IOP Publishing Inverse Problems

Inverse Problems **34** (2018) 115011 (21pp) https://doi.org/10.1088/1361-6420/aade1f

Semi-global inversion of v_p to v_s ratio for elastic wavefield inversion

Nuno V da Silva[1], Gang Yao and Michael Warner

Department of Earth Science and Engineering, Imperial College London, Prince Consort Road, SW7 2BP, London, United Kingdom

E-mail: n.vieira-da-silva@imperial.ac.uk

Received 5 March 2018, revised 20 August 2018
Accepted for publication 31 August 2018
Published 20 September 2018

Abstract

We introduce an approach to estimate the ratio between P- and S-wave velocities, v_p/v_s, in the scope of elastic full waveform inversion (FWI). Elastic FWI is generally implemented with local optimization methods relying on initial estimates of the long wavelengths of P- and S-wave models. However, successful inversions can be hindered if an accurate enough relation between v_p and v_s velocities is not used as a constraint. This relation can be estimated from empirical relations. Herein, we introduce an alternative approach based upon a semi-global inversion scheme. We observe that for a large number of cases, and particularly in the context of FWI, v_p/v_s can be represented on a sparse basis. This sparse basis has a much smaller dimension than that of the typical model space in elastic FWI. This creates the possibility of using global optimization methods. The optimal estimate of v_p/v_s is obtained with quantum particle swarm optimization (QPSO). This method probes a population of possible models. The assessment of each model of v_p/v_s in the population is obtained with nested local iterations updating for v_p only. Conventional elastic FWI is then carried out for jointly estimating high-resolution models of v_p and v_s. We demonstrate with synthetic examples that the estimates of v_p are relatively robust to errors in the estimated v_p/v_s, and that effectively a sparse representation of the model of v_p/v_s is feasible for the reconstruction of a model of v_s. We also demonstrate that the proposed approach performs better than constraining elastic FWI with an empirical relation between v_p and v_s, leading to improved estimates of models of v_p and v_s from seismic data.

Keywords: semi-global inversion, elastic inversion, full waveform inversion, global inversion, inverse problems

(Some figures may appear in colour only in the online journal)

[1] Author to whom any correspondence should be addressed.

Introduction

Full waveform inversion (FWI) was introduced more than 30 years ago (Lailly 1983, Tarantola 1984). Since its inception there has been a significant effort to make it practical for the inversion of seismic data, especially for the inversion of industrial datasets (Sirgue *et al* 2009, da Silva *et al* 2016, Routh *et al* 2017). FWI requires the solution of a wave equation multiple times. This can be carried out both in time- or frequency-domain (Tarantola 1984, Pratt and Shipp 1999).

It is also generally assumed that the Earth's response is acoustic in the scope of industrial FWI. It is common practice taking into account seismic anisotropy (Warner *et al* 2013), and anelastic losses (Operto *et al* 2015). The implementation of elastic FWI has also been reported (Tarantola 1986, Mora 1987). Vigh *et al* (2014) and Raknes *et al* (2015) demonstrated the application of elastic FWI to real industrial datasets. Queißer and Singh (2013) applied elastic FWI to time-lapse CO_2 storage monitoring. However, the application of elastic FWI to large-scale industrial datasets is less common than acoustic FWI. Several factors are hindering practical elastic FWI, particularly the increasing computational cost when compared to acoustic FWI.

A more critical aspect is that elastic FWI aims to estimate a much larger number of parameters from the data when compared to acoustic FWI for example. This introduces a larger dimension of the model space and also a larger null-space. The existence of the null-space introduces ambiguity in the estimates of the parameters. Cross-talk between parameters is a manifestation of that ambiguity when jointly estimating different classes of parameters (Köhn *et al* 2012), and it affects the joint estimation of v_p and v_s with elastic FWI. This issue can be addressed constraining the elastic inversion with the Poisson ratio (Vigh *et al* 2014), for example. However, constraining the inversion in this manner requires assuming an empirical relation between v_p and v_s. These relations have been compiled from rock-physics and well-logging data. Dvorkin *et al* (2014, chapter 2) gave a comprehensive overview of the different type of relations between v_p and v_s, depending on the type of lithology. Alternatively, one can also consider using a theoretical constraint derived from the laws of continuous media (Ikelle and Amundsen 2005). The Poisson ratio is one of such constraints as pointed out earlier.

We observe that FWI is generally carried out starting from models that contain only the long wavelength information. Hence, it is reasonable using only the long wavelength components of v_p/v_s to estimate a starting model of v_s from the starting model of v_p. The long-wavelength components have a smooth variation in space. Hence, they can be represented on a reduced, or sparse, basis. This reduced basis has a much lower dimension than that of a typical discretization grid used in FWI (when using the finite difference method for example). The models of parameters typically used in FWI are discretized over a full basis. The distribution of v_p/v_s in space can then be determined by a very small number of coefficients. The parameters to be estimated are the coefficients associated with each basis function of the reduced basis. Hence, a space defined by this reduced basis is much more suitable for using global optimization, as the number of parameters to be estimated is significantly reduced when compared to that of a typical discretization grid used in conventional FWI. For example, Diouane *et al* (2016) and Datta and Sen (2016) used global inversion methods for estimating background velocity models from seismic data. Such approaches lead to models containing only long-wavelength components, and they can be used as initial models for carrying out conventional FWI.

Herein, we investigate the issue of estimating models of v_p/v_s with a semi-global algorithm in the scope of elastic FWI. These models are estimated from an initial population of uniformly distributed random models. The semi-global algorithm uses outer global iterations of quantum particle swarm optimization (QPSO) with nested local iterations of FWI. The semi-global inversion is carried out at a frequency-band containing only low frequencies. Inverting

high-frequency components of the data is not necessary as only the long wavelengths of v_p/v_s are estimated. The v_p/v_s model estimated with the semi-global inversion is then used to constrain the iterative inversion with elastic FWI. The latter is then carried out over a wide bandwidth to estimate high-resolution models of v_p and v_s. Our method requires, as in the case of acoustic FWI, a starting v_p velocity field that predicts the seismic waveforms within half-cycle of the real data, and an initial population of v_p/v_s models. Such a population of models is generated randomly. We demonstrate throughout numerical examples, using the Marmousi 2 (Versteeg 1994, Martin *et al* 2002) model, that our approach is adequate even when the true model of v_p/v_s ratio changes rapidly in space.

This paper is structured as follows. First, we introduce the theory of elastic FWI and the semi-global inversion algorithm. Then, we demonstrate with numerical examples the effectiveness of our method followed by a discussion of the results.

Basis functions

The distribution of a model parameter, $m(\boldsymbol{x})$, in space can be formally represented as the linear combination

$$m(\boldsymbol{x}) = \sum_{k=1}^{N} m_k \psi_k(\boldsymbol{x}), \tag{1}$$

where $\psi_k(\boldsymbol{x})$ is a basis function and m_k its respective coefficient. The solution of the discrete inverse problem consists of estimating the coefficients of the basis functions, m_k. Equation (1) is a general expression valid for any physical property. The number of basis functions necessary to represent the distribution of a model parameter in space gives the dimension of that space. For example, a basis function for a discretisation with equidistant grid-nodes distributed along the oriented axis, in a 3D space, can be defined as $\psi_k(\boldsymbol{x}_k) = \delta(\boldsymbol{x} - \boldsymbol{x}_k) = \delta(x - x_k, y - y_k, z - z_k)$, where $\boldsymbol{x}_k$ is the position of the kth node and $\delta(\boldsymbol{x})$ is the Dirac delta function. In such case, the number of nodes in the grid gives the dimension of the model space.

The computational cost of QPSO increases with the increasing number of unknowns. In other words, the larger the dimension of the model parameter space, the larger the number of particles in the population needs to be. Consequently, using a space with a reduced order is paramount for a good performance of the global iterations. The wavelet transform (Press *et al* 2007) is a feasible approach to reduce the dimension of the space of basis functions. Such an approach is particularly suitable when using second-order optimization as it reduces the computational cost as a result of reducing the number of unknowns (Abubakar *et al* 2012).

Herein, we define the basis functions of the reduced space using straightforward assumptions on the geology. These can be, for example, a vertical gradient, vertical and horizontal gradients, defining regions in space that correlate with the variation of the long wavelengths of the model of v_p, or any combination of the above. It is important to note that we consider a reduced basis to represent v_p/v_s only. The models of v_p and v_s are represented on the full basis.

Theory

Elastic FWI with local optimization

In this section, we review elastic waveform inversion with local minimization. FWI aims to estimate physical properties of the subsurface, inverting seismic data recorded with a

multi-source experiment, $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots)$, where the subscript denotes each unique source experiment. The inverse problem for elastic waveform inversion is formulated as the minimization of the least squares norm of the data residuals (Tarantola 1984, 1986, Mora 1987)

$$J\left(\boldsymbol{p}\left(\boldsymbol{w}\right)\right) = \frac{1}{2}\left\|\boldsymbol{R}\boldsymbol{p}\left(\boldsymbol{w}\right) - \boldsymbol{d}\right\|_2^2, \tag{2}$$

where $\boldsymbol{R} = \mathrm{diag}\left(\boldsymbol{R}_1, \boldsymbol{R}_1, \ldots\right)$ interpolates the synthetically generated data, $\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots)$, into each receiver position. The indexes in subscript denote a unique source identifier. The P- and S-wave velocities discretised over a grid are denoted with $\boldsymbol{w} = (\boldsymbol{v}_\mathrm{p}, \boldsymbol{v}_\mathrm{s})$. Note that we formulate the inverse problem in a discretised form. Hence, both data and model parameters are denoted in bold. The simulated data is generated utilising the Cauchy's law of motion

$$\rho(\boldsymbol{x})\frac{\partial \boldsymbol{v}(\boldsymbol{x},t)}{\partial t} = \nabla\boldsymbol{\sigma}(\boldsymbol{x},t) + \boldsymbol{f}(\boldsymbol{x},t), \tag{3}$$

coupled with the constitutive law

$$\boldsymbol{\sigma}(\boldsymbol{x},t) = \boldsymbol{C}(\boldsymbol{x})\boldsymbol{\varepsilon}(\boldsymbol{x},t), \tag{4}$$

where $\boldsymbol{x} = (x, y, z)$ is the position vector, t denotes time, $\boldsymbol{v}$ is the particle velocity, $\boldsymbol{\sigma}$ is the stress tensor, $\boldsymbol{\varepsilon}$ is the strain tensor, $\boldsymbol{f}$ denotes the body forces, ρ is the density, and $\boldsymbol{C}$ is the stiffness tensor. All the symmetries and energy considerations of the tensors have been taken into account (Sedov 1994) hence

$$\boldsymbol{v} = \left(v_1, v_2, v_3\right), \tag{5a}$$

$$\boldsymbol{\sigma} = \left(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{23}, \sigma_{13}, \sigma_{12}\right), \tag{5b}$$

$$\boldsymbol{\varepsilon} = \left(\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \varepsilon_{23}, \varepsilon_{13}, \varepsilon_{12}\right), \tag{5c}$$

and the stiffness tensor $\boldsymbol{C}$ in elastic isotropic media is parameterized with v_p and v_s

$$\boldsymbol{C} = \begin{pmatrix} \rho v_\mathrm{p}^2 & \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & & & \\ \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & \rho v_\mathrm{p}^2 & \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & & & \\ \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & \rho\left(v_\mathrm{p}^2 - v_\mathrm{s}^2\right) & \rho v_\mathrm{p}^2 & & & \\ & & & \rho v_\mathrm{s}^2 & & \\ & & & & \rho v_\mathrm{s}^2 & \\ & & & & & \rho v_\mathrm{s}^2 \end{pmatrix}. \tag{6}$$

In our numerical implementation, we discretized the elastic wave equations with a staggered finite difference method, second-order accurate in time and fourth-order accurate in space (Levander 1988). The pressure field, $p(\boldsymbol{x}, t)$, is given at the receiver points by

$$p = -\frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3}. \tag{7}$$

The local iterations of elastic FWI are carried out with the preconditioned steepest-descent method. The estimates of the model parameters are given by

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \alpha_k \boldsymbol{H}_k^{-1}\nabla_{\boldsymbol{w}} J, \tag{8}$$

where k denotes the iteration number, α_k is the step-length, $\boldsymbol{H}_k$ is an approximation of the diagonal of the Hessian (Pratt *et al* 1998), and $\nabla_{\boldsymbol{w}}\boldsymbol{J}$ is the gradient of the objective function 1 with respect to the model parameters. The gradient $\nabla_{\boldsymbol{w}}\boldsymbol{J}$ is computed with the adjoint-state method (Fernández-Berdaguer 1998, Fichtner *et al* 2006, Vigh *et al* 2014). In appendix, we outline a brief description of the adjoint-state method for the computation of $\nabla_{\boldsymbol{w}}\boldsymbol{J}$.

Quantum particle swarm optimization

In this work, the outer global iterations are carried out with the QPSO algorithm. We point out that other meta-heuristic methods could have been chosen as for example genetic algorithms (GA) (Sen and Stoffa 2013). However, it was reported higher efficiency (Debens 2015) and suitability for continuous optimization (Kachitvichyanukul 2012) of QPSO over GA. QPSO belongs to a class of meta-heuristic optimization algorithms known as particle-swarm optimization (Kennedy and Eberhart 1995). The latter emulates the dynamics of a swarm of particles ruled by the laws of classical mechanics. On the other hand, QPSO emulates the dynamics of particles with quantum behavior. Hence, it does not need a velocity field associated to each particle, as in the case of particle swarm optimization. Instead, QPSO uses a quantum potential similarly to quantum systems (Kennedy and Eberhart 1995, Kennedy and Eberhart 2001, Sun *et al* 2004a, 2004b). An important aspect of particle swarm optimization algorithms, including QPSO, is that particles are assumed to be dimensionless. Hence, they do not have a volume.

In our particular application of QPSO, the position of the particles defines a model of v_p to v_s ratio. The initial population of models is generated randomly, and it has a uniform distribution. The level of misfit for the model represented by each particle is assessed at each iteration. The algorithm keeps track of the best model estimated for each particle, and the best model in the population, along the successive iterations. These quantities are then used statistically to update the models represented by each particle. Along successive iterations, the position of each one of the particles tends to cluster around the best overall position. This position is, at least in principle, if the algorithm converged, the minimum of the objective function.

Semi-global inversion method

Semi-global methods combine characteristics of global and local optimization methods. The former has the potential of searching the entire model space while not using information of the *topography* of the misfit function (Fernández Martínez *et al* 2012). Global optimization methods can be ineffective if realizations of the model space are computationally demanding. This happens for the vast majority of applications of geophysical inversion because the solution of the forward problem is obtained throughout the numerical solution of partial differential equations with the finite-difference or the finite element method, for example. On the other hand, local optimization methods require less model space searches. However, these searches are carried out in a smaller portion of that space. In addition, they rely upon the *topography* of the misfit function. The successful application of local optimization methods requires an initial estimate of the model parameters that places the successive searching iterations within the basin of attraction of the global minimum. Otherwise, the algorithm can converge into local minima and the model estimates are sub-optimal. Semi-global methods allow searching large portions of the model space while using some of the information on the *topography* of the misfit function to guide that model search (Afanasiev *et al* 2014).

Table 1. Comparison between the computational cost of local elastic FWI, $O(L)$, and that of semi-global inversion, $O(G)$. The first row shows the absolute computational cost for local and semi-global inversion. The second row relates $O(L)$ and $O(G)$ when the number of outer global iterations is of the same order as that of the number of frequency bands in local inversion, and the number of nested local iterations is the same as the number of local iterations per frequency-band. The third row compres $O(L)$, and $O(G)$ when only a subset of the whole set of sources is utilised at each local iteration; N_{skip} is the number of shots skipped.

Approximation	Local inversion	Semi-global inversion
None	$O(L) = N_{\text{f}}\, N_{\text{L}}$	$O(G) = N_{\text{p}}\, N_{\text{Ln}}\, N_{\text{g}}$
$N_{\text{g}} \sim N_{\text{f}}$; $N_{\text{L}} \sim N_{\text{Ln}}$	—	$O(G) = N_{\text{p}}\, O(L)$
Skip N_{skip} shots	—	$O(G) = N_{\text{p}}\, O(L)/N_{\text{skip}}$
$N_{\text{p}} \sim N_{\text{skip}}$	—	$O(G) \sim O(L)$

Herein is outlined a semi-global inversion method combining QPSO iterations with local gradient-descent iterations. This algorithm aims to recover long wavelengths of models of $v_{\text{p}}/v_{\text{s}}$ ratio. The estimated model then can be used for generating a starting v_{s} model and carrying out conventional elastic FWI.

The key entity associated to the QPSO algorithm is the concept of the particle. The whole set of particles forms a population of particles or *particle swarm*. Each particle k (the index identifying the particle) has associated a position $\boldsymbol{m}_k$ with dimension N. The position of a particle is effectively the model parameter(s) estimated over the successive iterations. In the scope of this paper, the position of each particle is the model of $v_{\text{p}}/v_{\text{s}}$. The set of all the position vectors for the N_{p} particles in the swarm is formally represented as $\mathcal{M} = \left\{\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_{N_{\text{p}}}\right\}$. Effectively, each element of $\mathcal{M}$ is a set of coefficients of the reduced basis. Another key aspect is that the elements of $\mathcal{M}_0$, the set of all initial positions, are uniformly distributed. It is important to note that this property is not met along the successive iterations, $\mathcal{M}_n$, as the models tend to cluster along with the convergence of the successive iterations. This means that there will be models that have a higher likelihood than others. Each one of the elements of $\mathcal{M}_n$ has a misfit associated. The set of misfits for the population at a given iteration n is $\mathcal{J}_n = \left\{J_{1,n}, J_{2,n}, \ldots, J_{N_{\text{p}},n}\right\}$. The element $J_{k,n}$ corresponds to the misfit of the kth particle at nth outer iteration. The elements of $\mathcal{J}_n$ are estimated with local nested iterations of FWI, updating only for v_{p}. At any given outer iteration, each particle has a best-estimated position. That position is the one corresponding to the lowest value of the objective function attained by that particle up to that iteration. This means that the best position for a given particle at any given outer iteration can be the current estimate, or it can be any of the previous ones. The set of the best estimates of each particle is defined as $\mathcal{M}^* = \left\{\boldsymbol{m}_1^*, \boldsymbol{m}_2^*, \ldots, \boldsymbol{m}_{N_{\text{p}}}^*\right\}$. Each one of its elements has a misfit associated which is in the set $\mathcal{J}^* = \left\{J_1^*, J_2^*, \ldots, J_{N_{\text{p}}}^*\right\}$. In addition, at any given outer iteration the population has a best overall position $\boldsymbol{m}_g^* = \boldsymbol{m}_k^*$. The best overall position has a corresponding misfit J_k^*, with the index k given by $\text{argmin}_{k\in\{1,\ldots,N_p\}}(\mathcal{J}^*)$. The chief assumption is that the position with the least value of misfit is also the position with the highest likelihood of being the optimal solution. The position with the smallest level of misfit is selected as the optimal solution upon completing the set of outer global iterations. We outline a pseudo-code of the semi-global inversion method in algorithm 1.

Algorithm 1. Semi-global inversion of v_p/v_s ratio; N_g is the number of global outer iterations, N_p is the number of particles, N_{Ln} is the number of nested local iterations, N_f is the number of frequency bands, n indicates global outer iterations, k indexes a particle, l indexes nested local iterations, f indexes the frequency band, ε is the threshold for the level of misfit, β is the contraction-expansion coefficient. For mono-modal solutions β is set equal to 0.75 (Sun 2012).

$n = 0$
Select a starting model of the velocity of compressional waves, $\boldsymbol{v}_P$
Generate a uniform population of starting models of v_p/v_s, $\mathcal{M}_n = \{\boldsymbol{m}_{1,n}, \ldots, \boldsymbol{m}_{N_p,n}\}$
while $\min_{\boldsymbol{m}} \{\mathcal{J}\} > \varepsilon$ and $n < N_g$ and $f < N_f$:
  for each particle $k \in \{1, \ldots, N_p\}$:
    compute $\boldsymbol{v}_{S,k}$ from $\boldsymbol{v}_P$ and $\boldsymbol{m}_{k,n} \in \mathcal{M}_n$
    $\boldsymbol{w}_0 = \{\boldsymbol{v}_P; \boldsymbol{v}_{S,k}\}$
    for $l \in \{0, \ldots N_{Ln} - 1\}$ (nested local iterations):
      compute $\nabla_{\boldsymbol{v}_p} J_{k,n}(\boldsymbol{w}_l)$ as outlined in appendix
      $\boldsymbol{w}_{l+1} = \boldsymbol{w}_l - \alpha_l \boldsymbol{H}_l^{-1} \nabla_{\boldsymbol{v}_p} J_{k,n}(\boldsymbol{w}_l)$
  store $\mathcal{J}_n = \{J_{1,n}, \ldots, J_{N_p,n}\}$ and $\mathcal{V}_n = \{\boldsymbol{v}_{P,1}, \ldots, \boldsymbol{v}_{P,N_p}\}$
  if $n > 0$ (ignore models that did not improve):
    for each particle $k \in \{1, \ldots, N_p\}$:
      if $J_{k,n} > J_{k,n-1}$:
        $\boldsymbol{m}_{k,n} = \boldsymbol{m}_{k,n-1}$
        $J_{k,n} = J_{k,n-1}$
  select the best model for each particle so far, $\mathcal{M}^* = \{\boldsymbol{m}_1^*, \boldsymbol{m}_2^*, \ldots, \boldsymbol{m}_M^*\}$
  select the best model in the population so far, $\boldsymbol{m}_g^*$
  select the coefficient β
  for each particle k:
    generate a diagonal matrix $\boldsymbol{\varphi}$ with random values uniformly distributed
    determine a local attractor $\boldsymbol{q}_k = \boldsymbol{\varphi}\boldsymbol{m}_k^* + (\mathbf{1} - \boldsymbol{\varphi})\boldsymbol{m}_g^*$
    $\boldsymbol{l}_k = 2\beta\,|\boldsymbol{m}_k^* - \boldsymbol{m}_{k,n}|$
    generate a diagonal matrix $\boldsymbol{\phi}$ with random values uniformly distributed
    generate random r with a normal distribution with zero mean and unit variance
    if $r < 0.5$ then:
      $\boldsymbol{m}_{k,n+1} = \boldsymbol{q}_k - 0.5\ln(\boldsymbol{\phi}^{-1})\boldsymbol{l}_k$
    else:
      $\boldsymbol{m}_{k,n+1} = \boldsymbol{q}_k + 0.5\ln(\boldsymbol{\phi}^{-1})\boldsymbol{l}_k$
    compute the weighted average of velocity models $\boldsymbol{v}_P = \frac{1}{\sum_k J_{k,n}} \sum_k J_{k,n}\boldsymbol{v}_{P,k}$
  $n \leftarrow n + 1$

Dependency on the starting model

The semi-global inversion algorithm introduced herein relies on local nested iterations of FWI within each outer global iteration. When carrying out FWI, the level of data misfit can be reduced even when the starting model of v_p is not adequate. In such a case, the inversion converges towards a local minimum. A criterion used for enhancing the possibility of converging towards the global minimum is guaranteeing that the time difference between observed and computed waveforms is less than half a cycle. When this condition is not met, the data is referred to as being cycle-skipped. In such a case, conventional FWI does not converge towards the global minimum. A comprehensive review on some of the latest developments on this topic can be found in da Silva and Yao (2018).

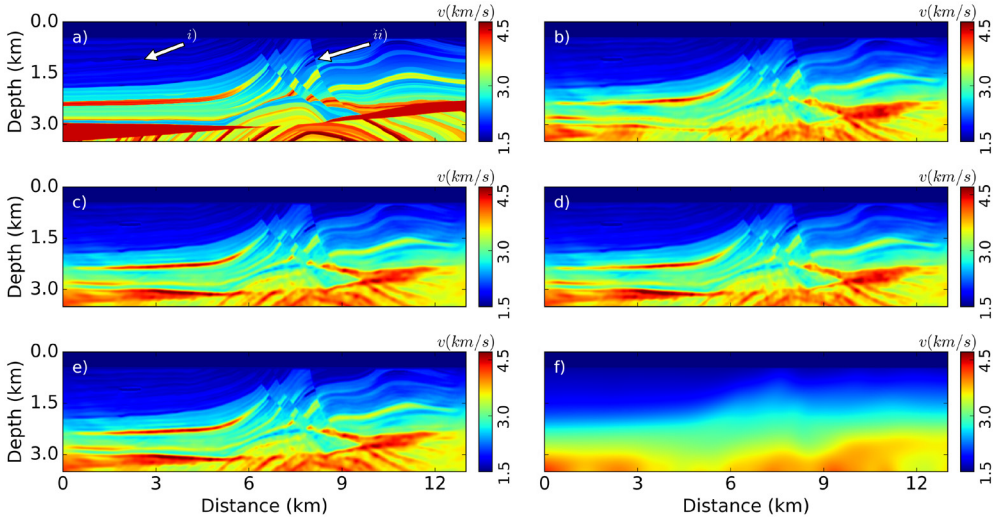

Figure 1. P-wave velocity model: (a) true model, (b) model inverted with the average v_p/v_s ratio of 2.87, (c) model inverted constraining with the 'mud-rock' model, (d) model inverted constraining with the vertical gradient of v_p/v_s, (e) model inverted with the v_p/v_s determined from the structure of the starting model of v_p and (f) starting v_p model.

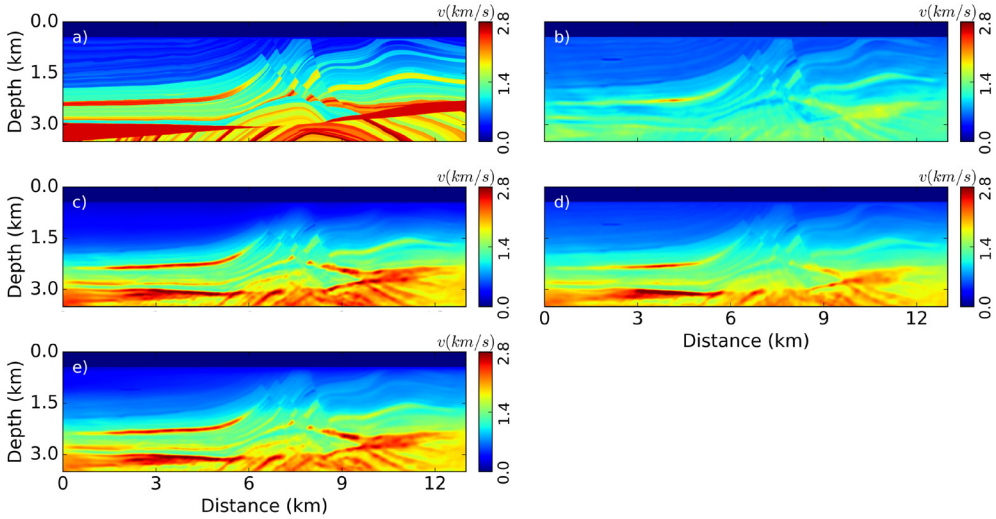

Figure 2. S-wave velocity model: (a) true model, (b) model inverted constraining the inversion with the average v_p/v_s ratio of 2.87, (c) model inverted constraining the inversion with the 'mud-rock' model, (d) model inverted constraining the inversion with a vertical gradient of v_p/v_s and (e) model inverted constraining the inversion with the v_p/v_s determined from the structure of the starting model of v_p/v_s.

In the approach introduced herein, the assessment of each model of v_p/v_s ratio at each outer global iteration is carried out with nested local iterations. Hence, the initial model of v_p must satisfy the cycle-skipping criterion for the nested local iterations. Not meeting this criterion along the successive global iterations is not critical, as long as, there are models in the population that generate data that is not cycle-skipped. The rationale for this statement is that models

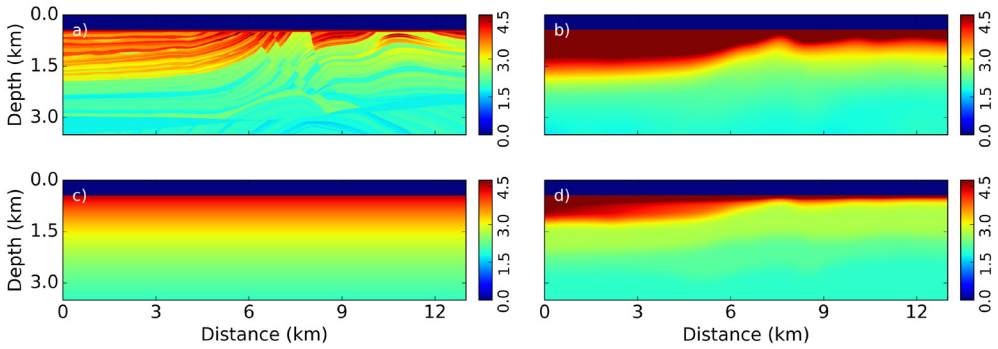

Figure 3. v_p/v_s ratio: (a) true ratio, (b) ratio determined from 'mud-rock' model, (c) ratio estimated with semi-global inversion and assuming a vertical gradient and (d) ratio estimated with semi-global inversion following the structure of the starting v_p model (figure 4).

that generate cycle-skipped data should have, at least in principle, a data misfit higher than that of data that is in phase. Hence, these models have a lesser role in the overall dynamics of the swarm. However, inversions should be carried out using adequate starting models. In such a case, the initial model of v_p has to predict the data within half of a cycle, and the successive estimates of v_p/v_s are obtained with small incremental steps. Then, it is very unlikely that the data generated over the successive iterations becomes cycle-skipped. The overall convergence of the algorithm depends principally on ensuring that the starting models generate data that is not cycle-skipped.

Marmousi 2—noise-free data

In this section, we demonstrate the application of the semi-global inversion method outlined herein. Our examples utilise the synthetic elastic model Marmousi 2 (Versteeg 1994, Martin *et al* 2002). This model includes both distributions of v_p and v_s, as depicted in figures 1(a) and 2(a), respectively. It is characterized by its complex structure and the existence of two low-velocity anomalies highlighted with (*i*) and (*ii*) in figure 1(a). The low-velocity anomalies labeled (*i*) and (*ii*) do not correlate with the low-velocity anomalies in the true v_s model (figure 2(a)). In addition, the Marmousi 2 model is also characterized by the existence of a reservoir in its central part with the top of the anticline at a depth of about 2.75 km.

The true v_p/v_s ratio is depicted in figure 3(a). One can observe a very detailed model of v_p/v_s with a smooth variation of the long-wavelength components along the vertical direction and with almost no variation along the horizontal direction. We refer to the long wavelengths components as the background model. The semi-global method is suitable for estimating the background model as it relies upon the use of a sparse basis. The rationale of our approach is based on the fact that elastic FWI starts from smooth models of v_p and v_s. Hence, one only has information on the long wavelength components prior to carrying out elastic FWI. Accordingly the same happens for the model of v_p/v_s.

Once the starting model of v_p and the constraining model of v_p/v_s are obtained, one can estimate high-resolution models for v_p and v_s with conventional elastic FWI utilising a local optimization method.

The models of velocity are discretized with a grid spacing of 12.5 m along the vertical and horizontal directions. The model of density is determined from the velocity of compressional

waves using Gardner's law (Gardner 1974), with the exception of the region corresponding to the seawater where the density is set equal to 1000 kg m^{-3}.

We generate a synthetic data set of pressure comprising 323 shots spaced of 50 m, and at a depth of 12.5 m. We apply a free-surface boundary condition at the top of the model, and we suppress the energy outgoing through the lateral and bottom boundaries with an absorbing layer (Yao *et al* 2018).

The time dependency of the source wavelet is defined with a Ricker wavelet with a peak frequency of 8 Hz. The synthetic data are generated recording the wavefield with a fixed receiver configuration. The array of receivers comprises 1290 receivers spaced of 12.5 m. The depth of all receivers is 12.5 m. In all the examples presented, the starting model of v_p (figure 1(f)) is obtained smoothing the true model of v_p (figure 1(a)) utilising a Gaussian filter. The kernel of the Gaussian filter has a length of 25 cells.

In all the examples, the semi-global inversions are carried out with four nested local steepest-descent iterations. We set a population size of ten particles. The initial models are selected randomly and follow a uniform distribution. We carried out ten outer global iterations for each particle. The semi-global inversion ran at a low-frequency band, filtering the data with a low-pass filter with cut-off at 2.5 Hz. We use only every fourth shot at each nested local iteration. This means that every three shots are skipped, and that each nested local iteration uses one-fourth of all the available data. After completing four local iterations, each shot has been used once. Hence, all the data has been used after completing each outer global iteration. The shot distribution does not need to be very dense as long as the receiver distribution is dense enough and only the long wavelengths of the anomalies are estimated. Effectively, the semi-global inversion is carried out at low frequencies and over a sparse basis. Hence, only anomalies of v_p/v_s with long wavelengths are estimated, and the condition for decimating the data, by skipping shots, is satisfied.

After completing the semi-global iterations, we jointly estimate v_p and v_s carrying out conventional elastic FWI. The local elastic FWI inversion starts from the same v_p model as the semi-global inversion. The starting v_s model is determined by the v_p/v_s ratio estimated with the semi-global inversion. Elastic FWI is carried out in blocks of 5 iterations for each frequency band. After completing each block of iterations the frequency band is widened. The first frequency band has a cut-off filter at 2.5 Hz, the second has a cut-off filter at 3 Hz and all the subsequent bands of frequency are widened by 1 Hz up to 10 Hz. All the data are used when carrying out local elastic FWI iterations. The local inversion algorithm runs 45 steepest-descent iterations in total.

Elastic FWI constrained with average v_p/v_s

We carry out this FWI example in order to demonstrate that in this particular case it is not sufficient to consider a simple relation between v_p and v_s. In the region corresponding to the sea-water the ratio is set equal to zero, and in the region corresponding to the geological formations we take the average value of v_p/v_s which is 2.87. This value is determined from the starting models of v_p and v_s and agrees with the relation $v_P \geqslant 1.4 v_S$ (Ikelle and Amundsen 2005), which is derived from the relations of continuous mechanics.

The model of v_p estimated with FWI is depicted in figure 1(b). The corresponding model of v_s is depicted in figure 2(b). One can observe that both inverted models are inaccurate. The shallower part of the v_p model shows the main features reconstructed with good accuracy. However, the accuracy of the inverted model degrades significantly with depth. The deeper high-velocity dipping layer in the left region of the model is poorly reconstructed.

Without complementary information, and if the true model is not known, one cannot determine whether such a structure is real or not. In addition, both the reservoir and the region below the reservoir show a poor reconstruction.

The inverted v_s model shows the spatial structure of the true model. However, it is quantitatively inaccurate and the estimated v_s is averagely lower than that of the true model. In addition, there is also evidence of crosstalk as the inverted model of v_s shows low-velocity anomalies correlating with the structures (*i*) and (*ii*) in the model of v_p. Suppressing this ambiguity requires either constraining the model of v_p/v_s ratio capturing these particular anomalies or constraining the inversion with complementary data.

One can conclude that the assumption made on a constant and average value of the v_p/v_s ratio is not accurate enough for this particular case. One can also observe from this test that the estimates of v_p are less sensitive to errors in v_p/v_s than the estimates of v_s.

Constraining with the v_p/v_s ratio determined with the mud-rock model

In this example, we carry out elastic FWI constraining the inversion with the 'mud-rock' model (Castagna 1985)

$$v_s = 0.862 v_p - 1172, \tag{9}$$

where the units of both v_p and v_s are in m s^{-1}. Figure 3(b) shows the v_p/v_s model obtained from the starting model of v_p and from equation (9). The resulting inverted v_p and v_s models are depicted in figures 1(c) and 2(c), respectively. The main features in the model of v_p are reconstructed accurately. The inverted model of v_s is also reasonably well reconstructed, both qualitatively and quantitatively. However, in the shallower region, the structures are poorly resolved. In addition, there is not a noticeable 'leakage' of structure (*i*) into the reconstructed v_s. The 'mud-rock' model demonstrated to be relatively effective as a constraint. This is particularly relevant considering that the 'mud-rock' model is derived from laboratory experiments, whereas the example outlined herein utilises synthetic data only. This suggests that the relation between v_p and v_s in the Marmousi 2 model obeys a realistic rock physics relation.

Semi-global inversion of v_p/v_s assuming vertical gradient variation

The previous examples demonstrated that it is necessary a relatively accurate v_p/v_s constraint in order to carry out elastic FWI successfully. In this example, we use the semi-global inversion method in order to determine a sensible v_p/v_s ratio. As demonstrated with the 'mud-rock' model, empirical relationships can be effective to achieve this goal. However, we aim to estimate these parameters exclusively from available seismic data. Then, the use of empirical relations can be avoided as these depend upon lithology, and lithology changes significantly with geographical location.

As pointed out, the semi-global inversion scheme requires the use of a space of basis functions with a small dimension. Hence, one needs to make assumptions in order to generate a suitable sparse basis. One such option is assuming that the v_p/v_s ratio changes linearly with depth. We find this assumption reasonable for this particular example. For example, figure 3(a) shows that the highest values of v_p/v_s ratio are closer to the sea-bottom. The v_p/v_s ratio decreases gradually with depth down to the deepest part of the model. On the other hand, the horizontal variation is very smooth. Then, the distribution in space of v_p/v_s ratio can be approximated with a vertical gradient and a constant

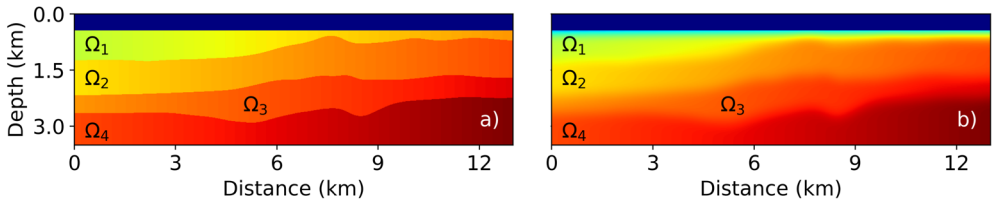

Figure 4. Domain of the basis functions determined from the starting v_p model: (a) without Gaussian smoothing applied, (b) with Gaussian smoothing applied. The range of v_p/v_s is just for illustrative purposes.

$$m(\boldsymbol{x}) = m_0 + \nabla_z m(\boldsymbol{x}) z. \tag{10}$$

The basis functions are then

$$\{\psi_i\} = \{1, z\}\,. \tag{11}$$

The vertical gradient of v_p/v_s ratio is defined as

$$\nabla_z m(\boldsymbol{x}) = \frac{m_1 - m_0}{L}, \tag{12}$$

where L is the vertical extension of the model excluding the seawater. Hence, effectively the semi-global algorithm estimates two coefficients: m_0, and $\nabla_z m(\boldsymbol{x})$. The model of v_p/v_s is determined over the whole domain of interest with the expression (12).

The lower and upper bounds of the searching space are set to 1.5 and 5, respectively. The ratio v_p/v_s is set to zero in the seawater region. This range encompasses the range of values of v_p/v_s of the true model. The model of v_p/v_s estimated with the semi-global inversion method (figure 3(c)) is then used to estimate an initial model of v_s prior to inverting the data with conventional elastic FWI. The local inversion is carried out as outlined above.

Figures 1(d) and 2(d) depict the inverted models of v_p and v_s, respectively. As in the previous cases, the inversion of v_p is more robust to errors in v_p/v_s than the inversion of v_s. One can observe that the inverted model of v_p is very similar to that obtained with the mud-rock constraint (figure 1(c)). However, the inverted v_s model is less accurate than that inverted with the 'mud-rock' model. In addition, it also shows evidence of crosstalk, as anomalies (*i*) and (*ii*) are present in the inverted model of v_s. Even though the inverted model of v_s (figure 2(d)) is not as accurate as the one obtained with the 'mud-rock' model, it still leads to a reasonably accurate estimate. This is relevant, considering that it is assumed a very simple distribution in space of the v_p/v_s ratio.

Semi-global inversion of v_p/v_s following the structure of the v_p model

In this example, we estimate a model of the long wavelengths of the v_p/v_s ratio using the structure of the starting v_p model as a prior. We use the structure of the starting model of v_p to define a sparse basis. This assumption is made as it seems reasonable to assume a correlation between the long wavelengths of the model of v_p and v_s. One can observe that the starting model of v_p (figure 1(f)) is essentially a vertical gradient of velocity with smooth lateral variation. We take advantage of this observation postulating that one can define a structure for the model of v_s (and v_p/v_s) that is identical in some degree to that of the v_p model. Figure 4(a) shows how these subdomains are distributed in space. We define these subdomains clipping regions within a pre-selected range of values of v_p. In each one of these subdomains, we also allow horizontal variation defining a horizontal gradient of v_p/v_s. Then, we can define for each

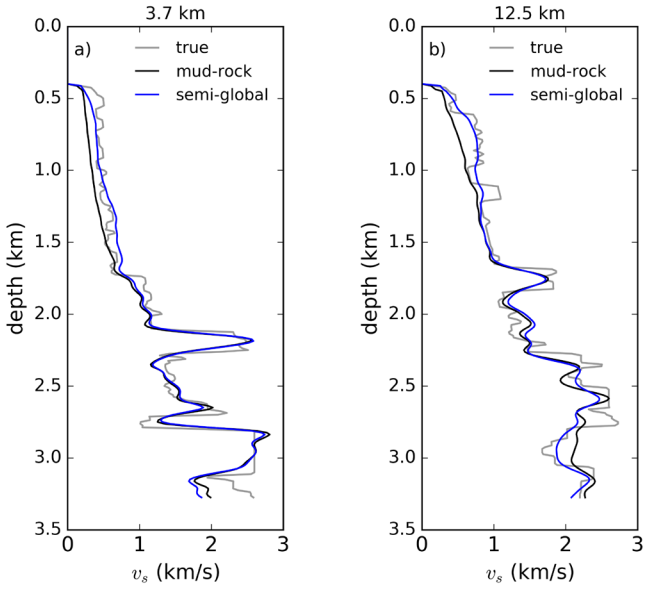

Figure 5. Comparison between the vertical log of the true v_s model (in grey), inverted v_s model constrained with the mud-rock model (in black), and inverted v_s model constrained with the v_p/v_s model (in blue) in figure 4(d) (using structure of the starting v_p model) at: (a) 3700 m and at (b) 12 500 m.

subdomain Ω_i a basis $\psi_{\Omega_i} = \{1, x\}_{\Omega_i}$. The sparse basis over the entire domain is defined as $\{\psi_i\} = \bigcup_{i=1}^{4} \psi_{\Omega_i}$. The distribution of v_p/v_s in each subdomain Ω_i is given by

$$m_i(\boldsymbol{x}) = \begin{cases} m_{0,i} + x\nabla_x m_i(\boldsymbol{x}),\ \boldsymbol{x} \in \Omega_i \\ 0,\ \boldsymbol{x} \notin \Omega_i \end{cases}. \tag{13}$$

Each subdomain Ω_i has two coefficients associated, and the coefficients of the basis functions $m_j \in \{m_{0,i}, \nabla_x m_i(\boldsymbol{x})\}_{i=1,4}$. Given that all the subdomains have the same horizontal extension, L, and defining $m_{0,i} = m_i(x = 0)$, and $m_{1,i} = m_i(x = L)$, then the horizontal gradient of $m_i(\boldsymbol{x})$ is given by

$$\nabla_x m_i(\boldsymbol{x}) = \frac{m_{1,i} - m_{0,i}}{L}. \tag{14}$$

Finally, the distribution of v_p/v_s in space is defined in the whole domain throughout

$$\widetilde{m}(\boldsymbol{x}) = \sum_{i=1}^{4} m_i(\boldsymbol{x}). \tag{15}$$

In this example, we also applied Gaussian smoothing in order to avoid discontinuity of the estimates across the boundaries of each subdomain Ω_i. These discontinuities are undesirable as they can create spurious reflections. The Gaussian smoothing is formally defined as

$$m(\boldsymbol{x}) = G_K\left[\widetilde{m}(\boldsymbol{x})\right] = G_K\left[\sum_{i=1}^{4} m_i(\boldsymbol{x})\right], \tag{16}$$

where K is the length of the kernel of the Gaussian filter. In this example, we selected the length of the kernel $K = 5$. Figure 4 illustrates how we construct the basis functions. Figure 4(a)

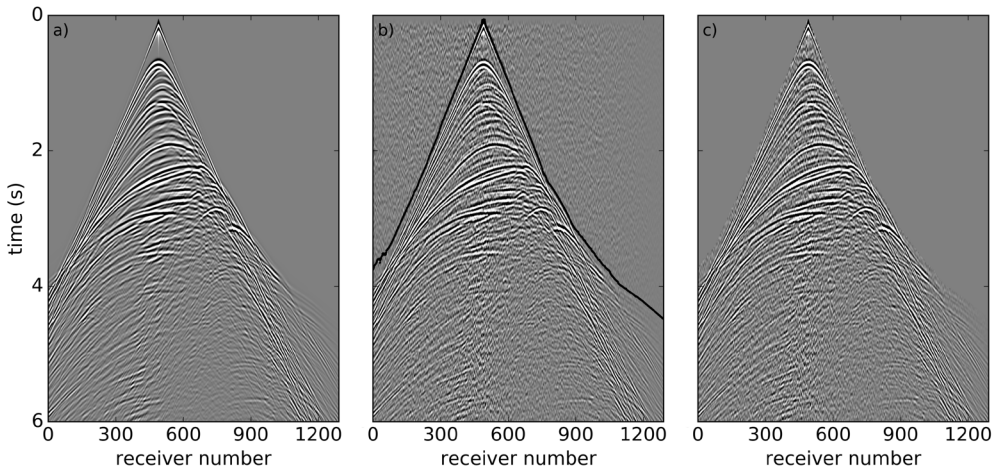

Figure 6. Shot gather generated placing a source at 6.25 km from the leftmost boundary of the Marmousi 2 model: (a) noise-free data, (b) data after adding random noise with SNR $= -0.2$ dB (the black line denotes the recording time defining the transition from noise-only to signal affected by noise) and (c) data affected by noise with SNR $= -0.2$ dB after applying a mute to the recording times earlier than those defined by the black line in (b).

shows a model defined using expression 15. One can observe smooth variation along the horizontal directions and the existence of a discontinuity across the border between each subdomain. The discontinuities are eliminated after applying Gaussian smoothing, as depicted in figure 4(b).

In this example, we also set an initial population of 10 uniformly distributed models of v_p/v_s. The lower and upper bounds of these models range between 1.5 and 5. The best estimated model of v_p/v_s (figure 3(d)) was then used to constrain elastic FWI. The local elastic FWI is carried out using the setting described at the beginning of this section. Figures 1(e) and 2(e) depict the final inverted models of v_p and v_s, respectively.

One can observe that the inverted v_p model is well reconstructed. There is not a noticeable 'leakage' of structure (*i*) into the reconstructed v_s. Furthermore, the inverted v_s model is well reconstructed both quantitatively and qualitatively. In fact, the reconstructed model is more accurate than that estimated using the 'mud-rock' constraint. This improvement is especially visible in the shallower region and at the central and rightmost regions. This statement is supported comparing vertical profiles of the different models. Figures 5(a) and (b) show vertical profiles of the true v_s model, inverted v_s model using the mud-rock model, and inverted v_s model with the semi-global inversion algorithm. These vertical profiles are extracted at the distance of 3.7 and 12.5 km from the leftmost boundary. One can observe an improved fitting of the model inverted with semi-global inversion.

Marmousi 2—noisy data

The inversion of real observations is affected by noise. The existence of noise in the data affects the stability of the solution of the inverse problem. In this example, we investigate the effect of noise in the robustness of the semi-global inversion method. The noise is generated randomly and follows a Gaussian distribution with zero mean and unit variance. We then convolve this randomly generated noise with the source wavelet in order to match its bandwidth

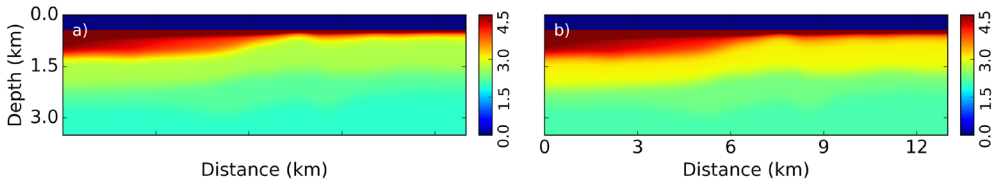

Figure 7. v_p/v_s ratio estimated with semi-global inversion following the structure of the starting v_p: (a) without noise in the data (same as figure 4(d)), and (b) with noise in the data.

to that of the recorded signal. The filtered noise is then added to the synthetically generated seismic data setting the Signal to Noise Ratio (SNR) to −0.22 dB. This corresponds to a loss of 5% of the power of useful signal. Useful signal means the component of the record that contains information on the physical properties of the medium.

Figure 6 depicts a shot gather generated using the Marmousi 2 model with different levels of noise. The shot gather is generated placing a source at 6.25 km from the leftmost boundary. The source depth and its time dependency is the same as in the previous examples. Figure 6(a) depicts the resulting noise-free modelled data. The shot gather in figure 6(b) results from adding random noise (as described above) to the shot gather in figure 6(a). The black line in figure 6(b) marks the transition between pure random noise and the recording of signal and noise. This line is used to set a mute to the region of the signal that only contains noise. Figure 6(c) shows the shot gather depicted in figure 6(b) after applying the mute. We carried out a similar procedure to all shot gathers in the dataset. We point out that muting noise from data is a common practice in real data applications.

We then carried out the semi-global inversion of the data affected by noise. We chose to estimate a model of v_p/v_s ratio using the structure of the initial model of v_p as a prior. We only considered this option in this case as it leads to the best estimates of v_p and v_s as demonstrated in the previous example. We used the same number of global iterations, the same number of local nested iterations, the same number of particles in the swarm, and the same frequency band, as in the case when the data is not affected by noise. After completing the semi-global inversion we then upscaled the inversion in frequency with local elastic FWI. We carried out elastic FWI with the same number of iterations and the same frequency bands as used in the previous example.

Figures 7(a) and (b) compare the estimated v_p/v_s ratio, when the data is not affected by noise and when the data is affected by noise, respectively. Figure 7(a), depicts the same model as that in figure 3(d). One can observe that both v_p/v_s models are very similar both in the shallower and deeper regions. The v_p/v_s ratio is slightly higher on the second layer in the case when the data is affected by noise. One can then observe that the semi-global inversion algorithm was robust for the level of noise added to the data. This statement is supported by the fact that the estimates remain bounded, and the inversion converged towards a model that keeps much of the same properties of the model inverted with noise-free data.

We then carried out elastic FWI generating the initial v_s model with the v_p/v_s model depicted in figure 7(b). Figures 8(a) and (b) show the true v_p and v_s models, respectively. Figures 8(c) and (d) show the inverted models of v_p and v_s, respectively, when the data is not affected by noise (they are the same as figures 1(e) and 2(e)). Figures 8(e) and (f) depict the inverted models of v_p and v_s, when the data is affected by noise. One can observe that the main features of the models are recovered accurately. The reconstructed models are more accurate in the central regions. Towards the edges and the bottom of the models, the reconstruction is less

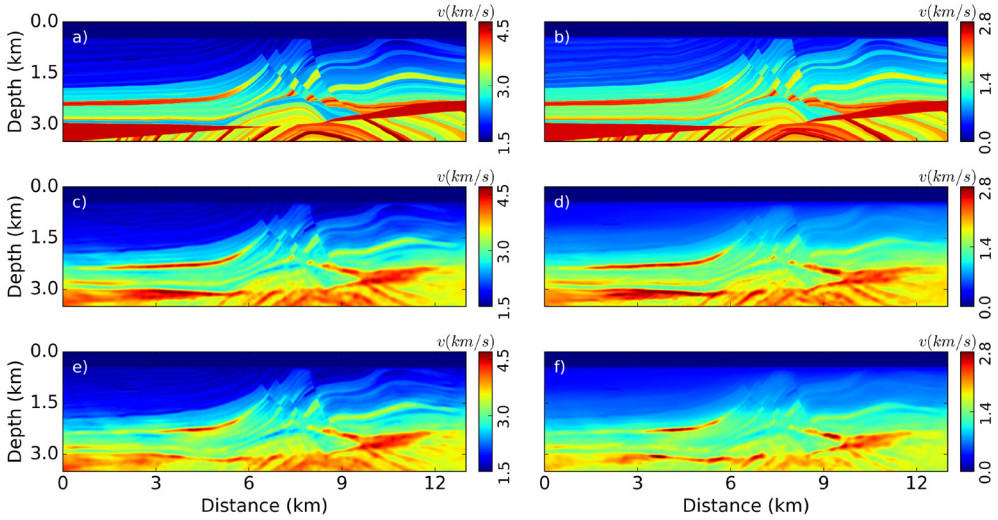

Figure 8. True models of (a) v_p and (b) v_s; inverted models of (c) v_p and (d) v_s when the data is noise-free; inverted models of (e) v_p and (f) v_s when the data is contaminated with noise with a SNR $= -0.2$ dB. Figures (c) and (d) are the same as figures 1(e) and 2(e), respectively.

accurate. The main reason for this is that these regions of the model are less illuminated than the central region of the model. The acquisition geometry imposes a limitation on the range of angles within which energy that has propagated close to the edges can be recorded. For this reason, there is less redundancy in the data carrying information on the physical properties in the regions closer to the edges. Then the estimates of the quantities within these regions become more sensitive to noise. This example demonstrates that the proposed semi-global inversion method is robust to the existence of noise in the data and it can estimate models of v_p/v_s ratio in realistic settings.

Computational aspects

In this section, we compare the computational cost of the semi-global inversion algorithm, $O(G)$, with that of conventional FWI, $O(L)$. Local iterations of FWI require computing wavefields numerically. Generally, this computation is carried out over grids with several hundreds of thousands of grid nodes, in 2D, up to several million or even thousands of millions of grid nodes, when dealing with large-scale 3D applications. These wavefields are computed several times per local iteration, as the adjoint-state method requires both the solution of the state-variable (wavefields) and of the adjoint-variable. In addition, a large amount of memory needs to be allocated in order to store these wavefields as well as the physical parameters (v_p v_s, ρ), and model updates. On the other hand, the key operations carried out at each outer global iteration are generating a set of random numbers, and carry out the update of a small set of parameters for each particle. The number of these parameters varies. However, it is in the order of a few tens. The computational intensity of operations carried out at each outer global iteration is much less than that of each nested local iteration. The most computationally intensive part of each semi-global iteration are the nested local iterations. Hence, our key estimator for the computational cost of the semi-global inversion is that of the local nested iterations. Then, the computational cost of the semi-global inversion can be directly compared

to that of conventional FWI using a single factor. Consequently, the relations we present here are general and independent of the constitutive law, e.g. acoustic, elastic or attenuating.

The computational cost of a semi-global inversion is quantified by the number of global iterations, N_g, the number of local nested iterations, N_{Ln}, and the number of particles, N_p, for each frequency band, and for each shot. Note that in the case of semi-global inversion, it suffices to carry out the inversion for one frequency-band only. As pointed out earlier, the fact that the model parameters are estimated with a sparse support means that its spatial resolution is very low. Hence, up-scaling the semi-global inversion in frequency is unnecessary. In addition, it would represent a very unfavorable trade-off between cost and benefit. The computational cost of elastic FWI is parameterized by the number of N_L local iterations per frequency band, and the number of frequency bands, N_f, for each shot. In table 1, we compare the computational cost of semi-global and local inversions per shot.

When comparing the computational cost of each type of inversion in a crude fashion, it may seem that $O(G)$ is much higher than $O(L)$. Nonetheless, when comparing all the factors contributing to each type of inversion one can obtain a relation that is, in fact, favorable to semi-global inversion.

First, the empirical experience obtained when applying this algorithm to the examples outlined before, shows that in general the number of outer global iterations is about the order of the number of frequency bands utilised in conventional FWI: $N_g \sim N_f$. In addition, the number of nested local iterations is of the same order as that of the number of local iterations, per frequency band, in conventional FWI: $N_L \sim N_{Ln}$. Then, $O(L)$ outperforms $O(G)$ by a factor given by the number of particles (second row of table 1), or $O(G) = N_p\, O(L)$, per shot. Second, we do not need to use all the data when carrying out each nested local iteration. What we do instead is skipping a portion of the data such that after completing the set of nested iterations, all the data has been used. This means that within each global iteration all the data have been used but at a fraction of the computational cost. We take advantage of the fact that the semi-global inversion is carried out at a relatively low-frequency and the parameters are estimated over a sparse basis. Hence, as long as the receiver spacing is dense enough, spatial aliasing due to shot skipping does not occur. Using this approach is very effective for alleviating the computational load. Contrarily to increasing the number of shots, increasing the number of receivers does not contribute noticeably to the computational load. When carrying out conventional FWI one can also apply the same strategy. However, as the inversion progresses in frequency less and less data can be skipped. The relation between $O(G)$ and $O(L)$ then becomes $O(G) = N_p\, O(L)/N_{skip}$ (third row of table 1), where N_{skip} is the number of shots skipped, or the proportion between the number of shots used in elastic FWI and that used in the semi-global inversion. Depending on the acquisition geometry, if the number of shots is very large and the number of skipped shots is comparable to that of the number of particles, then $O(G) \sim O(L)$ (fourth row of table 1). If the number of bands of frequency is larger than the number of global iterations and N_L is larger than N_{Ln}, then this relation becomes more favorable to semi-global inversion. Then one can run semi-global inversion for a larger number of particles, or a larger number of outer global iterations, at about the same cost as that of large bandwidth conventional elastic FWI. It is important to note that these relations do not translate into run-times. The run-time depends upon the availability of computational resources, and how the tasks are distributed. In our implementation we distribute the computations over the number of particles and shots. This allows having a quasi-linear scaling of computational resources with increasing number of particles and shots. The main obstacle to scaling are limited bandwidth of the network, and typical bottlenecks of message passing between working nodes, when using a distributed memory system. In conclusion, $O(G)$ per shot and per frequency is given by a

factor of $O(L)$ for one shot and upscaling in frequency. This factor is more or less favorable to semi-global inversion depending on the number of shots and bandwidth used in the inversions.

In the previous examples, we set $N_p = 10$, $N_g = 10$, $N_f = 9$, $N_{Ln} = 4$, and $N_L = 5$. Then, we can set the following relations: $N_g = 10N_f/9$ and $N_{Ln} = 4N_L/5$. As we use one-fourth of the data, at each nested local iteration, the computational cost is reduced by the same factor and $O(G) \approx 2.2O(L)$. The computational cost of estimating a model of v_p/v_s ratio for our examples is about a factor of 2.2 of that of running a local elastic FWI up to 10 Hz. We point out that increasing the bandwidth of the inversion requires decreasing the grid spacing in order to prevent numerical dispersion. Accordingly, the time-stepping of the numerical simulation also has to be decreased, in order to guarantee numerical stability. Hence that factor will become closer to the unity or even smaller very rapidly with increasing inversion bandwidth in elastic FWI.

Conclusion

We introduced a new approach for estimating v_p/v_s ratio from seismic data using a semi-global inversion algorithm. This new approach combines QPSO with a nested local optimization algorithm. The models of v_p/v_s are represented on a sparse basis. This sparse-basis can be chosen arbitrarily, however, it is most suited when a degree of structural information is introduced as a prior. The semi-global inversion algorithm estimates the coefficients of that sparse basis. The model of v_p/v_s is then determined in the full space of inversion (generally the grid where elastic FWI is carried out) from the mapping between the sparse basis and the inversion grid used in FWI.

Our synthetic examples demonstrated that this approach is suitable for determining a v_p/v_s constraint that is accurate enough for carrying out elastic FWI. Our method was more accurate than using an empirical relation derived from rock physics experiments. In addition, we also demonstrated that the semi-global inversion method is robust even when the data are contaminated by noise. This demonstrates its potential for inverting real data without the knowledge of an empirical relation between v_p and v_s. Our method can only determine the long wavelengths of v_p/v_s. However, the examples outlined herein demonstrated that the long wavelengths of v_p/v_s estimated with the semi-global inversion method are accurate for determining a suitable initial model of v_s, to be used as a prior in wideband elastic FWI, and to estimate high-resolution v_p and v_s models.

Our future research will be focused on applying this method to the inversion of real field data with elastic FWI.

Acknowledgments

We very gratefully acknowledge the contributions of two anonymous reviewers for reading our paper carefully and giving suggestions that improved its overall quality significantly. The authors are also thankful to the sponsors of the Fullwave consortium under the ITF research program.

Appendix. Adjoint-state method

The forward modelling operator is defined as

$$\boldsymbol{F}(\boldsymbol{p}, \boldsymbol{w}) = \text{diag}\left(\boldsymbol{A}(\boldsymbol{w})\boldsymbol{p}_1 - \boldsymbol{s}_1, \boldsymbol{A}(\boldsymbol{w})\boldsymbol{p}_2 - \boldsymbol{s}_2, \ldots\right) \ = \boldsymbol{0}, \tag{A.1}$$

where $\boldsymbol{A}(\boldsymbol{w})\boldsymbol{p}_e = \boldsymbol{s}_e$ is the discretized wave equation for a given source $\boldsymbol{s}_e$ term. The state-variable, $\boldsymbol{p} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots)$ contains the discretized wavefield, both in time and space, for all sources, and the source term $\boldsymbol{s} = (\boldsymbol{s}_1, \boldsymbol{s}_2, ...)$ contains the whole time history of a source function, and its respective distribution in space.

The gradient of the data misfit term (equation (2)), is computed in a discretize-then-optimize approach introducing the Lagrangian functional

$$\mathcal{L}(\boldsymbol{p}, \boldsymbol{\lambda}, \boldsymbol{w}) = J(\boldsymbol{p}) + \left\langle \boldsymbol{\lambda}^T, \boldsymbol{F}(\boldsymbol{p}, \boldsymbol{w}) \right\rangle_S, \tag{A.2}$$

where $\boldsymbol{w}$ is the control parameter, and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots)$ is the adjoint-variable. When carrying out local nested iterations in semi-global inversion the control parameter is $\boldsymbol{w} = \boldsymbol{v}_\mathrm{p}$, and when carrying out elastic FWI $\boldsymbol{w} = (\boldsymbol{v}_\mathrm{p}, \boldsymbol{v}_\mathrm{s})$. The angle brackets denote an inner product defined over the space of source functions S, and the condition $\langle \boldsymbol{a}, \boldsymbol{b} \rangle < \infty, \forall \boldsymbol{a}, \boldsymbol{b} \in S$ must hold. The superscript T denotes the transpose of a mathematical object. The relations for the adjoint, state, and control parameter variables, are determined from the first-order optimality conditions, $\nabla_{p,\lambda,m}\mathcal{L}(\boldsymbol{p}, \boldsymbol{\lambda}, \boldsymbol{w}) = \boldsymbol{0}$, yielding

$$\begin{cases} \boldsymbol{A}(\boldsymbol{w})\boldsymbol{p} = \boldsymbol{s} \\ \boldsymbol{A}^T\boldsymbol{\lambda} = -\nabla_p J(\boldsymbol{p}(\boldsymbol{w})). \\ \nabla_w J(\boldsymbol{p}(\boldsymbol{w})) = \langle \boldsymbol{\lambda}, [\nabla_w \boldsymbol{F}]\boldsymbol{p} \rangle_S \end{cases} \tag{A.3}$$

The first equation is the discrete forward modeling operator, the second equation determines the adjoint field and the third is the decision equation determining the update of the control parameters. The third expression in equation (A.3) determines the updates in equation (8). The mathematical operations defined in equation (A.3) are carried out at each preconditioned gradient descent iteration.

ORCID iDs

Nuno V da Silva https://orcid.org/0000-0002-3676-9634
Gang Yao https://orcid.org/0000-0002-0466-8484

References

Abubakar A, Habashy T M, Lin Y and Li M 2012 A model-compression scheme for nonlinear electromagnetic inversions *Geophysics* **77** E379–89

Afanasiev M V, Pratt R G, Kamei R and McDowell G 2014 Waveform-based simulated annealing of crosshole transmission data: a semi-global method for estimating seismic anisotropy *Geophys. J. Int.* **199** 1586–607

Castagna J, Batzle M and Eastwood R 1985 Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks *Geophysics* **50** 571–81

da Silva N V, Ratcliffe A, Vinje V and Conroy G 2016 A new parameter set for anisotropic multiparameter full-waveform inversion and application to a North Sea data set *Geophysics* **81** U25–38

da Silva N V and Yao G 2018 Wavefield reconstruction inversion with a multiplicative cost function *Inverse Problems* **34** 015004

Datta D and Sen M 2016 Estimating a starting model for full-waveform inversion using a global optimization method *Geophysics* **81** R211–23

Debens H 2015 Three-dimensional anisotropic full-waveform inversion *PhD Thesis* Imperial College London

Diouane Y, Gratton S, Vasseur X, Vicente L N and Calandra H 2016 A parallel evolution strategy for an earth imaging problem in geophysics *Optim. Eng.* **17** 3–26
Dvorkin J, Gutierrez M A and Grana D 2014 *Seismic Reflections of Rock Properties* (Cambridge: Cambridge University Press) p 352
Fernández-Berdaguer E 1998 Parameter estimation in acoustic media using the adjoint method *SIAM J. Control Optimization* **36** 1315–30
Fernández Martínez J L, Muñiz M Z F and Tompkins M J 2012 On the topography of the cost functional in linear and nonlinear inverse problems *Geophysics* **77** W1–15
Fichtner A, Bunge H-P and Igel H 2006 The adjoint method in seismology: I—theory *Phys. Earth Planet. Inter.* **157** 86–104
Gardner G H F, Gardner L W and Gregory A R 1974 Formation velocity and density the diagnostic basics for stratigraphic traps *Geophysics* **39** 770–80
Ikelle L T and Amundsen L 2005 *Introduction to Petroleum Seismology: 12* (*Investigations in Geophysics*) (Tulsa, OK: Society of Exploration Geophysicists) p 706
Kachitvichyanukul V 2012 Comparison of three evolutionary algorithms: GA, PSO, and DE, Industrial *Eng. Manage. Syst.* **11** 215–23
Kennedy J and Eberhart R C 1995 Particle swarm optimization *Proc. of IEEE Int. Conf. on Neural Networks* pp 1942–8
Kennedy J and Eberhart R C 2001 *Swarm Intelligence* (San Francisco, CA: Morgan Kaufmann Publishers)
Köhn D, De Nil D, Kurzmann A, Przebindowska A and Bohlen T 2012 On the influence of model parametrization in elastic full waveform tomography *Geophysical J. Int.* **191** 325–45
Lailly P 1983 The seismic inverse problem as a sequence of before stack migrations *Conference on Inverse Scattering: Theory and Application* ed J B Bednar (Philadelphia, PA: Society for Industrial and Applied Mathematics) pp 206–20
Levander A R 1988 Fourth-order finite-difference *P-SV* seismograms *Geophysics* **53** 1425–36
Martin G S, Marfurt K J and Larsen S 2002 Marmousi-2: an updated model for the investigation of AVO in structurally complex areas *72nd Annual Int. Meeting, SEG, Expanded Abstracts. 1979–1982* (https://doi.org/10.1190/1.1817083)
Mora P 1987 Nonlinear two-dimensional elastic inversion of multioffset seismic data *Geophysics* **52** 1211–28
Operto S, Miniussi A, Brossier R, Combe L, Métivier L, Monteiller V, Ribodetti A and Virieux J 2015 Efficient 3D frequency-domain multi-parameter full-waveform inversion of ocean bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation *Geophys. J. Int.* **202** 1362–91
Pratt R G, Shin C and Hick G J 1998 Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion *Geophys. J. Int.* **133** 341–62
Pratt R G and Shipp R M 1999 Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data *Geophysics* **64** 902–14
Press W, Teukolsky S, Vetterling W and Flannery B 2007 *Numerical Recipes* (*The Art of Scientific Computing*) 3rd edn (Cambridge: Cambridge University Press)
Queißer M and Singh S C 2013 Full waveform inversion in the time lapse mode applied to CO_2 storage at Sleipner *Geophys. Prospect.* **61** 537–55
Raknes E B, Arnsten B and Weibull W 2015 Three-dimensional elastic full waveform inversion using seismic data from the Sleipner area *Geophys. J. Int.* **202** 1877–94
Routh P *et al* 2017 Impact of high-resolution FWI in the Western Black Sea: revealing overburden and reservoir complexity *Leading Edge* **36** 6066
Sedov L I 1994 *Mechanics of Continuous Media Vol. 1* (Series in Theoretical and Applied 20 Mechanics) (Singapore: World Scientific Publishing)
Sen M and Stoffa P 2013 *Global Optimization Methods in Geophysical Inversion* (Cambridge: Cambridge University Press)
Sirgue L, Barkved O, Gestel J V, Askim O and Kommedal J 2009 3D waveform inversion on Valhall wide-azimuth obc *71st Annual Int. Conf. and Exhibition* (EAGE) (https://doi.org/10.3997/2214-4609.201400395)
Sun J, Fang W, Wu X, Palade V and Xu W 2012 Quantum-behaved particle swarm optimization: analysis of individual particle behavior and parameter selection *Evolutionary Comput.* **20** 349–93
Sun J, Feng B and Xu W 2004a Particle swarm optimization with particles having quantum behavior *Proc. of the 2004 Congress on Evolutionary Computation* (Institute of Electrical and Electronics Engineers) pp 325–31

Sun J, Feng B and Xu W 2004b A global search strategy of quantum-behaved particle swarm optimization *IEEE Conf. on Cybernetics and Intelligent Systems* pp 111–6

Tarantola A 1984 Inversion of seismic reflection data in the acoustic approximation *Geophysics* **49** 1259–66

Tarantola A 1986 A strategy for nonlinear elastic inversion of seismic reflection data *Geophysics* **51** 1893–903

Versteeg R 1994 The Marmousi experience: velocity model determination on a synthetic complex data set *Leading Edge* **13** 927–36

Vigh D, Jiao K, Watts D and Sun D 2014 Elastic full-waveform inversion application using multicomponent measurements of seismic data collection *Geophysics* **79** R63–77

Warner M A *et al* 2013 Anisotropic 3D full-waveform inversion *Geophysics* **78** R59–80

Yao G, da Silva N V and Wu D 2018 An effective absorbing layer for the boundary condition in acoustic seismic wave simulation *J. Geophys. Eng.* **15** 495–511