GEOPHYSICS

Volume 82, Issue 5 (September 2017)

Ke Chen and Mauricio D. Sacchi (2017). "Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning
https://doi.org/10.1190/geo2016-0613.1
Technical Papers

Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning

Article History
Received: 24 November 2016
Accepted: 31 May 2017
Published: 27 July 2017

Publication Data
ISSN (print): 0016-8033
ISSN (online): 1942-2156
CODEN: GPYSA7

Keywords
elastic, imaging, least-squares migration, multiparameter, full-waveform inversion

Ke Chen[1] **and Mauricio D. Sacchi**[1]
[1]**University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: ke7@ualberta.ca; msacchi@ualberta.ca.**

Sections: ChooseAbstractINTRODUCTIONTHEORY EXAMPLESDISCUSSIONCONCLUSION ACKNOWLEDGMENTSREFERENCESAP PENDIX ACONNECTIONS BET...CITING ARTICLES

Abstract <<INTRODUCTIONTHEORYEXAMPLESDISCUSSIONCONCLUSIONACKNOWLEDGMENTSREFERENCESAPPENDIX ACONNECTIONS BET...CITING ARTICLES Choose Top of pageAbstract

Time-domain elastic least-squares reverse time migration (LSRTM) is formulated as a linearized elastic full-waveform inversion problem. The elastic Born approximation and elastic reverse time migration (RTM) operators are derived from the time-domain continuous adjoint-state method. Our approach defines P- and S-wave impedance perturbations as unknown elastic images. Our algorithm is obtained using continuous functional analysis in which the problem is discretized at the final stage (optimize-before-discretize approach). The discretized numerical versions of the elastic Born operator and its adjoint (elastic RTM operator) pass the dot-product test. The conjugate gradient least-squares method is used to solve the least-squares migration quadratic optimization problem. In other words, the Hessian operator for elastic LSRTM is implicitly inverted via a matrix-free algorithm that only requires the action of forward and adjoint operators on vectors. The diagonal of the pseudo-Hessian operator is used to design a preconditioning operator to accelerate the convergence of the elastic LSRTM. The elastic LSRTM provides higher resolution images with fewer artifacts and a superior balance of amplitudes when compared with elastic RTM. More important, elastic LSRTM can mitigate crosstalk between the P- and S-wave impedance perturbations given that the off-diagonal elements of the Hessian are attenuated via the inversion.

Permalink: https://doi.org/10.1190/geo2016-0613.1
Keywords: elastic, imaging, least-squares migration, multiparameter, full-waveform inversion

INTRODUCTION

Acoustic reverse time migration (RTM) was initially proposed for poststack seismic data migration (Baysal et al., 1983; McMechan, 1983; Whitmore, 1983). Prestack RTM was implemented in the shot-profile domain by either applying the excitation-time imaging condition (Chang and McMechan, 1986) or the crosscorrelation imaging condition (Etgen, 1986). RTM uses the two-way wave equation for extrapolating wavefields into the interior of the earth. RTM can handle steep and complex geologic structures, such as sedimentary areas with salt inclusions (Etgen et al., 2009). Lailly (1983) points out that computing the gradient in full-waveform inversion (FWI) is equivalent to applying the prestack RTM operator on data residuals. From then, RTM has been connected to the adjoint-state method that is used in FWI (Tromp et al., 2005; Plessix, 2006; Douma et al., 2010).

Acoustic methods approximate the elastic solid earth by a fluid. However, S-waves also convey subsurface rock-property information. S-waves can be recorded by multicomponent sensors (Hardage et al., 2011). The multicomponent seismic data can be used for elastic seismic imaging. Elastic RTM was developed by Sun and McMechan (1986) and Chang and McMechan (1987) with the excitation-time imaging condition. The elastic data record is back-extrapolated using the elastic wave equation, and the image time is computed by ray tracing using a P-wave velocity model. In these methods, P- and S-images are not explicitly separated. Instead, they compute the so-called vertical and horizontal images. The latter could impede the proper interpretation of multicomponent images. Sun and McMechan (2001) and Sun et al. (2004) propose another method, in which the elastic data are first back-propagated by the elastic wave equation and then a Helmholtz decomposition (Dellinger and Etgen, 1990) is used to separate the P- and S-wave modes at a predefined datum. The separated P- and S-wave data are extrapolated upward to the surface of the earth via the acoustic wave equation. Then, the separated P- and S-wave data are injected into two acoustic RTM algorithms that adopt an excitation-time imaging condition. In a similar vein, Yan and Sava (2008) propose to perform forward and backward extrapolation of wavefields via elastic wave equation and to apply Helmholtz decomposition to separate P- and S-wavefields followed by the crosscorrelation imaging condition (Lu et al., 2009). One drawback of methods based on Helmholtz wave-mode decomposition is that the P-S images will have polarity reversals. The latter will prevent stacking of individual shot contributions. Polarity reversal correction strategies for elastic RTM methods that adopt Helmholtz decomposition have been proposed (Du et al., 2012; Duan and Sava, 2015). Early work in the field of FWI (Tarantola, 1986) recognized that the gradient in each iteration of an elastic FWI algorithm is indeed a prestack elastic RTM operator applied to data residuals. This idea is adopted by Luo et al. (2009) and Zhu et al. (2009), who propose elastic RTM algorithms based on FWI sensitivity kernels. In these methods, the elastic model perturbations on the background model are defined as “elastic images.”

A migration algorithm can be described as the adjoint of the forward Born modeling operator. In this case, the migrated image can be considered as a blurred version of subsurface model perturbations. The deblurring operator is the inverse of the Hessian of the imaging problem, which is defined, for linearized inversion, as the cascade of the forward and adjoint operators. Several strategies have been developed to approximate the Hessian to deblur seismic images. For instance, prestack least-squares Kirchhoff migration was initially formulated to retrieve acoustic velocity perturbation given a background velocity model (Tarantola, 1984b; Lambare et al., 1992). Nemeth et al. (1999) implement least-squares Kirchhoff migration for migrating incomplete reflection seismic data. Least-squares Fourier finite-difference one-way wave-equation migration was formulated by Rickett (2003) and then used by Tang (2009) for blended seismic data migration. Least-squares reverse time migration (LSRTM) was investigated for estimating P-wave impedance perturbation and P-wave velocity perturbation (Bourgeois et al., 1989) under the name “linearized inversion.” Ostmo et al. (2002) implement acoustic LSRTM in the frequency domain under the name of

"linearized waveform inversion." In recent years, LSRTM was further developed for high-resolution true amplitude imaging (Dong et al., 2012; Zhang et al., 2015; Yao and Jakubowicz, 2016), migration of multisource blended seismic data (Dai et al., 2012; Xue et al., 2016), and for imaging with multiples (Zhang and Schuster, 2014; Wong et al., 2015). The acoustic slowness perturbation or velocity perturbation represents the "image" or reflectivity model. We would like to classify the above-mentioned least-squares migration techniques as linearized waveform inversion. These methods, in general, invert for an image that is proportional to an averaged subsurface reflectivity. Least-squares migration can also be implemented in an extended domain to produce an image volume that depends on redundant parameters (Symes, 2008). This idea was implemented by Kuehl and Sacchi (2003) using the survey-sinking approach (Claerbout, 1985). The technique was modified to process 3D field data via the constant azimuth approximation by Wang et al. (2005) and to include sparsity constraints to increase vertical resolution by Wang and Sacchi (2007). Moreover, Kaplan et al. (2010) derive least-squares split-step migration for extended shot-domain image inversion. The latter was also applied for the migration of blended seismic data (Cheng et al., 2016). Similarly, Dai and Schuster (2013) implement LSRTM in the extended plane-wave domain for blended seismic data. Finally, we also mention that Hou and Symes (2016) and Huang et al. (2016) implement LSRTM in the extended subsurface offset domain and extended shot domain, respectively. We can classify this type of least-squares migration algorithms as extended least-squares migration (Symes, 2008), in which an extended reflectivity volume is produced. Least-squares migration has also been formulated in the image domain that the inverse of the Hessian is approximated via various strategies (Rickett, 2003; Guitton, 2004; Fletcher et al., 2016; Wang et al., 2016). Image-domain least-squares migration requires lower computational cost than data-domain least-squares migration methods.

The aforementioned least-squares migration methods are based on the acoustic approximation. Land data and ocean-bottom data record both P- and S-waves. The geophysical community has investigated several elastic least-squares migration algorithms. Elastic least-squares ray-Born migration/inversion is implemented by Beydoun and Mendes (1989) and Jin et al. (1992) in heterogeneous media for multicomponent seismic data. Tura and Johnson (1993) discuss an elastic least-squares migration/inversion method in the f-k domain for homogeneous background media. Anikiev et al. (2013) investigate the decoupling of parameters for frequency-domain elastic LSRTM for the case of a point scatterer in a homogeneous background model. In these studies, the elastic parameter perturbations are estimated and defined as elastic images. Recently, Stanton and Sacchi (2015, 2017) and Xu et al. (2016) use Helmholtz decomposition (Dellinger and Etgen, 1990) for elastic least-squares split-step and reverse time migration to estimate elastic reflectivity volumes in the extended domain, respectively.

This paper formulates time-domain elastic LSRTM as a linearized elastic FWI problem (Chen and Sacchi, 2016). We discussed the relationship between LSRTM and Gauss-Newton FWI in Appendix A. The elastic Born approximation and elastic RTM operators are derived via a time-domain continuous adjoint-state method. The adjoint-state equation system is equivalent to the state equation system with one difference: The explosive source is replaced by an adjoint source. In our work, P- and S-wave impedance perturbations are defined as elastic images. The two terms "model perturbation" and "image" are used interchangeably. All derivations are in a continuous functional form; the problem is discretized after developing the algorithm. This is the so-called optimize-before-discretize approach (Borzi and Schulz, 2012). The discretized numerical version of elastic Born operator and its adjoint (elastic RTM operator) pass the dot-product test (Mora, 1987; Claerbout, 1992). The latter allows the use of the conjugate gradient least-squares (CGLS) algorithm (Hestenes and Stiefel, 1952; Paige and Saunders, 1982) to solve the least-squares migration optimization problem. In other words, the Hessian operator is implicitly inverted via a matrix-free algorithm that requires only the forward and adjoint operator applied to vectors. The diagonal of the pseudo-Hessian operator (Shin et al., 2001a) was adapted for the elastic case and used as a preconditioning operator to accelerate the convergence of the elastic LSRTM algorithm. Our elastic LSRTM yields higher resolution images with fewer artifacts and more balanced amplitudes than elastic RTM. Moreover, elastic LSRTM can reduce the multiparameter crosstalk given that the off-diagonal terms of the Hessian operator are attenuated via the least-squares inversion. Finally, the adopted pseudo-Hessian preconditioning strategy accelerates the convergence of our algorithm and improves the amplitude responses of the P- and S-wave impedance perturbation images.

Our research originated from a detailed study of the time-domain elastic Gauss-Newton FWI and its sensitivity kernels (Epanomeritakis et al., 2008). Recently, we noticed that a similar line of work is also presented by Feng and Schuster (2016). Our research complements the work of Feng and Schuster (2016). We stress the following differences between our work and the aforementioned work and summarize our contributions in the next paragraph. First, we formally derived an adjoint-state equation system via the continuous adjoint-state method for the first-order velocity-stress elastic wave-equation system. The adjoint-state equation system is equivalent to the state equation system after a variable transformation (Vigh et al., 2014). This allows us to reuse our forward-modeling code to compute the receiver-side wavefield (adjoint-state variable). The differences are the replacement of the explosive source by an adjoint source and the back-propagation of the adjoint source from final time to zero time. Second, we carefully discretized the continuous functional forms of the elastic Born operator and the elastic RTM operator to guarantee that the operators pass the dot-product test (Mora, 1987). This allows us to adopt the CGLS algorithm (Hestenes and Stiefel, 1952; Paige and Saunders, 1982) to solve the least-squares migration optimization problem. The advantage of the CGLS algorithm is that the step size of the method is analytically calculated. In other words,

one does not need to compute the step size via line search (Dong et al., 2012) as it often done when the forward and adjoint operators do not satisfy the numerical condition for adjointness. Third, we have also investigated the structure of the multiparameter Hessian operator for elastic LSRTM to design a preconditioning strategy. We adapted the pseudo-Hessian (Shin et al., 2001a) to the time-domain elastic case and derived the equations of the Hessian and pseudo-Hessian for elastic parameters. We adopted the diagonal of the pseudo-Hessian to precondition the CGLS algorithm.

This paper is organized as follows: First, we describe the system of equations that we have adopted to forward model elastic wavefields. Then, we introduce linearized forward modeling by adopting the elastic Born approximation. Subsequently, we derive the adjoint operator (elastic RTM operator) of the linearized forward-modeling operator. We discuss the numerical adjointness of forward and adjoint operators and propose to solve the elastic LSRTM via the CGLS algorithm. Furthermore, we discuss preconditioning as a strategy to accelerate the convergence of the CGLS algorithm. In the last section, we provide numerical examples that permit us to evaluate the performance of the proposed algorithm. The first numerical example is a simple elastic Camembert model. Our second numerical example entails adopting the elastic Marmousi2 model for additional tests.

Heterogeneous, isotropic elastic wave equation

We assume a heterogeneous, isotropic elastic earth media. Propagation of the seismic wave is governed by first-order linear partial differential equations (Virieux, 1986; Vigh et al., 2014)

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \partial_t \mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \mathbf{u} = \mathbf{f}, \tag{1}$$

where

$$\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_x \\ v_z \end{pmatrix}, \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \end{pmatrix}, \mathbf{f} = \begin{pmatrix} \mathbf{0} \\ \mathbf{f}_\sigma \end{pmatrix}, \mathbf{C} = \begin{pmatrix} \lambda+2\mu & \lambda & 0 \\ \lambda & \lambda+2\mu & 0 \\ 0 & 0 & \mu \end{pmatrix}, \tag{2}$$

$$\mathbf{D} = \begin{pmatrix} \partial_x & 0 \\ 0 & \partial_z \\ \partial_z & \partial_x \end{pmatrix},$$

with zero initial condition $\mathbf{u}|^{t=0} = \mathbf{0}$ and appropriate boundary conditions. The vector $\mathbf{u}$ denotes the wavefield with $\mathbf{v}$ being the vector particle velocity field and $\boldsymbol{\sigma}$ being the stress field. Similarly, ρ indicates the density and $\mathbf{C}$ is being the isotropic elastic tensor in Voigt notation with λ and μ the Lamé parameters. The matrix $\mathbf{D}$ is a collection of spatial differential operators, and $\mathbf{f}^\sigma$ is the explosive source term. Finally, $\mathbf{I}$ is the identity matrix. In the wave equation, we dropped the dependence on spatial and temporal coordinates $\mathbf{x}$ and t of our variables to make the notations concise, but we understand that $\mathbf{v}=\mathbf{v}(\mathbf{x},t)$, $\lambda=\lambda(\mathbf{x})$, etc. The elastic wave equation is the state equation of the elastic parameter inversion problem when it is regarded as optimal control problem (Lions, 1971; Plessix, 2006). Abstractly, the elastic wave equation 1 can be written in functional form as follows:

$$\mathbf{S}\mathbf{u} = \mathbf{f}, \tag{3}$$

where $\mathbf{u}$ is the wavefield vector in space U, $\mathbf{f}$ is the source vector in space F, $\mathrm{S}=\mathrm{S}(\mathbf{m})$ $(\mathrm{S}:\mathrm{U}\rightarrow\mathrm{F})$ is the wave-equation operator with initial conditions and boundary conditions, and $\mathbf{m}=(\rho,\lambda,\mu)^T$ denotes the model parameter vector in space M. The solution of the wave equation can be abstractly written as

$$\mathbf{u} = \mathrm{S}^{-1}\mathbf{f} = \mathrm{G}\mathbf{f}, \tag{4}$$

where $\mathrm{G}=\mathrm{S}^{-1}$ is the inverse of wave equation operator S called the Green's operator (Tarantola, 1988). The Green's operator is an integral operator with the integration kernel given by the Green's function of the wave equation

(Tarantola, 1988). The wavefield $\mathbf{u}$ is linear in the source term $\mathbf{f}$, but it is nonlinear in the model $\mathbf{m}$. If the source term is assumed known, $\mathbf{u}$ can be regarded as a nonlinear function of the model parameters $\mathbf{m}$

$$\mathbf{u}=\mathbf{u}(\mathbf{m}). \tag{5}$$

In a general heterogeneous media, there is no analytic solution for $\mathbf{u}$ given $\mathbf{m}$. A numerical method must be used to solve the forward problem. In this paper, a time-domain staggered-grid finite-difference (FD) scheme (Virieux, 1986; Levander, 1988) is used to discretize the continuous form elastic wave equation 1. The unsplit convolutional perfectly matched layer (C-PML) method is used to absorb incident waves on artificial computational boundaries (Komatitsch and Martin, 2007). In our FD code, we adopted a second-order centerd difference scheme in time and a selectable order staggered difference scheme in space. In seismic exploration, the wavefield is sampled at the surface of the earth by a finite number of receivers

$$\mathbf{d}(\mathbf{m})=\mathrm{T}\mathbf{u}(\mathbf{m}), \tag{6}$$

where $\mathbf{d}$ is the recorded full-waveform seismic data and operator T represents the sampling operator that extracts the wavefield at the receivers positions.

The linearized forward problem: Elastic Born approximation

The relationship between the seismic data $\mathbf{d}$ and model parameters $\mathbf{m}$ is nonlinear as discussed above. An expansion in terms of Taylor series can be used to linearize the nonlinear forward problem $\mathbf{d}=\mathbf{d}(\mathbf{m})$. A perturbation of the model parameters $\mathbf{m}\rightarrow\mathbf{m}+\delta\mathbf{m}$ leads to a perturbation of the seismic data $\mathbf{d}(\mathbf{m})\rightarrow\mathbf{d}(\mathbf{m}+\delta\mathbf{m})$

$$\mathbf{d}(\mathbf{m}+\delta\mathbf{m})=\mathbf{d}(\mathbf{m})+\partial\mathbf{d}\partial\mathbf{m}\delta\mathbf{m}+\mathrm{O}(\|\delta\mathbf{m}\|^2), \tag{7}$$

where $\mathbf{m}$ is the background model, $\delta\mathbf{m}$ is the model perturbation, $\mathbf{d}(\mathbf{m})$ is the seismic data associated with propagation in the background model, $\mathbf{d}(\mathbf{m}+\delta\mathbf{m})$ is the seismic data associated with propagation in the perturbed model, and the linear operator $\partial\mathbf{d}/\partial\mathbf{m}=([\partial\mathbf{d}/\partial\rho],[\partial\mathbf{d}/\partial\lambda],[\partial\mathbf{d}/\partial\mu])$ is the Fréchet derivative or Jacobian matrix of $\mathbf{d}$. The second and higher order terms in the Taylor series are dropped resulting in the first-order Born approximation

$$\delta\mathbf{d}=\mathrm{L}\delta\mathbf{m}=\partial\mathbf{d}\partial\mathbf{m}\delta\mathbf{m}=\mathrm{T}\partial\mathbf{u}\partial\mathbf{m}\delta\mathbf{m}, \tag{8}$$

where Fréchet derivative $\mathrm{L}=\partial\mathbf{d}/\partial\mathbf{m}$ is the Born modeling operator, $\delta\mathbf{d}$ is the first-order scattered seismic data $\delta\mathbf{d}\approx\mathbf{d}(\mathbf{m}+\delta\mathbf{m})-\mathbf{d}(\mathbf{m})$, and the linear operator $\partial\mathbf{u}/\partial\mathbf{m}$ is the Fréchet derivative or Jacobian matrix of $\mathbf{u}$. The Fréchet derivative is prohibitively expensive to compute explicitly. Alternatively, the adjoint-state method is used to compute the action of the Fréchet derivative on vectors. For this purpose, we first differentiate the wave equation 3with respect to $\mathbf{m}$ (Fichtner, 2010; Fichtner and Trampert, 2011)

$$\partial\mathrm{S}\partial\mathbf{m}\mathbf{u}+\mathrm{S}\partial\mathbf{u}\partial\mathbf{m}=\mathbf{0}, \tag{9}$$

where linear operator $\partial\mathrm{S}/\partial\mathbf{m}=([\partial\mathrm{S}/\partial\rho],[\partial\mathrm{S}/\partial\lambda],[\partial\mathrm{S}/\partial\mu])$. The right side of equation becomes zero because the source term does not depend on the model. Then, the Fréchet derivative of $\mathbf{u}$ can be expressed as follows:

$$\partial\mathbf{u}\partial\mathbf{m}=-\mathrm{S}^{-1}\partial\mathrm{S}\partial\mathbf{m}\mathbf{u}=-\mathrm{S}^{-1}\partial\mathrm{S}\partial\mathbf{m}\mathrm{S}^{-1}\mathbf{f}, \tag{10}$$

where the multiplication of two operators follows the rule $\mathrm{AB}\mathbf{v}=\mathrm{A}(\mathrm{B}\mathbf{v})$ (Chen and Lee, 2015). The term $-(\partial\mathrm{S}/\partial\mathbf{m})\mathbf{u}$is the so-called “virtual secondary source,” which is the product of the incident wavefield and $\partial\mathrm{S}/\partial\mathbf{m}$. The operator $\partial\mathrm{S}/\partial\mathbf{m}$ represents the radiation pattern of the virtual secondary source (Pageot et al., 2013). Inserting the expression of Fréchet derivative of $\mathbf{u}$ into equation 8, the Born approximation can be written as

$$\delta\mathbf{d}=\mathrm{L}\delta\mathbf{m}=\partial\mathbf{d}\partial\mathbf{m}\delta\mathbf{m}=-\mathrm{TS}^{-1}\partial\mathrm{S}\partial\mathbf{m}\mathbf{u}\delta\mathbf{m}. \tag{11}$$

Equation 11 indicates that the incident wavefield $\mathbf{u}$ hits the scatterers $\delta\mathbf{m}$, acts as a secondary source and generates the scattered wavefield. The latter is sampled at the surface of the earth by receivers and generates the scattered seismic data. The linearized Born modeling operator L only depends on the smooth background model $\mathbf{m}$and the acquisition geometry.

One can apply the abstract linearized Born approximation equation 11 directly to the first-order velocity-stress elastic wave-equation system 1. Alternatively, perturbing the elastic wave equation will lead to the same result. A perturbation of the model parameters

$$\rho\rightarrow\rho+\delta\rho, \tag{12a}$$

$$\lambda\rightarrow\lambda+\delta\lambda, \tag{12b}$$

$$\mu\rightarrow\mu+\delta\mu, \tag{12c}$$

leads to a perturbation of the wavefield

$$\mathbf{u}\rightarrow\mathbf{u}+\delta\mathbf{u}. \tag{12d}$$

Inserting equation 12 into equation 1, subtracting equation 1, and dropping second and higher order terms leads to the Born approximation for the first-order velocity stress elastic wave-equation system

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \partial\partial t \delta\mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \delta\mathbf{u} = \begin{pmatrix} -\delta\rho\partial\partial t\mathbf{v} \\ \delta\mathbf{C}\mathbf{D}^T\mathbf{v} \end{pmatrix}, \tag{13}$$

where

$$\delta\mathbf{u} = \begin{pmatrix} \delta\mathbf{v} \\ \delta\boldsymbol{\sigma} \end{pmatrix}, \delta\mathbf{v} = \begin{pmatrix} \delta v_x \\ \delta v_z \end{pmatrix}, \delta\boldsymbol{\sigma} = \begin{pmatrix} \delta\sigma_{xx} \\ \delta\sigma_{zz} \\ \delta\sigma_{xz} \end{pmatrix}, \delta\mathbf{C} = \begin{pmatrix} \delta\lambda+2\delta\mu & \delta\lambda & 0 \\ \delta\lambda & \delta\lambda+2\delta\mu & 0 \\ 0 & 0 & \delta\mu \end{pmatrix}, \tag{14}$$

with zero initial condition $\delta\mathbf{u}|^{t=0}=\mathbf{0}$ and appropriate boundary conditions. The vector $\delta\mathbf{u}$ is the scattered wavefield, with $\delta\mathbf{v}$ and $\delta\boldsymbol{\sigma}$ being the scattered particle velocity field and scattered stress field due to model perturbations $\delta\rho$, $\delta\lambda$, and $\delta\mu$, and $\mathbf{v}$ is the incident particle velocity field. The right side of equation 13 is the so-called “secondary source.” The scattered wavefield can be computed using the same FD code that is adopted to compute the source-side incident wavefield in equation 1. The scattered data are obtained by sampling the scattered wavefield at the receiver positions $\delta\mathbf{d}=\mathrm{T}\delta\mathbf{u}$.

The linearized adjoint problem: Elastic reverse time migration

The migration operator is the adjoint of the Born modeling operator that maps from reflection data to model perturbation or image. The adjoint operator of the Born operator satisfies

$$\langle\delta\mathbf{d},\mathrm{L}\delta\mathbf{m}\rangle_\mathrm{D}=\langle\mathrm{L}\dagger\delta\mathbf{d},\delta\mathbf{m}\rangle_\mathrm{M}, \tag{15}$$

where $\dagger$ denotes the adjoint of an operator and $\langle\cdot,\cdot\rangle^\mathrm{D}$ and $\langle\cdot,\cdot\rangle^\mathrm{M}$ denote inner products in the data domain and model domain, respectively. The adjoint of Born operator (equation 11) applied to reflection data can be expressed as (Tarantola, 1984a)

$$\delta\mathbf{m}^*=\mathrm{L}\dagger\delta\mathbf{d}=\begin{pmatrix} \partial\mathbf{d}\partial\mathbf{m} \end{pmatrix}\dagger\delta\mathbf{d}=-\begin{pmatrix} \partial\mathrm{S}\partial\mathbf{m}\mathbf{u} \end{pmatrix}\dagger(\mathrm{S}^{-1})\dagger\mathrm{T}\dagger\delta\mathbf{d}=-\begin{pmatrix} \partial\mathrm{S}\partial\mathbf{m}\mathbf{u} \end{pmatrix}\dagger(\mathrm{S}\dagger)^{-1}\mathrm{T}\dagger\delta\mathbf{d}, \tag{16}$$

where $\mathrm{T}^\dagger$ is the adjoint of sampling operator T. Notice that we used the property $(\mathrm{S}^{-1})^\dagger=(\mathrm{S}^\dagger)^{-1}$ in equation 16(Tarantola, 1988). We adopted the symbol $\delta\mathbf{m}^*$ to represent the model perturbation that one can obtain by applying the adjoint operator to data perturbation $\delta\mathbf{d}$. Evidently, the adjoint operator is not equal to the inverse of the linearized forward operator and therefore, $\delta\mathbf{m}^*\neq\delta\mathbf{m}$. To continue with our analysis, we now introduce the adjoint-state variable $\mathbf{p}=(\mathrm{S}^\dagger)^{-1}\mathrm{T}^\dagger\delta\mathbf{d}$. The latter satisfies the “adjoint-state equation” corresponding to the state equation 3

$$\mathrm{S}\dagger\mathbf{p}=\mathrm{T}\dagger\delta\mathbf{d}, \tag{17}$$

where $\mathbf{p}$ is the adjoint-state variable of the state variable $\mathbf{u}$, $\mathrm{S}^\dagger$ is the adjoint wave-equation operator, and $\mathrm{T}^\dagger\delta\mathbf{d}$ is the adjoint source. The model perturbations or images (equation 16) can be expressed as follows:

$$\delta\mathbf{m}^*=-\begin{pmatrix} \partial\mathrm{S}\partial\mathbf{m}\mathbf{u} \end{pmatrix}\dagger\mathbf{p}, \tag{18}$$

or rewritten the implicit inner product over time explicitly

$$\delta\mathbf{m}(\mathbf{x})^*=-\int\left(\partial\mathrm{S}(\mathbf{x},t)\partial\mathbf{m}(\mathbf{x})\mathbf{u}(\mathbf{x},t)\right)\dagger\mathbf{p}(\mathbf{x},t)dt=-\int\mathbf{p}(\mathbf{x},t)^T\left(\partial\mathrm{S}(\mathbf{x},t)\partial\mathbf{m}(\mathbf{x})\mathbf{u}(\mathbf{x},t)\right)dt, \tag{19}$$

where $\mathbf{p}(\mathbf{x},t)^{\dagger}=\mathbf{p}(\mathbf{x},t)^{T}$ was applied. This is the formulation of RTM with the adjoint-state method (Lions, 1971; Tarantola, 1984a, 1988; Tromp et al., 2005). The estimated images for different shots are usually stacked to form a stacked image.

The abstract form adjoint-state equation 17 can be applied to the first-order velocity-stress elastic wave equation system 1

$$\begin{pmatrix}\rho\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{I}\end{pmatrix}\dagger\begin{pmatrix}\partial\partial t&\mathbf{0}\\\mathbf{0}&\partial\partial t\end{pmatrix}\dagger\mathbf{p}-\begin{pmatrix}\mathbf{0}&\mathbf{D}\\\mathbf{C}&\mathbf{D}^{T}\mathbf{0}\end{pmatrix}\dagger\mathbf{p}=\mathrm{T}\dagger\delta\mathbf{d}, \tag{20}$$

with zero final condition $\mathbf{p}|^{t=T}=\mathbf{0}$ and appropriate boundary conditions. The vector $\mathbf{p}=(\boldsymbol{\upsilon},\boldsymbol{\varsigma})^{T}$, where $\boldsymbol{\upsilon}=(\upsilon^{x},\upsilon^{z})^{T}$ is the adjoint-state particle velocity field and $\boldsymbol{\varsigma}=(\varsigma^{xx},\varsigma^{zz},\varsigma^{xz})^{T}$ is the adjoint-state stress field. The reflection data $\delta\mathbf{d}=(\delta\mathbf{d}^{v},\delta\mathbf{d}^{\sigma})^{T}$ act as the adjoint source to the adjoint-state equation. The differential operator is antiself-adjoint $(\partial/\partial x)^{\dagger}=-(\partial/\partial x)$. Consequently, one can write the expression $\mathbf{D}^{\dagger}=-\mathbf{D}^{T}$. Finally, the adjoint-state equation can be rewritten as follows:

$$\begin{pmatrix}\rho\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{I}\end{pmatrix}\begin{pmatrix}-\partial\partial t&\mathbf{0}\\\mathbf{0}&-\partial\partial t\end{pmatrix}\mathbf{p}+\begin{pmatrix}\mathbf{0}&\mathbf{D}\\\mathbf{C}&\mathbf{D}^{T}\mathbf{0}\end{pmatrix}\mathbf{p}=\mathrm{T}\dagger\delta\mathbf{d}. \tag{21}$$

The structure of the adjoint-state equation is slightly different to the structure of the state equation 1. However, the adjoint-state equation can be redefined into a form that resembles the state equation by adopting a transformation of variables (Vigh et al., 2014)

$$\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{C}\end{pmatrix}\mathbf{p}, \tag{22}$$

where $\tilde{\mathbf{p}}=(\tilde{\boldsymbol{\upsilon}},\tilde{\boldsymbol{\varsigma}})^{T}$ is the transformed adjoint-state variable, $\tilde{\boldsymbol{\upsilon}}=(\tilde{\upsilon}^{x},\tilde{\upsilon}^{z})^{T}$ and $\tilde{\boldsymbol{\varsigma}}=(\tilde{\varsigma}^{xx},\tilde{\varsigma}^{zz},\tilde{\varsigma}^{xz})^{T}$. If we multiply both sides of the adjoint-state equation 21 by the transformation matrix, the adjoint-state equation can be rewritten as follows:

$$\begin{pmatrix}\rho\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{I}\end{pmatrix}\begin{pmatrix}-\partial\partial t&\mathbf{0}\\\mathbf{0}&-\partial\partial t\end{pmatrix}\tilde{\mathbf{p}}+\begin{pmatrix}\mathbf{0}&\mathbf{D}\\\mathbf{C}&\mathbf{D}^{T}\mathbf{0}\end{pmatrix}\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{C}\end{pmatrix}\mathrm{T}\dagger\delta\mathbf{d}, \tag{23}$$

where the adjoint-state equation now has the same structure as the state equation 1. Consequently, the FD code adopted to solve the forward equation system 1 and the Born modeling equation system 13 can be reused to compute the adjoint wavefield in equation 23. The only difference is that the source term is replaced by an elastic tensor scaled adjoint source and the FD steps are in time-reversal mode. After computing the transformed adjoint-state variable $\tilde{\mathbf{p}}$, the original adjoint-state variable $\mathbf{p}$ can be retrieved by the inverse transformation

$$\mathbf{p}=\begin{pmatrix}\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{C}^{-1}\end{pmatrix}\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I}&\mathbf{0}\\\mathbf{0}&\mathbf{C}^{-1}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{\upsilon}}\\\tilde{\boldsymbol{\varsigma}}\end{pmatrix}=\begin{pmatrix}\tilde{\boldsymbol{\upsilon}}\\\tilde{\boldsymbol{\upsilon}}\mathbf{C}^{-1}\tilde{\boldsymbol{\varsigma}}\end{pmatrix}, \tag{24}$$

with

$$\mathbf{C}^{-1}=\begin{pmatrix}\lambda+2\mu&\frac{4\lambda\mu+4\mu^{2}}{-\lambda}&\frac{4\lambda\mu+4\mu^{2}}{0}\\\frac{-\lambda}{4\lambda\mu+4\mu^{2}}&\frac{\lambda+2\mu}{4\lambda\mu+4\mu^{2}}&0\\0&0&\frac{1}{\mu}\end{pmatrix}. \tag{25}$$

Using the migration operator formulation given by equation 19, elastic wave equation 1 and the adjoint wave equation 23, and variable transformation equation 24, the adjoint model perturbations or images ($\delta\mathbf{m}^*=(\delta\rho^*,\delta\lambda^*,\delta\mu^*)^T$) for a single shot can be expressed as

$$\delta\rho^*=-\int\dot{\mathbf{v}}\cdot\boldsymbol{\upsilon}dt=-\int\dot{\mathbf{v}}\cdot\tilde{\boldsymbol{\upsilon}}dt=-\int(\dot{v}_x\tilde{\upsilon}_x+\dot{v}_z\tilde{\upsilon}_z)dt,\delta\lambda^*=\int\left(\partial\mathbf{C}\partial\lambda\mathbf{D}^T\mathbf{v}\right)\cdot\boldsymbol{\varsigma}dt=\int\left(\partial\mathbf{C}\partial\lambda\mathbf{C}^{-1}\dot{\boldsymbol{\sigma}}\right)\cdot(\mathbf{C}^{-1}\tilde{\boldsymbol{\varsigma}})dt=\int(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})(\tilde{\varsigma}_{xx}+\tilde{\varsigma}_{zz})4(\lambda+\mu)^2dt,\delta\mu^*=\int\left(\partial\mathbf{C}\partial\mu\mathbf{D}^T\mathbf{v}\right)\cdot\boldsymbol{\varsigma}dt=\int\left(\partial\mathbf{C}\partial\mu\mathbf{C}^{-1}\dot{\boldsymbol{\sigma}}\right)\cdot(\mathbf{C}^{-1}\tilde{\boldsymbol{\varsigma}})dt=\int\Big[\dot{\sigma}_{xz}\tilde{\varsigma}_{xz}\mu^2+(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})(\tilde{\varsigma}_{xx}+\tilde{\varsigma}_{zz})4(\lambda+\mu)^2+(\dot{\sigma}_{xx}-\dot{\sigma}_{zz})(\tilde{\varsigma}_{xx}-\tilde{\varsigma}_{zz})4\mu^2\Big]dt, \quad (26)$$

where the over-dot means the time derivative. The interaction of the forward (state variable) and backward wavefields (adjoint-state variable) requires access to the two wavefields at the same time step. However, these two wavefields are computed in the reverse time direction. Naive methods, such as saving either forward or backward wavefield to disk can be used. However, I/O can degrade computational performance for large-scale problems. In our work, we have adopted the source-wavefield reconstruction method (Gauthier et al., 1986; Dussaud et al., 2008). During the forward simulation of state (source side) wavefield, only the wavefield within the depth of half of the spatial FD operator length on boundaries and the final time snapshots are saved in memory. Then, the state (source side) wavefield is recomputed from the saved wavefield by backward propagation while simultaneously computing the backward adjoint (receiver side) wavefield. This completes the derivation of the elastic Born operator and elastic RTM operator using continuous functional analysis and the time-domain adjoint-state method.

Numerical adjointness of the elastic Born and RTM programs

We derived the elastic Born operator and elastic RTM operator using continuous functional analysis first and then discretize them. This is the so-called optimize-before-discretize approach (Borzi and Schulz, 2012). One must be careful when discretizing the forward and adjoint operators to guarantee that they truly behave like a forward and adjoint pair. In particular, one needs to be attentive to scaling terms, source-injection strategies, and rules for updating particle velocities and stresses. We adopted the dot-product test to numerically evaluate how close the discretize adjoint operator is to the true adjoint of the forward operator (Mora, 1987; Claerbout, 1992; Le, 2016). This implies generating data and model vectors of random numbers $\mathbf{d}^1$ and $\mathbf{m}^2$. Then, we evaluate $\mathbf{d}^2=\mathrm{L}\delta\mathbf{m}^2$ and $\mathbf{m}^1=\mathrm{L}^\dagger\delta\mathbf{d}^1$ and finally, we compute the closeness of inner products via the following expression:

$$\mathrm{err}=\left|\langle\delta\mathbf{d}^1,\delta\mathbf{d}^2\rangle_\mathrm{D}-\langle\delta\mathbf{m}^1,\delta\mathbf{m}^2\rangle_\mathrm{M}\langle\delta\mathbf{d}^1,\delta\mathbf{d}^2\rangle_\mathrm{D}+\langle\delta\mathbf{m}^1,\delta\mathbf{m}^2\rangle_\mathrm{M}\right|. \quad (27)$$

Our code was written in single-precision float data type in C. The relative error of our dot-product test is 10^{-3} for a model that consists of 500×500 samples in space and a single shot that consists of 5000 samples in time and 500 receivers.

Elastic least-squares reverse time migration
CGLS with adjoint-state method

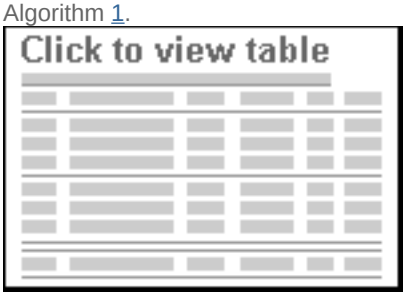
From the above derivation, a properly designed elastic RTM code can be considered equivalent to the adjoint operator of the elastic Born forward-modeling operator. The adjoint operator is an approximation to the inverse operator, in which the Hessian of the linearized inversion problem is replaced by an identity matrix. In other words, the migrated image obtained via the adjoint operator is a blurred version of the true subsurface image. The migrated image, in general, suffers from relative low resolution, unbalanced amplitudes due to geometric spreading and acquisition footprint. Moreover, multiparameter elastic migration will generate crosstalk among different components because different parameters are coupled. To estimate higher resolution images with properly balanced amplitudes and fewer crosstalk and artifacts, the elastic LSRTM is formulated as a quadratic optimization problem, in which one minimizes the following cost function:

$$J(\delta\mathbf{m})=12Ns\sum i=1\|Li\delta\mathbf{m}-\delta\mathbf{d}i\|22, \tag{28}$$

where L^i is the Born approximation operator for the ith shot, $\delta\mathbf{d}^i$ is the reflection data associated to the ith shot gather, $\delta\mathbf{m}$ denotes model perturbation (elastic images), N^s indicates the number of shots, and $\|\cdot\|^2$ indicates the ℓ^2norm of vector. The optimal solution satisfies the condition $[\partial J(\delta\mathbf{m})]/\partial\delta\mathbf{m}=0$. The latter leads to the normal equations

$$\left(Ns\sum i=1L\dagger iLi\right)\delta\mathbf{m}=Ns\sum i=1L\dagger i\delta\mathbf{d}i, \tag{29}$$

where $\sum^{Ns}i=1L^{\dagger i}\delta\mathbf{d}^i$ is the reverse time migrated image and $\sum^{Ns}i=1L^{\dagger i}L^i$ is the Hessian operator for least-squares migration. Last expression indicates that the migrated image is a blurred version of the true parameter perturbations. By applying the inverse of the Hessian operator, the raw migrated images can be deblurred. Explicit forming and inverting the Hessian is prohibitive expensive in terms of computational cost and memory requirements. Instead, we adopt an iterative method: the CGLS (Hestenes and Stiefel, 1952; Paige and Saunders, 1982). The CGLS only requires two operators L^i and $L^{\dagger i}$ that are applied "on the fly" to vectors. The operators are applied on vectors efficiently via the adjoint-state method. The CGLS algorithm can be safely used because our discretized numerical versions of L^i and $L^{\dagger i}$ passed the dot-product test. A series of steps (CGLS iterations) are required to solve the quadratic optimization problem given in equation 28. The CGLS algorithm for elastic LSRTM is summarized as Algorithm 1.



View Larger Version

Algorithm 1. CGLS algorithm.

Parameterization

For reflection data FWI, P-wave impedance, S-wave impedance, and density are the most suitable parameters to invert for. The latter was confirmed by a detailed radiation pattern analysis for different parameterizations of elastic FWI (Tarantola, 1986). We parameterized our elastic LSRTM in terms of P-wave impedance perturbation δI^p and S-wave impedance perturbation δI^s. We have preferred to omit the inversion of the density perturbation because it cannot be properly resolved from reflection data (Tarantola, 1986; Lebrun et al., 2001). We use the following relationships between elastic parameters:

$$\lambda=\rho V2p-2\rho V2s,\mu=\rho V2s,Ip=\rho Vp,Is=\rho Vs, \tag{30}$$

in conjunction with total derivatives to write down the following parameter perturbation transformation:

$$\left(\delta\lambda\delta\mu\right)=\left(2Vp-4Vs02Vs\right)\left(\delta Ip\delta Is\right). \tag{31}$$

We insert the above parameter transformation into the elastic LSRTM formulation (equation 28). The parameter transformation matrix and its adjoint are incorporated into the CGLS algorithm (Algorithm 1). The change of parameters is similar to adding preconditioning to our system of equations. In our experience, inverting for impedance perturbations leads to an algorithm with faster convergence than inverting for the Lamé parameter perturbations.

Preconditioning using multiparameter Hessian

The elastic LSRTM implicitly inverts the Hessian operator via CGLS iterations combined with adjoint-state method; i.e., we apply the Fréchet derivative and its adjoint on vectors via the adjoint-state method on the fly in each CGLS iteration. It does not need to form the Fréchet derivative or Hessian explicitly. However, the CGLS algorithm may need a relatively large number of iterations to converge to an optimal solution. Preconditioning of the gradients is important to accelerate the convergence of CGLS and save computational resources. We investigate the structure of the Hessian operator of the elastic LSRTM and use the diagonal of the pseudo-Hessian for preconditioning (Shin et al., 2001a). The Hessian operator of elastic LSRTM problem can be expressed as

$$\mathbf{H}=N_s\sum_{i=1}\mathbf{L}_i^{\dagger}\mathbf{L}_i=N_s\sum_{i=1}\left(\mathbf{T}\frac{\partial\mathbf{u}_i}{\partial\mathbf{m}}\right)^{\dagger}\left(\mathbf{T}\frac{\partial\mathbf{u}_i}{\partial\mathbf{m}}\right)=N_s\sum_{i=1}\left(\mathbf{T}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_i\right)^{\dagger}\left(\mathbf{T}\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_i\right), \tag{32}$$

where $\mathbf{u}_i$ is the source-side wavefield for ith shot, and recall that $\mathbf{S}^{-1}$ is the Green's operator. If the Green's functions are not saved, explicitly computing the Hessian needs $N_s \cdot N_m$ forward simulations with N_m as the number of model grid points. Using the reciprocity of the Green's function (Tarantola, 1988), the number of forward simulations needed for explicitly computing the Hessian reduces to $N_s \cdot N_g$, where N_g denotes the number of receivers. The diagonal element of the Hessian is the zero-lag autocorrelation of the Fréchet derivative. It accounts for the geometric spreading effect (Shin et al., 2001b). Explicitly computing the diagonal of the Hessian also requires $N_s \cdot N_g$ forward simulations. Shin et al. (2001a) propose to neglect the receiver Green's function to save computation cost. Under this assumption, the Hessian can be simplified to the so-called "pseudo-Hessian":

$$\mathbf{H}=N_s\sum_{i=1}\left(\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_i\right)^{\dagger}\left(\frac{\partial\mathbf{S}}{\partial\mathbf{m}}\mathbf{u}_i\right). \tag{33}$$

For multiparameter problem, the Hessian and pseudo-Hessian are blockwise. The pseudo-Hessian for Lamé parameters is given by

$$\mathbf{H}_L=\begin{pmatrix}\mathbf{H}_{\lambda\lambda} & \mathbf{H}_{\lambda\mu}\\ \mathbf{H}_{\mu\lambda} & \mathbf{H}_{\mu\mu}\end{pmatrix}. \tag{34}$$

Using the pseudo-Hessian operator expression (equation 33) and the elastic wave equation 1, the diagonal blocks can be expressed as

$$\mathbf{H}_{\lambda\lambda}(\mathbf{x},\mathbf{x}')=N_s\sum_{i=1}\int\left(\frac{\partial\mathbf{C}}{\partial\lambda}\mathbf{D}^T\mathbf{v}(\mathbf{x})\right)\cdot\left(\frac{\partial\mathbf{C}}{\partial\lambda}\mathbf{D}^T\mathbf{v}(\mathbf{x}')\right)dt,\mathbf{H}_{\mu\mu}(\mathbf{x},\mathbf{x}')=N_s\sum_{i=1}\int\left(\frac{\partial\mathbf{C}}{\partial\mu}\mathbf{D}^T\mathbf{v}(\mathbf{x})\right)\cdot\left(\frac{\partial\mathbf{C}}{\partial\mu}\mathbf{D}^T\mathbf{v}(\mathbf{x}')\right)dt, \tag{35}$$

where the dependence on shot index i of particle velocity field is omitted to avoid notation clutter. The diagonal terms of the pseudo-Hessian ($\mathbf{x}=\mathbf{x}'$) for Lamé parameters can be expressed as follows:

$$H_{\lambda\lambda}(\mathbf{x},\mathbf{x})=N_s\sum_{i=1}\int(\dot{}\sigma_{xx}+\dot{}\sigma_{zz})^2 2(\lambda+\mu)^2 dt, H_{\mu\mu}(\mathbf{x},\mathbf{x})=N_s\sum_{i=1}\int\Bigg[\dot{}\sigma_{xz}^2\mu^2+ \tag{36}$$

$$(\dot{}\sigma_{xx}+\dot{}\sigma_{zz})^2 2(\lambda+\mu)^2+(\dot{}\sigma_{xx}-\dot{}\sigma_{zz})^2 2\mu^2\Bigg] dt.$$

Using the relationships between elastic parameters and chain rule, the pseudo-Hessian for P- and S-wave impedances is given by

$$H_I=\begin{pmatrix} H_{I_pI_p} & H_{I_pI_s} \\ H_{I_sI_p} & H_{I_sI_s} \end{pmatrix}=\begin{pmatrix} 2V_p & 0 \\ -4V_s & 2V_s \end{pmatrix}\begin{pmatrix} H_{\lambda\lambda} & H_{\lambda\mu} \\ H_{\mu\lambda} & H_{\mu\mu} \end{pmatrix} \tag{37}$$

$$\begin{pmatrix} 2V_p & -4V_s \\ 0 & 2V_s \end{pmatrix}.$$

We use the inverse of the diagonal of the pseudo-Hessian for preconditioning. The diagonal terms of the pseudo-Hessian for P- and S-wave impedances can be expressed as follows:

$$H_{I_pI_p}(\mathbf{x},\mathbf{x})=N_s\sum_{i=1}\int 2V_{2p}(\dot{}\sigma_{xx}+\dot{}\sigma_{zz})^2(\lambda+\mu)^2 dt, H_{I_sI_s}(\mathbf{x},\mathbf{x})=N_s\sum_{i=1}\int 4V_{2s}\Bigg[\dot{}\sigma_{xz}^2\mu^2+ \tag{38}$$

$$(\dot{}\sigma_{xx}+\dot{}\sigma_{zz})^2 2(\lambda+\mu)^2+(\dot{}\sigma_{xx}-\dot{}\sigma_{zz})^2 2\mu^2\Bigg] dt.$$

The preconditioned version of elastic LSRTM minimizes

$$J(\delta\tilde{}\mathbf{m})=12N_s\sum_{i=1}\|L_i P^T P^H \delta\tilde{}\mathbf{m}-\delta\mathbf{d}_i\|_2^2, \tag{39}$$

where P^H denotes the inverse of the diagonal of pseudo-Hessian for P- and S-wave impedances, P^T denotes the parameter transformation matrix in equation 31. As discussed in the last section, the parameter transformation matrix P^T and its adjoint also play the role of preconditioning to our system of equations. The preconditioned CGLS algorithm (Bjorck, 1996) can be summarized as Algorithm 2. The output of Algorithm 2 is the inverted Lamé parameter perturbations $\delta\hat{}\mathbf{m}=(\delta\hat{}\lambda,\delta\hat{}\mu)^T=P^T P^H \delta\tilde{}\tilde{}\mathbf{m}$. And, the inverted P- and S-wave impedance perturbations can be retrieved by $(\delta\hat{}I_p,\delta\hat{}I_s)^T=P^{-1T}(\delta\hat{}\lambda,\delta\hat{}\mu)^T$.

Click to view table

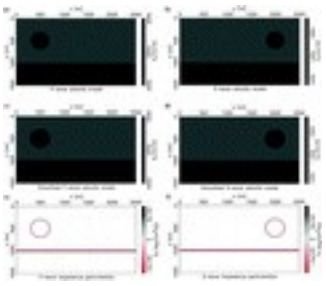
View Larger Version

Algorithm 2. Preconditioned CGLS algorithm.

The proposed method was tested on two synthetic models: the elastic version of the Camembert model (Gauthier et al., 1986) and the elastic Marmousi2 model (Martin et al., 2006). All the "observed data" are generated with the time-domain elastic staggered-grid FD method. In other words, we have not committed the so-called "inverse crime" that entails using the linearized Born modeling operator to generate data to test least-squares migration. The same staggered-grid FD code was used for the elastic LSRTM inversion. The C-PML boundary condition was applied on the top of the model. The observed data were assumed to be vector particle velocity fields. Only the direct wave was muted from the data. The data contain internal multiples that are not honored by the linearized Born modeling. In real data applications, the multiples can be attenuated from the data to only keep first-order scattered energy. The code for our numerical examples was written in C and parallelized with message passing interface (MPI).

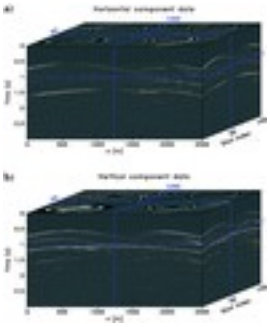
Elastic Camembert model

The elastic LSRTM is tested on a synthetic elastic version of the Camembert model. This test shows that elastic LSRTM can attenuate crosstalk between P- and S-wave impedance perturbations. Figure 1a and 1b shows the true P- and S-wave velocity models. The velocity anomalies are embedded in two layered models. The velocity anomalies for P and S are in different positions. Density is assumed to be constant ($1500\ kg/m^3$). The model has a dimension of 2.5 km in the horizontal axis and 1.5 km in depth with 501×301 grid points. There are 101 shots and 501 receivers that simulate a fixed-spread survey geometry. The shot interval is 25 m, and the receiver interval is 5 m. The shot depth is 5 m, and the receiver depth is 10 m. A 20 Hz central frequency Ricker wavelet is used to simulate an explosive source. The observed data are simulated using our elastic FD code. No other preprocessing was applied to the data except for muting the direct wave. The observed data are shown in Figure 2. Figure 1c and 1d shows the smoothed background velocity models for elastic RTM and elastic LSRTM. Smoothed models were obtained by convolving the true models with a 2D Gaussian function of 50 m width with standard deviation as half the width. The width of the 2D Gaussian function is approximately equal to the shortest P-wave wavelength. Figure 1e and 1f shows the true P- and S-wave impedance perturbations with respect to the background models.



View larger version (40K)

Figure 1. Elastic Camembert model. (a) P-wave velocity model. (b) S-wave velocity model. (c) Smoothed P-wave velocity model. (d) Smoothed S-wave velocity model. (e) True P-wave impedance perturbation. (f) True S-wave impedance perturbation.

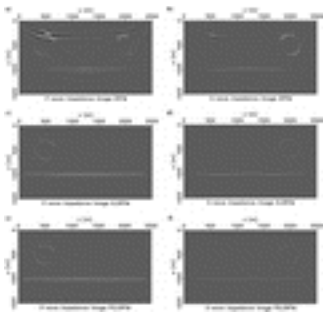


View larger version (87K)

Figure 2. Prestack multicomponent data for elastic Camembert model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.

The results of elastic RTM are shown in Figure 3a and 3b. The elastic RTM operator generates high-amplitude low-frequency artifacts caused by the crosscorrelation of the head wave, diving wave, and backscattered internal reflections. A Laplacian filter (Youn and Zhou, 2001) was used to attenuate the artifacts. As expected, there is crosstalk between the P- and S-wave impedance perturbations in the elastic RTM images. Elastic LSRTM (Figure 3c and 3d) not only reduces multiparameter crosstalk but also displays fewer artifacts, properly balanced amplitudes, and higher resolution. To make a fair comparison with the elastic RTM images (Figure 3a and 3b), the least-squares inverted images were postprocessed by Laplacian filtering. No filters were applied during the inversion process. These results were computed after 82 iterations of elastic LSRTM. The relative data-misfit percentage reduces to approximately 6%. The relative data misfit is defined as

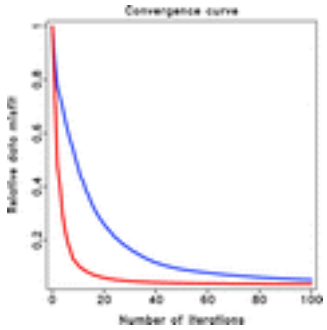
$$\sum_{si=1}^{Ns} \|L_i \delta\hat{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2 \Big/ \sum_{si=1}^{Ns} \|\delta\mathbf{d}_i\|_2^2. \quad (40)$$



View larger version (42K)

Figure 3. (a) P-wave impedance perturbation image estimated via elastic RTM. (b) S-wave impedance perturbation estimated via elastic RTM. (c) P-wave impedance perturbation image estimated via elastic LSRTM. (d) S-wave impedance perturbation image estimated via elastic LSRTM. (e) P-wave impedance perturbation image estimated via preconditioned elastic LSRTM. (f) S-wave impedance perturbation image estimated via preconditioned elastic LSRTM.

Figure 3e and 3f shows the pseudo-Hessian preconditioned elastic LSRTM after 20 iterations. The relative data misfit also reduces to approximately 6%. The results are similar to unpreconditioned elastic LSRTM. Figure 4compares the convergence curves of the unpreconditioned and preconditioned version of elastic LSRTM. From this figure, one can observe that preconditioned elastic LSRTM converges much faster than unpreconditioned LSRTM.

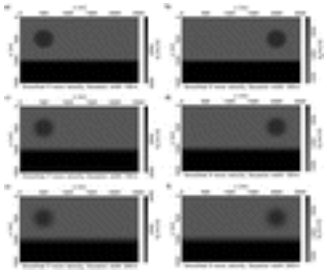


View larger version (24K)

Figure 4. Comparison of relative data-misfit convergence curves for elastic LSRTM (blue) and preconditioned elastic LSRTM (red) for Camembert model.

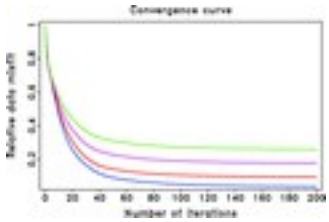
Sensitivity to background model error

We tested the proposed elastic LSRTM using background models with different degrees of spatial smoothing. The setup is equivalent to the setup adopted in the last section (the elastic Camembert model test). We have run our algorithm using background models with increasing degrees of smoothing. The background velocity models were smoothed via 2D Gaussian functions of width W 50, 100, 150, and 200 m. The standard derivation of the 2D Gaussian function equals to its half width. Results for W=50 m smoothing have already been shown in the previous section. We only plot the models and results for smoothing with 100, 150, and 200 m widths. Figure 5shows background models using different levels of smoothing. The comparison of the data-misfit convergence curves of elastic LSRTM using the four background models is shown in Figure 6. We observe that the level of smoothing of the background models influences the data misfit. The elastic LSRTM for different background models converge to different levels of data misfit. We have compared the results at fixed number of iterations (82 iterations) that all misfit curves have converged. The inverted results are shown in Figure 7. The quality of the inverted images degrades and more artifacts are present in the images as the level of smoothing increases.



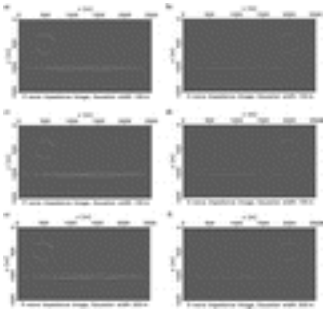
View larger version (36K)

Figure 5. Analysis of the influence of the background models on elastic LSRTM. (a and b) Smoothed P- and S-wave velocity models using a 2D Gaussian function of width W=100 m. (c and d) Smoothed P- and S-wave velocity models via 2D Gaussian function of width W=150 m. (e and f) Smoothed P- and S-wave velocity models using a 2D Gaussian function of width W=200 m.



View larger version (24K)

Figure 6. Comparison of relative data-misfit convergence curves for elastic LSRTM for background models with different degree of smoothing: W=50 m(blue), W=100 m (red), W=150 m (purple), and W=200 m (green).

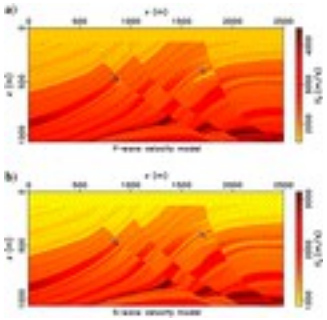


View larger version (42K)

Figure 7. Comparison of elastic LSRTM results for background models with increasing degree of smoothing. (a and b) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width W=100 m. (c and d) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width W=150 m. (e and f) Inverted P- and S-wave impedance perturbation images using background models smoothed with a 2D Gaussian function of width W=200 m.

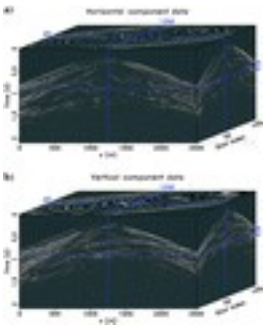
Elastic Marmousi2 model

The proposed method was also tested on a complex elastic model. To this end, we adopted the elastic Marmousi2 model (Martin et al., 2006). The model consists of a total of 199 layers with a steep anticline fault zone. The size of the original model was reduced to 1001×426 grid points to decrease the turnaround time of our tests. The water layer in the original Marmousi2 model was removed and replaced by a low-velocity layer to simulate a purely elastic model. Figure 8 shows the modified elastic Marmousi2 P- and S-wave velocity models. In the steep fault zone, there are two hydrocarbon reservoirs around depth 500 m that have decreased P-wave velocity and a small change in the S-wave velocity (indicated by the white triangles). This uncorrelated P- and S-wave structures will cause crosstalk in elastic RTM images. We will show that the elastic LSRTM can attenuate the crosstalk. Density is assumed to be constant ($2000\ \mathrm{kg/m^3}$). The model has dimensions of 2500 m in the horizontal distance and 1062.5 m in depth. A land-acquisition geometry is simulated with 101 shots and 1001 receivers distributed on the surface of the earth. The shot interval is 25 m, and the receiver interval is 2.5 m. The central frequency of the source function (Ricker wavelet) is 35 Hz. The observed data were simulated with our FD code, and the direct wave was removed from the observed data. The observed data are shown in Figure 9. Figure 10 shows the smoothed background velocity models for elastic RTM and elastic LSRTM. Smoothing was accomplished by convolving the true velocity models with a 35 m width 2D Gaussian function. Figure 11 shows the true P- and S-wave impedance perturbations with respect to the smoothed background models. From this figure, we can also observe that the P- and S-wave models are inconsistent in the two hydrocarbon reservoirs region at approximately 500 m depth in the steep fault zone (indicated by white triangles). Figure 12 shows the P- and S-wave impedance perturbation images obtained via elastic RTM. The elastic RTM algorithm has successfully imaged the geologic structures. However, the amplitudes of the elastic images are unbalanced. Uncollapsed energy artifacts caused by not having a dense distribution of sources and limited aperture are also visible. Most important, the elastic RTM operator has generated crosstalk between P and S images in the two hydrocarbon reservoir areas because the P- and S-wave velocity structures are different. These problems are caused by the fact that the elastic RTM operator is an adjoint operator as opposed to an ideal inverse operator. The elastic LSRTM (Figure 13) can solve these problems. Results for elastic LSRTM were computed after 98 iterations. The relative data misfit percentage reduces to approximately 40%. The geologic structure of the elastic Marmousi2 model is complex. The data generated by FD modeling contain internal multiples that are not honored by the linearized Born modeling operator. We believe that the latter explains the inability of the algorithm to reduce the data misfit further. The elastic LSRTM corrected the unbalanced amplitudes and suppressed the low-frequency RTM artifacts and artifacts caused by limited aperture. The elastic LSRTM also generates high-resolution images. More important, the elastic LSRTM can successfully decouple elastic parameters and suppress multiparameter crosstalk in areas with hydrocarbon traps in the P- and S-impedance perturbation images. These benefits are the result of the embedded de-blurring process that is associated with the inversion of the Hessian operator in elastic LSRTM.



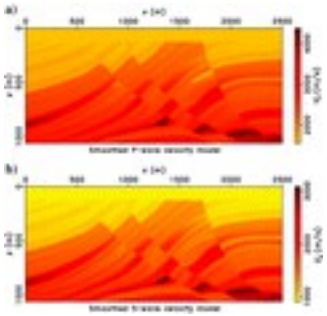
View larger version (82K)

Figure 8. Elastic Marmousi2 model. (a) P-wave velocity model. (b) S-wave velocity model.



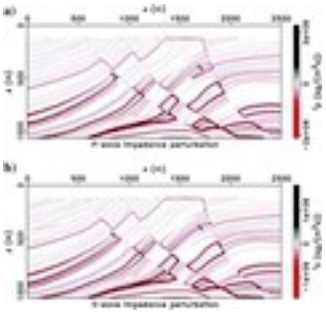
View larger version (111K)

Figure 9. Prestack multicomponent data for elastic Marmousi2 model. (a) Horizontal particle velocity data. (b) Vertical particle velocity data.



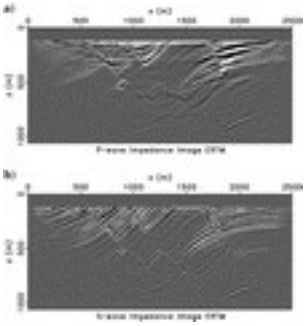
View larger version (77K)

Figure 10. Elastic Marmousi2 model. (a) Smoothed P-wave velocity model. (b) Smoothed S-wave velocity model.



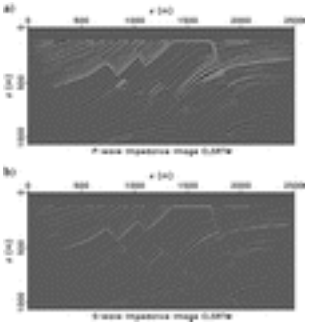
View larger version (118K)

Figure 11. Elastic Marmousi2 model. (a) True P-wave impedance perturbation. (b) True S-wave impedance perturbation.



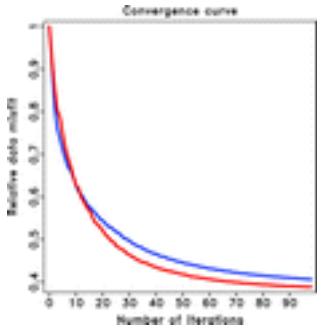
View larger version (69K)

Figure 12. Images obtained via elastic RTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.



View larger version (68K)

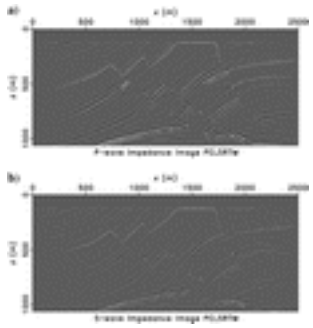
Figure 13. Images obtained via elastic LSRTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.



View larger version (25K)

Figure 15. Comparison of relative data-misfit convergence curves for elastic LSRTM (blue) and preconditioned elastic LSRTM (red) for the Marmousi2 model.

Figure 14 shows the pseudo-Hessian preconditioned elastic LSRTM after 58 iterations. The relative data misfit also reduces to approximately 40%. The preconditioned version of the elastic LSRTM yielded more amplitude balanced images than the unpreconditioned elastic LSRTM. To finalize our analysis, we also provide Figure 15 in which we compared convergence curves for the unpreconditioned and preconditioned elastic LSRTM. Preconditioning with the pseudo-Hessian has led to a visible improvement in convergence.



View larger version (64K)

Figure 14. Images obtained via preconditioned elastic LSRTM. (a) P-wave impedance perturbation image. (b) S-wave impedance perturbation image.

DISCUSSION

Choose Top of pageAbstractINTRODUCTIONTHEORYEXAMPLESDISCUSSION <<CONCLUSIONACKNOWLEDGMENTSREFERENCESAPPENDIX ACONNECTIONS BET...CITING ARTICLES

The source type adopted in our numerical examples is an explosive P-wave source that imitates an air gun or a dynamite source. A different type of source could have been implemented in the proposed elastic LSRTM. For example, a vertical force source term could have been implemented in our FD code to simulate a vibroseis source. For land seismic acquisition, the recorded data can contain multicomponent or single vertical component observations. In the single-component case, one should use the vertical component and use a sampling operator to exclude the horizontal component. Clearly, multicomponent vector data are preferable than the single vertical component data. The horizontal component will help to resolve the S-wave impedance perturbation because it primarily captures converted modes. Recording long-offset vertical component data containing converted S-wave energy will contribute to resolve the S-wave impedance perturbation as well. Applying elastic LSRTM to field land seismic data requires having high-quality data and access to accurate background models for P- and S-wave velocities. The proposed elastic LSRTM method also can be applied on marine seismic data. The staggered-grid FD modeling code used for elastic solid media can be used for acoustic fluid media as well. In this case, one can set the S-wave velocity to zero, and no particular treatment is needed for the solid-fluid interface (Virieux, 1986). In acoustic fluid media, the inverse elastic tensor $\mathbf{C}^{-1}$ in equation 24 is replaced by the pseudoinverse of $\mathbf{C}$ (Albertin et al., 2016). Conventional wide-aperture marine streamer pressure data can be used in the proposed elastic LSRTM. However,

the normal stress fields simulated by the elastic wave equation need to be transformed to pressure field (Vigh et al., 2014). It is clear that modern multicomponent ocean-bottom cable data can resolve the subsurface elastic images better than the streamer data.

CONCLUSION

Choose Top of pageAbstractINTRODUCTIONTHEORYEXAMPLESDISCUSSIONCONCLUSION <<ACKNOWLEDGMENTSREFERENCESAPPENDIXACONNECTIONS BET...CITING ARTICLES

Elastic LSRTM is formulated as a linearized elastic waveform inversion problem. The inversion is parameterized in terms of P- and S-wave impedance perturbations. The formulations of elastic Born approximation operator and elastic RTM operator are derived from the time-domain continuous adjoint-state method. The adjoint-state equation system is the same as the state-equation system. The only difference is the replacement of an explosive source to an adjoint source. After developing the functional formulations for our forward and adjoint operators, we have discretized the elastic Born and RTM operators. The numerical discretized versions of the two operators pass the dot-product test. This allows us to use the CGLS algorithm for solving the least-squares optimization problem. The Hessian is implicitly inverted via the adjoint-state method combined with CGLS algorithm. We investigated adopting the diagonal of the pseudo-Hessian operator to precondition the elastic LSRTM and thereby to accelerate its convergence. The elastic LSRTM produces high-resolution images with fewer artifacts and more balanced amplitudes than elastic RTM. More importantly, elastic LSRTM can reduce crosstalk artifacts between P- and S-wave impedance perturbations that are present in elastic RTM images. In essence, the off-diagonal elements of the Hessian operator are attenuated by the inversion process. The pseudo-Hessian preconditioning operator adopted in our work not only accelerates the convergence of the elastic LSRTM but also improves the overall amplitude response of our images.

ACKNOWLEDGMENTS

Choose Top of pageAbstractINTRODUCTIONTHEORYEXAMPLESDISCUSSIONCONCLUSIONACKNOWLEDGMENTS <<REFERENCESAPPENDIXACONNECTIONS BET...CITING ARTICLES

We thank the associate editor J. Schleicher, reviewer J. Rickett, and two anonymous reviewers for their valuable suggestions and comments that improved the manuscript. We thank the sponsors of the Signal Analysis and Imaging Group at the University of Alberta for financial support. High-performance clustered computing resources were provided by WestGrid Compute Canada. Figures in this paper were plotted using the Madagascar open-source software package.

REFERENCES

Choose Top of pageAbstractINTRODUCTIONTHEORYEXAMPLESDISCUSSIONCONCLUSIONACKNOWLEDGMENTSREFERENCES <<APPENDIXACONNECTIONS BET...CITING ARTICLES

1. Albertin, U., P. Shen, A. Sekar, T. Johnsen, C. Wu, K. Nihei, and K. Bube, 2016, 3D orthorhombic elastic full-waveform inversion in the reflection domain from hydrophone data: 86th Annual International Meeting, SEG, Expanded Abstracts, 1094–1098. Open URL [Google Scholar]

2. Alkhalifah, T., and R.-E. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: *Geophysics*, **79**, no. 3, R91–R101, doi:https://doi.org/10.1190/geo2013-0366.1. [Abstract] [Web of Science] Open URL [Google Scholar]

3. Anikiev, D., B. Kashtan, and W. A. Mulder, 2013, Decoupling of elastic parameters with iterative linearized inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3185–3190. Open URL [Google Scholar]

4. Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: *Geophysics*, **48**, 1514–1524, doi: https://doi.org/10.1190/1.1441434. [Abstract] [Web of Science] Open URL [Google Scholar]

5. Beydoun, W. B., and M. Mendes, 1989, Elastic ray-Born l2-migration/inversion: *Geophysical Journal International*, **97**, 151–160, doi: https://doi.org/10.1111/j.1365-246. [Crossref] Open URL [Google Scholar]

6. Bjorck, A., 1996, *Numerical methods for least squares problems*: SIAM. [Crossref] Open URL [Google Scholar]

7. Borzi, A., and V. Schulz, 2012, *Computational optimization of systems governed by partial differential equations*: SIAM. Open URL [Google Scholar]

8. Bourgeois, A., B. F. Jiang, and P. Lailly, 1989, Linearized inversion: A significant step beyond pre-stack migration: *Geophysical Journal International*, **99**, 435–445, doi: https://doi.org/10.1190/1.1442179. [Abstract] [Web of Science] Open URL [Google Scholar]

9. Chang, W., and G. A. McMechan, 1986, Reverse-time migration of offset vertical seismic profiling data using the excitation-time imaging condition: *Geophysics*, **51**, 67–84, doi: https://doi.org/10.1190/1.1442041. [Abstract] [Web of Science] Open URL [Google Scholar]

10. Chang, W., and G. A. McMechan, 1987, Elastic reverse-time migration: *Geophysics*, **52**, 1365–1375, doi:https://doi.org/10.1190/1.1442249. [Abstract] [Web of Science] Open URL [Google Scholar]

11. Chen, K., and M. D. Sacchi, 2016, Elastic least-squares reverse time migration via linearized elastic full waveform inversion with pseudo-Hessian preconditioning: Signal Analysis and Imaging Group, Report 17, University of Alberta, 274–310. Open URL [Google Scholar]

12. Chen, P., and E.-J. Lee, 2015, *Full-3D seismic waveform inversion*: Springer Geophysics. [Crossref] Open URL [Google Scholar]

13. Cheng, J., N. Kazemi, and M. Sacchi, 2016, Least-squares migration via a gradient projection method — Application to seismic data deblending: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: https://doi.org/10.3997/2214-4609.201601413. Open URL [Google Scholar]

14. Claerbout, J. F., 1985, *Imaging the earth's interior*: Blackwell Scientific Publications Inc. Open URL [Google Scholar]

15. Claerbout, J. F., 1992, *Earth soundings analysis: Processing versus inversion*: Blackwell Scientific Publications 6. Open URL [Google Scholar]

16. Dai, W., P. Fowler, and G. T. Schuster, 2012, Multi-source least-squares reverse time migration: *Geophysical Prospecting*, **60**, 681–695, doi: https://doi.org/10.1111/j.1365-2478.2012.01092. [Crossref] [Web of Science] Open URL [Google Scholar]

17. Dai, W., and G. T. Schuster, 2013, Plane-wave least-squares reverse-time migration: *Geophysics*, **78**, no. 4, S165–S177, doi: https://doi.org/10.1190/geo2012-0377.1. [Abstract] [Web of Science] Open URL [Google Scholar]

18. Dellinger, J., and J. Etgen, 1990, Wavefield separation in two-dimensional anisotropic media: *Geophysics*,**55**, 914–919, doi: https://doi.org/10.1190/1.1442906. [Abstract] [Web of Science] Open URL [Google Scholar]

19. Dong, S., J. Cai, M. Guo, S. Suh, Z. Zhang, B. Wang, and Z. Li, 2012, Least-squares reverse time migration: Towards true amplitude imaging and improving the resolution: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: https://doi.org/10.1190/segam2012-1488.1. Open URL [Google Scholar]

20. Douma, H., D. Yingst, I. Vasconcelos, and J. Tromp, 2010, On the connection between artifact filtering in reverse-time migration and adjoint tomography: *Geophysics*, **75**, no. 6, S219–S223, doi:https://doi.org/10.1190/1.3505124. [Abstract] [Web of Science] Open URL [Google Scholar]

21. Du, Q., Y. Zhu, and J. Ba, 2012, Polarity reversal correction for elastic reverse time migration: *Geophysics*,**77**, no. 2, S31–S41, doi: https://doi.org/10.1190/geo2011-0348.1. [Abstract] [Web of Science] Open URL [Google Scholar]

22. Duan, Y., and P. Sava, 2015, Scalar imaging condition for elastic reverse time migration: *Geophysics*, **80**, no. 4, S127–S136, doi: https://doi.org/10.1190/geo2014-0453.1. [Abstract] [Web of Science] Open URL [Google Scholar]

23. Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271. Open URL [Google Scholar]

24. Epanomeritakis, I., V. Akcelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: *Inverse Problems*, **24**, 034015, doi:https://doi.org/10.1088/0266-5611/24/3/034015. [Crossref] [Web of Science] Open URL [Google Scholar]

25. Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: *Geophysics*, **74**, no. 6, WCA5–WCA17, doi: https://doi.org/10.1190/1.3223188. [Abstract] [Web of Science] Open URL [Google Scholar]

26. Etgen, J. T., 1986, Prestack reverse time migration of shot profiles: *SEP Report*, **50**, 151169. Open URL [Google Scholar]

27. Feng, Z., and G. Schuster, 2016, Elastic least-squares reverse time migration: 86th Annual International Meeting, SEG, Expanded Abstracts, 4163–4167. Open URL [Google Scholar]

28. Fichtner, A., 2010, *Full seismic waveform modelling and inversion*: Springer. Open URL [Google Scholar]

29. Fichtner, A., and J. Trampert, 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: *Geophysical Journal International*, **185**, 775–798, doi: https://doi.org/10.1111/j.1365-246X.2011.04966. [Crossref] [Web of Science] Open URL [Google Scholar]

30. Fletcher, R. P., D. Nichols, R. Bloor, and R. T. Coates, 2016, Least-squares migration data domain versus image domain using point spread functions: *The Leading Edge*, **35**, 157–162, doi:https://doi.org/10.1190/tle35020157.1. [Abstract] Open URL [Google Scholar]

31. Gao, F., A. Levander, R. G. Pratt, C. A. Zelt, and G.-L. Fradelizio, 2007, Waveform tomography at a groundwater contamination site: Surface reflection data: *Geophysics*, **72**, no. 5, G45–G55, doi:https://doi.org/10.1190/1.2752744. [Abstract] [Web of Science] Open URL [Google Scholar]

32. Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: *Geophysics*, **51**, 1387–1403, doi: https://doi.org/10.1190/1.1442188. [Abstract] [Web of Science] Open URL [Google Scholar]

33. Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: *Geophysics*, **69**, 1017–1024, doi: https://doi.org/10.1190/1.1778244. [Abstract] [Web of Science] Open URL [Google Scholar]

34. Hardage, B., M. DeAngelo, P. Murray, and D. Sava, 2011, *Multicomponent seismic technology*: SEG. [Abstract] Open URL [Google Scholar]

35. Hestenes, M. R., and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: *Journal of Research of the National Bureau of Standards*, **49**, 409–436, doi: https://doi.org/10.6028/jres.049.044. [Crossref] [Web of Science] Open URL [Google Scholar]

36. Hou, J., and W. W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: *Geophysics*, **81**, no. 4, S165–S179, doi: https://doi.org/10.1190/geo2015-0499.1. [Abstract] [Web of Science] Open URL [Google Scholar]

37. Huang, Y., R. Nammour, and W. Symes, 2016, Flexibly preconditioned extended least-squares migration in shot-record domain: *Geophysics*, **81**, no. 5, S299–S315, doi: https://doi.org/10.1190/geo2016-0023.1. [Abstract] [Web of Science] Open URL [Google Scholar]

38. Innanen, K. A., 2014, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: *Geophysical Journal International*, **199**, 717–734, doi: https://doi.org/10.1093/gji/ggu291. [Crossref] [Web of Science] Open URL [Google Scholar]

39. Jin, S., R. Madariaga, J. Virieux, and G. Lambar, 1992, Two-dimensional asymptotic iterative elastic inversion: *Geophysical Journal International*, **108**, 575–588, doi: https://doi.org/10.1111/j.1365-246X.1992.tb04637. [Crossref] [Web of Science] Open URL [Google Scholar]

40. Kaplan, S. T., P. S. Routh, and M. D. Sacchi, 2010, Derivation of forward and adjoint operators for least-squares shot-profile split-step migration: *Geophysics*, **75**, no. 6, S225–S235, doi:https://doi.org/10.1190/1.3506146. [Abstract] [Web of Science] Open URL [Google Scholar]

41. Komatitsch, D., and R. Martin, 2007, An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation: *Geophysics*, **72**, no. 5, SM155–SM167, doi:https://doi.org/10.1190/1.2757586. [Abstract] [Web of Science] Open URL [Google Scholar]

42. Kuehl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262–273, doi: https://doi.org/10.1190/1.1543212. [Abstract] [Web of Science] Open URL [Google Scholar]

43. Lailly, P., 1983, *The seismic inverse problem as a sequence of before-stack migrations: Conference on inverse scattering: Theory and applications*: SIAM, 206–220. Open URL [Google Scholar]

44. Lambare, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: *Geophysics*, **57**, 1138–1154, doi: https://doi.org/10.1190/1.1443328. [Abstract] [Web of Science] Open URL [Google Scholar]

45. Le, H., 2016, Anisotropic full-waveform inversion using the second-order pseudo-acoustic wave equations: 86th Annual International Meeting, SEG, Expanded Abstracts, 408–412. Open URL [Google Scholar]

46. Lebrun, D., V. Richard, D. Mace, and M. Cuer, 2001, SVD for multioffset linearized inversion: Resolution analysis in multicomponent acquisition: *Geophysics*, **66**, 871–882, doi: https://doi.org/10.1190/1.1444976. [Abstract] [Web of Science] Open URL [Google Scholar]

47. Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: *Geophysics*, **53**, 1425–1436, doi:https://doi.org/10.1190/1.1442422. [Abstract] [Web of Science] Open URL [Google Scholar]

48. Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: *Geophysics*,**81**, no. 6, R339–R348, doi: https://doi.org/10.1190/geo2016-0038.1. [Abstract] [Web of Science] Open URL [Google Scholar]

49. Lions, J. L., 1971, *Optimal control of systems governed by partial differential equations*: Springer. [Crossref] Open URL [Google Scholar]

50. Lu, R., P. Traynin, and J. E. Anderson, 2009, Comparison of elastic and acoustic reverse time migration on the synthetic elastic Marmousi-II OBC dataset: 79th Annual International Meeting, SEG, Expanded Abstracts, 2799–2803. Open URL [Google Scholar]

51. Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: *Geophysics*, **56**, 645–653, doi:https://doi.org/10.1190/1.1443081. [Abstract] [Web of Science] Open URL [Google Scholar]

52. Luo, Y., H. Zhu, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Seismic modeling and imaging based upon spectral-element and adjoint methods: *The Leading Edge*, **28**, 568–574, doi:https://doi.org/10.1190/1.3124932. [Abstract] Open URL [Google Scholar]

53. Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for marmousi: *The Leading Edge*, **25**, 156–166, doi: https://doi.org/10.1190/1.2172306. [Abstract] Open URL [Google Scholar]

54. McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: *Geophysical Prospecting*, **31**, 413–420, doi: https://doi.org/10.1111/j.1365-2478.1983.tb01060. [Crossref] [Web of Science] Open URL [Google Scholar]

55. Mora, P., 1987, *Elastic wavefield inversion*: Ph.D. thesis, Stanford University. Open URL [Google Scholar]

56. Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: *Geophysics*, **64**, 208–221, doi: https://doi.org/10.1190/1.1444517. [Abstract] [Web of Science] Open URL [Google Scholar]

57. Ostmo, S., W. A. Mulder, and R-. Plessix, 2002, Finite-difference iterative migration by linearized waveform inversion in the frequency domain: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1384–1387. Open URL [Google Scholar]

58. Pageot, D., S. Operto, M. Valle, R. Brossier, and J. Virieux, 2013, A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging: *Geophysical Journal*

International, **193**, 1479–1505, doi: https://doi.org/10.1093/gji/ggs132. [Crossref] [Web of Science] Open URL [Google Scholar]

59. Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: *ACM Transaction on Mathematical Software*, **8**, 43–71, doi:https://doi.org/10.1145/355984.355989. [Crossref] [Web of Science] Open URL [Google Scholar]
60. 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**, 495–503, doi:https://doi.org/10.1111/j.1365-246X.2006.02978. [Crossref] [Web of Science] Open URL [Google Scholar]
61. Pratt, R. G., C. Shin, and G. J. Hick, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi:https://doi.org/10.1046/j.1365-246X.1998.00498. [Crossref] [Web of Science] Open URL [Google Scholar]
62. Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: *Geophysics*, **68**, 1371–1379, doi: https://doi.org/10.1190/1.1598130. [Abstract] [Web of Science] Open URL [Google Scholar]
63. Shin, C., S. Jang, and D.-J. Min, 2001a, Improved amplitude preservation for prestack depth migration by inverse scattering theory: *Geophysical Prospecting*, **49**, 592–606, doi: https://doi.org/10.1046/j.1365-2478.2001.00279. [Crossref] [Web of Science] Open URL [Google Scholar]
64. Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001b, Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion: *Geophysics*,**66**, 1856–1863, doi: https://doi.org/10.1190/1.1487129. [Abstract] [Web of Science] Open URL [Google Scholar]
65. Stanton, A., and M. Sacchi, 2015, Least squares wave equation migration of elastic data: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: https://doi.org/10.3997/2214-4609.201412706. Open URL [Google Scholar]
66. Stanton, A., and M. D. Sacchi, 2017, Elastic least-squares one-way wave equation migration: *Geophysics*,**82**, this issue, doi: https://doi.org/10.1190/geo2016-0391.1. [Abstract] [Web of Science] Open URL [Google Scholar]
67. Sun, R., and G. A. McMechan, 1986, Pre-stack reverse-time migration for elastic waves with application to synthetic offset vertical seismic profiles: *Proceedings of the IEEE*, **74**, 457–465, doi:https://doi.org/10.1109/PROC.1986.13486. [Crossref] [Web of Science] Open URL [Google Scholar]
68. Sun, R., and G. A. McMechan, 2001, Scalar reverse-time depth migration of prestack elastic seismic data: *Geophysics*, **66**, 1519–1527, doi: https://doi.org/10.1190/1.1487098. [Abstract] [Web of Science] Open URL [Google Scholar]
69. Sun, R., G. A. McMechan, H. Hsiao, and J. Chow, 2004, Separating P and S-waves in prestack 3D elastic seismograms using divergence and curl: *Geophysics*, **69**, 286–297, doi: https://doi.org/10.1190/1.1649396. [Abstract] [Web of Science] Open URL [Google Scholar]

70. Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790, doi: https://doi.org/10.1111/j.1365-2478.2008.00698. [Crossref] [Web of Science] Open URL [Google Scholar]

71. Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian: *Geophysics*, **74**, no. 6, WCA95–WCA107, doi: https://doi.org/10.1190/1.3204768. [Abstract] [Web of Science] Open URL [Google Scholar]

72. Tarantola, A., 1984a, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: https://doi.org/10.1190/1.1441754. [Abstract] [Web of Science] Open URL [Google Scholar]

73. Tarantola, A., 1984b, Linearized inversion of seismic reflection data: *Geophysical Prospecting*, **32**, 998–1015, doi: https://doi.org/10.1111/j.1365-2478.1984.tb00751. [Crossref] [Web of Science] Open URL [Google Scholar]

74. Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903, doi: https://doi.org/10.1190/1.1442046. [Abstract] [Web of Science] Open URL [Google Scholar]

75. Tarantola, A., 1988, *Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation*: Birkhäuser, 365–399. [Crossref] Open URL [Google Scholar]

76. Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: *Geophysical Journal International*, **160**, 195–216, doi: https://doi.org/10.1111/j.1365-246X.2004.02453. [Crossref] [Web of Science] Open URL [Google Scholar]

77. Tura, M. A. C., and L. R. Johnson, 1993, A stable method for linearized inversion of elastic parameters: *Geophysical Journal International*, **115**, 1–13, doi: https://doi.org/10.1111/j.1365-246X.1993.tb05585. [Crossref] [Web of Science] Open URL [Google Scholar]

78. Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: *Geophysics*, **79**, no. 2, R63–R77, doi:https://doi.org/10.1190/geo2013-0055.1. [Abstract] [Web of Science] Open URL [Google Scholar]

79. Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: *Geophysics*, **51**, 889–901, doi: https://doi.org/10.1190/1.1442147. [Abstract] [Web of Science] Open URL [Google Scholar]

80. Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: https://doi.org/10.1190/1.3238367. [Abstract] [Web of Science] Open URL [Google Scholar]

81. Wang, J., H. Kuehl, and M. D. Sacchi, 2005, High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the western Canadian sedimentary basin: *Geophysics*, **70**, no. 5, S91–S99, doi:https://doi.org/10.1190/1.2076748. [Abstract] [Web of Science] Open URL [Google Scholar]

82. Wang, J., and M. D. Sacchi, 2007, High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints: *Geophysics*, **72**, no. 1, S11–S18, doi:https://doi.org/10.1190/1.2387139. [Abstract] [Web of Science] Open URL [Google Scholar]

83. Wang, P., A. Gomes, Z. Zhang, and M. Wang, 2016, Least-squares RTM: Reality and possibilities for subsalt imaging: 86th Annual International Meeting, SEG, Expanded Abstracts, 4204–4209. Open URL [Google Scholar]

84. Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 53rd Annual International Meeting, SEG, Expanded Abstracts, 382–385. Open URL [Google Scholar]

85. Wong, M., B. L. Biondi, and S. Ronen, 2015, Imaging with primaries and free-surface multiples by joint least-squares reverse time migration: *Geophysics*, **80**, no. 6, S223–S235, doi:https://doi.org/10.1190/geo2015-0093.1. [Abstract] [Web of Science] Open URL [Google Scholar]

86. Xu, L., A. Stanton, and M. Sacchi, 2016, Elastic least-squares reverse time migration: 86th Annual International Meeting, SEG, Expanded Abstracts, 2289–2293. Open URL [Google Scholar]

87. Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2016, Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization: *Geophysics*, **81**, no. 1, S11–S20, doi: https://doi.org/10.1190/geo2014-0524.1. [Abstract] [Web of Science] Open URL [Google Scholar]

88. Yan, J., and P. Sava, 2008, Isotropic angle-domain elastic reverse-time migration: *Geophysics*, **73**, no. 6, S229–S239, doi: https://doi.org/10.1190/1.2981241. [Abstract] [Web of Science] Open URL [Google Scholar]

89. Yang, D., F. Liu, S. Morton, A. Malcolm, and M. Fehler, 2016, Time-lapse full-waveform inversion with ocean-bottom-cable data: Application on Valhall field: *Geophysics*, **81**, no. 4, R225–R235, doi:https://doi.org/10.1190/geo2015-0345.1. [Abstract] [Web of Science] Open URL [Google Scholar]

90. Yao, G., and H. Jakubowicz, 2016, Least-squares reverse-time migration in a matrix-based formulation: *Geophysical Prospecting*, **64**, 611–621, doi: https://doi.org/10.1111/1365-2478.12305. [Crossref] [Web of Science] Open URL [Google Scholar]

91. Youn, O. K., and H. Zhou, 2001, Depth imaging with multiples: *Geophysics*, **66**, 246–255, doi:https://doi.org/10.1190/1.1444901. [Abstract] [Web of Science] Open URL [Google Scholar]

92. Zhang, D., and G. T. Schuster, 2014, Least-squares reverse time migration of multiples: *Geophysics*, **79**, no. 1, S11–S21, doi: https://doi.org/10.1190/geo2013-0156.1. [Abstract] [Web of Science] Open URL [Google Scholar]

93. Zhang, Y., L. Duan, and Y. Xie, 2015, A stable and practical implementation of least-squares reverse time migration: *Geophysics*, **80**, no. 1, V23–V31, doi: https://doi.org/10.1190/geo2013-0461.1. [Abstract] [Web of Science] Open URL [Google Scholar]

94. Zhu, H., Y. Luo, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Elastic imaging and time-lapse migration based on adjoint methods: *Geophysics*, **74**, no. 6, WCA167–WCA177, doi:https://doi.org/10.1190/1.3261747. [Abstract] [Web of Science] Open URL [Google Scholar]

APPENDIX ACONNE Choose Top of pageAbstractINTRODUCTIONTHEORYEXAMPLESDISCUSSIONCONCLUSIONACKNOWLEDGMENTSREFERENCESAPPENDIX ACONNECTIONS

The FWI problem (Tarantola, 1984a; Luo and Schuster, 1991; Pratt et al., 1998; Gao et al., 2007; Alkhalifah and Plessix, 2014; Innanen, 2014; Li and Demanet, 2016; Yang et al., 2016) inverts for the spatially varying coefficients of the wave equation from the observed data on the surface of the earth. We will show that LSRTM is equivalent to one iteration of Gauss-Newton FWI. The difference is that LSRTM is applied to primary reflection data, whereas one iteration of Gauss-Newton FWI is applied to the data residual. The FWI minimizes the cost function (Tarantola, 1984a)

$$J(\mathbf{m})=12Ns\sum i=1\|\mathbf{T}\mathbf{u}i(\mathbf{m})-\mathbf{d}i\|22, \quad (A-1)$$

where $\mathbf{T}$ is the sampling operator, $\mathbf{u}^i(\mathbf{m})$ represents full wave mode forward modeled wavefield for ith shot, and $\mathbf{d}^i$ is the full wave mode observed data for ith shot. Gradient-based FWI is a local optimization that uses the Born approximation in each iteration (Virieux and Operto, 2009). Introducing a model perturbation $\mathbf{m}\rightarrow\mathbf{m}+\delta\mathbf{m}$ and a second-order Taylor series expansion, the cost function in the vicinity of $\mathbf{m}$ is given by the following expression:

$$J(\mathbf{m}+\delta\mathbf{m})=J(\mathbf{m})+\partial J(\mathbf{m})\partial\mathbf{m}\delta\mathbf{m}+12\delta\mathbf{m}T\partial 2J(\mathbf{m})\partial\mathbf{m}2\delta\mathbf{m}+O(\|\delta\mathbf{m}\|3). \quad (A-2)$$

In the vicinity of $\mathbf{m}$, the cost function is linearized and an optimal model update $\delta\mathbf{m}$ should satisfy $[\partial J(\mathbf{m}+\delta\mathbf{m})]/\partial\delta\mathbf{m}=0$. The latter results in the following expression:

$$\delta\mathbf{m}=-\left[\partial 2J(\mathbf{m})\partial\mathbf{m}2\right]-1\partial J(\mathbf{m})\partial\mathbf{m}=\left\{Ns\sum i=1\left[\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)\dagger\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)+\left(\mathbf{T}\partial 2\mathbf{u}i\partial\mathbf{m}2\right)\dagger(\mathbf{T}\mathbf{u}i-\mathbf{d}i)\right]\right\}-1\left[Ns\sum i=1\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)\dagger(\mathbf{d}i-\mathbf{T}\mathbf{u}i)\right], \quad (A-3)$$

where the term in the braces is called the full Hessian in the FWI problem. The second term in the braces corresponds to the second-order multiple scattering. This term is small when the cost function is close to a minimum. Dropping the second-order term results in the Gauss-Newton update (Pratt et al., 1998)

$$\delta\mathbf{m}=\left[Ns\sum i=1\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)\dagger\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)\right]-1\left[Ns\sum i=1\left(\mathbf{T}\partial\mathbf{u}i\partial\mathbf{m}\right)\dagger(\mathbf{d}i-\mathbf{T}\mathbf{u}i)\right]. \quad (A-4)$$

We recognize that the latter is same equation adopted in the LSRTM algorithm 29. The main difference is that the operator is applied on data residual at each iteration of the Gauss-Newton FWI. On the other hand, the operator is applied to primary reflection data when is used to solve the LSRTM problem. If the model is close to the optimal

solution and the reflection data are preprocessed, one iteration of the Gauss-Newton FWI is equivalent to solving the LSRTM problem.

Cited by

Guanchao Wang, Shangxu Wang, Jianyong Song, Chunhui Dong, Mingqiang Zhang. (2018) Elastic reflection traveltime inversion with decoupled wave equation. *GEOPHYSICS* **83**:5, R463-R474.
Online publication date: 21-Aug-2018.
Abstract | Full Text | PDF (2877 KB) | PDF w/Links (965 KB) | Permissions
Zongcai Feng, Bowen Guo, Gerard T. Schuster. (2018) Multiparameter deblurring filter and its application to elastic migration and inversion. *GEOPHYSICS* **83**:5, S421-S435.
Online publication date: 28-Aug-2018.
Abstract | Full Text | PDF (10190 KB) | PDF w/Links (3120 KB) | Permissions
Ke Chen, Mauricio D. Sacchi. 2018. Should we include the density perturbation in elastic least-squares reverse time migration?. SEG Technical Program Expanded Abstracts 2018, 4226-4230.
Abstract | PDF (2874 KB) | PDF w/Links (1129 KB) | Permissions
Ke Chen, Mauricio D. Sacchi. 2018. Time-domain elastic Gauss-Newton full-waveform inversion: A matrix-free approach. SEG Technical Program Expanded Abstracts 2018, 1208-1212.
Abstract | PDF (589 KB) | PDF w/Links (526 KB) | Permissions
Ke Chen, Mauricio Sacchi. 2017. Elastic least-squares reverse time migration via linearized elastic full-waveform inversion with pseudo-Hessian preconditioning. SEG Technical Program Expanded Abstracts 2017, 4364-4369.
Abstract | PDF (2464 KB) | PDF w/Links (1217 KB) | Supplemental Material | Permissions

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t}\mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix}\mathbf{u} = \mathbf{f},$$

where

$$\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \boldsymbol{\sigma} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} v_x \\ v_z \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} \\ \sigma_{zz} \\ \sigma_{xz} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{0} \\ \mathbf{f}_\sigma \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} \frac{\partial}{\partial x} & 0 & \frac{\partial}{\partial z} \\ 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial x} \end{pmatrix},$$

$$\mathbf{d}(\mathbf{m}+\delta\mathbf{m}) = \mathbf{d}(\mathbf{m}) + \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|^2),$$

where $\mathbf{m}$ is the background model, $\delta\mathbf{m}$ is the model perturbation, $\mathbf{d}(\mathbf{m})$ is the seismic data associat propagation in the background model, $\mathbf{d}(\mathbf{m}+\delta\mathbf{m})$ is the seismic data associated with propagation i perturbed model, and the linear operator $\partial\mathbf{d}/\partial\mathbf{m} = ([\partial\mathbf{d}/\partial\rho], [\partial\mathbf{d}/\partial\lambda], [\partial\mathbf{d}/\partial\mu])$ is the Fréchet deriv Jacobian matrix of $\mathbf{d}$. The second and higher order terms in the Taylor series are dropped resulting i Born approximation

$$\delta\mathbf{d} = \mathscr{L}\delta\mathbf{m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} = \mathscr{T}\frac{\partial \mathbf{u}}{\partial \mathbf{m}}\delta\mathbf{m},$$

where Fréchet derivative $\mathscr{L} = \partial\mathbf{d}/\partial\mathbf{m}$ is the Born modeling operator, $\delta\mathbf{d}$ is the first-order scattered $\delta\mathbf{d} \approx \mathbf{d}(\mathbf{m}+\delta\mathbf{m}) - \mathbf{d}(\mathbf{m})$, and the linear operator $\partial\mathbf{u}/\partial\mathbf{m}$ is the Fréchet derivative or Jacobian ma Fréchet derivative is prohibitively expensive to compute explicitly. Alternatively, the adjoint-state meth compute the action of the Fréchet derivative on vectors. For this purpose, we first differentiate the wa with respect to $\mathbf{m}$ (Fichtner, 2010; Fichtner and Trampert, 2011)

$$\frac{\partial \mathcal{S}}{\partial \mathbf{m}}\mathbf{u} + \mathcal{S}\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = \mathbf{0},$$

where linear operator $\partial\mathcal{S}/\partial\mathbf{m} = ([\partial\mathcal{S}/\partial\rho], [\partial\mathcal{S}/\partial\lambda], [\partial\mathcal{S}/\partial\mu])$. The right side of equation becomes the source term does not depend on the model. Then, the Fréchet derivative of $\mathbf{u}$ can be expressed

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\mathcal{S}^{-1}\frac{\partial \mathcal{S}}{\partial \mathbf{m}}\mathbf{u} = -\mathcal{S}^{-1}\frac{\partial \mathcal{S}}{\partial \mathbf{m}}\mathcal{S}^{-1}\mathbf{f},$$

where the multiplication of two operators follows the rule $\mathscr{A}\mathscr{B}\,\mathbf{v} = \mathscr{A}(\mathscr{B}\mathbf{v})$ (Chen and Lee, 2015). T $-(\partial\mathcal{S}/\partial\mathbf{m})\mathbf{u}$ is the so-called "virtual secondary source," which is the product of the incident wavefiel The operator $\partial\mathcal{S}/\partial\mathbf{m}$ represents the radiation pattern of the virtual secondary source (Pageot et al., Inserting the expression of Fréchet derivative of $\mathbf{u}$ into equation 8, the Born approximation can be wr

$$\delta\mathbf{d} = \mathscr{L}\delta\mathbf{m} = \frac{\partial \mathbf{d}}{\partial \mathbf{m}}\delta\mathbf{m} = -\mathscr{T}\mathcal{S}^{-1}\frac{\partial \mathcal{S}}{\partial \mathbf{m}}\mathbf{u}\delta\mathbf{m}.$$

$$\begin{pmatrix} \rho\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \frac{\partial}{\partial t} \delta\mathbf{u} - \begin{pmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{C}\mathbf{D}^T & \mathbf{0} \end{pmatrix} \delta\mathbf{u} = \begin{pmatrix} -\delta\rho \frac{\partial}{\partial t}\mathbf{v} \\ \delta\,\mathbf{C}\mathbf{D}^T\mathbf{v} \end{pmatrix},$$

where

$$\delta\mathbf{u} = \begin{pmatrix} \delta\mathbf{v} \\ \delta\boldsymbol{\sigma} \end{pmatrix}, \quad \delta\mathbf{v} = \begin{pmatrix} \delta v_x \\ \delta v_z \end{pmatrix}, \quad \delta\boldsymbol{\sigma} = \begin{pmatrix} \delta\sigma_{xx} \\ \delta\sigma_{zz} \\ \delta\sigma_{xz} \end{pmatrix}, \quad \delta\mathbf{C} = \begin{pmatrix} \delta\lambda + 2\delta\mu & \delta\lambda & 0 \\ \delta\lambda & \delta\lambda + 2\delta\mu & 0 \\ 0 & 0 & \delta\mu \end{pmatrix},$$

with zero initial condition $\delta\mathbf{u}|_{t=0} = \mathbf{0}$ and appropriate boundary conditions. The vector $\delta\mathbf{u}$ is the sca wavefield, with $\delta\mathbf{v}$ and $\delta\boldsymbol{\sigma}$ being the scattered particle velocity field and scattered stress field due to perturbations $\delta\rho$, $\delta\lambda$, and $\delta\mu$, and $\mathbf{v}$ is the incident particle velocity field. The right side of equation called "secondary source." The scattered wavefield can be computed using the same FD code that is compute the source-side incident wavefield in equation 1. The scattered data are obtained by sampli scattered wavefield at the receiver positions $\delta\mathbf{d} = \mathcal{T}\delta\mathbf{u}$.

The linearized adjoint problem: Elastic reverse time migration

The migration operator is the adjoint of the Born modeling operator that maps from reflection data to perturbation or image. The adjoint operator of the Born operator satisfies

$$\langle \delta\mathbf{d}, \mathscr{L}\delta\mathbf{m} \rangle_{\mathrm{D}} = \langle \mathscr{L}^{\dagger}\delta\mathbf{d}, \delta\mathbf{m} \rangle_{\mathrm{M}},$$

where $\dagger$ denotes the adjoint of an operator and $\langle \cdot, \cdot \rangle_{\mathrm{D}}$ and $\langle \cdot, \cdot \rangle_{\mathrm{M}}$ denote inner products in the data model domain, respectively. The adjoint of Born operator (equation 11) applied to reflection data can as (Tarantola, 1984a)

$$\delta\mathbf{m}^* = \mathscr{L}^{\dagger}\delta\mathbf{d} = \left(\frac{\partial\mathbf{d}}{\partial\mathbf{m}}\right)^{\dagger}\delta\mathbf{d} = -\left(\frac{\partial\mathcal{S}}{\partial\mathbf{m}}\mathbf{u}\right)^{\dagger}(\mathcal{S}^{-1})^{\dagger}\mathcal{T}^{\dagger}\delta\mathbf{d} = -\left(\frac{\partial\mathcal{S}}{\partial\mathbf{m}}\mathbf{u}\right)^{\dagger}(\mathcal{S}^{\dagger})^{-1}\mathcal{T}^{\dagger}\delta\mathbf{d},$$

where $\mathcal{T}^{\dagger}$ is the adjoint of sampling operator $\mathcal{T}$. Notice that we used the property $(\mathcal{S}^{-1})^{\dagger} = (\mathcal{S}^{\dagger})^{-}$ equation 16 (Tarantola, 1988). We adopted the symbol $\delta\mathbf{m}^*$ to represent the model perturbation that obtain by applying the adjoint operator to data perturbation $\delta\mathbf{d}$. Evidently, the adjoint operator is not e inverse of the linearized forward operator and therefore, $\delta\mathbf{m}^* \neq \delta\mathbf{m}$. To continue with our analysis, introduce the adjoint-state variable $\mathbf{p} = (\mathcal{S}^{\dagger})^{-1}\mathcal{T}^{\dagger}\delta\mathbf{d}$. The latter satisfies the "adjoint-state equation corresponding to the state equation 3

$$\mathcal{S}^{\dagger}\mathbf{p} = \mathcal{T}^{\dagger}\delta\mathbf{d},$$

where $\mathbf{p}$ is the adjoint-state variable of the state variable $\mathbf{u}$. $\mathcal{S}^{\dagger}$ is the adjoint wave-equation operato

$$\begin{pmatrix}\rho\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}^{\dagger}\left(\frac{\partial}{\partial t}\right)^{\dagger}\mathbf{p}-\begin{pmatrix}\mathbf{0} & \mathbf{D}\\ \mathbf{C}\mathbf{D}^{T} & \mathbf{0}\end{pmatrix}^{\dagger}\mathbf{p}=\mathcal{F}^{\dagger}\delta\mathbf{d},$$

with zero final condition $\mathbf{p}|_{t=T}=\mathbf{0}$ and appropriate boundary conditions. The vector $\mathbf{p}=(\boldsymbol{\upsilon},\boldsymbol{\varsigma})^{T}$, where $\boldsymbol{\upsilon}=(\upsilon_x,\upsilon_z)^{T}$ is the adjoint-state particle velocity field and $\boldsymbol{\varsigma}=(\varsigma_{xx},\varsigma_{zz},\varsigma_{xz})^{T}$ is the adjoint-state stress field. reflection data $\delta\mathbf{d}=(\delta\mathbf{d}_v,\delta\mathbf{d}_\sigma)^{T}$ act as the adjoint source to the adjoint-state equation. The differential operat antiself-adjoint $(\partial/\partial x)^{\dagger}=-(\partial/\partial x)$. Consequently, one can write the expression $\mathbf{D}^{\dagger}=-\mathbf{D}^{T}$. Finally, the adjoi state equation can be rewritten as follows:

$$\begin{pmatrix}\rho\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\left(-\frac{\partial}{\partial t}\right)\mathbf{p}+\begin{pmatrix}\mathbf{0} & \mathbf{D}\mathbf{C}\\ \mathbf{D}^{T} & \mathbf{0}\end{pmatrix}\mathbf{p}=\mathcal{F}^{\dagger}\delta\mathbf{d}.$$

The structure of the adjoint-state equation is slightly different to the structure of the state equation 1. However, t adjoint-state equation can be redefined into a form that resembles the state equation by adopting a transformati of variables (Vigh et al., 2014)

$$\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{pmatrix}\mathbf{p},$$

where $\tilde{\mathbf{p}}=(\tilde{\boldsymbol{\upsilon}},\tilde{\boldsymbol{\varsigma}})^{T}$ is the transformed adjoint-state variable, $\tilde{\boldsymbol{\upsilon}}=(\tilde{\upsilon}_x,\tilde{\upsilon}_z)^{T}$ and $\tilde{\boldsymbol{\varsigma}}=(\tilde{\varsigma}_{xx},\tilde{\varsigma}_{zz},\tilde{\varsigma}_{xz})^{T}$. If we mul both sides of the adjoint-state equation 21 by the transformation matrix, the adjoint-state equation can be rewrit as follows:

$$\begin{pmatrix}\rho\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\left(-\frac{\partial}{\partial t}\right)\tilde{\mathbf{p}}+\begin{pmatrix}\mathbf{0} & \mathbf{D}\\ \mathbf{C}\mathbf{D}^{T} & \mathbf{0}\end{pmatrix}\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{pmatrix}\mathcal{F}^{\dagger}\delta\mathbf{d},$$

where the adjoint-state equation now has the same structure as the state equation 1. Consequently, the FD cod adopted to solve the forward equation system 1 and the Born modeling equation system 13 can be reused to compute the adjoint wavefield in equation 23. The only difference is that the source term is replaced by an elast tensor scaled adjoint source and the FD steps are in time-reversal mode. After computing the transformed adjoi state variable $\tilde{\mathbf{p}}$, the original adjoint-state variable $\mathbf{p}$ can be retrieved by the inverse transformation

$$\mathbf{p}=\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{-1}\end{pmatrix}\tilde{\mathbf{p}}=\begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{-1}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{\upsilon}}\\ \tilde{\boldsymbol{\varsigma}}\end{pmatrix}=\begin{pmatrix}\tilde{\boldsymbol{\upsilon}}\\ \mathbf{C}^{-1}\tilde{\boldsymbol{\varsigma}}\end{pmatrix},$$

with

$$\mathbf{C}^{-1}=\begin{pmatrix}\frac{\lambda+2\mu}{4\lambda\mu+4\mu^2} & -\frac{\lambda}{4\lambda\mu+4\mu^2} & 0\\ -\frac{\lambda}{4\lambda\mu+4\mu^2} & \frac{\lambda+2\mu}{4\lambda\mu+4\mu^2} & 0\\ 0 & 0 & \frac{1}{\mu}\end{pmatrix}.$$

Using the migration operator formulation given by equation 19, elastic wave equation 1 and the adjoint wave

$$\text{err} = \left| \frac{\langle \delta\mathbf{d}_1, \delta\mathbf{d}_2 \rangle_{\text{D}} - \langle \delta\mathbf{m}_1, \delta\mathbf{m}_2 \rangle_{\text{M}}}{\langle \delta\mathbf{d}_1, \delta\mathbf{d}_2 \rangle_{\text{D}} + \langle \delta\mathbf{m}_1, \delta\mathbf{m}_2 \rangle_{\text{M}}} \right|. \quad (2$$

Our code was written in single-precision float data type in C. The relative error of our dot-product test is 10^{-3} for model that consists of 500×500 samples in space and a single shot that consists of 5000 samples in time and 500 receivers.

Elastic least-squares reverse time migration
CGLS with adjoint-state method

From the above derivation, a properly designed elastic RTM code can be considered equivalent to the adjoint operator of the elastic Born forward-modeling operator. The adjoint operator is an approximation to the inverse operator, in which the Hessian of the linearized inversion problem is replaced by an identity matrix. In other word the migrated image obtained via the adjoint operator is a blurred version of the true subsurface image. The migrated image, in general, suffers from relative low resolution, unbalanced amplitudes due to geometric spreadi and acquisition footprint. Moreover, multiparameter elastic migration will generate crosstalk among different components because different parameters are coupled. To estimate higher resolution images with properly balanced amplitudes and fewer crosstalk and artifacts, the elastic LSRTM is formulated as a quadratic optimizati problem, in which one minimizes the following cost function:

$$\mathcal{J}(\delta\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathcal{L}_i \delta\mathbf{m} - \delta\mathbf{d}_i\|_2^2, \quad (2$$

where $\mathcal{L}_i$ is the Born approximation operator for the ith shot, $\delta\mathbf{d}_i$ is the reflection data associated to the ith sho gather, $\delta\mathbf{m}$ denotes model perturbation (elastic images), N_s indicates the number of shots, and $\|\cdot\|_2$ indicates t ℓ_2 norm of vector. The optimal solution satisfies the condition $[\partial\mathcal{J}(\delta\mathbf{m})]/\partial\delta\mathbf{m} = 0$. The latter leads to the norm equations

$$\left(\sum_{i=1}^{N_s} \mathcal{L}_i^\dagger \mathcal{L}_i \right) \delta\mathbf{m} = \sum_{i=1}^{N_s} \mathcal{L}_i^\dagger \delta\mathbf{d}_i, \quad (2$$

$$\lambda = \rho V_p^2 - 2\rho V_s^2, \quad \mu = \rho V_s^2,$$
$$I_p = \rho V_p, \quad I_s = \rho V_s,$$

in conjunction with total derivatives to write down the following parameter perturbation transformation:

$$\begin{pmatrix} \delta\lambda \\ \delta\mu \end{pmatrix} = \begin{pmatrix} 2V_p & -4V_s \\ 0 & 2V_s \end{pmatrix} \begin{pmatrix} \delta I_p \\ \delta I_s \end{pmatrix}.$$

We insert the above parameter transformation into the elastic LSRTM formulation (equation 28). The parameter transformation matrix and its adjoint are incorporated into the CGLS algorithm (Algorithm 1). The change of parameters is similar to adding preconditioning to our system of equations. In our experience, inverting for impedance perturbations leads to an algorithm with faster convergence than inverting for the Lamé parameter perturbations.

Preconditioning using multiparameter Hessian

The elastic LSRTM implicitly inverts the Hessian operator via CGLS iterations combined with adjoint-state metho i.e., we apply the Fréchet derivative and its adjoint on vectors via the adjoint-state method on the fly in each CGL iteration. It does not need to form the Fréchet derivative or Hessian explicitly. However, the CGLS algorithm may need a relatively large number of iterations to converge to an optimal solution. Preconditioning of the gradients is important to accelerate the convergence of CGLS and save computational resources. We investigate the structur of the Hessian operator of the elastic LSRTM and use the diagonal of the pseudo-Hessian for preconditioning (Sh et al., 2001a). The Hessian operator of elastic LSRTM problem can be expressed as

$$\mathscr{H} = \sum_{i=1}^{N_s} \mathscr{L}_i^\dagger \mathscr{L}_i = \sum_{i=1}^{N_s} \left(\mathscr{T} \frac{\partial \mathbf{u}_i}{\partial \mathbf{m}} \right)^\dagger \left(\mathscr{T} \frac{\partial \mathbf{u}_i}{\partial \mathbf{m}} \right) = \sum_{i=1}^{N_s} \left(\mathscr{T} \mathscr{S}^{-1} \frac{\partial \mathscr{S}}{\partial \mathbf{m}} \mathbf{u}_i \right)^\dagger \left(\mathscr{T} \mathscr{S}^{-1} \frac{\partial \mathscr{S}}{\partial \mathbf{m}} \mathbf{u}_i \right),$$

where $\mathbf{u}_i$ is the source-side wavefield for ith shot, and recall that $\mathscr{S}^{-1}$ is the Green's operator. If the Green's functions are not saved, explicitly computing the Hessian needs $N_s \cdot N_m$ forward simulations with N_m as the number of model grid points. Using the reciprocity of the Green's function (Tarantola, 1988), the number of forwa simulations needed for explicitly computing the Hessian reduces to $N_s \cdot N_g$, where N_g denotes the number of receivers. The diagonal element of the Hessian is the zero-lag autocorrelation of the Fréchet derivative. It accoun for the geometric spreading effect (Shin et al., 2001b). Explicitly computing the diagonal of the Hessian also requires $N_s \cdot N_g$ forward simulations. Shin et al. (2001a) propose to neglect the receiver Green's function to sav computation cost. Under this assumption, the Hessian can be simplified to the so-called "pseudo-Hessian":

$$\mathbb{H} = \sum_{i=1}^{N_s} \left(\frac{\partial \mathscr{S}}{\partial \mathbf{m}} \mathbf{u}_i \right)^\dagger \left(\frac{\partial \mathscr{S}}{\partial \mathbf{m}} \mathbf{u}_i \right).$$

For multiparameter problem, the Hessian and pseudo-Hessian are blockwise. The pseudo-Hessian for Lamé parameters is given by

$$\mathbb{H}_L = \begin{pmatrix} \mathbb{H}^{\lambda\lambda} & \mathbb{H}^{\lambda\mu} \\ \cdots & \cdots \end{pmatrix}.$$

$$\mathbb{H}^{\lambda\lambda}(\mathbf{x},\mathbf{x}) = \sum_{i=1}^{N_s} \int \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{2(\lambda+\mu)^2} dt,$$

$$\mathbb{H}^{\mu\mu}(\mathbf{x},\mathbf{x}) = \sum_{i=1}^{N_s} \int \left[\frac{\dot{\sigma}_{xz}^2}{\mu^2} + \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{2(\lambda+\mu)^2} + \frac{(\dot{\sigma}_{xx}-\dot{\sigma}_{zz})^2}{2\mu^2} \right] dt.$$

Using the relationships between elastic parameters and chain rule, the pseudo-Hessian for P- and S-wave impedances is given by

$$\mathbb{H}_I = \begin{pmatrix} \mathbb{H}^{I_p I_p} & \mathbb{H}^{I_p I_s} \\ \mathbb{H}^{I_s I_p} & \mathbb{H}^{I_s I_s} \end{pmatrix} = \begin{pmatrix} 2V_p & 0 \\ -4V_s & 2V_s \end{pmatrix} \begin{pmatrix} \mathbb{H}^{\lambda\lambda} & \mathbb{H}^{\lambda\mu} \\ \mathbb{H}^{\mu\lambda} & \mathbb{H}^{\mu\mu} \end{pmatrix} \begin{pmatrix} 2V_p & -4V_s \\ 0 & 2V_s \end{pmatrix}.$$

We use the inverse of the diagonal of the pseudo-Hessian for preconditioning. The diagonal terms of the pseudo Hessian for P- and S-wave impedances can be expressed as follows:

$$\mathbb{H}^{I_p I_p}(\mathbf{x},\mathbf{x}) = \sum_{i=1}^{N_s} \int 2V_p^2 \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{(\lambda+\mu)^2} dt,$$

$$\mathbb{H}^{I_s I_s}(\mathbf{x},\mathbf{x}) = \sum_{i=1}^{N_s} \int 4V_s^2 \left[\frac{\dot{\sigma}_{xz}^2}{\mu^2} + \frac{(\dot{\sigma}_{xx}+\dot{\sigma}_{zz})^2}{2(\lambda+\mu)^2} + \frac{(\dot{\sigma}_{xx}-\dot{\sigma}_{zz})^2}{2\mu^2} \right] dt.$$

The preconditioned version of elastic LSRTM minimizes

$$\mathcal{J}(\delta\tilde{\mathbf{m}}) = \frac{1}{2} \sum_{i=1}^{N_s} \| \mathcal{L}_i \mathcal{P}_T \mathcal{P}_H \delta\tilde{\mathbf{m}} - \delta\mathbf{d}_i \|_2^2,$$

where $\mathcal{P}_H$ denotes the inverse of the diagonal of pseudo-Hessian for P- and S-wave impedances, $\mathcal{P}_T$ denotes the parameter transformation matrix in equation 31. As discussed in the last section, the parameter transformatio matrix $\mathcal{P}_T$ and its adjoint also play the role of preconditioning to our system of equations. The preconditioned CGLS algorithm (Bjorck, 1996) can be summarized as Algorithm 2. The output of Algorithm 2 is the inverted Lam parameter perturbations $\delta\hat{\mathbf{m}} = (\delta\hat{\lambda}, \delta\hat{\mu})^T = \mathcal{P}_T \mathcal{P}_H \delta\hat{\tilde{\mathbf{m}}}$. And, the inverted P- and S-wave impedance perturbations can be retrieved by $(\delta\hat{I}_p, \delta\hat{I}_s)^T = \mathcal{P}_T^{-1}(\delta\hat{\lambda}, \delta\hat{\mu})^T$.

$$\frac{\sum_{i=1}^{N_s} \|\mathscr{L}_i \delta\hat{\mathbf{m}} - \delta\mathbf{d}_i\|_2^2}{\sum_{i=1}^{N_s} \|\delta\mathbf{d}_i\|_2^2}. \tag{4}$$