{2}
$$

where $A(m)$ is a linear wave operator, depending on the medium parameters $m \in \{\rho, \kappa\}$. Note that the P-wave impedance is defined as $I_p = \rho V_p$, the wave operator can therefore be parameterized by different combinations of the parameters, using the set (V_p, ρ) or another (V_p, I_p).

Nonlinear full waveform inversion

Full waveform inversion aims to iteratively reconstruct the true subsurface parameters by minimizing the error between the observed seismogram d and the synthetic data u predicted according to the wave equation and the given parameters. The misfit functional for FWI is defined in least-squares sense

$$J(m) = \frac{1}{2}\|Ru - d\|^2 = \frac{1}{2}\sum_s \sum_r \int_0^T (u(\mathbf{x}_r, t; \mathbf{x}_s) - d(\mathbf{x}_r, t; \mathbf{x}_s))^2 \mathrm{d}t, \tag{3}$$

where R is a restriction operator projecting the simulated wavefield to the receiver locations, i.e., $Ru := \int_X u(\mathbf{x}, t)\delta(\mathbf{x} - \mathbf{x}_r)\mathrm{d}\mathbf{x} = u(\mathbf{x}_r, t)$. To do minimize the misfit $J(m)$, FWI iteratively updates the model

$$m^{k+1} = m^k + \alpha\delta m, \tag{4}$$

where α is a proper step length; δm is the descent direction satisfying the normal equation

$$H\delta m = -\nabla J(m), \tag{5}$$

where H denotes the Hessian matrix which could be prohibitively large to store in memory for 3D large scale problems. Solution of the normal equation (5) via conjugate gradient iterations leads to the so-called Newton-CG method (Yang et al., 2018), which involves two nested loops. The the convergence rate of the outer loop can be very fast but each outer loop requires many inner CG iterations, therefore every expensive. In this work, we mainly rely on limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm so that $\delta m = -H^{-1}\nabla J(m)$ by approximating the inverse Hessian H^{-1} with stored gradients from previous iterations.

Using standard adjoint state method (Tromp et al., 2005; Plessix, 2006), we can easily find the gradient of the misfit functional with respect to the model parameter m in the time

7

domain

$$\nabla J(m) = \int_0^T \mathrm{d}t \lambda(\mathbf{x}, t)^{\mathrm{H}} \frac{\partial A(m)}{\partial m} u(\mathbf{x}, t) \tag{6}$$

where the stacking over sources and receivers are omitted henceforth; λ is the so-called adjoint field satisfying the equation

$$A^{\mathrm{H}}(m)\lambda = -\frac{\partial J(m)}{\partial u} = R^{\mathrm{H}}(d - Ru). \tag{7}$$

Let us split adjoint field and data residual into particle velocity and pressure, say, $\lambda = [\lambda_{\mathbf{v}}, \lambda_p]^{\mathrm{H}}$ and $R^{\mathrm{H}}(d - Ru) = [\delta d_{\mathbf{v}}, \delta d_p]^{\mathrm{H}}$. Since $(\partial_t)^{\mathrm{H}} = -\partial_t$, $(\nabla)^{\mathrm{H}} = -\nabla\cdot$ and $(\kappa\nabla\cdot)^{\mathrm{H}} = -\nabla\kappa$, the adjoint equation in equation (7) is

$$\begin{bmatrix} -\rho\partial_t & -\nabla\kappa \\ -\nabla\cdot & -\partial_t \end{bmatrix} \begin{bmatrix} \lambda_{\mathbf{v}} \\ \lambda_p \end{bmatrix} = \begin{bmatrix} \delta d_{\mathbf{v}} \\ \delta d_p \end{bmatrix}. \tag{8}$$

where we have tacitly assumed zero boundary conditions, initial and final conditions for forward and adjoint wavefields. The adjoint wave operator in equation (8) is different from the forward operator. Define $\bar{\mathbf{v}} = \lambda_{\mathbf{v}}$ and $\bar{p} = \kappa\lambda_p$. We obtain

$$\begin{cases} \rho\partial_t \bar{\mathbf{v}} + \nabla\bar{p} = -\delta d_{\mathbf{v}}, \\ \partial_t \bar{p} + \kappa\nabla \cdot \bar{\mathbf{v}} = -\kappa\delta d_p, \end{cases} \tag{9}$$

where the wave operator becomes the same as the forward operator thanks to a compliance formulation (Yang et al., 2016a, 2018). This variant of the adjoint equation implies that the same forward modelling code can be reused for adjoint simulation, as long as the adjoint source is properly scaled during the backward injection of the weighted data residual. The

gradient of the misfit with respect to density and bulk modulus are

$$\frac{\partial J(m)}{\partial \rho} = \int_0^T \mathrm{d}t \begin{bmatrix} \lambda_{\mathbf{v}}^{\mathrm{H}} & \lambda_p \end{bmatrix} \begin{bmatrix} \partial_t & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ p \end{bmatrix} = \int_0^T \lambda_{\mathbf{v}}^{\mathrm{H}} \partial_t \mathbf{v} \mathrm{d}t = \int_0^T \bar{\mathbf{v}}^{\mathrm{H}} \partial_t \mathbf{v} \mathrm{d}t \tag{10a}$$

$$\frac{\partial J(m)}{\partial \kappa} = \int_0^T \mathrm{d}t \begin{bmatrix} \lambda_{\mathbf{v}}^{\mathrm{H}} & \lambda_p \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \nabla \cdot & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ p \end{bmatrix} = \int_0^T \lambda_p \nabla \cdot \mathbf{v} \mathrm{d}t = \frac{1}{\kappa} \int_0^T \bar{p} \nabla \cdot \mathbf{v} \mathrm{d}t. \tag{10b}$$

Because model parameters are never updated at source locations, we can make the substitution $\nabla \cdot \mathbf{v} = -\partial_t p / \kappa$ in equation (10b) so that $\partial J(m)/\partial \kappa = -\int_0^T \bar{p} \partial_t p \mathrm{d}t / \kappa^2$, which is the gradient formulae presented in Yang et al. (2018). At each time step, we have to access the forward background field for the calculation of the gradient. This creates a key computational challenging. Therefore, a number of computing schemes have been proposed, such as lossy data compression (Sun and Fu, 2013; Boehm et al., 2016), wavefield reconstruction via stored boundaries (Yang et al., 2016c) and optimal checkpointing strategy (Symes, 2007; Yang et al., 2016b).

To manage the significant magnitude difference of various parameters, a commonly used strategy is to take the transformation of the physical parameters as the inversion parameters. One of the most well known tricks is unity-based min-max normalization (also known feature scaling in machine learning) (Yang et al., 2018): the inversion parameters are dimensionless and rescaled within 0 and 1 using minimum and maximum bounds of the physical parameters. Another popular choice for positive parameters is to take the logarithmic scaling so that the magnitude of different inversion parameters are comparable. Under the parameterization of $m_1 = \ln \rho, m_2 = \ln \kappa$ we obtain

$$\frac{\partial J(m)}{\partial \ln \rho} = \frac{\partial J(m)}{\partial \rho} \frac{\partial \rho}{\partial \ln \rho} = \rho \int_0^T \bar{\mathbf{v}}^{\mathrm{H}} \partial_t \mathbf{v} \mathrm{d}t, \frac{\partial J(m)}{\partial \ln \kappa} = \frac{\partial J(m)}{\partial \kappa} \frac{\partial \kappa}{\partial \ln \kappa} = \int_0^T \bar{p} \nabla \cdot \mathbf{v} \mathrm{d}t. \tag{11}$$

It has become trendy to use tools of automatic differentiation (AD) to derive the gradient of the misfit function, thanks to the increasingly advanced machine learning tools.

9

$_{153}$ Admittedly, automatic differentiation alleviates the job of the developer by getting rid of $_{154}$ comprehensive mathematical derivation, but requires significantly larger amount of com- $_{155}$ puter memory for computation. This is because AD relies on the chain rule, which requires $_{156}$ heavy floating point operations and storage of all intermediate variables to derive a correct $_{157}$ computing procedure. Constructing the gradient based on explicit expressions are therefore $_{158}$ preferred in 3D large scale applications to avoid these overheads.

$_{159}$ Gradient expressions under different parameterizations

The chain rule allows us to easily switch from one specific parameterization to any other combination of medium parameters. Note that the logarithmic scaling of the bulk modulus $(\kappa = \rho V_p^2)$ is $\ln \kappa = \ln \rho + 2 \ln V_p$. Under $V_p - \rho'$ parameterization, the chain rule gives

$$\frac{\partial A(m)}{\partial \ln \rho'} = \frac{\partial A(m)}{\partial \ln \rho} \frac{\partial \ln \rho}{\partial \ln \rho'} + \frac{\partial A(m)}{\partial \ln \kappa} \frac{\partial \ln \kappa}{\partial \ln \rho'} = \frac{\partial A(m)}{\partial \ln \rho} + \frac{\partial A(m)}{\partial \ln \kappa}, \tag{12a}$$

$$\frac{\partial A(m)}{\partial \ln V_p} = \frac{\partial A(m)}{\partial \ln \kappa} \frac{\partial \ln \kappa}{\partial \ln V_p} = 2 \frac{\partial A(m)}{\partial \ln \kappa}. \tag{12b}$$

$_{160}$ where we introduce a prime $(\rho' := \rho)$ to distinguish the same parameter under different $_{161}$ parameterization. As a result, the gradient of the misfit with respect to $m_1 = \ln \rho'$ and $_{162}$ $m_2 = \ln V_p$ can be computed

$$\frac{\partial J(m)}{\partial \ln \rho'} = \frac{\partial J(m)}{\partial \ln \rho} + \frac{\partial J(m)}{\partial \ln \kappa}, \frac{\partial J(m)}{\partial \ln V_p} = 2 \frac{\partial J(m)}{\partial \ln \kappa}. \tag{13}$$

$_{163}$ The logarithmic transformation of density and bulk modulus are related to impedance $_{164}$ and wavespeed through

$$\begin{cases} \ln \rho = \ln I_p - \ln V_p \\ \ln \kappa = \ln I_p + \ln V_p \end{cases} \tag{14}$$

10

Under $I_p - V_p$ parameterization, the chain rule leads to

$$\frac{\partial A(m)}{\partial \ln I_p} = \frac{\partial A(m)}{\partial \ln \rho} \frac{\partial \ln \rho}{\partial \ln I_p} + \frac{\partial A(m)}{\partial \ln \kappa} \frac{\partial \ln \kappa}{\partial \ln I_p} = \frac{\partial A(m)}{\partial \ln \rho} + \frac{\partial A(m)}{\partial \ln \kappa}, \tag{15a}$$

$$\frac{\partial A(m)}{\partial \ln V_p} = \frac{\partial A(m)}{\partial \ln \rho} \frac{\partial \ln \rho}{\partial \ln V_p} + \frac{\partial A(m)}{\partial \ln \kappa} \frac{\partial \ln \kappa}{\partial \ln V_p} = -\frac{\partial A(m)}{\partial \ln \rho} + \frac{\partial A(m)}{\partial \ln \kappa}. \tag{15b}$$

Consequently, the gradient of the misfit with respect to $m_1 = \ln I_p$ and $m_2 = \ln V_p$ are

$$\frac{\partial J(m)}{\partial \ln I_p} = \frac{\partial J(m)}{\partial \ln \rho} + \frac{\partial J(m)}{\partial \ln \kappa}, \frac{\partial J(m)}{\partial \ln V_p} = -\frac{\partial J(m)}{\partial \ln \rho} + \frac{\partial J(m)}{\partial \ln \kappa}. \tag{16}$$

Clearly, the multiplicative relations between various physical parameters becomes additive after log scaling, therefore very convenient and essential to reduce floating point computations.

Multiparameter pseudo-Hessian preconditioning

To improve the convergence rate of the inversion, pseudo-Hessian preconditioner proposed in Shin et al. (2001) approximates the diagonal elements of the Hessian up to a scaling factor

$$\tilde{H}_{m,m} = \int_0^T \mathrm{d}t \left(\frac{\partial A}{\partial m} u \right)^{\mathrm{H}} \left(\frac{\partial A}{\partial m} u \right) \tag{17}$$

by replacing the receiver field with the source field based on the assumption of zero-offset geometry. In view of the fact that

$$\frac{\partial A}{\partial \ln \rho} u = \frac{\partial \rho}{\partial \ln \rho} \frac{\partial A}{\partial \rho} u = \rho \begin{bmatrix} \partial_t & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ p \end{bmatrix} = \begin{bmatrix} \rho \partial_t \mathbf{v} \\ 0 \end{bmatrix}, \tag{18a}$$

$$\frac{\partial A}{\partial \ln \kappa} u = \frac{\partial \kappa}{\partial \ln \kappa} \frac{\partial A}{\partial \kappa} u = \kappa \begin{bmatrix} 0 & 0 \\ \nabla \cdot & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ \kappa \nabla \cdot \mathbf{v} \end{bmatrix} \tag{18b}$$

11

we arrive at

$$\tilde{H}_{\ln\rho,\ln\rho} = \int_0^T dt \left(\frac{\partial A}{\partial \ln\rho}u\right)^{\mathrm{H}} \left(\frac{\partial A}{\partial \ln\rho}u\right) = \rho^2 \int_0^T dt(\partial_t \mathbf{v})\cdot(\partial_t \mathbf{v}), \tag{19a}$$

$$\tilde{H}_{\ln\kappa,\ln\kappa} = \int_0^T dt \left(\frac{\partial A}{\partial \ln\kappa}u\right)^{\mathrm{H}} \left(\frac{\partial A}{\partial \ln\kappa}u\right) = \kappa^2 \int_0^T dt(\nabla\cdot\mathbf{v})^2. \tag{19b}$$

Using the relation in equation (12), the diagonal elements of the pseudo-Hessian under $V_p - \rho'$ parameterization are

$$\tilde{H}_{\ln\rho',\ln\rho'} = \tilde{H}_{\ln\rho,\ln\rho} + \tilde{H}_{\ln\kappa,\ln\kappa}, \tag{20a}$$

$$\tilde{H}_{\ln V_p,\ln V_p} = 4\tilde{H}_{\ln\kappa,\ln\kappa}. \tag{20b}$$

In case of $I_p - V_p$ parameterization, equation (15) leads to

$$\tilde{H}_{\ln I_p,\ln I_p} = \tilde{H}_{\ln\rho,\ln\rho} + \tilde{H}_{\ln\kappa,\ln\kappa}, \tag{21a}$$

$$\tilde{H}_{\ln V_p,\ln V_p} = \tilde{H}_{\ln\rho,\ln\rho} + \tilde{H}_{\ln\kappa,\ln\kappa}. \tag{21b}$$

Despite the fact that no parameter coupling has been taken into account, the multiparameter pseudo-Hessian derived here is much easier to compute than the one in Yang et al. (2018).

Scattering equation via Born linearization

Based on the perturbation theory, the same wave equation should be satisfied

$$A(m + \delta m)(u + \delta u) = f, \tag{22}$$

where δu is the wavefield perturbation excited by a perturbation of the model δm. The linearization of the wave operator using Taylor series up to the first order gives

$$A(m + \delta m) \approx A(m) + \delta m \cdot \partial_m A, \tag{23}$$

yielding

$$A(m)u + A(m)\delta u + \delta m \cdot \partial_m A(m)u + \delta m \cdot \partial_m A \delta u = f. \tag{24}$$

Since $A(m)u = f$ holds for background field, we obtain a linear system for the scattering field δu after neglecting the 2nd order perturbation $\delta m \cdot \partial_m A \delta u$

$$A(m)\delta u = -\delta m \cdot \partial_m A(m)u. \tag{25}$$

This reveals that (1) the scattering field δu has a linear dependence on the model perturbation δm; (2) modelling the scattering field δu requires modelling background field u, which is used to construct the scattering source. This process is often coined Born modelling or reverse time demigration (RTDM).

For the 1st order acoustic wave equation, we define the model parameterized by $\delta m_1 = \delta \ln V_p = \delta V_p / V_p$ and $\delta m_2 = \delta \ln \rho = \delta \rho / \rho$. Using the equations (12) and (18a), we have

$$
\begin{aligned}
\delta m \cdot \frac{\partial A(m)}{\partial m} u &= \delta \ln V_p \frac{\partial A}{\partial \ln V_p} u + \delta \ln \rho' \frac{\partial A}{\partial \ln \rho'} u \\
&= 2\delta m_1 \frac{\partial A}{\partial \ln \kappa} u + \delta m_2 \left(\frac{\partial A}{\partial \ln \rho} u + \frac{\partial A}{\partial \ln \kappa} u \right) \\
&= (2\delta m_1 + \delta m_2) \frac{\partial A}{\partial \ln \kappa} u + \delta m_2 \frac{\partial A}{\partial \ln \rho} u \\
&= \begin{bmatrix} \delta m_2 \rho \partial_t \mathbf{v} \\ (2\delta m_1 + \delta m_2) \kappa \nabla \cdot \mathbf{v} \end{bmatrix},
\end{aligned}
\tag{26}
$$

13

the scattering equation translates into

$$
\begin{cases}
\rho\partial_t\delta\mathbf{v} + \nabla\delta p = -\delta m_2\rho\partial_t\mathbf{v} \\
\partial_t\delta p + \kappa\nabla\cdot\delta\mathbf{v} = -(2\delta m_1 + \delta m_2)\kappa\nabla\cdot\mathbf{v}
\end{cases}
\tag{27}
$$

If $m_1 = \ln V_p$ and $m_2 = \ln I_p$, we have $\delta m_1 = \delta V_p/V_p$ and $\delta m_2 = \delta I_p/I_p$. Using equations (15) and (18b), we obtain

$$
\begin{aligned}
\delta m \cdot \frac{\partial A(m)}{\partial m}u &= \delta m_1 \frac{\partial A(m)}{\partial \ln V_p}u + \delta m_2 \frac{\partial A(m)}{\partial \ln I_p}u \\
&= \delta m_1(-\frac{\partial A}{\partial \ln \rho}u + \frac{\partial A}{\partial \ln \kappa}u) + \delta m_2(\frac{\partial A}{\partial \ln \rho}u + \frac{\partial A}{\partial \ln \kappa}u) \\
&= \begin{bmatrix} (\delta m_2 - \delta m_1)\rho\partial_t\mathbf{v} \\ (\delta m_2 + \delta m_1)\kappa\nabla\cdot\mathbf{v} \end{bmatrix},
\end{aligned}
$$

yielding the scattering equation

$$
\begin{cases}
\rho\partial_t\delta\mathbf{v} + \nabla\delta p = -(\delta m_2 - \delta m_1)\rho\partial_t\mathbf{v} \\
\partial_t\delta p + \kappa\nabla\cdot\delta\mathbf{v} = -(\delta m_2 + \delta m_1)\kappa\nabla\cdot\mathbf{v}
\end{cases}
\tag{28}
$$

The above manipulation of linearized wave equation gives a concise and novel derivation of the scattering equation under various kinds of parameterization, which is much simpler than the usual derivations done in the literature.

By perturbing one parameter while fixing the other, we can investigate the scattering/radiation pattern of each parameter under a specific model parameterization. Figure 1 provides a graphic display of the scattering pattern of each parameter under three families of model parameterization. Figures 1a and 1b show that the scatterings of κ and ρ are overlapped in most of the radiation angles, indicating that the indistinguishable responses of them at the same location. Figures 1c and 1d show that the responses from V_p and ρ are coupled at small angles, while V_p diffracts in the full aperture. Figures 1e and 1f reveal

a better decoupling between V_p and I_p. Compared with $V_p - I_p$ parameterization, the reconstruction of V_p in an inversion under $V_p - \rho$ parameterization may lack of resolution due to weak scattering energy at small angles. This is in consistent with Douma et al. (2010): elimination of the so-called low frequency artifacts in conventional RTM image by Laplacian filtering is equivalent to using the impedenace kernel as RTM image, since the scattering of I_p mainly comes from small angles related to near offset reflections, as indicated in Figure 1f.

Linearized waveform inversion

The seismic data d can be split into background refractions (first arrivals and the diving waves) d_0 and the reflections δd, $d(\mathbf{x}_r, t) = d_0(\mathbf{x}_r, t) + \delta d(\mathbf{x}_r, t)$. Assume the background model parameters m are known. The synthetic data modelled by the wave equation $A(m)u = f$ using m should reproduce the first arrivals and the diving waves $d_0(\mathbf{x}_r, t)$ precisely. In this case, the waveform misfit in equation (3) reduces to

$$J(\delta m) = \frac{1}{2}\|R\delta u - \delta d\|^2. \tag{29}$$

As a reliable background model m is known, the goal of the linearized waveform inversion (Symes, 2008a,b) is to find the dimensionless image δm by minimizing (29) through linear optimization. The following Lagrangian functional translates the linear minimization with scattering equation constraint into an unconstrained optimization

$$\mathcal{L}(\delta m, \delta u, \mu) = \frac{1}{2}\|R\delta u - \delta d\|^2 + \langle \mu, A\delta u + \delta m \cdot \partial_m Au \rangle_{X \times [0,T]} \tag{30}$$

where we have equipped with the inner product defined as follows:

$$\langle h, g \rangle_{X \times [0,T]} = \int_X \int_0^T h(\mathbf{x}, t)^{\mathrm{H}} g(\mathbf{x}, t) \mathrm{d}\mathbf{x}\mathrm{d}t, \quad \mathbf{x} \in X, \quad t \in [0, T]. \tag{31}$$

According to the chain rule, the gradient of the Lagrangian with respect to the reflectivity is given by

$$\frac{\partial \mathcal{L}}{\partial \delta m} = \frac{\partial \mathcal{L}}{\partial \mu}\frac{\partial \mu}{\partial \delta m} + \frac{\partial \mathcal{L}}{\partial \delta u}\frac{\partial \delta u}{\partial \delta m} + \langle \mu, \frac{\partial \delta m \cdot \partial_m A(m)}{\partial \delta m}u \rangle_{[0,T]} \tag{32}$$

Zeroing the partial derivatives of the Lagrangian with respect to μ ($\partial \mathcal{L}/\partial \mu = 0$) gives exactly the forward scattering equation (25). Zeroing the partial derivatives of the Lagrangian with respect to δu ($\partial \mathcal{L}/\partial \delta u = 0$) gives the adjoint equation

$$A^{\mathrm{H}}\mu = R^{\mathrm{H}}(\delta d - R\delta u). \tag{33}$$

At the saddle point, both the forward scattering equation and the adjoint equation are satisfied. In terms of (32), the gradient of the objective with respect to δm is then

$$\frac{\partial J(\delta m)}{\partial \delta m} = \langle \mu, \partial_m A(m)u \rangle_{[0,T]} = \int_0^T \mathrm{d}t \mu^{\mathrm{H}}(\mathbf{x},t)\frac{\partial A(m)}{\partial m}u(\mathbf{x},t), \tag{34}$$

Obviously, the expressions in (33) and (34) are highly similar to the adjoint equation (7) and the gradient (6) in nonlinear FWI. The key difference is the adjoint field μ, which must be simulated using the residual of reflection data.

In (34), the cross-correlation between forward background field u and adjoint field μ resembles RTM imaging. Thus it is often called 'migration', while its adjoint (Born modelling) is referred to as demigration. Let denote Born modelling $R\delta u := L\delta m$ where the operator L highlights the linear dependence between synthetic data $R\delta u$ and the image δm. The misfit then reads $J(\delta m) = 1/2\|L\delta m - \delta d\|^2$, so that the gradient becomes

$$\frac{\partial J(\delta m)}{\partial \delta m} = L^{\mathrm{H}}(L\delta m - \delta d), \tag{35}$$

where L^{H} is then the migration operator mapping the reflection data residual $(L\delta m - \delta d)$ back to the image domain. The expression in (35) is abstract and formal. To be concrete, we use (34). Let us point out that the adjoint equation must be performed backwards in

16

236 time.

237 In fact, the above derivation is quite generic, applicable to all kinds of wave equations
238 (acoustics, viscous acoustics and elastics). The optimality condition of this linear inverse
239 problem requires $\partial J(\delta m)/\partial \delta m = 0$. In terms of equation (35), it is

$$L^{\mathrm{H}}L\delta m = L^{\mathrm{H}}\delta d. \tag{36}$$

240 The Hessian matrix of the above system $L^{\mathrm{H}}L$ is symmetric semi-positive definite, allowing us
241 to use classic conjugate gradient method to find the solution. The convergence rate of this
242 linear inversion may be accelerated using pseudo-Hessian preconditioning. Let us remark
243 that the pseudo-Hessian preconditioner needs to be built only once and can then be reused
244 through all iterations, because the background field remains the same in linearized waveform
245 inversion. The final algorithm to do linearized waveform inversion will be preconditioned
246 conjugate gradient algorithm (Saad, 2003, algorithm 9.1).

247 It is worth noting that linearized waveform inversion is a specific form of least-squares
248 migration (LSM). However, LSM does not necessarily follow the same linearization as in this
249 work. In classic LSM approaches, one can invert a migration image in least-squares sense
250 using any different definitions of the image, as long as a linear dependence between the image
251 and the scatter field is established, while the adjoint operator of the scattering equation can
252 be constructed, see for example (Zhang et al., 2015). Another typical example is the prestack
253 LSRTM code released by the author a decade ago in `Madagascar`[1] open software, which relies
254 on an image defined via cross-correlation between source and receiver wavefields. However,
255 this kind of LSRTM does not bridge nonlinear FWI and LSRTM in a consistent manner: the
256 inverted image is only a structural representation of the subsurface interfaces, inconsistent
257 with definition of reflectivity. The scattering equation of above linearized waveform inversion
258 defines the image as the relative change of the physical parameters. This makes the image of
259 impedance $\delta(\ln I_p) = \delta I_p/I_p$ quantitatively consistent with the definition of the reflectivity

---

[1] `https://github.com/ahay/src/blob/master/user/pyang/Mlsprtm2d.c`

$_{260}$ $(I_2 - I_1)/(I_2 + I_1)$, where I_1 and I_2 are the impedance of two layers around an interface

$_{261}$ ($\delta I_p = I_2 - I_1$ and $I_p = (I_1 + I_2)/2$ are the jump and the average/background impedance at

$_{262}$ the interface, respectively).

$_{263}$ Reflection waveform inversion and its extension

$_{264}$ Reflection waveform inversion

$_{265}$ A practical scenario is that only near offset seismic data are acquired. This challenges

$_{266}$ conventional least-squares FWI to work, as it mainly relies on the diving waves at far offset

$_{267}$ to update the velocity model at depth. RWI has been proposed to deal with such difficulties.

$_{268}$ The key point of RWI is to invert a background model using reflection data, where the low

$_{269}$ wavenumber information is coming from the reflected waves along the wavepath thanks to

$_{270}$ the embedding of interfaces. The RWI misfit functional reads

$$J(m, \delta m) = \frac{1}{2}\|R\delta u - \delta d\|^2, \tag{37}$$

under the constraint of two equations (one for background field and the other for total field)

$$A(m)u = f, \tag{38a}$$

$$A(m + \delta m)(u + \delta u) = f. \tag{38b}$$

$_{271}$ The Lagrangian functional again translated the above constrained optimization into the

$_{272}$ following unconstrained optimization

$$\mathcal{L}(m, \delta m, \lambda_1, \lambda_2, u, \delta u) = \frac{1}{2}\|R\delta u - \delta d\|^2$$
$$+ \langle \lambda_1, A(m)u - f \rangle_{X \times [0,T]} + \langle \lambda_2, A(m + \delta m)(u + \delta u) - f \rangle_{X \times [0,T]} \tag{39}$$

18

where λ_1 and λ_2 are Lagrange multipliers. The chain rule yields

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial m} =& \frac{\partial \mathcal{L}}{\partial \lambda_1}\frac{\partial \lambda_1}{\partial m} + \frac{\partial \mathcal{L}}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial m} + \frac{\partial \mathcal{L}}{\partial u}\frac{\partial u}{\partial m} + \frac{\partial \mathcal{L}}{\partial \delta u}\frac{\partial \delta u}{\partial m} \\
&+ \langle \lambda_1, \frac{\partial A(m)}{\partial m} u \rangle_{[0,T]} + \langle \lambda_2, \frac{\partial A(m+\delta m)}{\partial m}(u+\delta u) \rangle_{[0,T]}
\end{aligned}
\tag{40a}
$$

$$
\frac{\partial \mathcal{L}}{\partial \delta m} = \frac{\partial \mathcal{L}}{\partial \lambda_2}\frac{\partial \lambda_2}{\partial \delta m} + \frac{\partial \mathcal{L}}{\partial \delta u}\frac{\partial \delta u}{\partial \delta m} + \langle \lambda_2, \frac{\partial A(m+\delta m)}{\partial \delta m} u \rangle_{[0,T]}
\tag{40b}
$$

At the saddle point, we require all the first derivative of $\mathcal{L}$ with respect to state variables vanish ($\partial \mathcal{L}/\partial u = \partial \mathcal{L}/\partial \delta u = \partial \mathcal{L}/\partial \lambda_1 = \partial \mathcal{L}/\partial \lambda_2 = 0$). This imply the following two adjoint equations must be satisfied:

$$
A^{\mathrm{H}}(m)\lambda_1 = - R^{\mathrm{H}}(\delta d - R\delta u),
\tag{41a}
$$

$$
A^{\mathrm{H}}(m+\delta m)\lambda_2 = R^{\mathrm{H}}(\delta d - R\delta u).
\tag{41b}
$$

where the source terms for simulating λ_1 and λ_2 are the same except an opposite sign. The gradient of the misfit functional with respect to m and δm are respectively

$$
\frac{\partial J}{\partial m} = \langle \lambda_1, \frac{\partial A(m)}{\partial m} u \rangle_{[0,T]} + \langle \lambda_2, \frac{\partial A(m+\delta m)}{\partial m}(u+\delta u) \rangle_{[0,T]},
\tag{42a}
$$

$$
\frac{\partial J}{\partial \delta m} = \langle \lambda_2, \frac{\partial A(m+\delta m)}{\partial \delta m} u \rangle_{[0,T]}.
\tag{42b}
$$

Joint full waveform inversion

Most of the modern seismic acquisition system acquires both near and far offset data. This motivates Zhou et al. (2015) to develop a joint full waveform inversion approach, to combine both diving waves and reflections. The misfit functional of JFWI is written as

$$
J(m, \delta m) = \frac{1}{2}(\|R\delta u - \delta d\|^2 + \|W(Ru - d_0)\|^2),
\tag{43}
$$

where both the mismatch of reflection data δd and the early arrivals d_0 are added into the misfit. An additional weighting matrix W is introduced to balance the weights of the two parts. This weighting may be adhoc but is necessary in that the energy of early arrivals is much larger than the one of reflections. The Lagrangian of JFWI becomes

$$
\begin{aligned}
\mathcal{L}(m, \delta m, \lambda_1, \lambda_2, u, \delta u) =& \frac{1}{2}\|R\delta u - \delta d\|^2 + \frac{1}{2}\|W(Ru - d_0)\|^2 \\
& + \langle \lambda_1, A(m)u - f \rangle_{X \times [0,T]} + \langle \lambda_2, A(m + \delta m)(u + \delta u) - f \rangle_{X \times [0,T]}
\end{aligned}
\tag{44}
$$

Application of first order optimality condition leads to the adjoint equation

$$
A^{\mathrm{H}}(m)\lambda_1 = - R^{\mathrm{H}}(\delta d - R\delta u) + R^{\mathrm{H}}W^{\mathrm{H}}W(d_0 - Ru),
\tag{45a}
$$

$$
A^{\mathrm{H}}(m + \delta m)\lambda_2 = R^{\mathrm{H}}(\delta d - R\delta u).
\tag{45b}
$$

and the gradient of the misfit with respect to m and δm of the same form as (42a) and (42b).

Equations (45b) and (41b) are of exactly the same form. Compared with equation (41a), the only difference in equation (45a) is the additional contribution of diving wave mismatch in the misfit functional. These similarities imply that JFWI and RWI should follow a similar computer implementation. In order to compute $\partial J / \partial m$, two different adjoint simulations must be performed: the one for λ_1 must be done within the unperturbed media m, while the other for λ_2 must be done within the perturbed media $m + \delta m$.

The above derivation of RWI and JFWI leads to a multiparameter nonlinear inversion strategy to simultaneously update the background media m and its perturbation δm. Such a scheme is very likely to be instable due to dramatic difference of the data sensitivity between m and δm. Researchers therefore turn to a variable projection approach, to fix one parameter when updating the other, so that the parameters are updated in an alternating mode.

Note that no linearization has been introduced in the total field equation (38b). Under the approximation of Born linearization in equation (23), equation (42b) reduces to equation

295 (34) because

$$\frac{\partial A(m + \delta m)}{\partial \delta m} \approx \frac{\partial}{\partial \delta m}(A(m) + \delta m \cdot \partial_m A(m)) = \frac{\partial A(m)}{\partial m}. \quad (46)$$

296 The inversion of model perturbation δm under a fixed background m is nothing else than

297 least-squares RTM. Since the optimization is now linear instead of nonlinear, thus the stepsize

298 in each iteration can be determined explicitly and line search is no longer needed.

299   Even so, both RWI and JFWI still require two nested loops for optimization, the outer

300 loop updates the background model while the inner loop updates the model perturbation via

301 LSM. Such a scheme can be expensive and may be unaffordable for 3D field data applications.

302 A direction in the future will be the efficient implementation of RWI and JFWI, combined

303 with good preconditioner to reduce the cost of inner loops.

Sensitivity kernels

305 To understand the similarities and distinctions among these methods, we perform a study

306 of monoparameter sensitivity kernel analysis. Figure 2b displays a 2D homogeneous velocity

307 model with one interface perturbation for this experiment. We deploy the source and the

308 receiver at 30 m depth. Assume a constant density. The seismic data simulated from such

309 a model has been plotted in Figure 2b: both the direct arrivals and the reflections due to

310 the interface can be clearly identified. It can be seen that the early direct waves have larger

311 amplitude than the reflected waves arrived later.

312   We consider the source and receiver pair placed at 100 m and 400 m in horizontal direc-

313 tion. Figure 3a is the sensitivity kernel of velocity by injecting the data residual of both early

314 arrivals and reflections at the receiver location, starting from a smooth model of a different

315 velocity from the background as in Figure 2a. The first Fresnel zone and the banana dough-

316 nut (the migration isochrone) can be clearly identified in Figure 2a. Muting the reflection

317 when injecting the data residual leads to the sensitivity kernel including only the first Fresnel

318 zone in Figure 3b. Injecting only data residual of the reflections gives a sensitivity kernel

including only the migration isochrone in Figure 3c. Embedding a reflector in the starting model gives the sensitivity kernel in Figure 3d, in which the low-frequency rabbit ears along the wavepath connecting the source, the reflector and the receiver can be clearly seen.

Comparing Figures 3b and 3c, we also see that the low frequency energy from first Fresnel zone demands a long offset between source and receiver in order to reach deep target. Figure 3d reveals that in the absence of long offset data, the presence of reflector in the deep part of the model can still generate low frequency rabbit ears, to build up the low frequency macromodel. The low frequency energy built from the reflector goes beyond the penetration depth that the first Fresnel zone can reach.

Compared with classic FWI sensitivities in Figure 3, we also plotted the sensitivities of RWI and JFWI in Figure 4 and 5. Both of them remove the migration isochrone. RWI leaves only the low-frequency rabbit ears connecting the source, the reflector and the receiver, while JFWI reserves also the first Fresnel zone.

Conlcusion

The adjoint state method supplies a generic framework to formulate nonlinear FWI, LSM, RWI and JFWI into a unified formulation. Thanks to such a consistent mathematical treatment, the gradient expression of FWI and LSM turns out to be of the same form, despite the linearization of the wave equation based on a smooth background. RWI takes advantages of the low-frequency sensitivity generated due to the presence of reflector to provide low-wavenumber update of the model, while JFWI combines also the first Fresnel zone, using both reflection and refractions.

The chain rule allows us to switch between various parameterizations. The pseudo-Hessian preconditioning has been developed under acoustic multiparameter settings. The same preconditioning operator applies to both nonlinear FWI and LSM. The use of logarithmic transformation of medium parameters as the inversion parameters is particularly

important to get rid of artificial ill-conditioning induced by various levels of magnitude. The adoption of log scaling for model perturbation makes the seismic image computed through least-squares migration consistent with reflectivity in physics.

Current formulation illustrates the idea based on acoustic wave equation, which may have a direct usage for FWI to processing marine seismic recordings for exploration purposes. Comprehensive effort is being made to integrate these methodological developments into an open source software for seismic imaging. The extension of this work from acoustic to elastic regime is straightforward, which can be of interest for seismic tomography in global seismology.

Data and Resources

No data were used in this paper.

Declaration of Competing Interests

The author declares no competing interests.

Acknowledgements

References

Bamberger, A., Chavent, G., Hemon, C., and Lailly, P. (1982). Inversion of normal incidence seismograms. *Geophysics*, 47(5):757–770.

Boehm, C., Hanzich, M., de la Puente, J., and Fichtner, A. (2016). Wavefield compression for adjoint methods in full-waveform inversion. *Geophysics*, 81(6):R385–R397.

Bozdağ, E., Trampert, J., and Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2):845–870.

Brossier, R., Operto, S., and Virieux, J. (2015). Velocity model building from seismic reflection data by full waveform inversion. *Geophysical Prospecting*, 63:354–367.

Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.

Chavent, G., Clément, F., and Gòmez, S. (1994). Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example. *SEG Technical Program Expanded Abstracts 1994*, pages 1179–1182.

Clément, F., Chavent, G., and Gómez, S. (2001). Migration-based traveltime waveform inversion of 2-D simple structures: A synthetic example. *Geophysics*, 66:845–860.

Douma, H., Yingst, D., Vasconcelos, I., and Tromp, J. (2010). On the connection between artifact filtering in reverse-time migration and adjoint tomography. *Geophysics*, 75(6):S219–S223.

Engquist, B. and Froese, B. D. (2014). Application of the wasserstein metric to seismic signals. *Communications in Mathematical Science*, 12(5):979–988.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, 175:665–685.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2009). Full seismic waveform

tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. P. (2010). Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. *Earth and Planetary Science Lettters*, 290(3-4):270–280.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pages 206–220.

Luo, S. and Sava, P. (2011). A deconvolution-based objective function for wave-equation inversion. *SEG Technical Program Expanded Abstracts*, 30(1):2788–2792.

Luo, Y. and Schuster, G. T. (1991). Wave-equation traveltime inversion. *Geophysics*, 56(5):645–653.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, 205:345–377.

Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.

Nemeth, T., Wu, C., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation. *Geophysical Journal International*, 202(2):1362–1391.

Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

Plessix, R. E., Chavent, G., and De Roeck, Y. (1995). Automatic and simultaneous migration velocity analysis and waveform inversion of real data using a MBTT/WBKBJ formulation. In *Expanded Abstracts*, pages 1099–1101. Society of Exploration Geophysics.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, Part I: theory and verification in a physical scale model. *Geophysics*, 64:888–901.

Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia.

Shin, C., Jang, S., and Min, D. J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49:592–606.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.

Sun, W. and Fu, L.-Y. (2013). Two effective approaches to reduce data storage in reverse time migration. *Computers & Geosciences*, 56:69–75.

Symes, W. W. (2007). Reverse time migration with optimal checkpointing. *Geophysics*, 72(5):SM213–SM221.

Symes, W. W. (2008a). Approximate linearized inversion by optimal scaling of prestack depth migration. *Geophysics*, 73(2):R23–R35.

Symes, W. W. (2008b). Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790.

Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.

Tarantola, A. and Valette, B. (1982). Inverse problems = quest for information. *Journal of Geophysics*, 50:159–170.

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.

van Leeuwen, T. and Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation traveltime tomography. *Geophysical Journal International*, 182(3):1383–1394.

Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

Warner, M. and Guasch, L. (2016). Adaptive waveform inversion: Theory. *Geophysics*, 81(6):R429–R445.

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

Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016a). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207(1):129–149.

Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016b). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method. *Geophysics*, 81(6):R349–R362.

Yang, P., Brossier, R., Métivier, L., Virieux, J., and Zhou, W. (2018). A Time-Domain Pre-conditioned Truncated Newton Approach to Visco-acoustic Multiparameter Full Waveform Inversion. *SIAM Journal on Scientific Computing*, 40(4):B1101–B1130.

Yang, P., Brossier, R., and Virieux, J. (2016c). Wavefield reconstruction from significantly decimated boundaries. *Geophysics*, 80(5):T197–T209.

Zhang, Y., Duan, L., and Xie, Y. (2015). A stable and practical implementation of least-squares reverse time migration. *Geophysics*, 80(1):V23–V31.

Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2015). Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, 202(3):1535–1554.

Zhou, W., Brossier, R., Operto, S., Virieux, J., and Yang, P. (2018). Velocity model building by waveform inversion of early arrivals and reflections: a 2d ocean-bottom-cable study with gas cloud effects. *Geophysics*, 83(2):R141–R157.
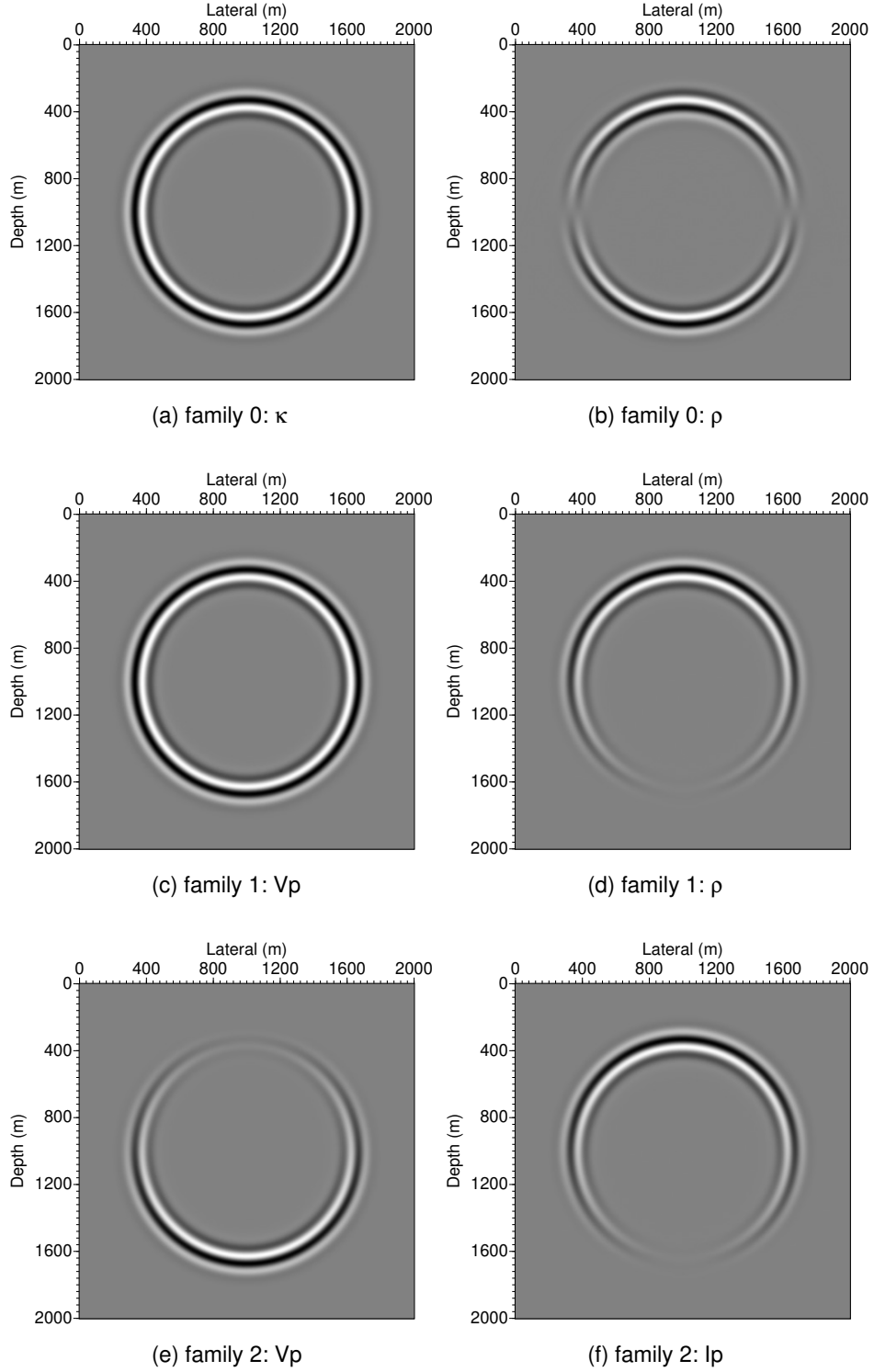
List of Figures

29

Figure 1: The scattering pattern of each parameter under three families of model parameterization: (a) κ and (b) ρ in family 0; (c) V_p and (d) ρ in family 1; (e) V_p and (f) I_p in family 2.
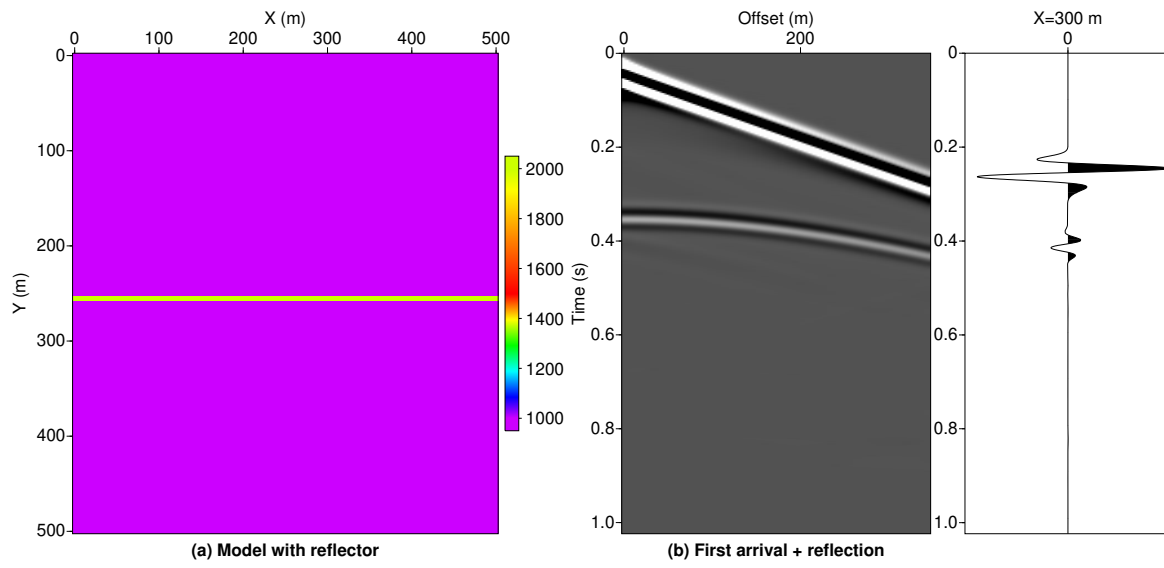
Figure 2: (a) Homogeneous velocity model including a reflector; (b) the simulated seismic data including both direct waves and reflections.
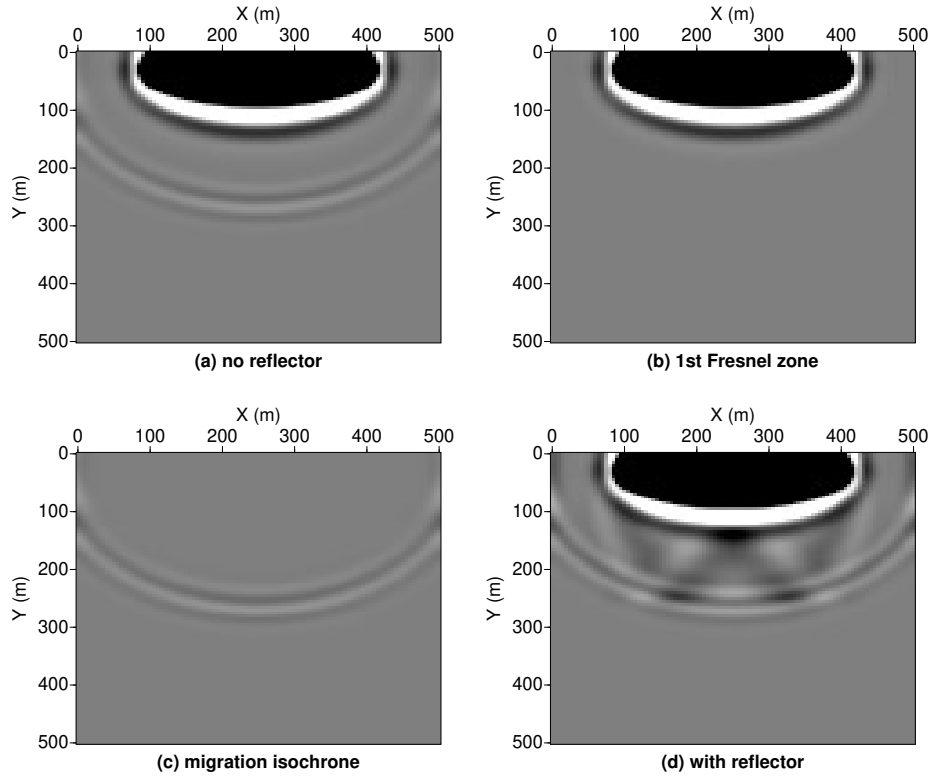
Figure 3: FWI sensitivity kernels corresponding to (a) injecting both early arrival and reflections using a smooth starting model; (b) muting data residual from reflections using smooth background model; (c) muting data residual from early arrivals using smooth initial model; (d) including a reflector in the starting model.
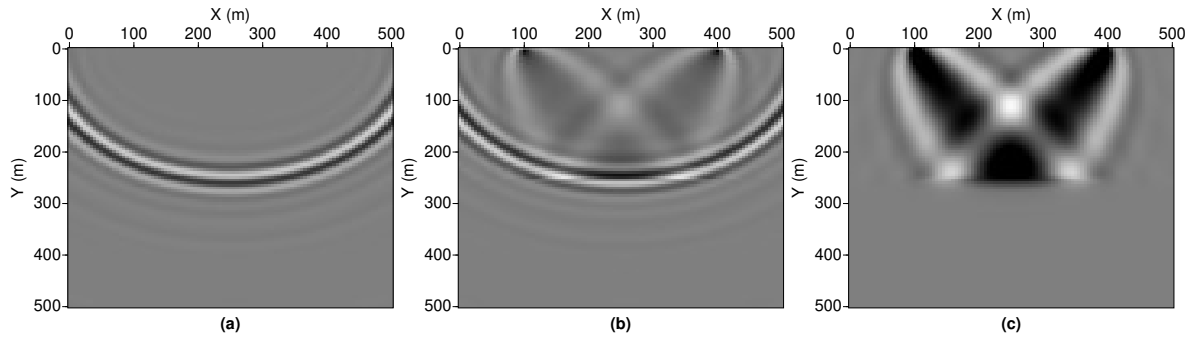
Figure 4: RWI sensitivity kernels corresponding to (a) the first term in equation (41a); (b) the second term in equation (41a); (c) the summation of the first and second terms in equation (41a). Due to the opposite polarity, the migration isochrones in (a) and (b) are cancelled in (c) after summation.
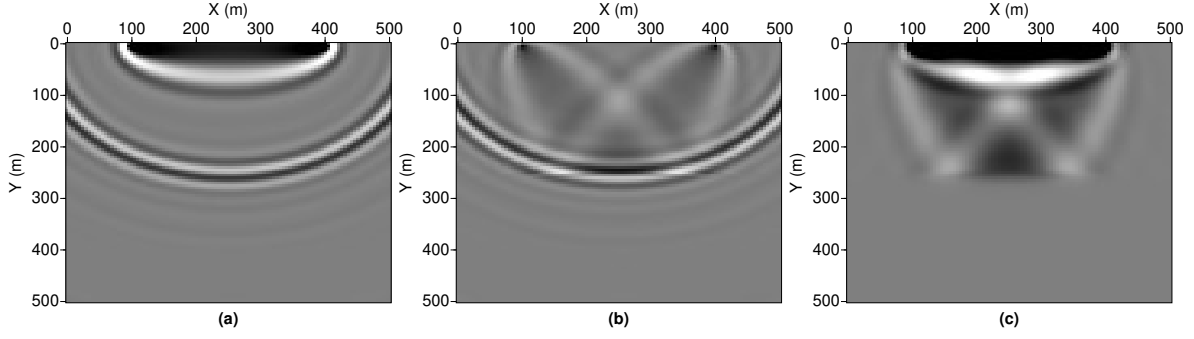
Figure 5: JFWI sensitivity kernels corresponding to (a) the first term in equation (45a); (b) the second term in equation (45a); (c) the summation of the first and second terms in equation (45a). Compared to the first term of RWI, JFWI adds the first Fresnel zone in (a) due to the contribution of early arrivals.