JOINT MIGRATION INVERSION: CONTINUOUS EQUATIONS AND DISCRETIZED SOLUTION VIA MULTIPARAMETER GAUSS–NEWTON METHOD

C. A. M. Assis, Y. Pan, J. Schleicher, T. Bohlen, and H. B. Santos

email: *carlosmartinsx@gmail.com, js@ime.unicamp.br*
keywords: *Imaging, inversion, one-way, Bremmer series*

ABSTRACT

Joint Migration Inversion (JMI) is an independent approach to solve the seismic inverse problem, based on decoupled imaging and inversion operators. Here, we review the background equations in their continuous form. Furthermore, the general expression for the migration gradient seems to provide a new imaging condition. We then proceed to test the JMI methodology using the multiparameter Gauss–Newton method to estimate simultaneously image and slowness updates, and we compare the results to those of the conventionally used steepest-descent method. Our numerical results show that the Gauss–Newton method can provide velocity models with improved resolution, albeit at a higher cost. These results also demonstrate that the JMI implementation under the assumptions discussed here can provide a good depth migrated image and a satisfying initial velocity model for a subsequent Full Waveform Inversion.

INTRODUCTION

Joint Migration Inversion (JMI) has been proposed as a new approach to the seismic inverse problem (Berkhout, 2014b). The method's distinguishing characteristics are the ability to reproduce intrabed multiples even in a smooth velocity model and a computational cost which is lower than that of finite-difference based methods. JMI is built upon the seismic wavefield decomposition into its downgoing and upgoing components. This approach intrinsically decouples high spatial frequency information, represented by scattering operators, from low spatial frequency information, represented by the velocity model (Wapenaar, 1996).

By means of the migration part, the velocity model allows for the positioning of events. In turn, from the resulting image, scattering operators are built to update the velocity model. Since the methodology deals mainly with reflections below the critical angle, it is conceptually analogous to Reflection Waveform Inversion (Xu et al., 2012).

Because of the higher cost of wave-based methods, ray-based tomography methods are still routinely used in oil and gas exploration (Jones, 2010; Santos, 2015). In terms of cost and quality, JMI offers an intermediate solution between ray-based methods and Full Waveform Inversion (FWI), accounting for finite frequency effects, but not demanding event picking and with lower computational cost than finite-difference solutions of the wave equation.

This work aims at presenting continuous versions of the JMI equations and at solving the inverse problem using the Gauss–Newton method to estimate simultaneously image and slowness updates. Usually, the methodology is presented in the discrete matrix form, and the parameters are updated by alternating between imaging and tomography (Masaya and Verschuur, 2018). In the continuous form, the equations may be slightly more cumbersome, but they help to state the method and its assumptions as clearly as possible. The results of our implementation indicate that JMI is a promising methodology for velocity model

building with the potential to assist in reducing the number of necessary iterations in a subsequent FWI application, allowing it to be integrated within a robust initial-velocity-model building workflow (see, e.g., Santos et al., 2015, 2016b,a,c).

FORWARD PROBLEM

In this section, we discuss the two-dimensional acoustic constant-density modeling equations. First, we show the differential formulation to provide a general understanding about the problem under consideration. Then, we present the integral formulation, being the approach adopted in the implementation. Finally, using the integral formulation, we discuss the discrete recursive modeling algorithm. Note that we adopt the following Fourier convention:

$$F(\omega) = \int_{\mathbb{R}} f(t)e^{-i\omega t}dt \,, \tag{1}$$

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} F(\omega)e^{i\omega t}d\omega \,. \tag{2}$$

Differential modeling equations

Consider the constant-density acoustic wave equation decomposed into its downgoing and upgoing components. Similar to Ursin et al. (2012) and Wapenaar (1996), the downgoing $P^+(\mathbf{x}, \omega)$ and upgoing wavefields $P^-(\mathbf{x}, \omega)$ in the space-frequency domain $(\mathbf{x}, \omega)$ for a downgoing source S^+ must satisfy

$$\frac{\partial P^+}{\partial z} = -i\hat{\mathcal{H}}_1 P^+ + \hat{\mathcal{T}}_c^+ P^+ + \hat{\mathcal{R}}_c^- P^- + S^+ \,, \tag{3}$$

$$\frac{\partial P^-}{\partial z} = i\hat{\mathcal{H}}_1 P^- - \hat{\mathcal{T}}_c^- P^- - \hat{\mathcal{R}}_c^+ P^+ \,, \tag{4}$$

where i is the imaginary unit and $\mathbf{x} = (x, z)$ is the observation point, with x the horizontal axis and z the depth axis increasing downward. Moreover, $\hat{\mathcal{H}}_1$ is the square-root operator, $\hat{\mathcal{R}}_c$ and $\hat{\mathcal{T}}_c$ are the reflectivity the transmissivity operators, respectively, with the superscript $+$ indicating incidence from above a model position and the superscript $-$ indicating incidence from below. Note that as a consequence of the directional decoupling, horizontally propagating waves are not defined (Ursin et al., 2012).

The square-root operator $\hat{\mathcal{H}}_1$ is defined such that its twofold application results in the lateral Helmholtz operator $\hat{\mathcal{H}}_2$, i.e.,

$$\hat{\mathcal{H}}_2 = \hat{\mathcal{H}}_1 \hat{\mathcal{H}}_1 \,, \tag{5}$$

and it is formally defined as

$$\hat{\mathcal{H}}_1 = \left[\omega^2 \sigma^2 + \frac{\partial^2}{\partial x^2} \right]^{1/2} \,, \tag{6}$$

in which σ is the slowness.

In a laterally invariant model, the square-root operator in the frequency-wavenumber domain is equivalent to the vertical wavenumber

$$k_z = \omega \sqrt{\sigma^2 - k_x^2/\omega^2} \,, \tag{7}$$

where k_x is the horizontal wavenumber. The square-root operator and the scattering operators represent convolutions in the horizontal coordinate, viz.,

$$\left(\hat{\mathcal{H}}_1 P^\pm \right)(\mathbf{x}, \omega) = \int_{\mathbb{R}} \mathcal{H}_1(x, z, \omega; x') P^\pm(x', z, \omega)dx' \,, \tag{8}$$

$$(\hat{\mathcal{R}}_c^+ P^+)(\mathbf{x}, \omega) = \int_{\mathbb{R}} \hat{\mathcal{R}}_c^+(x, z, \omega; x') P^+(x', z, \omega)dx' \,. \tag{9}$$

A detailed discussion about the square-root operator can be found in Grimbergen et al. (1998).

Scattering operators and coefficients

The scattering operators in a constant-density model are defined as

$$\hat{\mathcal{R}}_c^{\pm} = \hat{\mathcal{T}}_c^{\pm} = \mp \frac{1}{2} \hat{\mathcal{H}}_1^{-1} \frac{\partial \hat{\mathcal{H}}_1}{\partial z} \,. \tag{10}$$

More details about these equations can be found in Wapenaar (1996), in which the coupled one-way wave equations are developed using a flux-normalized decomposition of the two-way wave operator. Here, we consider pressure normalization, so that the sum $P^+ + P^-$ provides the total two-way wavefield P.

The scattering operators act on a wavefield via the convolution integral as detailed in equation (9). In the angle-independent case, the scattering operators action on a wavefield is performed by a direct multiplication and they are approximately given by

$$\hat{\mathcal{R}}_c^{\pm} = \hat{\mathcal{T}}_c^{\pm} \approx \mp \frac{1}{2c} \frac{\partial c}{\partial z} \,. \tag{11}$$

where $c = 1/\sigma$ is the acoustic wavespeed. This approximation to the reflectivity $\hat{\mathcal{R}}_c^+$ is well-known in the literature, see, e.g., Berteussen and Ursin (1983). The reflectivity and transmissivity operators relate to the reflection and transmission operators as

$$\hat{\mathcal{R}}_c^{\pm} \Delta z \approx \hat{\mathcal{R}}^{\pm} \qquad \text{and} \qquad \hat{\mathcal{T}}_c^{\pm} \Delta z \approx \mathcal{I} + \hat{\mathcal{T}}^{\pm} \,. \tag{12}$$

In a discontinuous model we recover the conventional acoustic reflection and transmission coefficients, e.g.,

$$\mathcal{R}^+ \rightarrow R^+ = \frac{c(x, z + \Delta z) - c(x, z - \Delta z)}{c(x, z + \Delta z) + c(x, z - \Delta z)} \,, \tag{13}$$

which satisfy (Berkhout, 2014b)

$$T^+ = 1 + R^+ \,, \tag{14}$$

$$R^- = -R^+ \,, \tag{15}$$

$$T^- = 1 - R^+ \,, \tag{16}$$

where $T^{\pm}$ are the transmission coefficients from above and below and $R^{\pm}$ are the corresponding reflection coefficients.

Integral modeling equations

The downgoing G^+ and upgoing G^- Green's functions for our modeling equations, in a model vertically homogeneous, must satisfy

$$\frac{\partial G^+}{\partial z} = -i \hat{\mathcal{H}}_1 G^+ - \Delta z \delta(\mathbf{x} - \mathbf{x}') \,, \tag{17}$$

$$\frac{\partial G^-}{\partial z} = i \hat{\mathcal{H}}_1 G^- + \Delta z \delta(\mathbf{x} - \mathbf{x}') \,. \tag{18}$$

Using Green's theorem (see, e.g., Schleicher et al., 2007), we can build integral representations for the differential equations (3) and (4). The resulting equations are

$$P^+(\mathbf{x}, \omega) = -\frac{1}{\Delta z} \int_{\mathbb{R}} G^+(x, z, \omega; x', z') \left(T^+ P^+ + R^- P^- + S^+ \Delta z \right) (x', z', \omega) dx' \,, \tag{19}$$

$$P^-(\mathbf{x}, \omega) = -\frac{1}{\Delta z} \int_{\mathbb{R}} G^-(x, z, \omega; x', z') \left(T^- P^- + R^+ P^+ \right) (x', z', \omega) dx' \,. \tag{20}$$

These equations allow us to extrapolate wavefields from a depth level z' to another depth level z. Between depth levels, we consider the velocity model vertically homogeneous. At the boundary between

two layers, vertical discontinuities in the velocity model may occur and are accounted for by the scattering coefficients.

Introducing the operators $\hat{\mathcal{G}}^+$ and $\hat{\mathcal{G}}^-$ that represent the integrals over the lateral coordinate, we obtain

$$P^+(z) = \hat{\mathcal{G}}^+(z; z') \left(T^+ P^+ + R^- P^- + S^+ \Delta z\right)(z'), \tag{21}$$

$$P^-(z) = \hat{\mathcal{G}}^-(z; z') \left(T^- P^- + R^+ P^+\right)(z'), \tag{22}$$

where we exhibit only the depth coordinate to simplify notation. These equations account for full scattering.

Discrete recursive modeling

Using a procedure similar to the Born expansion, it is possible to obtain a recursive relationship between the wavefields that accounts for higher order scattering. This can be done by defining two downgoing wavefields. One is related to the downgoing transmission of the physical source term S^+, namely the incident downgoing wavefield. The other one is the perturbed downgoing wavefield resulting from the reflection of the total upgoing wavefield. The corresponding quantities for the upgoing wavefield can be defined upon reflection of the incident and scattered downgoing wavefields. Applying this procedure to the integral formulation, equations (21) and (22), we obtain the recursive equations

$$P_{j+1}^+(z_{n+1}) = \hat{\mathcal{G}}^+(z_{n+1}; z_n) \left(R^- P_j^- + T^+ P_{j+1}^+ + S^+ \Delta z\right)(z_n), \tag{23}$$

$$P_{j+1}^-(z_{n-1}) = \hat{\mathcal{G}}^-(z_{n-1}; z_n) \left(R^+ P_{j+1}^+ + T^- P_{j+1}^-\right)(z_n), \tag{24}$$

where the subscript j indicates the scattering order and z_n denotes the depth of the nth boundary. Moreover, $z_{n\pm 1} = z_n \pm \Delta z$. The zero-order terms, equivalent to our definition of incident wavefields, are given by

$$P_0^+(z_{n+1}) = \hat{\mathcal{G}}^+(z_{n+1}; z_n) \left(T^+ P_0^+ + S^+ \Delta z\right)(z_n), \tag{25}$$

$$P_0^-(z_{n-1}) = \hat{\mathcal{G}}^-(z_{n-1}; z_n) \left(R^+ P_0^+ + T^- P_0^-\right)(z_n). \tag{26}$$

The source term in equation (23) is calculated as

$$S^+(z_n) = \begin{cases} -\frac{i}{2\Delta z} \hat{\mathcal{H}}_1^{-1} C(\omega), \, z_n = z_s, \\ 0, \, z_n \neq z_s, \end{cases} \tag{27}$$

where $C(\omega)$ is the spectrum of the source wavelet, e.g., a Ricker wavelet.

Therefore, P_0^+ denotes the downgoing transmitted wavefield, P_0^- represents the upgoing waves reflected once, P_1^+ are all twice reflected downgoing waves, P_1^- stands for upgoing waves that have bounced three times, and so on. This recursive modeling procedure is closely related to the work of Bremmer (1951) and explored in detail by Berkhout (2014a). Note that the only necessary information to account for higher-order scattering at iteration $j + 1$ is the source wavefield S^+ and the downgoing wavefield from the last iteration P_j^-. To simplify the presentation, we assume a two-dimensional acoustic model, but the extension to 3D is straightforward.

Still for simplicity, we assume the model parameters to be locally homogeneous, and the scattering operators/coefficients to be angle independent. This last assumption leads to scattering operators that are independent of frequency and allows us to substitute the convolution in equation (9) by a direct multiplication (de Bruin et al., 1990).

Concerning propagation, the locality assumption allows us to approximately implement the Green's function for laterally varying models using the approach of Thorbecke et al. (2004). However, any other one-way extrapolation technique could be used.

INVERSE PROBLEM

In this section, we introduce the inverse problem of estimating the model parameters. The developments made here are the main contributions of this work.

First, we discuss the Gauss–Newton method ands apply it to the simultaneous estimation of model parameters, i.e., image and slowness. Next, we return to the differential formulation in order to obtain

the most general expressions for the imaging and inversion gradients. To reduce computational cost, we implement only the receiver side of the derived expressions for the gradients. In other words, the variations in the downgoing wavefield are neglected. In practice, this means that upon progress of the iterative model-parameter estimation, the upgoing wavefield is kept the same between iterations. From now on, we will omit the subscript of the wavefields related to the scattering order so as to simplify notation.

Gauss–Newton method

The Gauss–Newton method is an approximate approach to account for the effect of the inverse Hessian of the objective function on the gradient. Some benefits of the method are its capability to focus the gradient by reducing band-limiting effects from acquisition and, consequently, it may speed up the convergence of the inversion process (Pratt et al., 1998). Additionally, the Gauss–Newton approach can reduce cross-talk between parameter classes in multiparameter inversion (Pan et al., 2018).

Under the hypotheses outlined above, we want to solve the least-square problem

$$E(R^+, \sigma_L) = \frac{1}{2} \sum_{s=1}^{N_s} \int_{\omega_i}^{\omega_f} \int_{\Omega} \left| D_s^- - \mathcal{S}_s P_s^- (R^+, \sigma_L) \right|^2 d\mathbf{x} d\omega \,, \tag{28}$$

where N_s is the number of shots, the subscript s denotes one shot, D_s^- is the measured data, and' P_s^- is the upgoing wavefield at all spatial positions in the model, parameterized in dependence on the reflection coefficient R^+ and the logarithmic slowness $\sigma_L = \ln(\sigma/\sigma_0)$. Moreover, Ω is the two-dimensioonal domain in the horizontal coordinates under investigation, and $\mathcal{S}_s$ samples the wavefield at the receivers position.

Using the LSMR algorithm (Fong and Saunders, 2011), we implemented the multiparameter Gauss–Newton method by solving the least-squares subproblem

$$\min_{\Delta \sigma_{L_k}, \Delta R_k^+} \sum_{s=1}^{N_s} \int_{\omega_i}^{\omega_f} \int_{\Omega} \left| \mathcal{S}_s \left[\left. \frac{\partial P_s^-}{\partial \sigma_L} \right|_{\sigma_{L_k}} \quad \left. \frac{\partial P_s^-}{\partial R^+} \right|_{R_k^+} \right] \begin{bmatrix} \Delta \sigma_{L_k} \\ \Delta R_k^+ \end{bmatrix} - (D_s^- - \mathcal{S}_s \bar{P}_s^-) \right|^2 d\mathbf{x} d\omega \,, \tag{29}$$

where $\Delta \sigma_{L_k}$ and ΔR_k^+ are the update directions at iteration k. After estimating the update directions, the parameters are updated according to

$$\sigma_{L_{k+1}} = \sigma_{L_k} + \alpha_k \Delta \sigma_{L_k} \qquad \text{and} \qquad R_{k+1}^+ = R_k^+ + \beta_k \Delta R_k^+ \,, \tag{30}$$

where the step-lengths α_k and β_k are calculated with the subspace method (Kennett et al., 1988). At each inner iteration of the LSMR algorithm to solve equation (29), our procedure performs one linearized modeling and one adjoint modeling.

In connection with the first-order Taylor expansion (see, e.g., Camargo, 2019), the action of the wavefield partial derivative on a vector Δm_k is equivalent to the calculation of a perturbed wavefield given by

$$\Delta P^-(m_k) = P^-(m_k + \Delta m_k) - P^-(m_k) \approx \left. \frac{\partial P^-}{\partial m} \right|_{m_k} \Delta m_k \,, \tag{31}$$

where the model parameter m is R^+ or σ_L. Using the upgoing modeling equation (24) and, as an example, considering the model parameter R_k^+, we obtain

$$\left(\left. \frac{\partial P^-}{\partial R^+} \right|_{R_k^+} \Delta R_k^+ \right) (z_{n-1}) = \hat{\mathcal{G}}^-(z_{n-1}; z_n) \left(\Delta R_k^+ P^+ + T_k^- \Delta P^- \right) (z_n) \,, \tag{32}$$

where we have neglected variations in the downgoing wavefield and in the transmission coefficient.

We now take into account that the down- and upgoing Green's functions approximately relate to each other as

$$\left[\hat{\mathcal{G}}^-(z_{n-1}; z_n) \right]^* \approx \hat{\mathcal{G}}^+(z_n; z_{n-1}) \,, \tag{33}$$

where $*$ denotes complex conjugate. With this relationship, the adjoint of equation (32) can be expressed as

$$\left[\left(\frac{\partial P^-}{\partial R^+}\right)^*\bigg|_{R_k^+}\Delta P^-\right](z_n) \approx \left(P^+\right)^*(z_n)\,\hat{\mathcal{G}}^+(z_n;z_{n-1})\Delta P^-(z_{n-1})\,, \tag{34}$$

where we have also negected second-order terms in ΔP^-.

In the next section, we discuss the method of Lagrangian multipliers in more detail in order to define more precisely what we mean by adjoint modeling.

Lagrange multipliers

Our aim is to simultaneously estimate the angle-independent reflection coefficient R^+ and the logarithmic slowness σ_L. After inversion of these parameters, the other scattering coefficients are estimated using equations (14), (15), and (16). For this purpose, we adapt the the method of Lagrangian multipliers presented by Askan et al. (2007) and Métivier et al. (2017) to our problem. This methodology provides the means of calculating the partial derivatives required by the Gauss–Newton method.

For our derivations, we make use of the following relationships between the reflectivity and transmissivity operators:

$$\hat{\mathcal{R}}_c^- = -\hat{\mathcal{R}}_c^+\,, \tag{35}$$

$$\hat{\mathcal{T}}_c^\pm = \pm\hat{\mathcal{R}}_c^+\,. \tag{36}$$

In preparation to build the Lagrangian function we define the forward modeling functionals

$$F^+(P^+,P^-,\hat{\mathcal{R}}_c^+,\sigma_L) = \frac{\partial P^+}{\partial z} + i\hat{\mathcal{H}}_1 P^+ - \hat{\mathcal{R}}_c^+\left(P^+ - P^-\right)\,, \tag{37}$$

$$F^-(P^+,P^-,\hat{\mathcal{R}}_c^+,\sigma_L) = \frac{\partial P^-}{\partial z} - i\hat{\mathcal{H}}_1 P^- + \hat{\mathcal{R}}_c^+\left(P^+ - P^-\right)\,, \tag{38}$$

where we have separated the dependency on $P^\pm$ from that on the model parameters $\hat{\mathcal{R}}_c^+$ and σ_L. In the case that the wavefields $P^\pm$ satisfy the one-way wave equations (3) and (4), we have $F^\pm = 0$ and the dependency restored.

Considering the above least-squares objective function E and wave-equation functionals $F\pm$, we define, for a single shot gather, the augmented Lagrangian function

$$L(P^+,P^-,\hat{\mathcal{R}}_c^+,\sigma_L,\Lambda^+,\Lambda^-) = E(P^-) + \boldsymbol{Re}\left\{\left\langle F^+(P^+,P^-,\hat{\mathcal{R}}_c^+,\sigma_L),\Lambda^+\right\rangle\right\}$$
$$+ \boldsymbol{Re}\left\{\left\langle F^-(P^+,P^-,\hat{\mathcal{R}}_c^+,\sigma_L),\Lambda^-\right\rangle\right\}\,, \tag{39}$$

where $\Lambda^\pm(\mathbf{x},\omega)$ are Lagrange multipliers or adjoint-state variables, $\boldsymbol{Re}\{.\}$ denotes the real-part operator, and $\langle.,.\rangle$ denotes inner product. Morever, the inner product $\langle A,B\rangle$ of two complex quantities $A(\mathbf{x},\omega)$ and $B(\mathbf{x},\omega)$ is defined as

$$\langle A,B\rangle = \int_\Omega\int_{\omega_i}^{\omega_f} A^* B\,d\omega d\mathbf{x}\,, \tag{40}$$

where ω_i and ω_f are initial and final angular frequencies, respectively, and Ω represents the two-dimensional model parameters spatial domain.

Our forward modeling equations evolves in the depth coordinate, this way, we define the following boundary conditions

$$P^+(x,z=z_f,\omega) = 0\,, \tag{41}$$

$$P^-(x,z=0,\omega) = 0\,, \tag{42}$$

where z_f is the depth position at the bottom of the model.

Considering now wavefields $\bar{P}^{\pm}$ that satisfy $F^{\pm} = 0$, we conclude that for these wavefields, the Lagrangian reduces to

$$L(\bar{P}^{+}, \bar{P}^{-}, \hat{\mathcal{R}}_c^{+}, \sigma_L, \Lambda^{+}, \Lambda^{-}) = E(\bar{P}^{-}(\hat{\mathcal{R}}_c^{+}, \sigma_L)) \,. \tag{43}$$

This equivalence between L (equation (43)) and E (equation (28)) will be used in the following sections. Furthermore, the Lagrangian must be stationary at the optimum solution. Therefore, the first variation δL must vanish, i.e.,

$$\delta L = \frac{\partial L}{\partial P^{+}} \delta P^{+} + \frac{\partial L}{\partial P^{-}} \delta P^{-} + \frac{\partial L}{\partial \hat{\mathcal{R}}_c^{+}} \delta \hat{\mathcal{R}}_c^{+} + \frac{\partial L}{\partial \sigma_L} \delta \sigma_L + \frac{\partial L}{\partial \Lambda^{+}} \delta \Lambda^{+} + \frac{\partial L}{\partial \Lambda^{-}} \delta \Lambda^{-} = 0 \,. \tag{44}$$

Adjoint wavefields

The variation of the Lagrangian with respect to P^{+} is given by

$$\delta_{P^{+}} L = \boldsymbol{Re} \left\{ \left\langle -\frac{\partial \Lambda^{+}}{\partial z} + i\hat{\mathcal{H}}_1 \Lambda^{+} - \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \Lambda^{+}, \delta P^{+} \right\rangle \right\}$$
$$+ \boldsymbol{Re} \left\{ \left\langle \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \Lambda^{-}, \delta P^{+} \right\rangle \right\} \,, \tag{45}$$

where superscript $\dagger$ denotes transpose and complex conjugate. Here, the transposition of a scattering operator means a correlation between the operator and the wavefield along the lateral coordinate. Moreover, we have also taken into account that under neglection of evanescent waves (Wapenaar and Grimbergen, 1996),

$$\hat{\mathcal{H}}_1^{t} = \hat{\mathcal{H}}_1 \qquad \text{and} \qquad \hat{\mathcal{H}}_1^{*} \approx \hat{\mathcal{H}}_1 \,, \tag{46}$$

where t denotes transpose and $*$ denotes complex conjugate. Additionally, we carried out an integration by parts to move the partial derivative with respect to depth to the adjoint wavefield. For this purpose, we used the boundary condition

$$\Lambda^{+}(x, z = 0, \omega) = 0 \,. \tag{47}$$

Imposing $\delta_{P^{+}} L = 0$, we obtain

$$\frac{\partial \Lambda^{+}}{\partial z} = i\hat{\mathcal{H}}_1 \Lambda^{+} - \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \left(\Lambda^{+} - \Lambda^{-}\right) \,. \tag{48}$$

Similarly to the variation with respect to P^{+} in equation (45), the variation of the Lagrangian with respect to P^{-} is given by

$$\delta_{P^{-}} L = \left\langle -\mathcal{S}^{\dagger} \left(D^{-} - \mathcal{S} P^{-}\right), \delta P^{-} \right\rangle + \boldsymbol{Re} \left\{ \left\langle -\frac{\partial \Lambda^{-}}{\partial z} - i\hat{\mathcal{H}}_1 \Lambda^{-} - \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \Lambda^{-}, \delta P^{-} \right\rangle \right\}$$
$$+ \boldsymbol{Re} \left\{ \left\langle \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \Lambda^{+}, \delta P^{-} \right\rangle \right\} \,, \tag{49}$$

with the associated boundary condition

$$\Lambda^{-}(x, z = z_f, \omega) = 0 \,. \tag{50}$$

Imposing $\delta_{P^{-}} L = 0$, we obtain

$$\frac{\partial \Lambda^{-}}{\partial z} = -i\hat{\mathcal{H}}_1 \Lambda^{-} + \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \left(\Lambda^{+} - \Lambda^{-}\right) - \mathcal{S}^{\dagger} \left(D^{-} - \mathcal{S} P^{-}\right) \,. \tag{51}$$

Hence, equations (48) and (51) govern the adjoint modeling procedure. Using relationships (35) and (36) between scattering operators, the adjoint modeling equations can be recast into the form

$$\frac{\partial \Lambda^{-}}{\partial z} = -i\hat{\mathcal{H}}_1 \Lambda^{-} + \left(\hat{\mathcal{T}}_c^{-}\right)^{\dagger} \Lambda^{-} + \left(\hat{\mathcal{R}}_c^{+}\right)^{\dagger} \Lambda^{+} - \mathcal{S}^{\dagger} \left(D^{-} - \mathcal{S} P^{-}\right) \,, \tag{52}$$

$$\frac{\partial \Lambda^{+}}{\partial z} = i\hat{\mathcal{H}}_1 \Lambda^{+} - \left(\hat{\mathcal{T}}_c^{+}\right)^{\dagger} \Lambda^{+} - \left(\hat{\mathcal{R}}_c^{-}\right)^{\dagger} \Lambda^{-} \,. \tag{53}$$

In the integral representation of these equations, the transmissivity operator will turn into the transmission operator/coefficient. Therefore, we substitute the adjoint of the transmission by its inverse. We believe that this modification will lead to improved amplitudes of deep reflectors.

Finally, using the integral formulation and discretizing the depth coordinate, we can write the zero-order terms as

$$\Lambda^-(z_{n+1}) = \hat{\mathcal{G}}^+(z_{n+1}; z_n) \left[\frac{1}{T^-}\Lambda^- - \Delta z \mathcal{S}^\dagger \left(D^- - \mathcal{S}P^- \right) \right](z_n), \tag{54}$$

$$\Lambda^+(z_{n-1}) = \hat{\mathcal{G}}^-(z_{n-1}; z_n) \left(\frac{1}{T^+}\Lambda^+ + R^-\Lambda^- \right)(z_n), \tag{55}$$

where we have already considered the inverse of the angle-independent transmission coefficient. Note that these equations are similar to the modeling equations (25) and (26), but in the adjoint modeling case, we start from the data residuals at the receivers position and extrapolate the data residual down to the bottom of the parameter space (model). Then, we calculate the adjoint wavefield Λ^+ from the model bottom back to the top.

Migration

The variation of the Lagrangian, equation (39), with respect to $\hat{\mathcal{R}}_c^+$ is given by

$$\delta_{\hat{\mathcal{R}}_c^+} L = \boldsymbol{Re} \left\{ \left\langle -\left(\hat{\mathcal{P}}^+ - \hat{\mathcal{P}}^- \right)^\dagger \Lambda^+, \delta \mathcal{R}_c^+ \right\rangle \right\} + \boldsymbol{Re} \left\{ \left\langle \left(\hat{\mathcal{P}}^+ - \hat{\mathcal{P}}^- \right)^\dagger \Lambda^-, \delta \mathcal{R}_c^+ \right\rangle \right\}, \tag{56}$$

where the adjoint of the wavefield operator is defined as

$$\left(\hat{\mathcal{P}}^+ \right)^\dagger \Lambda^+ = \int_{\mathbb{R}} \left[P^+(x, z, \omega; x') \right]^* \Lambda^+(x, z, \omega) dx, \tag{57}$$

with $*$ denoting the complex conjugate.

Using identity (43), the partial derivative of the objective function with respect to the reflectivity is given by

$$\frac{\partial E}{\partial \hat{\mathcal{R}}_c^+} = -\left(\hat{\bar{\mathcal{P}}}^+ - \hat{\bar{\mathcal{P}}}^- \right)^\dagger \left(\bar{\Lambda}^+ - \bar{\Lambda}^- \right), \tag{58}$$

where $\bar{\Lambda}^\pm$ are the adjoint wavefields calculated using the forward wavefields $\bar{P}^\pm$ that satisfy $F^\pm = 0$.

In the angle-independent and discrete model case, using equation (12) and the chain rule, we obtain

$$\frac{\partial E}{\partial \hat{\mathcal{R}}_c^+} \frac{\partial \hat{\mathcal{R}}_c^+}{\partial R^+} = \frac{1}{\Delta z} \sum_{s=1}^{N_s} \boldsymbol{Re} \left\{ \int_{\omega_i}^{\omega_f} \left[\bar{P}_s^+ - \bar{P}_s^- \right]^* \left[\bar{\Lambda}_s^+ - \bar{\Lambda}_s^- \right] d\omega \right\}, \tag{59}$$

where we have made explicit the sum over shot-gathers. Moreover, we have multiplied the expression by -1 to obtain a descent direction. Note that the integration over angular frequency is kept because the angle-independent reflection coefficient is also independent of angular frequency. This new imaging condition for seismic imaging needs to be further investigated in the future.

In this work, we discard variations in the downgoing wavefield, which means neglecting the adjoint wavefield Λ^+ in the Lagrangian (equation (39)). Upon also neglecting the upgoing wavefield in expression (59), it simplifies to

$$\frac{\partial E}{\partial R^+} \approx -\frac{1}{\Delta z} \sum_{s=1}^{N_s} \boldsymbol{Re} \left\{ \int_{\omega_i}^{\omega_f} \left[\bar{P}_s^+ \right]^* \left[\bar{\Lambda}_s^- \right] d\omega \right\}. \tag{60}$$

Note that this result is basically the conventional imaging condition, i.e., the zero-lag cross-correlation between the complex conjugate of the downgoing wavefield with the backpropagated residuals.

On the other hand, direct derivation of the objective function in equation (28) with respect to R^+ provides

$$-\frac{\partial E}{\partial R^+} = \sum_{s=1}^{N_s} \boldsymbol{Re}\left\{\int_{\omega_i}^{\omega_f} \left(\mathcal{S}_s\frac{\partial P_s^-}{\partial R^+}\right)^\dagger \left(D_s^- - \mathcal{S}_s P_s^-\right) d\omega\right\}. \tag{61}$$

A comparison of equations (60) and (61) leads to an interpretation of the gradient by means of the adjoint wavefield partial derivative. We conclude that the action of the adjoint of the upgoing-wavefield partial derivative on a residual wavefield ΔP is calculated as a product of the complex conjugate of the downgoing wavefield and the backpropagated data residual ΔP. This conclusion agrees with the adjoint equation (34) derived by other means.

Inversion

The variation of the Lagrangian in equation (39) with respect to σ_L is given by

$$\delta_{\sigma_L} L = \left\langle \Lambda^+, i\frac{\partial\hat{\mathcal{H}}_1}{\partial\sigma_L}P^+\delta\sigma_L \right\rangle + \left\langle \Lambda^-, -i\frac{\partial\hat{\mathcal{H}}_1}{\partial\sigma_L}P^-\delta\sigma_L \right\rangle. \tag{62}$$

Using identity (43), the partial derivative of the objective function with respect to σ_L can be written as

$$\frac{\partial E}{\partial\sigma_L} = -\sum_{s=1}^{N_s}\int_{\omega_i}^{\omega_f} \boldsymbol{Re}\left\{i\frac{\partial\hat{\mathcal{H}}_1}{\partial\sigma_L}\left[\bar{P}_s^+\left(\bar{\Lambda}_s^+\right)^* - \bar{P}_s^-\left(\bar{\Lambda}_s^-\right)^*\right]d\omega\right\}, \tag{63}$$

which we multiplied again by -1 to obtain a descent direction.

Discarding variations in the downgoing wavefield, i.e., neglecting Λ^+, we obtain

$$\frac{\partial E}{\partial\sigma_L} \approx \sum_{s=1}^{N_s}\int_{\omega_i}^{\omega_f} \boldsymbol{Re}\left\{i\frac{\partial\hat{\mathcal{H}}_1}{\partial\sigma_L}\left[\bar{P}_s^-\left(\bar{\Lambda}_s^-\right)^*\right]d\omega\right\}. \tag{64}$$

In analogy to the interpretation of equation (61), the action of the partial derivative of the upgoing-wavefield adjoint on a wavefield ΔP is given by equation (64) with the data residual in the adjoint wavefield substituted by the ΔP.

We highlight that even after discarding variation in the downgoing wavefield, the partial derivatives of the objective function with respect to R^+ and σ_L, still depend on wavefields propagating in different directions. The imaging gradient, equation (60), depends on the downgoing wavefield P^+, whereas the inversion gradient, equation (64), depends on the upgoing wavefield P^-. These expressions for the partial derivatives of the objective function represent the so-called receiver side and, in order to reduce computational cost, that is how we implement the derivatives involved in the Gauss–Newton method.

NUMERICAL TESTS

We have applied our formulation of JMI to two synthetic data sets. One model is similar to the lens model used in Masaya and Verschuur (2018), and the second model is the modified Marmousi2 model (Pan et al., 2018). The source wavelet in both synthetic data sets was a 20 Hz Ricker wavelet. We used the same algorithm for modeling and inversion with a multiscale approach. In the discussion below, when we indicate the Gauss–Newton computational cost, we consider that one linearized modeling step has approximately the same computational cost of one adjoint modeling step. For comparison, we include the results of the conventionally used steepest-descent method.

Lens model

The acquisition geometry for the first test consisted of 40 shots spaced by 50 m and receivers at all surface grid points. The vertical and lateral spatial sampling of the model was 12.5 m. Figure 1 shows the initial velocity model for the first test, a linear vertical gradient model. The initial image was a null array. Figures 2(a) and 2(b) show the exact velocity and image, respectively.
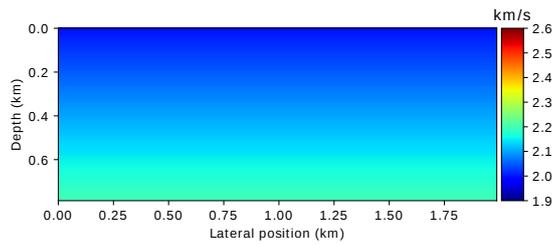
Figure 1: Lens model: Initial velocity model for JMI.

(a)

(b)
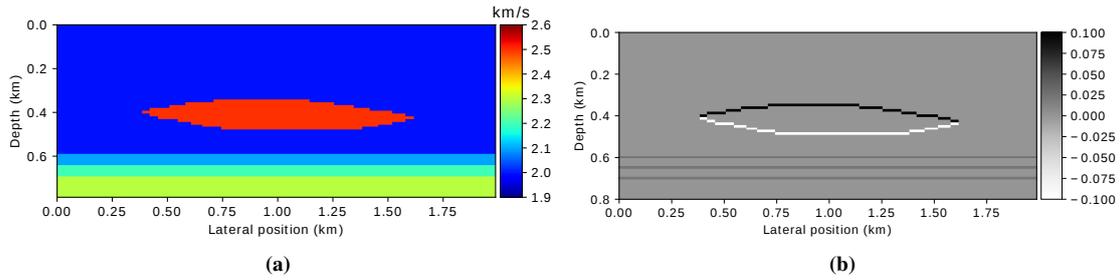
Figure 2: Lens model: (a) Exact velocity model; (b) Exact image.

(a)

(b)

(c)

(d)
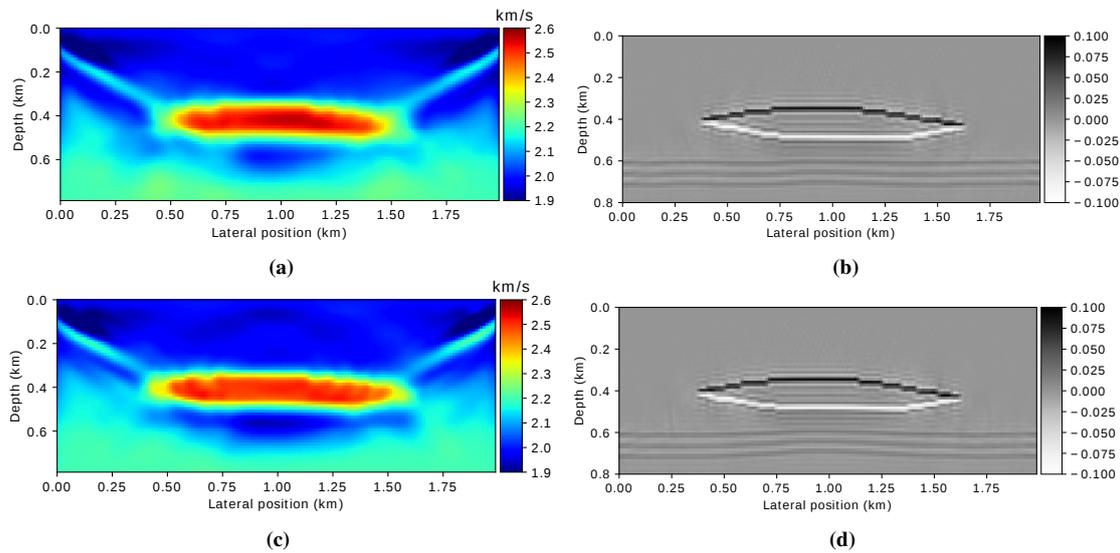
Figure 3: Results from the Lens model. Tomography with steepest descent: (a) Velocity; (b) Image from tomography with steepest descent; (c) Velocity; (d) Image from tomography with Gauss–Newton.

(a)

(b)
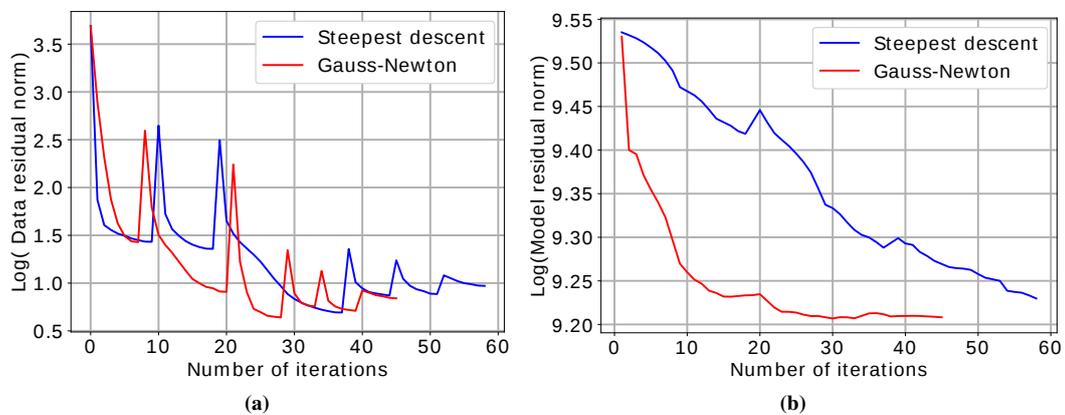
Figure 4: Results from the Lens model, residual over iterations. (a) Data residual; (b) Model residual.

For the inversion, the frequency stages were divided into four equal intervals with fixed minimum frequency. The minimum frequency was set to the first sample after 0 Hz and the first maximum frequency was 15 Hz. The number of iterations per stage was set to a maximum of 20 and a jump to a new stage was imposed if the relative decrease of the objective function was less than 1%.

The results of the inversion with the steepest-descent and Gauss–Newton methods are depicted in Figure 3. Overall, the results were acceptably close to the exact model (compare with Figure 2). The Gauss–Newton solution, Figure 3(c), exhibits a slightly higher resolution and more precision in reproducing the length and positioning of the lens than the steepest-descent result, Figure 3(a). Moreover, the fine layers right below the lens are better positioned in the Gauss–Newton image, Figure 3(d) than in the steepest-descent image, Figure 3(b). The apparent pull-up of these layers in Figure 3(d), which is slightly stronger than in Figure 3(b), is a result of more inaccurate positioning of the horizontal reflectors to the sides of the image, where the illumination was poorer. The area of full illumination can be inferred by the light blue boundary effect connecting the corners of the lens to the upper corners of the image.

Other indicators for the quality of the optimization strategies are the data and model residuals. Figure 4 shows these residuals in plots over iterations. The data and model residuals show that the Gauss–Newton method provided model updates which converged considerably faster and lead to a better final velocity model. The average number of inner iterations in the Gauss–Newton approach was 4.3. Therefore, the computational cost of each iteration was approximately 8.6 times that of the steepest-descent method.

Marmousi2 model

Figures 5 to 9 show the same sequence of figures for a modified version of the Marmousi2 model. For better visualization and comparison, we have separated the model from the image figures. Figure 5 depicts the initial velocity model, a strongly smoothed version of the exact velocity model, shown in Figure 6(a). The initial image was again a null array.

In this case, the vertical and lateral spatial sampling was 5 m. The acquisition geometry consisted of 22 shots spaced at 150 m and receivers at all surface grid points. The frequency stages were divided into four intervals with an increment of 10 Hz, and the first stage was defined as 0-10 Hz. The maximum number of iterations per stage was set to 20, skipping to a new stage if the relative decrease of the objective function was less than 5%. Both the steepest-descent and Gauss–Newton methods, Figures 7(a) and 7(b), were capable of introducing several details into the initial velocity model, Figure 5. The steepest-descent result makes an overall somewhat smoother impression than the Gauss–Newton velocity model. A number of details, e.g. the low-velocity layers in the lower part of the model, are better resolved in the Gauss–Newton velocity model.

Figure 8 compares the resulting images to the true reflectivity. The resulting images of the steepest-descent, Figure 8(b) and Gauss–Newton inversion, Figure 8(c), exhibit approximately the same quality. Differences in reflector positioning or continuity are hard to spot. Note, however, that the gas lens is visibly better resolved in the Gauss–Newton image.

Figure 9 compares the convergence behavior of the data and model residuals. We recognize that Gauss–Newton converged slower in the first stage than steepest descent but subsequently provided a more significant decrease of the data and model residuals, Figures 9(a) and 9(b). The average number of inner iterations to solve the Gauss–Newton approximation was 6.2. Therefore the computational cost of each iteration was approximately 12.4 times that of steepest descent.

CONCLUSION

In this paper, we have studied Joint Migration Inversion (JMI). Differently from what is usually done in the literature on JMI, we have reviewed the continuous forms of the differential and integral equations upon which the JMI methodology is built. Considering the inverse problem, in general, the JMI gradients are derived from the integral equations after discretization. We obtained the gradients using the original underlying differential equations together with the continuous form of the Lagrange multipliers.

The complete formulation of the Lagrangian with respect to variations in the downgoing and upgoing wavefields leads to a new imaging condition. Even upon neglection of variations in the downgoing wavefield, this result suggests that the imaging condition should be composed of the difference between
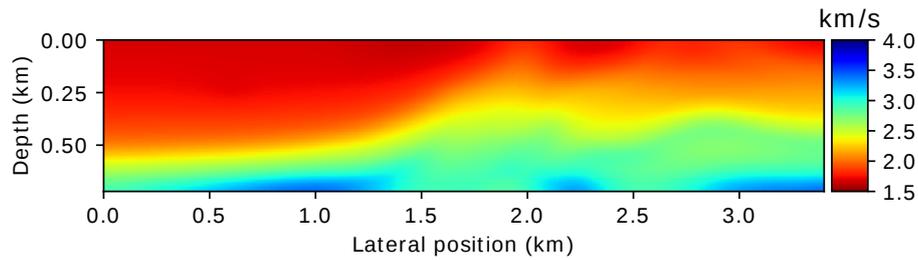
Figure 5: Marmousi2 model: Initial velocity model for JMI.
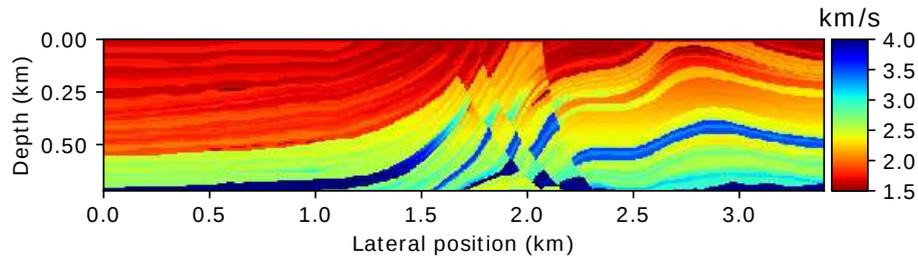
(a)

Figure 6: Marmousi2 model: Exact velocity model; (a) Exact image.
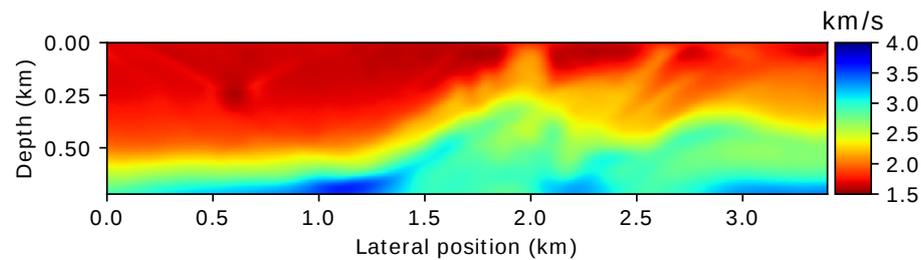
(a)
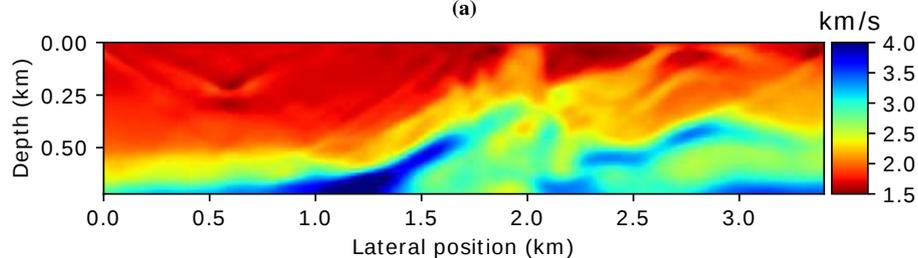
(b)

Figure 7: Marmousi2 model. (a) Exact velocity model; (a) Velocity from tomography with steepest descent; (b) Velocity from tomography with Gauss–Newton.

downgoing and upgoing wavefields correlated with the backpropagated seismic data from the receivers. The effectiveness of this modified imaging condition is the subject of ongoing research.

We have also discussed the implementation of the multiparameter Gauss–Newton method to simultaneously estimate updates for the scattering operator and the medium slowness. Our numerical tests on two synthetic models of different degrees of geologic complexity indicate that the computationally more expensive Gauss–Newton method can provide higher resolution in the resulting velocity models than the steepest-descent method. However, although the inverted velocity models showed some significant differences, the quality of the estimated reflectivity images turned out to be somewhat independent of the tested optimization algorithm. Note that in order to fairly compare these optimization methods, we chose identical stopping criteria for both inversion algorithms.

In conclusion, JMI is a promising methodology for imaging and velocity model building. However,

(a)

(b)

(c)
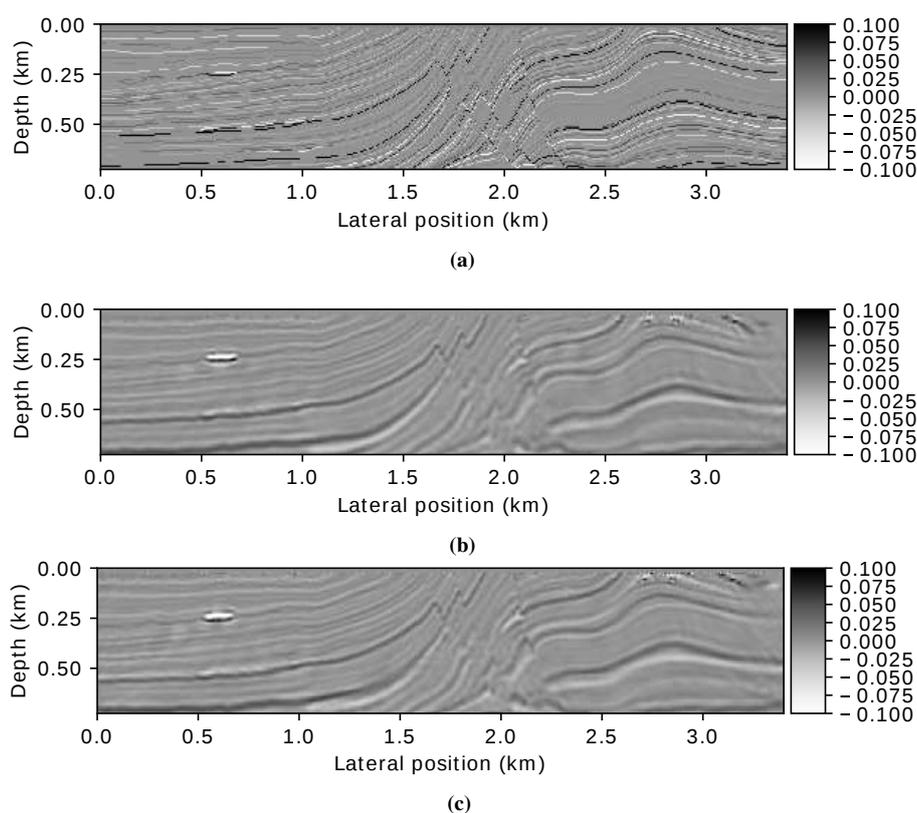
Figure 8: Marmousi2 model: (a) Exact image. (b) Image from tomography with steepest descent; (c) Image from tomography with Gauss–Newton.

the assumptions considered here, mainly the angle independence of the scattering operators, should be removed in order to reduce probable restrictions when applied to real data. Overall, the obtained velocity models seem to be of sufficient quality to serve as initial models for a subsequent FWI.

ACKNOWLEDGMENTS

REFERENCES

Askan, A., Akcelik, V., Bielak, J., and Ghattas, O. (2007). Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures. *Bulletin of the Seismological Society of America*, 97(6):1990–2008.

Berkhout, A. G. (2014a). Review paper: An outlook on the future of seismic imaging, Part I: forward and reverse modelling. *Geophysical Prospecting*, 62(5):911–930.

Berkhout, A. G. (2014b). Review paper: An outlook on the future of seismic imaging, Part III: Joint migration inversion. *Geophysical Prospecting*, 62(5):950–971.

Berteussen, K. A. and Ursin, B. (1983). Approximate computation of the acoustic impedance from seismic data. *GEOPHYSICS*, 48(10):1351–1358.
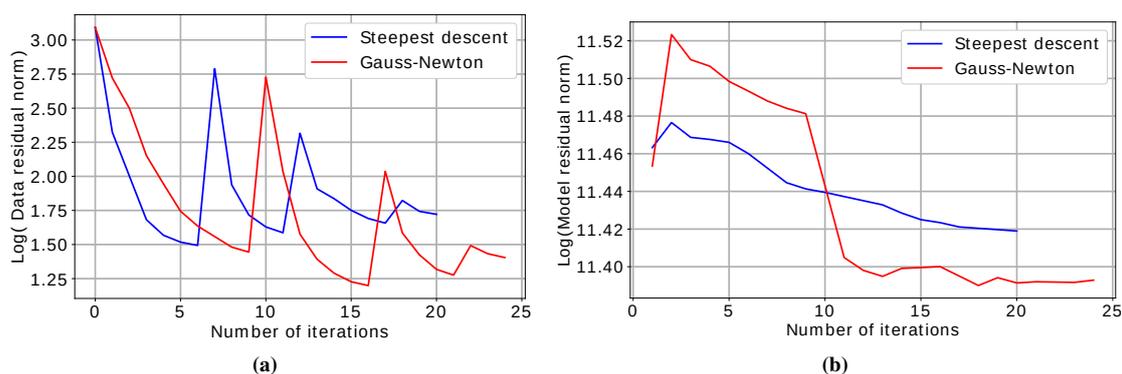
Figure 9: Results from the Marmousi2 model, residual over iterations. (a) Data residual; (b) Model residual.

Bremmer, H. (1951). The W.K.B. approximation as the first term of a geometric-optical series. *Communications on Pure and Applied Mathematics*, 4(1):105–115.

Camargo, A. W. (2019). *Analysis of the full waveform inversion using the augmented lagrangian method.* PhD thesis, University of Campinas, São Paulo, Brazil.

de Bruin, C. G. M., Wapenaar, C. P. A., and Berkhout, A. J. (1990). Angle-dependent reflectivity by means of prestack migration. *GEOPHYSICS*, 55(9):1223–1234.

Fong, D. C.-L. and Saunders, M. (2011). LSMR: An iterative algorithm for sparse least-squares problems. *SIAM Journal on Scientific Computing*, 33(5):2950–2971.

Grimbergen, J. L. T., Dessing, F. J., and Wapenaar, K. (1998). Modal expansion of one-way operators in laterally varying media. *GEOPHYSICS*, 63(3):995–1005.

Jones, I. (2010). Tutorial: Velocity estimation via ray-based tomography. *First Break*, 28(1719).

Kennett, B. L. N., Sambridge, M. S., and Williamson, P. R. (1988). Subspace methods for large inverse problems with multiple parameter classes. *Geophysical Journal International*, 94(2):237–247.

Masaya, S. and Verschuur, D. J. E. (2018). Iterative reflectivity-constrained velocity estimation for seismic imaging. *Geophysical Journal International*.

Métivier, L., Brossier, R., Operto, S., and Virieux, J. (2017). Full waveform inversion and the truncated newton method. *SIAM Review*, 59(1):153–195.

Pan, W., Geng, Y., and Innanen, K. A. (2018). Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and hussar land data set application. *Geophysical Journal International*, 213(2):1305–1333.

Pratt, G., Shin, C., and Hicks (1998). Gauss-newton and full newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362.

Santos, H. B. (2015). *Initial depth model building using robust time-migration velocity analysis methods.* Ph.D.'s dissertation, Petroleum Science and Engineering, University of Campinas, Campinas, São Paulo.

Santos, H. B., Schleicher, J., Novais, A., Kurzmann, A., and Bohlen, T. (2015). Robust time-domain migration velocity analysis methods for initial-model building in a full waveform tomography workflow. Annual WIT Report No. 19.

Santos, H. B., Schleicher, J., Novais, A., Kurzmann, A., and Bohlen, T. (2016a). Robust time-domain migration velocity analysis for initial-model building in a full-waveform tomography workflow. In *SEG Technical Program Expanded Abstracts 2016*, pages 5307–5312, Dallas. Society of Exploration Geophysicists.

Santos, H. B., Schleicher, J., Novais, A., Kurzmann, A., and Bohlen, T. (2016b). Robust time-domain migration velocity analysis methods for initial-model building in a full waveform tomography workflow. In *78th EAGE Conference and Exhibition 2016*, Vienna. EAGE Publications.

Santos, H. B., Schleicher, J., Novais, A., Kurzmann, A., and Bohlen, T. (2016c). Robust time-domain migration velocity analysis methods for initial-model building in a full waveform tomography workflow. In *VII Simpósio Brasileiro de Geofísica (SimBGf)*, Ouro Preto. Brazilian Geophysical Society (SBGf).

Schleicher, J., Tygel, M., and Hubral, P. (2007). *Seismic True-Amplitude Imaging*. Society of Exploration Geophysicists.

Thorbecke, J. W., Wapenaar, K., and Swinnen, G. (2004). Design of one-way wavefield extrapolation operators, using smooth functions in WLSQ optimization. *GEOPHYSICS*, 69(4):1037–1045.

Ursin, B., Pedersen, Ø., and Arntsen, B. (2012). Flux-normalized wavefield decomposition and migration of seismic data. *GEOPHYSICS*, 77(3):S83–S92.

Wapenaar, C. P. A. (1996). One-way representations of seismic data. *Geophysical Journal International*, 127(1):178–188.

Wapenaar, C. P. A. and Grimbergen, J. L. T. (1996). Reciprocity theorems for one-way wavefields. *Geophysical Journal International*, 127(1):169–177.

Xu, S., Wang, D., Chen, F., Zhang, Y., and Lambare, G. (2012). Full waveform inversion for reflected seismic data. In *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*. EAGE Publications BV.