Journal of Computational Physics 230 (2011) 7502–7518

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Interlocked optimization and fast gradient algorithm for a seismic inverse problem

Ludovic Métivier [1]

LAGA, Université Paris XIII, 99 Avenue Jean-Baptiste Clément, 93000 Epinay-Villetaneuse, France

ARTICLE INFO

Article history:
Received 25 March 2011
Received in revised form 15 June 2011
Accepted 20 June 2011
Available online 30 June 2011

Keywords:
Nonlinear inverse problem
Optimization
Adjoint state method
High performance computing
Seismic imaging
High resolution

ABSTRACT

We give a nonlinear inverse method for seismic data recorded in a well from sources at several offsets from the borehole in a 2D acoustic framework. Given the velocity field, approximate values of the impedance are recovered. This is a 2D extension of the 1D inversion of vertical seismic profiles [18]. The inverse problem generates a large scale undetermined ill-conditioned problem. Appropriate regularization terms render the problem well-determined. An interlocked optimization algorithm yields an efficient preconditioning. A gradient algorithm based on the adjoint state method and domain decomposition gives a fast parallel numerical method. For a realistic test case, convergence is attained in an acceptable time with 128 processors.

1. Introduction

Recovering structural and quantitative information about the crust composition is a fundamental problem in industrial applications including seismic prospecting and CO_2 storage monitoring. Since direct measurements only give very local information, seismic imaging methods are based upon indirect measures of the subsurface parameters obtained with measurements of acoustic probes. One measures the subsurface response to an impulse at the ground surface. The impulse is generated by seismic sources called the illuminating wave. This source field propagates through the subsurface and undergoes reflection, refraction, diffraction, transmission and attenuation. These interactions between waves and subsurface are governed by the subsurface physical properties. Traditionally, receivers located hundreds of meters from the sources record the reflected signal at the ground surface. The recordings constitute the seismic data. Seismic imaging inverse methods process this data in order to estimate the geophysical subsurface parameters. Given a seismic wave propagation model, the estimated parameters must reproduce as accurately as possible the data that have been recorded. Typical lengths of investigation can reach several kilometers in depth and in lateral directions.

Such seismic imaging methods have been intensively used and improved during the last three decades. The most frequently used ones are "linear" inverse methods. Denote by m the subsurface parameters that have to be estimated. The wave propagation model is linearized around a smooth approximation m_0 of m, and an estimation of a perturbation dm is sought, such that $m \approx m_0 + dm$. This linearization relies on the assumption that the part of the recorded signal that depends nonlinearly on dm can be neglected. Given a prior estimation of the seismic source, these methods result in a linear least squares problem which can be efficiently solved. They provide quantitative estimates of the subsurface parameters with a spatial resolution (that is to say the smallest length scale resolved) reaching about 25 m [6].

E-mail address: ludovic.metivier@gmail.com
[1] IFPEN, Rueil-Malmaison, and LAGA, Université de Paris XIII, France.

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.06.019

These linear inverse methods are faced with two important difficulties:

- determining a sufficiently accurate background estimation m_0 can be difficult or impossible;
- for strongly heterogeneous media, the part of the seismic signal depending nonlinearly on dm increases, rendering the linearization inappropriate.

In these two cases, the wave propagation model is not likely to give a good interpretation of the seismic data. As a result, the approximation for dm is not satisfactory [9,10].

This leads one to try to solve the inverse problem without the linearization approximation. There is evidence that this can alleviate these two difficulties, and even increase the resolution of the resulting method [11,12]. Unfortunately, in this case, another difficulty arises, related to the accuracy of the estimation of the seismic source. While the linearized inverse problem is not sensitive to an error on the source estimation, the nonlinear inverse problem is. This results in the impossibility of inverting ground surface seismic data with the nonlinear method.

Another type of imaging method, called *vertical seismic profile inversion*, was introduced in the 80s [18]. For this method, receivers are placed in a vertical well instead of being located at the ground surface. This configuration provides an interesting advantage. When receivers are located at the ground surface, they record only reflected signals. When they are located in a well below the sources, the receivers record both direct and reflected signals. The resulting data contain non negligible information from the illuminating wavefield that has directly passed through the substrate. This method can simultaneously recover the illuminating wavefield and the subsurface parameters from the data. The illuminating wavefield thus becomes an unknown. This opens the door to attacking the fully nonlinear problem.

Such a method was first implemented by Macé and Lailly in a 1D aoustic wave propagation problem [18]. The subsurface is described by two parameters: the acoustic impedance $I(z)$ and the seismic wave velocity $c(z)$. The latter is first estimated from arrival times. The method then recovers approximate acoustic impedance values. For the 1D problem, the shallowest receiver records a signal $h(t)$ corresponding to a Dirichlet boundary condition for a 1D wave propagation model starting at the depth z_1 of the first receiver. The method consists in finding the acoustic impedance model $I(z)$ from the depth z_1 and the Dirichlet condition $h(t)$ that minimizes the misfit between the data and the model prediction. The problem is proved to be well-posed, and the impedance model is uniquely determined up to a multiplicative constant [3,4]. The method is stable and robust even with highly noise-corrupted data. In addition, the impedance estimation shows an excellent vertical resolution, which suggests the interest of solving the fully nonlinear problem.

However, the 1D propagation assumption is not realistic, and the acoustic impedance estimation in only valid in the vicinity of the well. Therefore, extending this 1D nonlinear seismic imaging method to a multidimensional framework is an attractive idea. Such an extension could lead to the estimation of subsurface parameters within a zone around the well. The present paper proposes an implementation of such an extension. A 2D acoustic wave propagation problem is investigated. As in 1D, the subsurface is described by the wave velocity and the acoustic impedance. On the basis of an assumed known velocity field, the method aims at recovering the 2D acoustic impedance.[2] Walkaway data is used. Seismic sources are placed at different offsets from the borehole in order to illuminate horizontal variations of the medium. For each location of the source, the particle displacement velocities are recorded by an array of receivers in a well. This extension to a 2D framework of the originally 1D method raises two main difficulties.

- The 2D inverse problem is undetermined and requires additional *a priori* information on the unknowns.
- For realistic applications, the 2D inverse problem results in a large scale ill-conditioned nonlinear minimization problem for which an appropriate numerical method has to be constructed.

In [21], we mainly focused on the first of these two problems and proposed suitable regularization techniques. Here we focus on the second problem. The numerical method involves an interlocked optimization algorithm based upon an external quasi-Newton loop used to estimate the impedance, together with a conjugate gradient inner loop, used for the computation of the Dirichlet condition. The algorithm requires a fast determination of the gradient of the misfit function. The adjoint state approach is used together with a domain decomposition method to yield a parallelizable algorithm. The efficiency of this numerical method is evaluated on a realistic test case, with synthetic data. The results are encouraging.

- They show satisfactory convergence of the method in a reasonable time.
- The 2D acoustic impedance distribution is estimated with a high resolution within a 200 m and 3 km depth rectangle around the well.

The article is organised as follows. In Section 1, the 2D inverse problem is described in detail. In Section 2, the numerical method used to solve this inverse problem is presented. In Section 3 the results of the test case are exposed.

[2] This assumption is important. The determination of the wave velocity field from seismic data is a fundamental problem with its own methods [28]. We assume an approximation for $c(x)$ is available and focus on determining $I(x)$.

2. Inverse problem definition

2.1. Forward problem

2.1.1. Definition of the walkaway data

An array of Q receivers is placed in a vertical well drilled from the ground surface $z = 0$. For the sake of simplicity, the horizontal location of the receivers is set to $x = 0$. Their vertical locations are denoted by z_q, $1 \leqslant q \leqslant Q$. In order to collect information about lateral variations, a seismic source is placed at different offsets from the borehole. For each offset, the receivers record the displacement velocity vector, during a time $0 \leqslant t \leqslant T$. The total number of offsets is denoted by S. An example of such a configuration is displayed on Fig. 3. In a 2D framework, the total dataset (also called walkaway dataset in geophysics) is thus composed of $2 \times S \times Q$ functions $D_x^{s,q}(t)$ and $D_z^{s,q}(t)$, giving the vertical and the horizontal velocity displacement, for the shot number $1 \leqslant s \leqslant S$ and the receiver number $1 \leqslant q \leqslant Q$.

Datasets corresponding to a single shot s are denoted by $D^s(t)$ for example

$$D^s(t) = \left\{ D_x^{s,q}(t),\ D_z^{s,q}(t),\ q = 1, \ldots, Q \right\} \tag{1}$$

and the total walkaway dataset is denoted by $D(t)$,

$$D(t) := \left\{ D^s(t),\ 1 \leqslant s \leqslant S \right\} \tag{2}$$

2.1.2. Acoustic wave equations

Is is assumed that the seismic wave propagation can be described by the 2D acoustic wave equations. This amounts to neglecting surface waves and shear-waves (S-waves) propagating in the subsurface, and only retaining pressure waves (P-waves). The subsurface is described by the acoustic impedance $I(\mathbf{x})$ and the wave velocity $c(\mathbf{x})$, where $\mathbf{x} = (x,z)$ is the 2D vector of spatial coordinates. It is also assumed that a good approximate value of $c(\mathbf{x})$ is available. The aim of the method is to recover an approximation for the acoustic impedance $I(\mathbf{x})$. Then, the modelisation problem consists in computing the signal recorded by the receivers for a given acoustic impedance model $I(\mathbf{x})$ and an illuminating wavefield corresponding to a Dirichlet pressure condition at the depth of the shallowest receiver, denoted by $h(x,t)$. It should be emphasized that the propagation of seismic waves is not simulated from the ground surface $z = 0$, for a given source, but from the depth of the shallowest receiver $z = z_1$, for a given pressure wavefield at this depth. Thus, the entries of the modelisation problem are an acoustic impedance model, and a Dirichlet boundary condition corresponding to the illuminating wavefield at depth $z = z_1$. These boundary conditions shall be referred to as "pressure conditions" in the sequel.

The corresponding equations for the pressure wavefield $p(\mathbf{x},t)$ in the half-space $\mathbb{R} \times [z_1, +\infty[\times [0,T]$ are

$$\begin{cases} \partial_{tt} p(\mathbf{x},t) - c(\mathbf{x}) I(\mathbf{x}) \mathrm{div}\left(\frac{c(\mathbf{x})}{I(\mathbf{x})} \nabla p(\mathbf{x},t) \right) = 0 \\ p(x, z_1, t) = h(x,t) \end{cases} \tag{3}$$

2.1.3. Absorbing boundary conditions

Although these equations are defined in the half-space $\mathbb{R} \times [z_1, +\infty[$, the computational domain is necessarily bounded. Avoiding fictitious reflections implies to define absorbing boundary conditions at the boundaries of the computational domain. The PML (Perfectly Matched Layers), introduced by Bérenger for the 2D Maxwell equations [5], are chosen. The principle of this method is to surround the computational domain with a layer in which incident waves decay. Let $\mathcal{O}$ be the initial computational domain:

$$\mathcal{O} = [-a, a] \times [z_1, z_{max}] \tag{4}$$

Let $l \in \mathbb{R}$ be the layer width. The actual computational domain then becomes

$$\Omega = [-X; X] \times [z_1, Z], \quad \text{with } X = a + l, \quad Z = z_{max} + l \tag{5}$$

The Bérenger PML method is designed for hyperbolic systems. The Eq. (3) are thus rewritten as an order 1 in time system involving the pressure wavefield $p(\mathbf{x},t)$ and the displacement velocity vector $\mathbf{u}(\mathbf{x},t)$. Then the pressure field $p(\mathbf{x},t)$ is split into two fields $p_x(\mathbf{x},t)$, $p_z(\mathbf{x},t)$ such as $p = p_x + p_z$ and absorption coefficients $\sigma_x(x)$, and $\sigma_z(z)$ are introduced (respectively in directions x and z). The corresponding PML system is:

$$\begin{cases} \frac{I(\mathbf{x})}{c(\mathbf{x})} \partial_t u_x(\mathbf{x},t) - \partial_x p(\mathbf{x},t) + \sigma_x(x) \frac{I(\mathbf{x})}{c(\mathbf{x})} u_x(\mathbf{x},t) = 0 \\ \frac{I(\mathbf{x})}{c(\mathbf{x})} \partial_t u_z(\mathbf{x},t) - \partial_z p(\mathbf{x},t) + \sigma_z(z) \frac{I(\mathbf{x})}{c(\mathbf{x})} u_z(\mathbf{x},t) = 0 \\ \frac{1}{I(\mathbf{x})c(\mathbf{x})} \partial_t p_x(\mathbf{x},t) - \partial_x u_x(\mathbf{x},t) + \frac{\sigma_x(x)}{I(\mathbf{x})c(\mathbf{x})} p_x(\mathbf{x},t) = 0 \\ \frac{1}{I(\mathbf{x})c(\mathbf{x})} \partial_t p_z(\mathbf{x},t) - \partial_z u_z(\mathbf{x},t) + \frac{\sigma_z(z)}{I(\mathbf{x})c(\mathbf{x})} p_z(\mathbf{x},t) = 0 \end{cases} \tag{6}$$

with the boundary conditions

$$\begin{cases} (p_x + p_z)(x, z_1, t) = h(x,t), & (p_x + p_z)(-X, z, t) = 0 \\ (p_x + p_z)(X, z, t) = 0, & (p_x + p_z)(x, Z, t) = 0 \end{cases} \tag{7}$$

and null initial conditions.

The absorption coefficients $\sigma_x(x)$, $\sigma_z(z)$ are set to zero in the interest domain $\mathcal{O}$ and grow polynomially from the interface between $\mathcal{O}$ and the PML layer, toward the external border of the PML layer. They are chosen as in Hu [15].

A mathematical study of these equations has been achieved in [19]. An augmented system composed of the initial variables and their first order spatial derivatives is shown to be well-posed. Based on this system, energy estimates for the initial variables are computed. They show that the stability of the solution is ensured under regularity assumption over initial and boundary conditions. In the sequel, the pressure condition $h(x,t)$ is considered to be in $H^q(\Sigma)$, with $\Sigma = [-X;X] \times [0,T]$, and $q \in \mathbb{N}$ large enough for this regularity assumption to be met.

2.1.4. Definition of the forward problem

The forward problem is defined on the basis of Eqs. (6), (7). A first operator $\mathcal{M}_1(I,h)$ associates to one pressure condition $h(x,t)$ and one acoustic impedance model $I(\mathbf{x})$ the solution of these equations:

$$\begin{aligned} &\mathcal{M}_1 : (I,h) \;\rightarrow\; U(\mathbf{x},t) = (u_x(\mathbf{x},t), u_z(\mathbf{x},t), p_x(\mathbf{x},t), p_z(\mathbf{x},t)) \\ &H^1(\Omega) \times H^q(\Sigma) \;\rightarrow\; H^1(\Omega \times [0,T])^4 \end{aligned} \tag{8}$$

with $u_x(\mathbf{x},t)$, $u_z(\mathbf{x},t)$, $p_x(\mathbf{x},t)$, $p_z(\mathbf{x},t)$ are solutions of (6), (7).

Let $\mathcal{M}_2(U)$ be the so-called observation operator, mapping an arbitrary $U(\mathbf{x},t)$ into the data space:

$$\begin{aligned} &\mathcal{M}_2 : U(\mathbf{x},t) \;\rightarrow\; \big(u_x(0,z_q,t), u_z(0,z_q,t),\; q = 1,\dots,Q\big) \\ &H^1(\Omega \times [0,T])^4 \rightarrow H^1(0,T)^{2Q} \end{aligned} \tag{9}$$

The modelisation operator $\mathcal{M}(I,h)$ is defined as the composition of these two operators:

$$\begin{aligned} &\mathcal{M} : (I,h) \;\rightarrow\; (\mathcal{M}_2 \circ \mathcal{M}_1)(I,h) \\ &H^1(\Omega) \times H^q(\Sigma) \;\rightarrow\; H^1(0,T)^{2Q} \end{aligned} \tag{10}$$

From this modelisation operator, it is possible to define the 2D inverse problem.

2.2. Inverse problem

2.2.1. An undetermined inverse problem

Let $\mathcal{H}(x,t)$ be the vector containing the S pressure conditions associated to the S datasets acquired for each offset:

$$\mathcal{H}(x,t) = \left\{ h^1(x,t), \dots, h^S(x,t) \right\} \tag{11}$$

The 2D inverse problem is the following: for a given walkaway dataset $D(t)$, find a couple $\big(\overline{I}(\mathbf{x}), \overline{\mathcal{H}}(x,t)\big)$ such as

$$\big(\overline{I}(\mathbf{x}), \overline{\mathcal{H}}(x,t)\big) = \arg\min_{I,\mathcal{H}} J(I,\mathcal{H}), \tag{12}$$

where $J(I,\mathcal{H}) = \frac{1}{2}\sum_{s=1}^{S} \|M(I,h^s) - D^s\|^2$.

The function $J(I,\mathcal{H})$ is the so-called seismic misfit function. Note that $J(I,\mathcal{H})$ is defined as a sum of S misfits functions, each one corresponding to data acquired for one source position. The choice of the norm $\|\cdot\|$ involved in the definition of $J(I,\mathcal{H})$ reflects the prior knowledge about uncertainties on the dataset $D(t)$. In the present case, a L^2 norm may simply be chosen, which suggests that a Gaussian noise corrupts the data. More relevant choices could possibly be made depending on applications and available data [27].

In 1D, the inverse problem (12) is well-posed. The dataset is composed of a single shot, and due to the specificity of well-seismic data acquisition, it contains all relevant information about both the pressure condition $h(t)$, and the acoustic impedance $I(z)$ [3,4]. However, in 2D, this problem becomes severely undetermined, notably because the data contain information about $h(x,t)$ only in $x = 0$, where it is recorded. The number of unknowns and coupling effects between $h(x,t)$ and $I(\mathbf{x})$ on the seismic response make the inverse problem (12) ill-posed. Consequently, regularization terms should be included in the function $J(I,\mathcal{H})$, in order to take into account all the prior information that can be provided to the solution.

2.2.2. Introducing regularization terms

First, the method is restrained to heterogeneous subsurfaces varying slowly along the x direction. A regularization term of the form $\frac{\varepsilon}{2}\|\partial_x I(\mathbf{x})\|^2$ can thus be used.

However, this information about the impedance model regularity is not sufficient: regularity information about the pressure conditions is also needed. As it has already been mentioned, each of these pressure conditions is associated to a dataset acquired for a particular offset of the source. Consequently, they present hyperbolic patterns in their definition space (x,t) (Fig. 1 left). Since prior determination of these hyperbolic patterns is a difficult issue, an alternative method has to be found

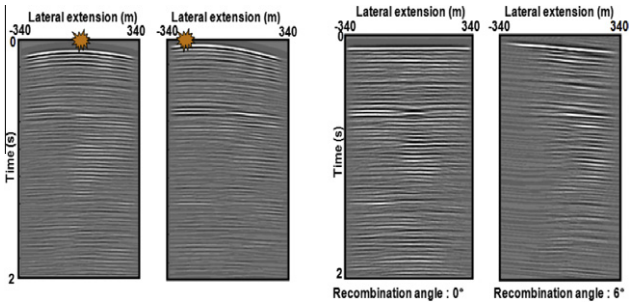

Fig. 1. Pressure conditions corresponding to data acquired with a single source at the ground surface (left) and after slant stacking (right).

to get prior information on pressure condition regularity. Instead of processing directly the walkaway data, a Radon transform is applied to the data [24]. This process has been introduced by Claerbout in geophysics as slant-stacking [8]. It transforms the datasets associated to each offset to datasets corresponding to the propagation of quasi-plane waves, through appropriate linear combination of time shifted data. The time shift governs the propagation angles of the quasi-plane waves. For these latter, the pressure conditions present regularity along a direction γ_s which can be computed from the propagation angle of the quasi-plane waves (Fig. 1 right).

A modified walkaway dataset is thus obtained $\widehat{D}(t) = \left\{\widehat{D}^s(t),\ s = 1, \widehat{S}\right\}$. It corresponds to the propagation of $\widehat{S}$ quasi-plane waves.[3] Consequently, the regularization terms $\frac{\varepsilon}{2}\|\gamma_s \nabla h^s(x,t)\|^2,\ s = 1, \widehat{S}$ are included in the misfit function.

This point is particularly crucial for defining a well-determined inverse problem. The interested reader is referred to [21] for more details.

2.2.3. The regularized inverse problem

The regularized function $J^\varepsilon(I, \mathcal{H})$ is defined by

$$J^\varepsilon(I,\mathcal{H}) = \frac{1}{2}\sum_{s=1}^{\widehat{S}} \|M(I,h^s) - \widehat{D}^s\|^2 + \frac{\varepsilon_1}{2}\|\partial_x I(\mathbf{x})\|^2 + \frac{\varepsilon_2}{2}\sum_{s=1}^{\widehat{S}} \|\gamma_s.\nabla h^s(x,t)\|^2, \tag{13}$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$ are the regularization weights. These weights reflect the relative confidence over the three following information sources:

- seismic information;
- *a priori* regularity information over acoustic impedance $I(\mathbf{x})$;
- *a priori* regularity information over pressure conditions $h^s(x,t)$.

In practice, these weights are determined by a trial and error approach. An acceptable balance between the three different sources of information at the optimum is sought. This corresponds to the assumption that an uncorrelated Gaussian noise corrupts the three sources of information in a comparable way.

The final nonlinear inverse problem to be solved is:

$$\min_{I,\mathcal{H}} J^\varepsilon(I,\mathcal{H}). \tag{14}$$

Thanks to the regularization terms, this problem is expected to be well-determined. No mathematical proof is currently available: this a difficult issue to address. However, a theoretical justification is proposed by the authors in [21]. Moreover, numerical experiments show that meaningful solutions to problem (14) can be computed. The numerical method set up to solve this problem is presented in the next section.

3. Numerical method

3.1. Main difficulties

Some elements about the discretization of problem (14) are first discussed. From the geophysical application point of view, the main motivation of this work is to propose an inverse method which yields a high resolution impedance estimation. This leads naturally to use fine discretization grids. This has two major consequences.

[3] Choosing $\widehat{S} = S$ should ensure no information loss in the slant stacking process.

- Second order discretization schemes suffice to reach acceptable accuracy.
- The number of discrete unknowns becomes rapidly tremendous when the method is applied to real problems.

Therefore, the discretization of the forward problem Eqs. (6), (7) is made through a second order Yee scheme [29]. For the test case presented in the next section, this turns out in the definition of approximately two millions discrete data and several hundred thousand discrete unknowns. Thus, an efficient solving numerical method is required.

Another difficulty also arises: the inverse problem (14), though well-determined, is ill-conditioned. This mainly comes from the fact that two different kind of unknowns are estimated simultaneously: pressure conditions and acoustic impedance. The considerable differences in numerical sensitivities of these two kind of unknowns result in an ill-conditioned problem, for which conventional optimization methods fail to converge.

The method presented here is set up to overcome these two difficulties. It takes advantage of the quadratic dependence of function $J(I,\mathcal{H})$ with respect to $\mathcal{H}(x,t)$, and acts as an efficient pre-conditioning of problem (14). In addition, a fast algorithm for the computation of the gradient is set up. Convergence can thus be reached within an acceptable time on a cluster based computer. These two aspects of the numerical method are presented in the following subsections.

3.2. Optimization algorithm

3.2.1. Choice of an optimizer

Newton methods. For large scale differentiable minimization problems, the most commonly used methods are the quasi-Newton algorithms, deriving from the well-known Newton algorithm. Their principle is recalled briefly. The Newton algorithm is defined as follows: for a positive real valued function $f(x)$ the problem $\min_{x\in\mathbb{R}^n} f(x)$ is solved by generating a sequence of iterates x_k, from a starting point x_0, such as $x_{k+1} = x_k - \alpha(M(x_k))^{-1}\nabla f(x_k)$, where $\alpha \in \mathbb{R}$, and $M(x_k) = f'(x_k)$.

The matrix $M(x)$ is known as the Hessian matrix of function $f(x)$, and the vector $M^{-1}(x_k)\nabla f(x_k)$ is the Newton descent direction. The role of the coefficient α is to guarantee theglobal convergence of the algorithm. Different types of linesearch or trust region methods lead to as many different definitions [13].

Quasi-Newton methods. For large scale problems, direct computation of $M(x_k)$ should be avoided. Quasi-Newton methods handle this difficulty replacing $(M(x_k))^{-1}$ by an approximation $C(x_k)$. This approximation is built from a starting matrix (which can be identity) and an update formula based on the gradient $\nabla f(x_k)$. Different update formulas lead to different quasi-Newton methods. Among them, the BFGS method (from the name of its conceivers Broyden, Fletcher, Goldfard, Shanno) [25] has proved to be very efficient, for numerous applications. In addition, this method has been adapted to very-large scale problem, in order to reduce storage and computation cost of the approximation matrix $C(x_k)$. This defines the l-BFGS algorithm, for which only the last l previous steps are used and stored to compute the approximation matrix $C(x_k)$.

Although this algorithm seems to be adequate to solve problem (14), it turns out that it is quite inefficient when applied to the minimization of function $J(I,\mathcal{H})$. The different numerical sensitivities of pressure conditions and acoustic impedance makes problem (14) ill-conditioned, and the l-BFGS algorithm fails to converge. Therefore, another method has to be defined.

3.2.2. Interlocked optimization algorithm

Basic ideas. The interlocked optimization method proposed here is based upon the decoupling of the minimization process over the two kinds of unknowns (pressure conditions $\mathcal{H}(x,t)$ and impedance model $I(\mathbf{x})$). This decoupling is motivated by the fact that $J^\varepsilon(I,\mathcal{H})$ depends quadratically on the pressure conditions $\mathcal{H}(x,t)$, what may be called a "weakly nonlinear" dependency, while it depends "strongly nonlinearly" on the impedance model $I(\mathbf{x})$. From there comes the idea to use a conjugate gradient algorithm to estimate the optimal pressure conditions, while the l-BFGS algorithm may only be used for the optimal acoustic impedance model. Moreover, the interlocked algorithm shall not require any further computations than the previous one: the computation of $\nabla J^\varepsilon(I,\mathcal{H})$ should be sufficient to ensure the convergence of the interlocked algorithm.

Method definition. On the basis of these observations, problem (14) is rewritten as

$$\min_I \widetilde{J}^\varepsilon(I), \quad \text{where } \widetilde{J}^\varepsilon(I) = J^\varepsilon(I,\overline{\mathcal{H}}(I)), \quad \text{and } \overline{\mathcal{H}}(I) = \arg\min_{\mathcal{H}} J^\varepsilon(I,\mathcal{H}) \tag{15}$$

The gradient of $\widetilde{J}^\varepsilon(I)$ is obtained straightforwardly from the gradient of the function $J^\varepsilon(I,\mathcal{H})$: $\widetilde{J}'^\varepsilon(I) = \partial_I J^\varepsilon(I,\overline{\mathcal{H}}(I)) + \partial_{\mathcal{H}} J^\varepsilon(I,\overline{\mathcal{H}}(I)) \cdot \partial_I \overline{\mathcal{H}}(I)$.

Since $\partial_{\mathcal{H}} J^\varepsilon(I,\overline{\mathcal{H}}(I)) = 0$ from the definition of $\overline{\mathcal{H}}(I)$, it comes that

$$\widetilde{J}'^\varepsilon(I) = \partial_I J^\varepsilon(I,\overline{\mathcal{H}}(I)). \tag{16}$$

In addition, if the function $J^\varepsilon(I,\mathcal{H})$ is rewritten formally as

$$J^\varepsilon(I,\mathcal{H}) = \sum_{s=1}^{\widehat{S}} \|F(I) \cdot h^s - \widehat{D}^s\|^2 + \varepsilon_1\|\partial_x I\|^2 + \varepsilon_2 \sum_{s=1}^{S} \|G^s \cdot h^s\|^2, \tag{17}$$

where $F(I)$ and G^s are linear applications acting on h^s, the computation of $\overline{\mathcal{H}}(I)$ for a given $I(\mathbf{x})$ consists in solving the $\widehat{S}$ normal equations

$$\left(F(I)^T F(I) + \varepsilon_2 (G^s)^T G^s\right) h^s = F^T(I) \widehat{D}^s, \quad s = 1, \widehat{S}. \tag{18}$$

For a given $s \in \mathbb{N}$, the application of the conjugate gradient method to solve the linear system (18) requires the computation of the right-hand side

$$F^T(I) \widehat{D}_s = \partial_{h^s} J^\varepsilon(I, 0) \tag{19}$$

and the matrix-vector products

$$\left(F(I)^T F(I) + \varepsilon_2 (G^s)^T G^s\right) h^s = \partial_{h^s} J^\varepsilon(I, \mathcal{H}) - \partial_{h^s} J^\varepsilon(I, 0) \tag{20}$$

Thus, the computation of $\partial_{h_s} J^\varepsilon(I, \mathcal{H}),\ s = 1, \widehat{S}$ is sufficient to obtain $\overline{\mathcal{H}}(I)$ through the conjugate gradient algorithms. Therefore, the interlocked optimization algorithm may only require the computation of $\nabla J^\varepsilon(I, \mathcal{H})$ at each iteration. It is conceived as follows: the function $\widetilde{J}^\varepsilon(I)$ is minimized over the impedance model $I(\mathbf{x})$ with an l-BFGS method, which requires the computation of $\partial_I J^\varepsilon(I, \overline{\mathcal{H}}(I))$. At each iteration of the l-BFGS method, the computation of $\overline{\mathcal{H}}(I)$ is performed by $\widehat{S}$ successive conjugate gradient runs.

Preconditioning the conjugate gradient. It is possible to improve and speed-up the convergence of this algorithm by preconditioning the linear systems solved in the inner conjugate gradient loop. The preconditioners are computed together with the solutions of the linear systems, each time the conjugate gradient is applied to the optimal pressure conditions $\overline{\mathcal{H}}(I)$, and then used for the preconditioning of the linear systems arising on the next iteration of the l-BFGS outer loop. These preconditioners are computed as the l-BFGS approximations of the inverse Hessian matrix of the quadratic function associated to the linear system. The software N1CG1 provides an implementation of such a conjugate gradient algorithm, which yields simultaneously the solution of a linear system and a preconditioner for a similar linear system [14]. Concerning the l-BFGS method, among numerous available l-BFGS algorithms, we chose the one proposed by Nocedal et al. [22], for its robustness, and its ability to propose bound constraints. These latter are important in this problem: impedance values have to remain strictly positive, unless the numerical solution of the forward problem diverges.

The interlocked algorithm is summarized below (Algorithm 1). This algorithm ensures convergence to a solution of problem (14). The decoupling between pressure conditions and impedance model acts as a preconditioning of the global problem. However, the convergence must be reached within an acceptable computational time. The most time-consuming part of the interlocked algorithm is the computation of the gradient $\nabla J^\varepsilon(I, \mathcal{H})$, thus it is fundamental to design a fast algorithm to compute it. This is presented in the next section.

Algorithm 1: Interlocked optimization algorithm

Data: $I^0, \mathcal{H}^0, \rho, N_{iter}$
Result: $\arg\min_{I,\mathcal{H}} J^\varepsilon(I, \mathcal{H})$
$I = I^0$
for $s = 1, \widehat{S}$ **do**
  Compute $\overline{h^s}$ using conjugate gradient without preconditoning
  Compute precodonitioner B^s
end
Compute $g = \partial_I J^\varepsilon(I, \overline{\mathcal{H}})$
Initialize inverse Hessian approximation $C = Id$
$k = 0$
while $\widehat{J}^\varepsilon(I, \overline{\mathcal{H}}(I)) > \rho$ and $k < N_{iter}$ **do**
  Compute quasi-Newton descent direction $d = -Cg$
  Compute coefficent $\alpha \in \mathbb{R}_+^*$ (linesearch)
  Compute new impedance iterate $I = I + \alpha d$
  for $s = 1, \widehat{S}$ **do**
    Compute $\overline{h^s}(I)$ with conjugate gradient preconditioned by B^s
    Compute new preconditioner and store it in B^s
  end
  Compute $g = \partial_I J^\varepsilon(I, \overline{\mathcal{H}})$
  Update inverse Hessian approximation C with BFGS formula
  $k = k + 1$
end

3.3. Fast gradient algorithm

3.3.1. Adjoint state method

Definition of the Lagrangian operator. The function $J^\varepsilon(I,\mathcal{H})$ is composed of three terms. Fast and reliable computation of the gradient of the regularization terms $\frac{\varepsilon_1}{2}\|\partial_x I\|^2$ and $\frac{\varepsilon_2}{2}\|\gamma_s \cdot \nabla h_s\|^2,\ s=1,\widehat{S}$ is straightforward. The main difficulty comes from the computation of the gradient of the seismic misfit function $J(I,\mathcal{H})$.

Direct computation of $\nabla J(I,\mathcal{H})$ implies computation of the forward operator jacobian matrixes $\partial_I \mathcal{M}(I,h^s)$ and $\partial_{h^s}\mathcal{M}(I,h^s)$. Due to the tremendous number of discretes unknowns, the storage and the computation of these matrices should be strictly avoided.

The adjoint state method provides a useful and efficient alternative gradient computation method. This method has been introduced by Lions for the optimal control of systems governed by PDEs [17], and Chavent for identification of distributed parameters [7]. An overview of its application to geophysics is given in [23]. This technique has been also introduced by Le Dimet and Talagrand in data assimilation for oceanography and climatology [16]. An application of this method to problem (14) is presented in this section.

Let $J^s(I,h^s)$ be the function measuring the seismic misfit over the sth dataset,

$$J^s(I,h^s)=\|\mathcal{M}(I,h^s)-\widehat{D}^s\|^2,\quad J(I,\mathcal{H})=\sum_{s=1}^{\widehat{S}}J^s(I,h^s) \tag{21}$$

The adjoint state method is applied to the computation of the gradient $\nabla J^s(I,h^s)$. Then $\nabla J(I,\mathcal{H})$ is easily obtained by summing the $\widehat{S}$ resulting gradients. In what follows, the subscript s is dropped in order to alleviate notations. Eqs. (6), (7) are put into the form $\mathcal{F}(U,I,h)=0$. Define $\mathcal{G}(U)$ as

$$\mathcal{G}(U)=\frac{1}{2}\|\mathcal{M}_2(U)-D\|^2,\quad \text{for } U\in L^2(\Omega\times[0,T])^4 \tag{22}$$

and introduce the Lagrangian variables $v(\mathbf{x},t)$

$$v(\mathbf{x},t)=\{\lambda(\mathbf{x},t),\quad \mu(\mathbf{x},t)\}\quad \text{with}\ \begin{cases}\lambda(\mathbf{x},t)=\{\lambda_i(\mathbf{x},t),\ i=1,4\},\\ \mu(\mathbf{x},t)=\{\mu_1(x,t),\mu_2(x,t),\mu_3(z,t),\mu_4(z,t),\mu_5(\mathbf{x})\}.\end{cases} \tag{23}$$

The Lagrangian operator associated to problem (14) can be defined as:

$$\mathcal{L}(U,I,h,v)=(v,\mathcal{F}(U,I,h))+\mathcal{G}(U). \tag{24}$$

where $(\cdot,\cdot)$ denotes the L^2 scalar product.

Derivation of the gradient formulas. Let $\overline{U}(I,h)$ be the solution of Eqs. (6), (7), thus

$$\mathcal{L}(\overline{U}(I,h),I,h,v)=J(I,h) \tag{25}$$

and

$$\begin{cases}\partial_I\mathcal{L}(\overline{U}(I,h),I,h,v)=\partial_I J(I,h)\\ \partial_h\mathcal{L}(\overline{U}(I,h),I,h,v)=\partial_h J(I,h)\end{cases} \tag{26}$$

Denote $\partial_i, i=1,4$ the partial derivatives of the Lagrangian $\mathcal{L}(U,I,h,v)$, then

$$\begin{cases}\partial_I\mathcal{L}(\overline{U}(I,h),I,h,v)=\partial_1\mathcal{L}(\overline{U}(I,h),I,h,v)\partial_I\overline{U}(I,h)+\partial_2\mathcal{L}(\overline{U}(I,h),I,h,v)\\ \partial_h\mathcal{L}(\overline{U}(I,h),I,h,v)=\partial_1\mathcal{L}(\overline{U}(I,h),I,h,v)\partial_h\overline{U}(I,h)+\partial_3\mathcal{L}(\overline{U}(I,h),I,h,v)\end{cases} \tag{27}$$

Thus, for $v(\mathbf{x},t)=\bar{v}(\mathbf{x},t)$ such that

$$\forall(U(\mathbf{x},t),I(\mathbf{x}),h(x,t)),\quad \partial_1\mathcal{L}(U,I,h,\bar{v})=0 \tag{28}$$

the following gradient formulas come

$$\begin{cases}\partial_I J(I,h)=\partial_2\mathcal{L}(\overline{U}(I,h),I,h,\bar{v})\\ \partial_h J(I,h)=\partial_1\mathcal{L}(\overline{U}(I,h),I,h,\bar{v})\end{cases} \tag{29}$$

Note that expression (29) does not involve the terms $\partial_I\overline{U}(I,h)$ an $\partial_h\overline{U}(I,h)$. Recalling $\overline{U}(I,h)=\mathcal{M}_1(I,h)$, it is obvious these terms are precisely the Jacobian matrixes of the forward operator whose computation must be avoided. The adjoint state $\bar{v}(\mathbf{x},t)=\{\bar{\lambda}(\mathbf{x},t),\bar{\mu}(\mathbf{x},t)\}$, solution of Eq. (28), has to be computed instead. Expliciting this equation, it turns out that $\bar{\lambda}(\mathbf{x},t)$ is actually the solution of acoustic wave equations to be solved backward in time. Compatibility equations between $\bar{\lambda}(\mathbf{x},t)$ and $\bar{\mu}_i,\ i=1,5$ complete the adjoint state computation. The equations verified by $\bar{\lambda}(\mathbf{x},t)$ are

$$\begin{cases} \frac{I(\mathbf{x})}{c(\mathbf{x})}\partial_t\lambda_1(\mathbf{x},t) - \partial_x\lambda_3(\mathbf{x},t) + \sigma_x(x)\frac{I(\mathbf{x})}{c(\mathbf{x})}\lambda_1(\mathbf{x},t) = f_x(\mathbf{x},t) \\ \frac{I(\mathbf{x})}{c(\mathbf{x})}\partial_t\lambda_2(\mathbf{x},t) - \partial_z\lambda_4(\mathbf{x},t) + \sigma_z(z)\frac{I(\mathbf{x})}{c(\mathbf{x})}\lambda_2(\mathbf{x},t) = f_z(\mathbf{x},t) \\ \frac{1}{I(\mathbf{x})c(\mathbf{x})}\partial_t\lambda_3(\mathbf{x},t) - \partial_x(\lambda_1+\lambda_2)(\mathbf{x},t) + \frac{\sigma_x(x)}{I(\mathbf{x})c(\mathbf{x})}\lambda_3(\mathbf{x},t) = 0 \\ \frac{1}{I(\mathbf{x})c(\mathbf{x})}\partial_t\lambda_4(\mathbf{x},t) - \partial_z(\lambda_1+\lambda_2)(\mathbf{x},t) + \frac{\sigma_z(z)}{I(\mathbf{x})c(\mathbf{x})}\lambda_4(\mathbf{x},t) = 0 \end{cases} \tag{30}$$

where

$$\begin{cases} f_x(\mathbf{x},t) = 2\sum_{q=1}^{Q}\left(u_x(0,z_q,t) - D_x^q\right)\delta_{M_q} \\ f_z(\mathbf{x},t) = 2\sum_{q=1}^{Q}\left(u_z(0,z_q,t) - D_z^q\right)\delta_{M_q} \end{cases} \tag{31}$$

with

$$\begin{cases} \lambda_3(-X,z,t) = 0 \quad \lambda_3(X,z,t) = 0 \\ \lambda_4(x,z_1,t) = 0 \quad \lambda_4(x,Z,t) = 0 \end{cases} \tag{32}$$

and null final condition in $t = T$.

In Eq. (31), δ_{M_q} is the delta Dirac function centered at the coordinates $(0,z_q)$ of receiver q. Functions $f_x(\mathbf{x},t)$ and $f_z(\mathbf{x},t)$ are, respectively, the horizontal and vertical seismic residuals. More details on how to obtain Eqs. (30)–(32) are available in [20].

According to the geophysicist vocabulary, the computation of $\bar{\lambda}(\mathbf{x},t)$ amounts to backpropagate the seismic residuals. The nonlinear inversion method presented here can thus be interpreted as an iterative process of time reversal acoustic inversions [1,12]. Indeed, time reversal imaging amounts to backpropagate perturbations of the field to detect anomalies of the investigated medium. Here, the gradient which is computed can be seen as the result of a time reversal inversion, used to improve iteratively the impedance estimation.

3.3.2. Gradient computation

Semi-discrete gradient formulas. Once the adjoint state is defined, the computation of the total gradient $\nabla J(I,\mathcal{H})$ is performed through Eq. (29). Reintroducing the subscript s, define $\bar{v}^s(\mathbf{x},t)$ as the adjoint state associated to the misfit function $J^s(I,h^s)$ and $U^s(\mathbf{x},t)$ as the solution of the seismic wave equations for impedance $I(\mathbf{x})$ and pressure condition $h_s(x,t) : U^s(\mathbf{x},t) = \mathcal{M}_1(I,h^s)$. The gradient equations follow:

$$\partial_I J(I,\mathcal{H}) = \sum_{s=1}^{\widehat{S}}\int_0^T \frac{1}{I}\left(\lambda_1^s\partial_x(p_x^s+p_z^s) + \lambda_2^s\partial_z(p_x^s+p_z^s)\right)dt - \sum_{s=1}^{\widehat{S}}\int_0^T \frac{1}{I}\left(\lambda_3^s\partial_x u_x^s + \lambda_4^s\partial_z u_z^s\right)dt \tag{33}$$

and

$$\partial_{h_s} J(I,\mathcal{H}) = -\frac{I}{c}\lambda_2^s(x,z_1,t), \quad 1 \leqslant s \leqslant \widehat{S} \tag{34}$$

The computation of the gradient $\nabla J(I,\mathcal{H})$ through Eqs. (33) and (34) amounts to solve $2\widehat{S}$ wave propagation problems: $\widehat{S}$ forward in time for what we call the "forward state" $U^s(\mathbf{x},t)$, $s = 1,\widehat{S}$, plus $\widehat{S}$ backward in time for the adjoint state $\lambda^s(\mathbf{x},t)$, $s = 1,\widehat{S}$. However, these two problems are coupled through the source terms functions $f_x^s(\mathbf{x},t)$ and $f_z^s(\mathbf{x},t)$ of Eq. (30). Thereupon, an appropriate algorithm has to be designed to minimize the gradient computation complexity.

For a better understanding, semi-discretizated in time notations are adopted for Eqs. (33) and (34):

$$\partial_I J(I,\mathcal{H}) = \sum_{s=1}^{\widehat{S}}\sum_{n=0}^{N}\frac{1}{I}\left([\lambda_1^s]^n\partial_x(p_x^s+p_z^s)^n + [\lambda_2^s]^n\partial_z(p_x^s+p_z^s)^n\right)dt - \sum_{s=1}^{\widehat{S}}\sum_{n=0}^{N}\frac{1}{I}\left([\lambda_3^s]^n\partial_x[u_x^s]^n + [\lambda_4^s]^n\partial_z[u_z^s]^n\right)dt \tag{35}$$

and

$$\partial_{h_s} J(I,\mathcal{H}) = -\frac{I}{c}[\lambda_2^s]^n(x,z_1), \quad 1 \leqslant s \leqslant \widehat{S} \tag{36}$$

where N is the total number of time steps, and the superscript n refers to the time step iteration in the Yee scheme used to solve numerically the direct and adjoint wave equations.

An efficient algorithm. Advancing $[U^s]^n$ of one time step using the Yee scheme only requires $[U^s]^{n-1}$ to be given. Conversely, backpropagating $[\lambda^s]^n$ from $[\lambda^s]^{n+1}$ requires $[\lambda^s]^{n+1}$ and $[U^s]^{n+1}$, in order to compute source terms function of Eq. (30). As a consequence, an usual strategy for gradient computation through Eqs. (35) and (36) is as follows: for each shot s, propagate $(U^s)^N$ from the initial condition $(U^s)^0$, and store only $(U^s)^N$. Then, backpropagate conjointly $(U^s)^N$ and $(\lambda^s)^N$ until the initial time $n = 0$, and update $\partial_I J(I,\mathcal{H})$ and $\partial_{h_s} J(I,\mathcal{H})$ at each time step. This strategy provides a suitable trade-off between the required computer memory, and the computational complexity of the algorithm. Indeed, the total number of computation steps is equal

to $3\widehat{S}N$ ($\widehat{S}N$ for the first propagation of $(U^s)^N$ from $(U^s)^0$, plus $2\widehat{S}N$ for the joint backpropagation of the forward and the adjoint state from final time toward the initial time), and only two vectors ($(U^s)^n$ and $(\lambda^s)^n$) are stored.

However, this strategy assumes that the equations describing the wave propagation for the forward problem can be numerically backpropagated. Unfortunately, this is not the case for Eqs. (6), (7): the introduction of absorbing terms to design the PML makes the backpropagation numerically unstable. Another strategy has to be investigated.

Three kind of general strategies are usually distinguished: the store-all strategy, the recompute-all strategy, and trade-off between these two strategies. The first one begins with computation and storage of $(U^s)^n$ at each time step $n = 1, N$. Then the adjoint state is backpropagated. This takes only $2\widehat{S}N$ computation steps, but depends on the ability to store the $N + 1$ vectors $(U^s)^n$, $n = 0, N$.

On the other hand, recompute-all strategy implies to compute $(U^s)^n$ at each time step from initial step $(U^s)^0$: thus the number of computation steps reaches $\widehat{S}(N \times (N+1)/2)$ while it requires only the storage of the two vectors $(U^s)^n$ and $(\lambda^s)^n$.

Finally, an interesting trade-off between these two strategies consists in computing $(U^s)^n$ from an intermediate starting point stored at the previous computation step. This is the so-called "checkpointing" strategy [26]. An implementation of this strategy consists in the following: initially, compute $(U^s)^N$ from $(U^s)^0$ and store $(U^s)^{N/2}$. Then compute $(\lambda^s)^{N-1}$ from $(\lambda^s)^N$ and $(U^s)^N$. At next iteration, compute $(U^s)^{N-1}$ from $(U^s)^{N/2}$ and store $(U^s)^{3N/4}$, and iterate. This produces an efficient algorithm requiring less than $N\log_2 N$ computation steps and $\log_2 N$ vectors to be stored.

The choice of the most appropriate strategy depends on the available computer memory and the computational resources. Generally, recompute-all strategy should be avoided, because of its complexity in $O(N^2)$, unless computation cost of one time step is very low. This is not the case for the problem of interest. Consequently, the choice lies between store-all and checkpointing strategies, and depends mainly on the memory resources available. As is explained in the next paragraph, the method is designed to be used on a cluster of several computation cores. This increases the memory available and makes the store-all strategy practicable for a suitable number of computation cores. Conversely, if the method is used on one or few computations cores, the checkpointing strategy will be preferred.

Domain decomposition and parallelization. Even with an efficient algorithm, the computation time of the gradient must be reduced as much as possible to reach convergence within an acceptable time. This is achieved by introducing parallel computation for the forward state propagation and the adjoint state backpropagation. This can be performed through the use of a domain decomposition method. The principle is as follows: the initial space domain Ω is subdivised into P regular rectangles subdomains Ω_p, $p = 1, P$. Each subdomain is associated to a computation core. In each one, a subproblem of propagation and backpropagation is defined. Thanks to the use of a second order Yee scheme, solving these suproblems only require the definition of pressure $p_x(\mathbf{x},t) + p_z(\mathbf{x},t)$ boundary conditions at the artificial boundaries appearing between subdomains. Defining overlaps one mesh wide between subdomains, subdomains can exchange these pressure boundary condition from one to another at each iteration in time. This produces an efficient parallel algorithm that improves gradient time computation.

Summary of the numerical method. Finally, the overall numerical strategy can be summarized as follows

- apply the interlocked optimization algorithm defined in Section 3.2.2 to solve problem (14);
- at each iteration, compute the gradient $\nabla J^e(I,\mathcal{H})$ through Eqs. (33), (34) derived from the adjoint state method;
- use the store-all algorithm and a domain decomposition method to solve these equations on a cluster of computation cores;

Numerical experiments presented in the next section illustrate the efficiency of this numerical method.

4. Numerical results on synthetic data

4.1. Presentation of the test case

The exact impedance model is defined over a 720 m wide and 3400 m depth rectangle, displayed in Fig. 4. An array of 100 receivers is disposed each 8 m from a depth $z_1 = 1000$ m, at the center of the model ($x = 0$). Synthetic data is acquired using a Ricker type source (see Fig. 2) located at the ground surface. A walkaway synthetic data is built, the source being located at 30 different offsets from $x = -360$ m to $x = 360$ m each 24 m. The geometry of the acquisition system is presented in Fig. 3.

The Radon transform applied to this data (following the regularization technic introduced in Section 2) produces 13 datasets corresponding to an illumination of the medium by 13 quasi-plane waves with incidence angles going from $-6°$ to $6°$ with a $1°$ sampling.

From this data, the method aims at recovering the impedance model from the depth of the shallowest receiver $z = 1000$ m. Pressure conditions corresponding to the $\widehat{S}$ modified datasets are also recovered, to be considered as intermediate unknowns. The initial guess for the impedance is an uniform model whose value is taken equal to the exact impedance at the shallowest receiver location. The initial pressure conditions are taken equal to zero. The exact wave velocity field displayed in Fig. 4 is assumed to be known.

A fine spatial discretization grid is used to reach high resolution: vertical and horizontal steps are $\Delta z = 8$ m and $\Delta x = 12$ m, respectively. The total recording time is $T = 2$ s. According to the CFL condition associated to the second order Yee scheme

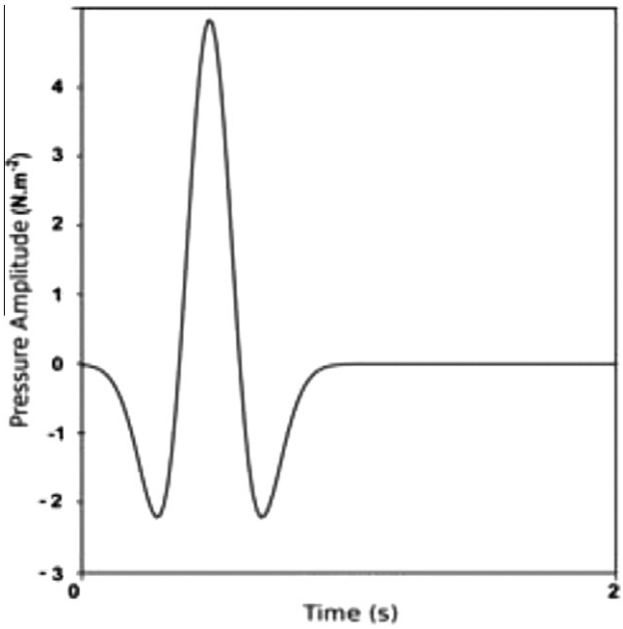

Fig. 2. Seismic source profile.

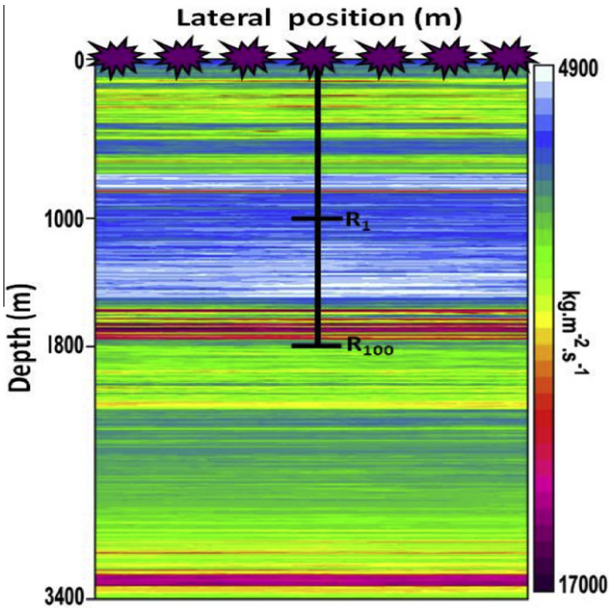

Fig. 3. Geometry of the acquisition system.

used for Eqs. (6) and (7), the time step Δt is taken equal to 9×10^{-4} s. This leads to a problem involving approximately 2×10^6 discrete data and 4×10^5 discrete unknowns.

4.2. Numerical performances

4.2.1. Interlocked algorithm

The convergence of the interlocked optimization algorithm is compared to the convergence of the l-BFGS quasi-Newton method applied to all the unknowns. The convergence curves displayed Fig. 5 illustrate what have already been mentioned: the quasi-Newton algorithm fails to reduce the function $J^e(I,H)$ beyond one order of magnitude (black curve), while the

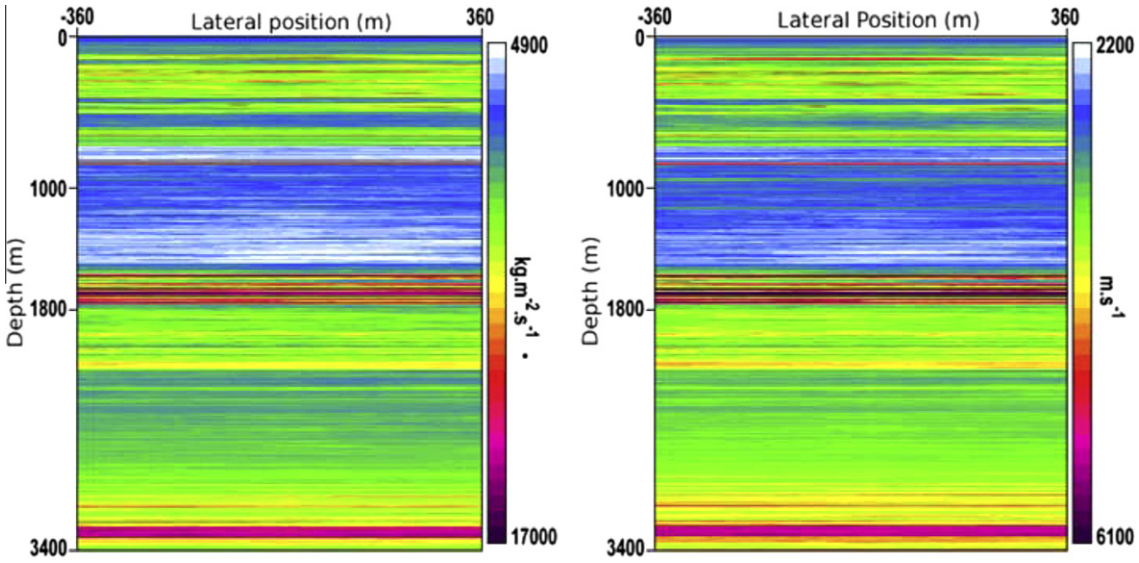

Fig. 4. Exact impedance model (left) and wave velocity model (right).

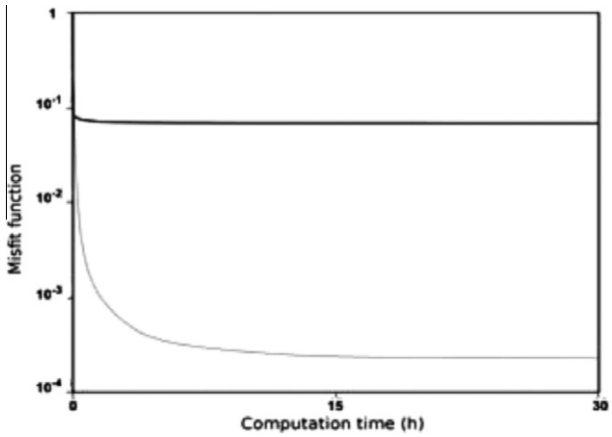

Fig. 5. Convergence profile comparison between the l-BFGS algorithm (black) and the interlocked optimization algorithm (gray).

interlocked optimization algorithm decreases the function of nearly four order of magnitudes (gray curve). This is an illustration of the preconditioning action of the interlocked algorithm over the ill-conditioned minimization problem (14).

4.2.2. Scaling

A scaling test is performed to evaluate the efficiency of the parallelization which has been introduced through the domain decomposition method. This test is made at constant domain size with an increasing number of subdomains, from 1 to 128. Since the numerical scheme used to solve the wave propagation problems involved in the gradient computation is explicit, the complexity of the gradient computation is linear with respect to the number of unknowns. Thus, the computation time is expected to decrease linearly with the number of subdomains.

Theoretical computation time (black) is compared to the measured computation time (gray) for the first ten iterations of the interlocked optimization algorithm (Fig. 7). For 1 to 4 computation cores, theoretical and observed computation times show a good agreement. For 4 to 16 computation cores, the measured time is lower than expected one, and we denote a superlinear convergence. Conversely, for the use of 16 to 128 computation cores, the convergence rate is inferior to the theoretical one.

These observations are typical of cache memory effects on scaling properties of parallel algorithms. The decrease of the number of unknowns by computation cores reduce the number of cache misses. Thus, access to the data is considerably sped up. This explains the superlinear convergence observed between 4 to 16 computation cores. For the use of 16 to 128 computation cores, the decrease of the number of unknowns still goes on, but this has no longer an effect on cache misses, since

Number of subdomains	2	4	8	16	32	64	128
Rate	150	25	18.75	12.5	6.81	5	3.26

Fig. 6. Rate between the number of computed unknowns and the number of exchanged unknowns for 2–128 subdomains.

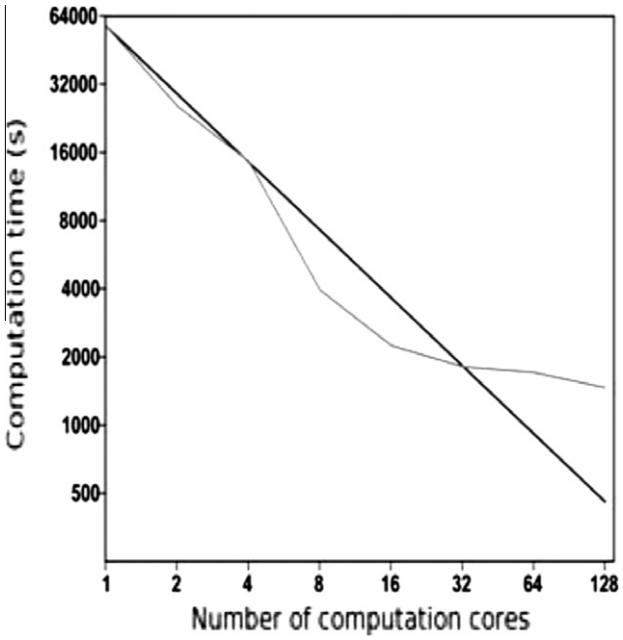

Fig. 7. Theoretical scaling (black) and observed scaling (gray) of the overall numerical method on the 10 first iteration of the interlocked optimization algorithm.

the data to be stored is already small enough to be stored in the cache memory. Conversely, when the number of subdomains is larger to 16, their sizes become sufficiently small for the rate between computed and exchanged unknowns reaches critical values (Fig. 6). In particular, this rate becomes inferior to 10 for more than 16 computation cores. Even on modern and efficient cluster of computers, the communication time between two cores is at least 5 times longer than the access to data in the cache memory. This explains the decrease of the speed-up for more than 16 computation cores.

The global scaling of the method is satisfactory: it should be underlined that the computation time for the 10 first iterations of the interlocked optimization algorithm goes from 17 h to 30 mn for 1 to 128 computation cores. In other words, the computation time is reduced of a factor 35. Therefore, the method that has been set up is efficient on a numerical point of view.

4.3. High resolution estimation of the acoustic impedance

The estimated impedance model obtained after 1000 iterations of our algorithm is displayed in Fig. 6 together with its difference with the exact model. At first sight:

- in the receiver zone (1000 m $< z <$ 1800 m), the estimation is nearly exact, until approximately 200 m away from the well;
- below the receivers ($z >$ 1800 m), the estimation seems to be far from the exact model, even if reproducing some of its main variations.

The vertical impedance traces displayed in Fig. 9 confirm the first point: the estimation is almost perfect in the receiver zone. It provides also additional information about the estimation below the receivers. Actually, the traces show that the vertical variations of the acoustic impedance are accurately recovered (the vertical resolution reaches the vertical stepsize $\Delta z = 8$ m). Only the low frequency part of the exact impedance is not recovered in the zone below the receivers. This shall not be surprising: the lowest frequencies of the impedance variations are not contained in the data. This is classical in seismic imaging. Therefore, it is impossible to recover the low frequency component of the impedance, unless it is given as an *a priori* information. This is not the case here, and that explains why this low frequency component is not recovered (Fig. 8).

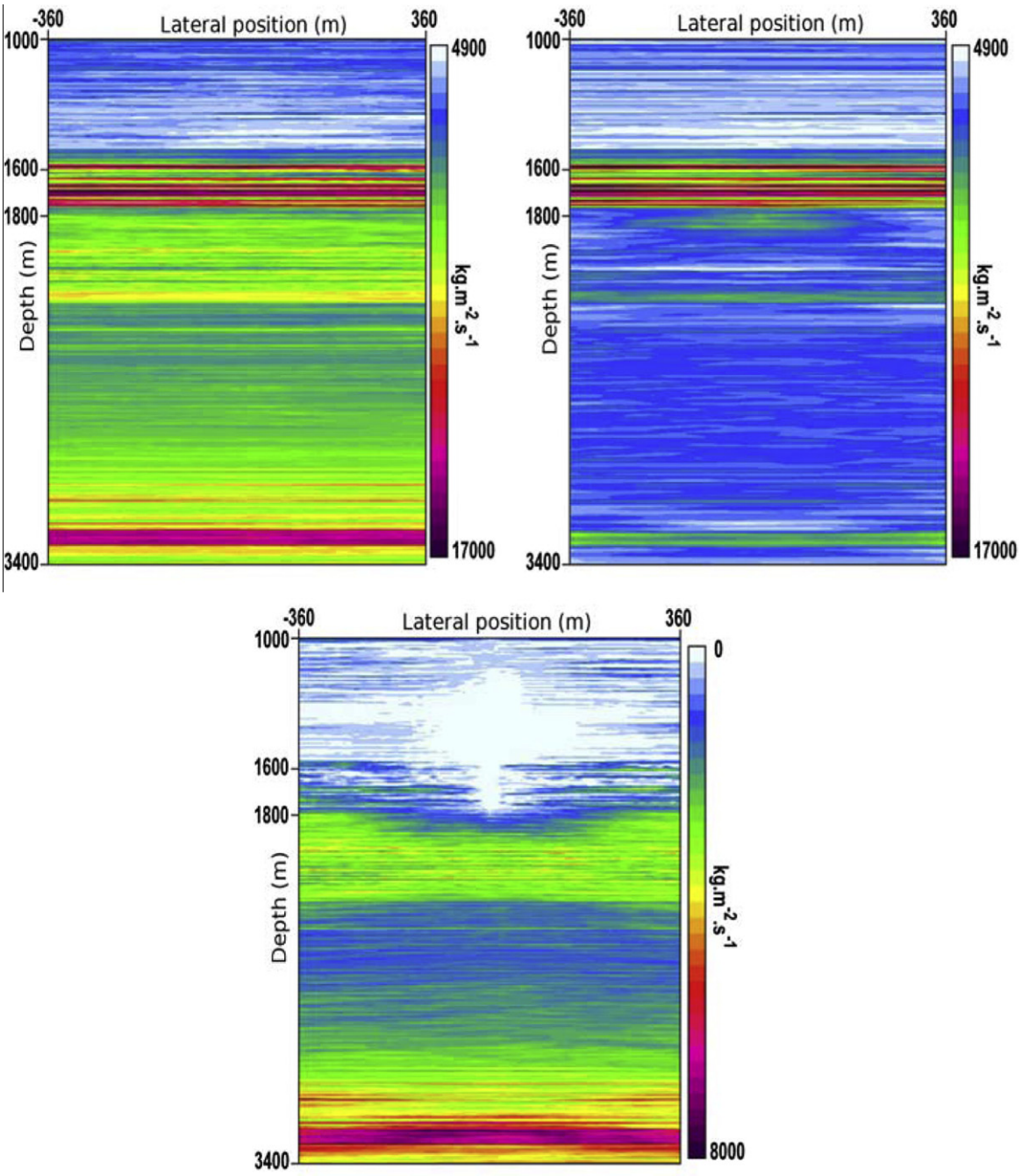

Fig. 8. Exact impedance model (top left), estimated impedance model (top right), difference between exact and estimated impedance model (center).

The impedance estimation is thus very accurate in the receivers zone, and the vertical variations are finely recovered until almost 3.5 km depth. This depth of investigation depends mainly on the recording time T. Taking $T = 4$ s for instance should provide a deeper length of investigation, even if a degradation of the resolution can be expected. Other numerical experiments show that the method is sensitive to small perturbations of the acoustic impedance within the receiver zone [20,21]. Therefore, application of this method to gas leak detection in CO_2 storage monitoring can be considered: the storage would be located below the receivers and the detection zone would be located at the receivers level. The detection of changes in the impedance estimation should indicate the presence of gas.

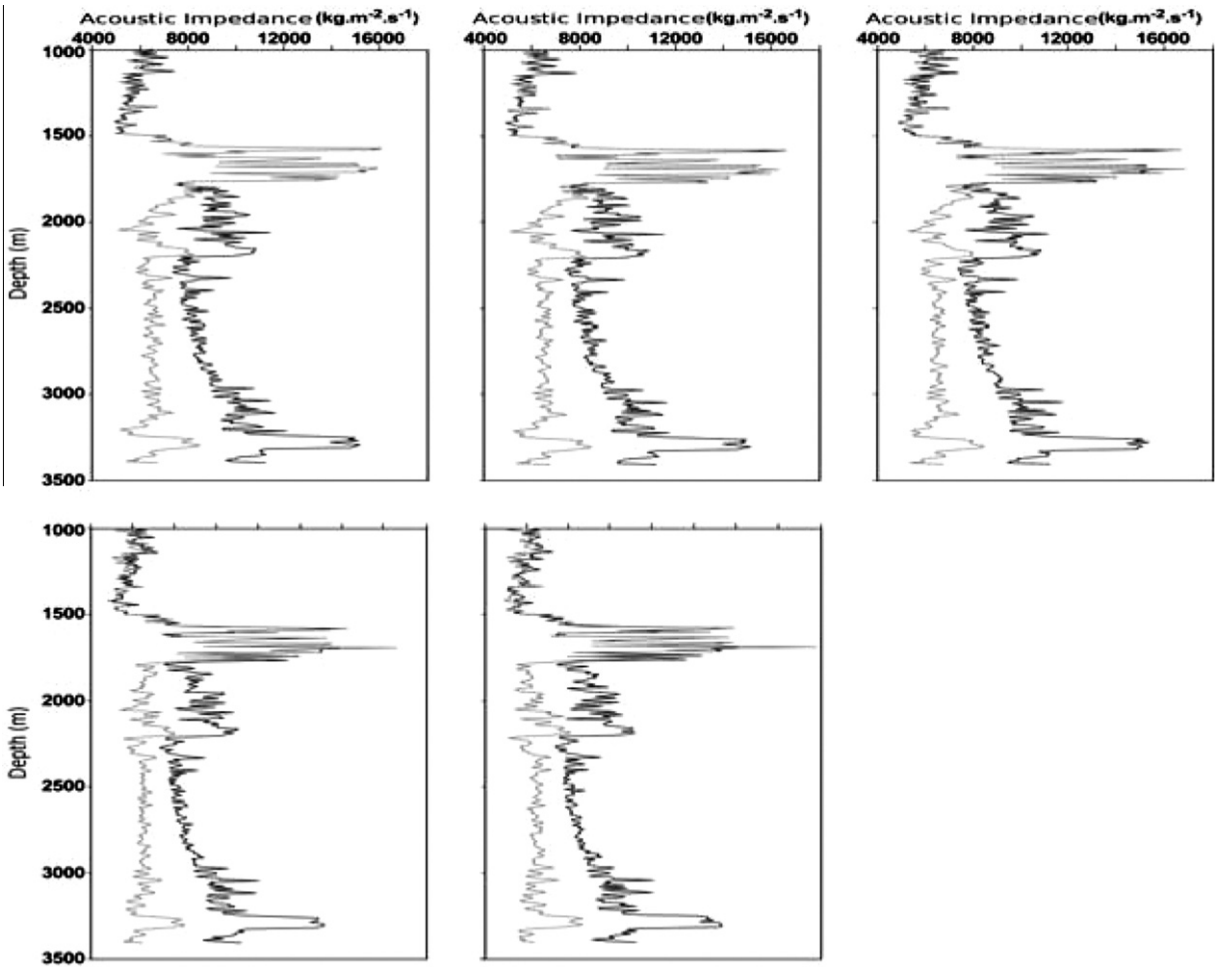

Fig. 9. Exact (black) and estimated (gray) vertical impedance profiles at different offsets from the borehole: 0 m, 40 m, 100 m, 200 m, 300 m.

5. Conclusion and perspectives

The quest for high resolution methods in seismic imaging leads to study the multidimensional extension of the 1D well-seismic data inversion. As a first step of this extension, a 2D inverse problem is defined as the minimization of the distance between walkaway data, and solutions of a 2D acoustic wave propagation model. Assuming an estimation of the wave velocity is available, the methods aims at recovering the acoustic impedance model from the data. In this framework, the previously well-posed 1D inverse problem becomes largely undetermined, and involves tremendous number of discrete unknowns. To face with the undetermination, the approach is restrained to media that vary slowly laterally. In addition, the seismic data is recombined through a Radon transform so as to obtain datasets corresponding to an illumination of the subsurface by quasi-plane waves. Thus, *a priori* information about the regularity of the unknowns can be introduced. This reduces the inherent indetermination of the inverse problem. However, special care must be taken to design an efficient numerical method able to solve the regularized problem.

The work presented here focuses on this aspect. First, the numerical minimization of the regularized function is an ill-conditioned problem. An interlocked optimization algorithm involving both quasi-Newton and conjugate gradient algorithms is designed. This method acts as a preconditioner of the problem, and the misfit function is efficiently decreased. Nevertheless, it requires to compute the gradient of the misfit function at each iteration. In order to reach convergence within an acceptable time, a fast gradient algorithm is designed, based upon adjoint state method and parallelization by domain decomposition. Using the adjoint state method, the computation of the gradient amounts to solve two coupled wave propagation problem, forward and backward in time. These two problems are parallelized by a domain decomposition technique. In addition, the use of a cluster based computer offers memory capacities that enables to use an efficient algorithm for solving the coupled propagation problems.

The results obtained on a realistic test case demonstrate the efficiency of the interlocked optimization algorithm and the good scaling of the overall numerical method. Whereas a traditional l-BFGS method fails to converge, the interlocked optimization algorithm decreases the objective function of approximately four orders of magnitude. The scaling test, performed at constant domain size, gives classical scaling results, enlightening the importance of memory accesses and cache misses in parallel computing. The use of 128 computation cores reduces the overall computation time by a factor 35.

In addition, the method shows its ability to determine a high resolution quantitative estimation of the impedance in a zone of 400 m wide around the receivers. Below the receivers, the impedance variations are recovered with a high vertical resolution until 1600 m below the receivers (to the order of the vertical step 8 m) and at some hundred meters from the well laterally. Consequently, the method seems to be very promising for high resolution seismic imaging.

These first results open the way to multiple developments. In particular, these 2D results shall be extended to a 3D acoustic framework. From a geophysical point of view, the 3D extension of the methodology set up for a 2D acoustic framework is straightforward. The same *a priori* information over the regularity of the unknowns should be introduced by restriction to 3D media varying slowly laterally and recombination of the walkaway data. Nevertheless, the numerical method should be adapted to this larger scale problem. The interlocked optimization loop could be conserved, but the fast gradient computation algorithm should be probably improved. For instance, the extension to 3D should require to switch from the store-all strategy to the checkpointing strategy for solving the coupled wave propagation problems. In addition, the parallelization scheme could be improved. Indeed, parallelization could be introduced at a physical level. As the function $J^c(I,\mathcal{H})$ can be rewritten as the sum of $\widehat{S}$ functions $J^s(I,h^s)$, each one corresponding to one dataset, it is possible to define $\widehat{S}$ groups of computation cores, each of these groups being in charge of the computation of $\nabla J^s(I,h^s)$, using the fast gradient computation algorithm designed in Section 2. This could result in a reduction of the overall computation time by a factor $\widehat{S}$. Moreover, the gradient computation time would become independent of the datasets number. As many datasets as available should be added without deteriorating the computation time. In turn, this would of course increase the number of computation cores required.

Other developments should also be considered. A careful study about the best choices of angles in the recombination of data by Radon transform could be carried on. Another important work should concern the study of the sensitivity of the method regarding the wave velocity estimation $c(\mathbf{x})$. An adaptation of the methodology presented here to medias varying rapidly in horizontal and transversal directions could also be considered. Other regularization techniques should be used to that purpose.

Using more realistic wave propagation model should also be considered. As a first step, accounting for attenuation in the wave propagation model would yield an interesting improvement. The difficulty raised by the attenuation is that the forward problem becomes unstable when solved backward in time. In the framework of the method developed here, this however does not raise additional problems: indeed, due to the attenuation term introduced to define the PML around the interest domain, the algorithm that has been set up to compute the gradient of the method avoid solving the direct problem backward in time. Consequently, attenuation terms could be taken into account easily. This is an important advantage regarding time reversal methods, for which important adaptation have to be made to come to the same result [2]. Moreover, the technic that has been designed here accounts for spatially variable attenuation coefficients, whereas only constant attenuation coefficients are considered in [2].

Finally, the extension of this methodology to more complex models such as elastic wave propagation still remains an open problem. Indeed, these models involve a larger number of subsurface parameters, hence, the first difficulty would be to define the parameters which could be reasonably estimated from the data. Then, the resulting inverse problem could be severely undetermined, and, again, appropriate regularization methods would have to be defined.

Acknowledgements

We would like to thank Laurence Halpern, Florence Delprat-Jannaud, Patrick Lailly and Jeffrey Rauch for their constant support and fruitful discussions.

References

[1] H. Ammari, An Introduction to Mathematics of Emerging Biomedical Imaging, Mathématiques et Applications, vol. 62, Springer-Verlag, Berlin, 2008.
[2] H. Ammari, E. Bretin, J. Garnier, A. Wahab, Time reversal in attenuating acoustic media, Contemporary Mathematics 548 (2011) 151–163.
[3] A. Bamberger, Analyse, contrôle et identification de certains systèmes, Thèse de doctorat d'état es-sciences mathématiques soutenue à l'université Pierre et Marie-Curie Paris VI, 1978.
[4] A. Bamberger, G.Chavent, P.Lailly, Etude mathématique et numérique d'un problème inverse pour l'équation des ondes à une dimension, rapport interne 14, école Polytechnique, CMA, 1977.
[5] J.P. Bérenger, A perfectly matched layer for the absorption of electromagnetic waves, Journal of Computational Physics 114 (1994) 185–200.
[6] N. Bleistein, J.K. Cohen, J.W. Stockwell Jr., Mathematics of Multidimensional Seismic Imaging, Migration and Inversion, Springer Verlag, New York, 2000. ISBN 0-387-95061-3.
[7] G. Chavent, Analyse fonctionnelle et identification de coefficients répartis dans les équations aux dérivées partielles, PhD Thesis, Université de Paris, France, 1971.
[8] J.F. Claerbout, Fundamental of Geophysical Data Processing, Mc Graw-Hill Book Co., 1976.
[9] F. Delprat-Jannaud, P. Lailly, The insidious effects of fine-scale heterogeneity in reflection seismology, Journal of Seismic Exploration 13 (2004) 39–84.
[10] F. Delprat-Jannaud, P. Lailly, A fundamental limitation for the reconstruction of impedance profiles from seismic data, Geophysics 70 (1) (2005) R1–R14.

[11] F. Delprat-Jannaud, P. Lailly, Waves propagation in heterogeneous media, Geophysics 73 (3) (2008) T37–T49.
[12] M. Fink, Time-reversed acoustics, Physics Today 50 (1997) 34–40.
[13] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, Claudia A. Safastizábal, Numerical Optimization Theoretical and Practical Aspects, second ed., Series Universitext, Springer, 2006.
[14] J.Ch. Gilbert, Module N1CG1, 1999. <http://www-rocq.inria.fr/estime/modulopt/modulopt.html>.
[15] F.Q. Hu, On absorbing boundary conditions for linearized Euler equations by a perfectly matched layer, Journal of Computational Physics 129 (1996) 1996.
[16] F.X. Le Dimet, O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus 38A (1986) 97–110.
[17] J.L. Lions, Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles, Dunod, Paris, 1968.
[18] D. Macé, P. Lailly, Solution of the VSP one-dimensional inverse problem, Geophysical Prospecting 34 (1986) 1002–1021.
[19] L. Métivier, Utilisation des équations Euler-PML en milieu hétérogène borné pour la résolution d'un problème inverse en géophysique, ESAIM: Proceedings 27 (2009) 156–170.
[20] L. Métivier, Une méthode d'inversion non linéaire pour l'imagerie sismique haute résolution, PhD Thesis, Université de Paris XIII, France, 2010. <http://tel.archives-ouvertes.fr/tel-00471293_v1/>.
[21] L. Métivier, P. Lailly, F. Delprat-Jannaud, L. Halpern, A 2D Nonlinear inversion of well seismic data, Inverse Problems 27 (2011) 055005.
[22] R.H. Byrd, P. Lu, J. Nocedal, A limited memory algorithm for bound constrained optimization, SIAM Journal on Scientific and Statistical Computing 16 (5) (1995) 1190–1208.
[23] R.E. Plessix, A review of the adjoint state method for computing of a functional with geophysical applications, Geophysical Journal International 167 (2006) 495–503.
[24] J. Radon, Ueber die Bestimmung von Funktionen durch uhre Integralwerte längs gewisser Mannigfaltigkeiten, Berichte Sächsische Akademie der Wissenschaften 69 (1917) 262–277.
[25] D.F. Shanno, Conditioning of quasi-newton methods for function minimization, Mathematics of Computation 24 (1970) 647–656.
[26] M.H.Tber, L.Hascoët, A.Vidard, B.Dauvergne, Building the Tangent and Adjoint codes of the Ocean General Circulation Model OPA with the automatic Differentiation tool TAPENADE, Rapport de recherche INRIA n °6372, 2007.
[27] A. Tarantola, Inverse Problem Theory and Model Parameter Estimation, SIAM book, 2005. ISBN 0-89871-572-5.
[28] J. Virieux, S. Operto, An overview of full waveform inversion in exploration geophysics, Geophysics 74 (6) (2009) WCC127–WCC152.
[29] K.S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, IEEE Transactions on Antennas and Propagation (1966) 302–307.