



# Wavefield separation methods for gradient filtering in time-domain full-waveform inversion

Leonardo Gómez, CPGG/UFBA & Reynam C. Pestana\*, CPGG/UFBA

Copyright 2017, SBGf - Sociedade Brasileira de Geofísica.

This paper was prepared for presentation at the 15<sup>th</sup> International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, July 31 to August 3, 2017.

Contents of this paper were reviewed by the Technical Committee of the 15<sup>th</sup> International Congress of The Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of The Brazilian Geophysical Society is prohibited.

## Abstract

**In this work, we review multiscale approach based on the direction of wavefield propagation when applied to full-waveform inversion (FWI). This strategy states that low-wavenumber model updates can be obtained by selective correlation of source and residuals wavefields in gradient vector construction. By enhancing this component the convergence to global minimum is improved and cycle-skipping artifacts are avoided.**

**Using synthetic data, the numerical implementation of this hierarchical strategy is performed by adapting three different wavefield separation methodologies: implicit and explicit separation in the vertical wavenumber domain and via Poynting vector. We state explicit equations for normalizing these components, in order to ensure a descent direction for the model update, and for computing gradient components when one of these methods is used. For the inversion algorithm, we adopt a scheme in which the step length is estimated via quadratic interpolation, being adjusted at each iteration as solution approaches the global minimum. Alternative non-quadratic functionals for objective function are also implemented, showing the applicability of the method in those cases.**

## Introduction

The problem of estimation of subsurface model parameters through seismic data constitute the problem that full-waveform inversion (FWI) attempts to solve. Since the works of Tarantola (1984) and Lailly (1983) FWI becomes a realizable technique thanks to the reformulation done by them to the reverse-time migration principle as a velocity model building approach. In the inversion process, an objective function that measures the mismatch between observed and synthetic data obtained for the estimated model is minimized.

Since the relation between model parameters and objective function is highly nonlinear and local optimization methods for solving the minimization problem are used, the convergence to global minimum is affected by secondary minima. Several multiscale approaches has been proposed as alternatives for dealing with this problem, which use several elements of seismic data and gradient

vector that are less sensitive to local minima.

In the domain of the gradient, the multiscale approach stated by Mora (1989) has been object of considerable attention due mainly to its ability for recovering low-wavenumber updates from any frequency component in seismic data. In this work we explore the applicability of three wavefield separation methods for extracting transmission and reflection model updates. Those methodologies are implement in the vertical wavenumber domain ( $k_z$ ) and by evaluating Poynting vector. We also study how this hierarchical approach is compatible with alternative non-quadratic objective functions. Equations for computing and normalizing gradient components are given.

In the inversion algorithm, a variation of the objective function perturbation method is implemented for estimating the step-length of the updates via quadratic interpolation. The algorithm adjust the amount in which current model is perturbed, as it approaches global minimum.

## Theory

Considering an isotropic and homogeneous physical medium, with constant density, the propagation of an acoustic oscillatory perturbation is completely describe, in both space an time, by the two-way wave equation:

$$\frac{1}{c(\mathbf{x})^2} \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} - \nabla^2 u(\mathbf{x}, t) = s(\mathbf{x}, t) \delta(\mathbf{x} - \mathbf{x}_s), \quad (1)$$

where  $u(\mathbf{x}, t)$  represents the seismic wavefield at time  $t$  and for position  $\mathbf{x} = (x, y, z)$ .  $c(\mathbf{x})$  is the velocity of propagation of the acoustic wave. Source term, acting on the position  $\mathbf{x}_s$ , is denoted by  $s(\mathbf{x}, t)$  and  $\nabla^2$  corresponds to Laplacian operator in Cartesian coordinates. In this work, we use the rapid expansion method (REM) for our extrapolation operator (Pestana and Stoffa, 2010).

The non-linear inverse problem that FWI attempts to solve is to estimate the seismic wavefield  $u(\mathbf{x}, t)$  and the distribution of model parameters in subsurface  $\mathbf{m}$  such that, satisfying the modeling operator (1), the distance between calculated data  $\mathbf{d}_{\text{cal}}$  for the estimated model and observed data  $\mathbf{d}_{\text{obs}}$  be minimum. The concept of distance is defined by a functional called objective function, that is minimized in the iterative algorithm.

The least-squares norm, or  $l_2$  norm, is widely used for defining the FWI objective function. It is given by (Tarantola, 1984)

$$E_{l_2}(\mathbf{m}) = \frac{1}{2} \|\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}\|_2^2. \quad (2)$$

where a sum over source  $s$ , receivers  $r$  and time samples until total record length  $T$  is performed. Alternatively, other

forms for  $E(\mathbf{m})$  can be adopted, such as the least-absolute value norm, or  $l_1$ -norm (Brossier et al., 2010)

$$E_{l_1}(\mathbf{m}) = \|\mathbf{d}_{\text{cal}} - \mathbf{d}_{\text{obs}}\|_1, \quad (3)$$

and the cross-correlation based objective function (Zhang et al., 2015; Klimm, 2013)

$$E_c(\mathbf{m}) = -\sum_s \frac{\langle \mathbf{d}_{\text{cal}}, \mathbf{d}_{\text{obs}} \rangle}{\|\mathbf{d}_{\text{cal}}\|_2 \|\mathbf{d}_{\text{obs}}\|_2}, \quad (4)$$

which measures the similarity between data vectors.  $\langle \cdot, \cdot \rangle$  represents inner product.

### Inversion scheme

A local gradient-based iterative scheme is given by (Nocedal and Wright, 2006)

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k, \quad (5)$$

where,  $\mathbf{p}_k$  stands for the search vector in the direction of the minimum of the objective function in the neighborhood of the current model  $\mathbf{m}_k$ . The step-length  $\alpha_k$  scales the search direction before updating the model. We use steepest-descent method for minimizing the FWI objective function, so the updated model is searched in the opposite direction given by the gradient vector evaluated at the position of the current model, i.e.,  $\mathbf{p}_k = -\mathbf{g}(\mathbf{m}_k)$ .

Gradient vector contains the local sensibility information of objective function with respect to each model parameter. Conventionally, it is estimated via (Tarantola, 1984)

$$\nabla E(\mathbf{m}) = \mathbf{g}(\mathbf{m}) = \frac{2}{c(\mathbf{x})^3} \sum_s \int_0^T \lambda(\mathbf{x}, t) \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} dt, \quad (6)$$

assuming a point-collocation scheme for model parametrization. The wavefield  $u(\mathbf{x}, t)$  corresponds to the state variable, and  $\lambda(\mathbf{x}, t)$  to the adjoint-state variable, obtained by solving the reverse-time modeling problem

$$\frac{1}{c(\mathbf{x})^2} \frac{\partial^2 \lambda(\mathbf{x}, t)}{\partial t^2} - \nabla^2 \lambda(\mathbf{x}, t) = \frac{\partial E}{\partial \mathbf{u}}(\mathbf{x}, t), \quad (7)$$

with final conditions  $\lambda(\mathbf{x}, t = T) = 0$  and  $\partial \lambda(\mathbf{x}, t = T) / \partial t = 0$ .

Step length is estimated by quadratic interpolation as depicted in Figure 1. We use a modified version of objective function perturbation method proposed by Pica et al. (1990) in which two additional forward modeling problems need to be solved. Firstly, the objective function is evaluated for the model perturbed in a proportion given by  $\varepsilon_1$ . Then, the amount  $\varepsilon_2$ , used for the second forward modeling problem, is determined based on the relation between  $E(0)$  and  $E(\varepsilon_1)$  as described in Algorithm 1. The coefficients of the parabola defined by the points  $(0, E(0))$ ,  $(\varepsilon_1, E(\varepsilon_1))$  and  $(\varepsilon_2, E(\varepsilon_2))$  are obtained by the solution of the least-squares problem stated by them, letting to compute the minimum of the parabola. The reduction in the perturbation proportions with iterations makes this scheme stable, even when current model is close to global minimum.

Finally, the step length is validated using first Wolfe condition in a backtracking line-search iterative method (Nocedal and Wright, 2006), for ensuring decrease in the value of the objective function.

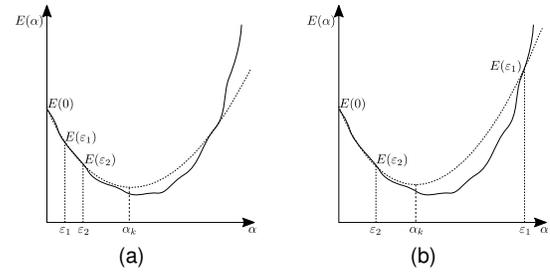


Figure 1: Step-length estimation by quadratic interpolation. Both cases described in Algorithm 1 are depicted: for (a)  $E(\mathbf{m}_k + \varepsilon_1 \mathbf{p}_k) < E(\mathbf{m}_k)$  and (b)  $E(\mathbf{m}_k + \varepsilon_1 \mathbf{p}_k) > E(\mathbf{m}_k)$ .

### Algorithm 1 Decreasing perturbation method for step-length estimation

- 1: solve  $\max(\varepsilon_1 |\mathbf{p}_k|) \leq \frac{\max(|\mathbf{m}_k|)}{100\sigma_k}$ ;
- 2: compute  $E(\mathbf{m}_k + \varepsilon_1 \mathbf{p}_k)$ ;
- 3:  $\psi_\varepsilon \leftarrow r[0, 1]$ ;
- 4:  $\psi_\sigma \leftarrow r[0, 1]$ ;
- 5: **if**  $E(\mathbf{m}_k + \varepsilon_1 \mathbf{p}_k) < E(\mathbf{m}_k)$  **then**
- 6:      $\varepsilon_2 \leftarrow 10\psi_\varepsilon \varepsilon_1$ ;
- 7:      $\sigma_{k+1} \leftarrow \sigma_k$ ;
- 8: **else**
- 9:      $\varepsilon_2 \leftarrow \psi_\varepsilon \varepsilon_1$ ;
- 10:     $\sigma_{k+1} \leftarrow \psi_\sigma \sigma_k$ ;
- 11: **end if**
- 12: compute  $E(\mathbf{m}_k + \varepsilon_2 \mathbf{p}_k)$ ;

### Gradient-domain multiscale approach

Multiscale approach consists on the restriction of the solution space of the problem in several scales such that, for the longest scales, a linear-like version of the objective function  $\mathcal{L}(\mathbf{m})$  is created. The solution estimated at each inversion stage is used as starting point for the next one, until the original problem is solved completely, thus avoiding convergence to local minima in the objective function.

Two main classes of multiscale strategies can be identified: those defined in the data domain, for which the restriction operator acts on seismograms before evaluating objective function and gradient vector, and those formulated in the domain of the gradient, which use a selective cross-correlation between source and residuals wavefields in gradient construction.

The data-domain multiscale approaches use elements of seismic data, such as frequency content (Bunks et al., 1995), time arrival (Chen et al., 2015) and offset between sources and receivers (Klimm, 2013), for extracting updates that improve the convergence to the global minimum of the objective function.

Following Tang et al. (2013), the wavefields  $u(\mathbf{x}, t)$  and  $\lambda(\mathbf{x}, t)$  can be considered as locally separable using plane waves and that this decomposition is reversible. Thus, the conventional gradient (6) can be viewed as the sum of contributions due to wavefields that propagate in a specific direction

$$\mathbf{g}(\mathbf{m}) = \sum_{\theta_u} \sum_{\theta_\lambda} \mathbf{g}(\mathbf{m}, \theta_u, \theta_\lambda), \quad (8)$$

being  $\theta_u$  and  $\theta_\lambda$  the local propagation angles for source and residuals wavefield, respectively. An attractive characteristic of gradient-domain multiscale approaches is that low-frequency components are not required for obtaining low-wavenumber updates, since they can be recovered for any frequency (Tang et al., 2013), given that seismic data recorded for sufficient long-offset is available.

Equation (8) derives in two ways for recovering low-wavenumber components of the model, as proposed in the literature. One of them states that the background model can be updated by considering only high-aperture angles in gradient construction (Alkhalifah, 2015). The second strategy follows the conclusions pointed out by Mora (1989), who demonstrated that gradient vector for FWI is formed by two components, a low-wavenumber or tomographic component  $\mathbf{g}_r(\mathbf{m})$  and a high-wavenumber or migration component  $\mathbf{g}_m(\mathbf{m})$ , i.e.,  $\mathbf{g}(\mathbf{m}) = \mathbf{g}_r(\mathbf{m}) + \mathbf{g}_m(\mathbf{m})$ . Low-wavenumber updates are due to transmission phenomena in the physical medium. Reflected waves, on the other hand, lead to recovering of high-wavenumber regions in the spectrum (Mora, 1989). In practice, the tomographic component is computed by correlating source and residuals wavefields selectively, only in those points in which they propagate in similar directions. On the other hand, the migration component is the result of the correlation of  $u(\mathbf{x}, t)$  and  $\lambda(\mathbf{x}, t)$  when they travel in opposite directions (Tang et al., 2013)

$$\mathbf{g}_r(\mathbf{m}) = \sum_{\theta_u} \sum_{\theta_\lambda} \mathbf{g}(\mathbf{m}, \theta_u, \theta_\lambda), \quad \text{for } \langle \hat{\mathbf{n}}(\theta_u), \hat{\mathbf{n}}(\theta_\lambda) \rangle > 0 \quad (9a)$$

$$\mathbf{g}_m(\mathbf{m}) = \sum_{\theta_u} \sum_{\theta_\lambda} \mathbf{g}(\mathbf{m}, \theta_u, \theta_\lambda), \quad \text{for } \langle \hat{\mathbf{n}}(\theta_u), \hat{\mathbf{n}}(\theta_\lambda) \rangle \leq 0, \quad (9b)$$

where,  $\hat{\mathbf{n}}(\theta_u)$  and  $\hat{\mathbf{n}}(\theta_\lambda)$  correspond to the unitary vectors in the direction of  $\theta_u$  and  $\theta_\lambda$ , respectively. Several methods can be used for extracting unidirectional wavefield components and any direction of propagation can be considered when computing (9). Here, we consider upward and downward distinction for wave propagation direction, which is the approach most widely used.

The conventional FWI gradient assumes that transmission and reflection components are equally weighted. Tomographic component can be enhanced, in order to improve low-wavenumber updates in the model, by using a given coefficient  $\beta$  (Tang et al., 2013),

$$\nabla \mathcal{L}(\mathbf{m}) = \beta \mathbf{g}_r(\mathbf{m}) + \mathbf{g}_m(\mathbf{m}), \quad (10)$$

which can be constant or variable for different model parameters. Any of the gradient components can be favored by using different values for  $\beta$ . The hierarchical inversion strategy states that high values for the weighting coefficient should be used at first iterations in order to update low-wavenumber components in the model. As this value is diminished, high-wavenumber updates are included in the model.

Since only partial seismic information is used for building transmission and reflection components separately, these can differ significantly from the conventional gradient vector or even lead to updates in a direction opposite to the decreasing path in the objective function (Wu and Alkhalifah, 2014). Thus, it is necessary to normalize the tomographic and migration components with respect to the

total gradient by applying the following coefficients

$$\beta_t = \frac{(\mathbf{g} \cdot \mathbf{g})[(\mathbf{g}_m \cdot \mathbf{g}_m)(\mathbf{g} \cdot \mathbf{g}_t) - (\mathbf{g}_t \cdot \mathbf{g}_m)(\mathbf{g} \cdot \mathbf{g}_m)]}{B} \quad (11a)$$

$$\beta_m = \frac{(\mathbf{g} \cdot \mathbf{g})[(\mathbf{g}_t \cdot \mathbf{g}_t)(\mathbf{g} \cdot \mathbf{g}_m) - (\mathbf{g}_t \cdot \mathbf{g}_m)(\mathbf{g} \cdot \mathbf{g}_t)]}{B}, \quad (11b)$$

respectively, with  $B = (\mathbf{g}_t \cdot \mathbf{g}_t)(\mathbf{g} \cdot \mathbf{g}_m)^2 - 2(\mathbf{g}_t \cdot \mathbf{g}_m)(\mathbf{g} \cdot \mathbf{g}_t)(\mathbf{g} \cdot \mathbf{g}_m) + (\mathbf{g}_m \cdot \mathbf{g}_m)(\mathbf{g} \cdot \mathbf{g}_t)^2$ . These coefficients are obtained from the minimization of the function  $f(\beta_t, \beta_m) = \|\beta_t \mathbf{g}_t + \beta_m \mathbf{g}_m\|_2^2$ , subject to  $g(\beta_t, \beta_m) = \mathbf{g} \cdot (\beta_t \mathbf{g}_t + \beta_m \mathbf{g}_m) > \mathbf{g} \cdot \mathbf{g} = \text{cte}$  (Wu and Alkhalifah, 2014), solved by using, e.g., Lagrange multipliers.

## Methods for wavefield separation

### Implicit separation in vertical wavenumber domain

Wavefield separation is required at each subsurface position. Wavefield evolution in time can be considered as a set of vertical seismic profiles (VSP),  $u(t, z)$ , for fixed positions in surface  $(x, y)$ . Let  $U(\omega, k_z)$  be the 2D Fourier transform of  $u(t, z)$ , with  $\omega$  stands for the temporal frequency and  $k_z$  for the wavenumber in  $z$  direction. Since the direction of wave propagation is determined by phase velocity,  $v = \omega/k$ , wavefield can be separated in Fourier domain as (Hu and McMechan, 1987)

$$U_u(\omega, k_z) = \begin{cases} U(\omega, k_z), & \omega k_z \geq 0 \\ 0, & \omega k_z < 0 \end{cases} \quad (12a)$$

$$U_d(\omega, k_z) = \begin{cases} U(\omega, k_z), & \omega k_z < 0 \\ 0, & \omega k_z \geq 0 \end{cases}, \quad (12b)$$

where  $U_u$  e  $U_d$  correspond to 2D Fourier transform of isolated upward  $u_u(t, z)$  and downward  $u_d(t, z)$  real-valued components, respectively. Implementation of equation (12) requires storing wavefield for all time steps, being impractical for 3D problems.

Considering a particular frequency  $\omega$  in wave spectrum, monochromatic wavefield elements propagating in opposite directions can be defined. Since the stack of unidirectional components is still a wavefield that propagates in a single direction, upward  $\bar{u}_+(t, k_z)$  and downward  $\bar{u}_-(t, k_z)$  elements of the wavefield  $\bar{u}(t, k_z)$ , which corresponds to the 1D Fourier transform of  $u(t, z)$  in  $z$  dimension, can be obtained, considering the contributions of all frequency components at time  $t$ , through (Liu et al., 2011)

$$\bar{u}_+(t, k_z) = \begin{cases} \bar{u}(t, k_z), & k_z \geq 0 \\ 0, & k_z < 0 \end{cases} \quad (13a)$$

$$\bar{u}_-(t, k_z) = \begin{cases} \bar{u}(t, k_z), & k_z < 0 \\ 0, & k_z \geq 0 \end{cases}. \quad (13b)$$

The components  $u_+(t, z)$  and  $u_-(t, z)$  are complex-valued functions and are related to the analytic signal of the wavefield in the direction of  $z$  (see equation 13) as

$$\mathbf{u}_+(z) = \frac{1}{2}[\mathbf{u}(z) + i\mathcal{H}_z\{\mathbf{u}(z)\}] \quad (14a)$$

$$\mathbf{u}_-(z) = \frac{1}{2}[\mathbf{u}(z) - i\mathcal{H}_z\{\mathbf{u}(z)\}], \quad (14b)$$

where,  $\mathcal{H}_z$  denotes the Hilbert transform with respect to  $z$  dimension and the wavefield  $\mathbf{u}_\pm(z)$  corresponding to a particular position at the surface  $(x, y)$  an time  $t$ , such that the real part of the sum reconstructs the propagated wavefield and its imaginary part is zero,

$$u(\mathbf{x}, t) = \Re\{u_+(\mathbf{x}, t) + u_-(\mathbf{x}, t)\} \quad (15a)$$

$$\mathbf{0} = \Im\{u_+(\mathbf{x}, t) + u_-(\mathbf{x}, t)\}. \quad (15b)$$

The upward  $\lambda_+(\mathbf{x}, t)$  and downward  $\lambda_-(\mathbf{x}, t)$  elements of the adjoint wavefield can be obtained analogously.

When applying the filter (13) in Fourier domain, the numerical discontinuities at origin and Nyquist wavenumber should be avoided. The discontinuity at  $k_z = 0$  is circumvented by incorporating a damping function with a form  $f(k_z) = 0.5[1 - \cos(\pi k_z / N_{k_z})]$ , where  $N_{k_z}$  defines the length of the window in which  $f(k_z)$  is applied (Dellinger and Etgen, 1990). The discontinuity at Nyquist wavenumber, due to the periodicity of the signal in Fourier domain, is treated by including the absorbing boundaries, used for modeling, in the wavefield separation.

When used this methodology of separation, the transmission and reflection components of the gradient are obtained through

$$\mathbf{g}_t(\mathbf{m}) = \frac{4}{c(\mathbf{x})^3} \Re \sum_s \int_0^T \lambda_-(\mathbf{x}, t) \frac{\partial^2 u_+(\mathbf{x}, t)}{\partial t^2} dt \quad (16a)$$

$$\mathbf{g}_m(\mathbf{m}) = \frac{4}{c(\mathbf{x})^3} \Re \sum_s \int_0^T \lambda_+(\mathbf{x}, t) \frac{\partial^2 u_+(\mathbf{x}, t)}{\partial t^2} dt. \quad (16b)$$

#### Explicit separation in vertical wavenumber domain

Computation of the upward wavefield component  $u_u(t, z)$  through (12) can be written as (Shen and Albertin, 2015)

$$u_u(t, z) = \frac{1}{4\pi^2} \Re \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} 2U(\omega, k_z) e^{i(\omega t + k_z z)} \Omega(\omega) \kappa(k_z) d\omega dk_z, \quad (17)$$

where,

$$\Omega(\omega) = \begin{cases} 1, & \omega \geq 0 \\ 0, & \omega < 0 \end{cases} \quad \text{and} \quad \kappa(k_z) = \begin{cases} 1, & k_z \geq 0 \\ 0, & k_z < 0 \end{cases}, \quad (18)$$

considering the symmetry of Fourier domain and that  $u_u(t, z)$  is also obtained when the conditions  $\omega < 0$  and  $k_z < 0$  are satisfied. Noticing that

$$\begin{aligned} \hat{u}(t, z) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} 2u(t', z) e^{i\omega(t-t')} \Omega(\omega) dt' d\omega \\ &= u(t, z) + i\mathcal{H}_t\{u(t, z)\}, \end{aligned} \quad (19)$$

the upward and downward components can be separated perfectly by applying the filters  $\kappa(k_z)$  and  $1 - \kappa(k_z)$  in Fourier domain, respectively, to the wavefield  $\hat{u}(t, z)$  (Shen and Albertin, 2015)

$$\begin{aligned} \bar{u}_u(t, k_z) &= \begin{cases} \hat{u}(t, k_z), & k_z \geq 0 \\ 0, & k_z < 0 \end{cases} \\ \bar{u}_d(t, k_z) &= \begin{cases} \hat{u}(t, k_z), & k_z < 0 \\ 0, & k_z \geq 0 \end{cases} \end{aligned} \quad (20)$$

where,  $\bar{u}(t, k_z)$  denotes 1D Fourier transform in the direction of  $z$  of the wavefield  $\hat{u}(t, z)$ , considering a specific position

in surface  $(x, y)$  and time  $t$ . Equation (19) describes the methodology used for estimate the analytic signal of  $u(t, z)$  in Fourier domain. Thus,  $\hat{u}(t, z)$  is a complex-valued field whose real part is the conventional wavefield  $u(\mathbf{x}, t)$  and the imaginary part correspond to its Hilbert transform in the direction of time  $\mathcal{H}_t\{u(\mathbf{x}, t)\}$ . It is interpreted as the positive-frequency component of the wavefield, such that, knowing the sign of temporal frequency, the extraction of upward and downward components through (20) is explicit.

Since wave equation operator (1) is commutative with Hilbert transform in time, the imaginary part of the analytic wavefield  $\hat{u}(\mathbf{x}, t)$  can be calculated by solving a forward modeling problem using  $\mathcal{H}_t\{s(\mathbf{x}, t)\}$  as source term. Equivalently, upward and downward components of residuals wavefield can be computed. In this case, virtual secondary source Hilbert transformed forward in time is used for obtaining the imaginary part of  $\hat{\lambda}(\mathbf{x}, t)$  (Shen and Albertin, 2015).

#### Separation using Poynting vector

Poynting vector can be computed for estimating the direction in which energy propagates. Considering an acoustic wave, the Poynting vector  $\mathbf{S}$  is given by (Bonomi et al., 1998)

$$\mathbf{S} = -\nabla u(\mathbf{x}, t) \frac{\partial u(\mathbf{x}, t)}{\partial t}, \quad (21)$$

where, the operator  $\nabla$  represents the gradient with respect to spacial variables and corresponds to the displacement vector of the particles. The propagation angle of the wavefield, at each position  $\mathbf{x}$  of modeling space and for each time step  $t$ , can be computing from Poynting vector (Araujo et al., 2014)

$$\theta = \arctan\left(\frac{S_z}{S_x}\right), \quad (22)$$

being  $S_x$  and  $S_z$  the components in the direction of  $x$  and  $z$  of the Poynting vector, respectively. Upward and downward elements are then obtained as

$$u_u(\mathbf{x}, t) = \begin{cases} u(\mathbf{x}, t), & \theta \in [0, \pi] \\ 0, & \theta \in [-\pi, 0] \end{cases} \quad (23a)$$

$$u_d(\mathbf{x}, t) = \begin{cases} u(\mathbf{x}, t), & \theta \in [-\pi, 0] \\ 0, & \theta \in [0, \pi] \end{cases}. \quad (23b)$$

This separation procedure does not alter seismic amplitude of the wave in their respective components.

When computing Poynting vector, time-derivative of wavefield is required. It can be estimated using an extension of REM operator, with high accuracy and negligible computational cost (Tessmer, 2011).

In these cases, in which wavefield components are obtained in an explicit manner, in  $k_z$  domain or through Poynting vector, tomographic and migration modes of gradient vector are calculated as

$$\mathbf{g}_t(\mathbf{m}) = \frac{2}{c(\mathbf{x})^3} \sum_s \int_0^T \left[ \lambda_u(\mathbf{x}, t) \frac{\partial^2 u_u(\mathbf{x}, t)}{\partial t^2} + \lambda_d(\mathbf{x}, t) \frac{\partial^2 u_d(\mathbf{x}, t)}{\partial t^2} \right] dt \quad (24a)$$

$$\mathbf{g}_m(\mathbf{m}) = \frac{2}{c(\mathbf{x})^3} \sum_s \int_0^T \left[ \lambda_u(\mathbf{x}, t) \frac{\partial^2 u_u(\mathbf{x}, t)}{\partial t^2} + \lambda_d(\mathbf{x}, t) \frac{\partial^2 u_d(\mathbf{x}, t)}{\partial t^2} \right] dt. \quad (24b)$$

## Numerical Results

We test the performance of wavefield separation methods in the scatter model shown in Figure 2(a) (Tang et al., 2013). It consists on a grid of  $101 \times 201$  points located each 10 m. Observed data corresponds to the record of 11 shots, separated each 200 m. Record length is 1.2 s, with interval sampling of 4 ms. Seismic source is a Ricker wavelet with a dominant frequency of 25 Hz. The initial model is shown in Figure 2(b). The inversion problem is solved in four stages, each one defined by a particular tomographic weight:  $\beta = 15.0, 10.0, 5.0$  and  $1.0$ . At each scale, 10 iterations are performed. In Algorithm 1 we use  $\sigma_0 = 1.0$  and  $\psi_\epsilon = \psi_\sigma = 0.2$ .

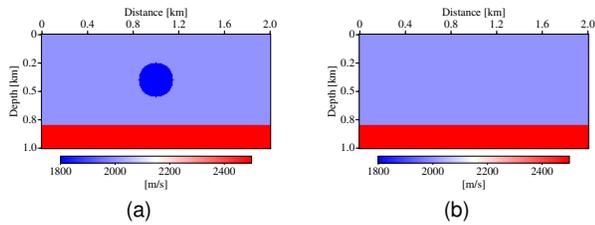


Figure 2: Scatter model: (a) true and (b) initial model used for inversion.

The inversion results are compared in the velocity profiles, located at  $x = 1000$  m, depicted in Figure 3. While the results obtained when used objective functions based on  $l_2$ -norm and cross-correlation between data vectors are satisfactory for all cases and significantly similar, the reconstruction is poor for the  $l_1$ -norm. This is caused by the contributions of multiple scattering events in gradient computation. Due to virtual secondary source is a sign function, it reinforces the amplitudes of second-order reflections not considered in initial model, which are spuriously correlated with source wavefield as first-order events in gradient construction. Also, it can be seen that the recovered profiles are closer to true parameter values in the cases in which Poynting vector is used as wavefield separation parameter.

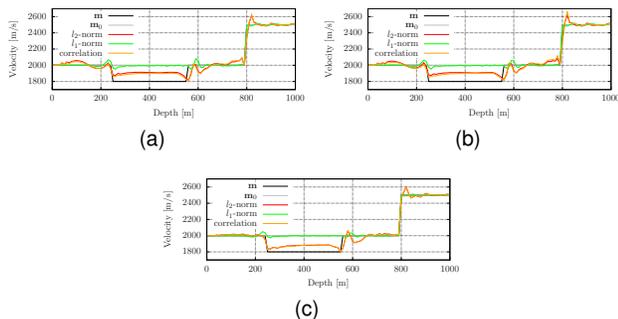


Figure 3: Velocity profile in depth of estimated scatterer model after inversion. Corresponding wavefield separation methods are (a) implicit in  $k_z$  domain, (b) explicit in  $k_z$  domain and (c) Poynting vector.

The computational performance was measured. The execution times required for each algorithm are summarized in Figure 4. In general, separation via Poynting presents the lowest computational cost. The

requirements of Fourier transforming for implicit separation method make it slightly more demanding. The highest computation times correspond to explicit separation methodology in  $k_z$  domain, given that modeling problem has to be solved twice for each analytic wavefield  $\hat{u}(\mathbf{x}, t)$  and  $\hat{\lambda}(\mathbf{x}, t)$ . When used an explicit method for wavefield separation, the algorithms guided by an objective function based on  $l_1$  norm shows anomalous high times for execution. In these cases, the estimated step length by quadratic interpolation do not satisfy first Wolfe condition and several iterations of the backtracking algorithm need to be performed before accepting the updated model.

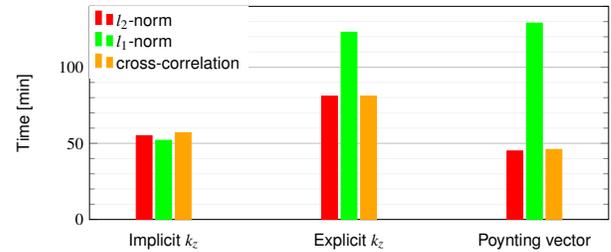


Figure 4: Algorithm-execution time for scatter model using alternative objective functions and methods for wavefield separation.

The multiscale inversion method was also applied to Marmousi model (Figure 5(a)). The model consist of  $425 \times 369$  grid points with 8 m space interval in  $z$  and 25 m in  $x$  directions. It contains an upper water layer (not shown) of 400 m in thickness, for avoiding direct wave effects in model updates. 62 sources spaced each 150 m are considered for inversion. The record length is 3.4 s sampled each 4 ms. The source has a peak frequency of approximately 15 Hz. All the model nodes at  $z = 0$  m act as receivers. Inversion is carried out in six stages in which  $\beta$  is defined as 6.0, 5.0, 4.0, 3.0, 2.0 and 1.0. At each scale, 30 FWI iterations are performed.

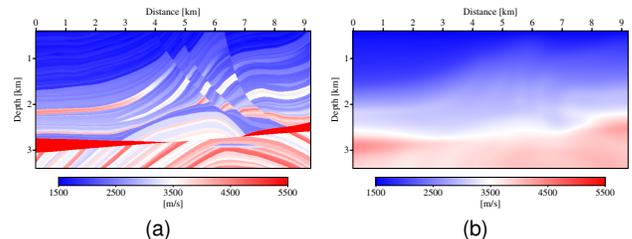


Figure 5: Marmousi model: (a) true and (b) initial model used for inversion.

Estimated models for each objective function definition are shown in Figure 6. The results obtained are satisfactory regardless of the functional used for  $E(\mathbf{m})$ . We confirm that, since initial model does not contain any reflector, they are approximately located in the first iterations. Afterwards, along the inversion process, reflector positions and interval velocities are updated simultaneously as established by Mora (1989). The updated reflectors in the model at the beginning of the iterative process let to alleviate the influence of second-order events in the search directions obtained when objective function was based on the  $l_1$ -norm.

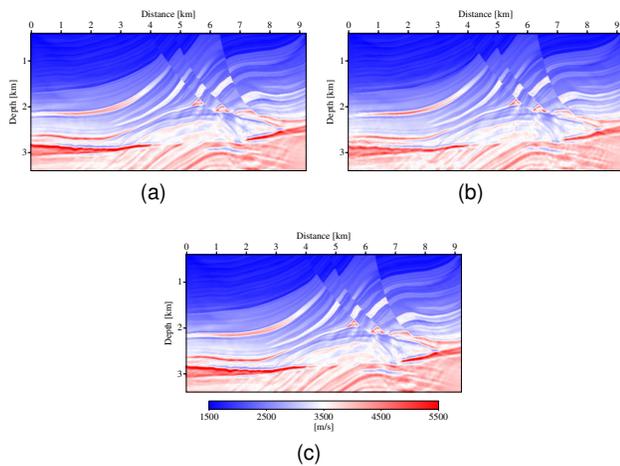


Figure 6: Estimated Marmousi model after inversion. Objective functions are based on (a)  $l_2$ -norm, (b)  $l_1$ -norm and (c) cross-correlation functionals.

## Conclusions

Different separation methods are applied for extracting directional wavefield elements and then used for estimating low- and high-wavenumber updates in time-domain full-waveform inversion. The equations used for computing these components are also given for those cases in which wavefield separation is carried out by implicit schemes in  $k_z$  domain or by explicit strategies in  $k_z$  domain or using Poynting vector. Numerical experiments indicate that gradient components obtained via Poynting vector scheme are more suitable when these are combined by using a constant coefficient at each inversion stage, showing superior inversion resolution and less computation time.

A strategy for choosing model perturbations in step-length computation via quadratic interpolation is described. This methodology let us to modify the proportion in which model is perturbed with iterations, making the inversion scheme stable even when the estimated model is approaching the global minimum of objective function. Due to gradient filtering performed at each iteration, the computation of search direction using model and gradient information from past iterations is prohibitive.

## Acknowledgements

This research was supported by CNPq, CAPES and INCT-GP/CNPq. The facility support from CPGG/UFBA is also acknowledged. Authors thank to Adriano W. G. dos Santos, Dr. Edvaldo S. Araujo and Daniel E. Revelo for providing useful computer algorithms. To Dr. Tariq Alkhalifah for worthy comments on numerical implementation issues.

## References

- Alkhalifah, T., 2015, Scattering-angle based filtering of the waveform inversion gradients: *Geophysical Journal International*, **200**, 363–373.
- Araujo, E. S., R. C. Pestana, and A. W. G. dos Santos, 2014, Symplectic scheme and the Poynting vector in reverse-time migration: *Geophysics*, **79**, S163–S172.
- Bonomi, E., L. Brieger, C. Nardone, and E. Pieroni, 1998, 3D spectral reverse time migration with no-

- wraparound absorbing conditions: SEG Technical Program Expanded Abstracts, 1925–1928.
- Brossier, R., S. Operto, and J. Virieux, 2010, Which data residual norm for robust elastic frequency-domain full waveform inversion?: *Geophysics*, **75**, R37–R46.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473.
- Chen, G., R.-S. Wu, and S. Chen, 2015, Full waveform inversion in time domain using time-damping filters: SEG Technical Program Expanded Abstracts, 1451–1455, Society of Exploration Geophysicists.
- Dellinger, J. and J. Etgen, 1990, Wave-field separation in two-dimensional anisotropic media: *Geophysics*, **55**, 914–919.
- Hu, L.-Z. and G. A. McMechan, 1987, Wave-field transformations of vertical seismic profiles: *Geophysics*, **52**, 307–321.
- Klimm, B., 2013, Time domain full waveform inversion using ADI modeling: PhD thesis, Technischen Universität Kaiserslautern.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: theory and application, 206–220, SIAM.
- Liu, F., G. Zhang, S. A. Morton, and J. P. Leveille, 2011, An effective imaging condition for reverse-time migration using wavefield decomposition: *Geophysics*, **76**, S29–S39.
- Mora, P., 1989, Inversion = migration + tomography: *Geophysics*, **54**, 1575–1586.
- Nocedal, J. and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.
- Pestana, R. C. and P. L. Stoffa, 2010, Time evolution of the wave equation using rapid expansion method: *Geophysics*, **75**, T121–T131.
- Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: *Geophysics*, **55**, 284–292.
- Shen, P. and U. Albertin, 2015, Up-down separation using Hilbert transformed source for causal imaging condition: SEG Technical Program Expanded Abstracts, 4175–4179.
- Tang, Y., S. Lee, A. Baumstein, and D. Hinkley, 2013, Tomographically enhanced full wavefield inversion: SEG Technical Program Expanded Abstracts, 1037–1041, Society of Exploration Geophysicists.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- Tessmer, E., 2011, Using the rapid expansion method for accurate time-stepping in modeling and reverse-time migration: *Geophysics*, **76**, S177–S185.
- Wu, Z. and T. Alkhalifah, 2014, Full waveform inversion based on the optimized gradient and its spectral implementation: EAGE 76th Conference & Exhibition, Tu E106 07.
- Zhang, Y., L. Duan, and Y. Xie, 2015, A stable and practical implementation of least-squares reverse time migration: *Geophysics*, **80**, V23–V31.