MOX–Report No. 60/2014

CMFWI: Coupled Multiscenario Full Waveform Inversion for seismic inversion

Signorini, M.; Micheletti, S.; Perotto, S.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

CMFWI: Coupled Multiscenario Full Waveform Inversion for seismic inversion *

Marianna Signorini, Stefano Micheletti and Simona Perotto

December 12, 2014

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica
Politecnico di Milano
Piazza L. da Vinci 32, I-20133 Milano, Italy
{marianna.signorini,stefano.micheletti,simona.perotto}@polimi.it

Keywords: seismic inversion, Full Waveform Inversion, inverse problem, TV regularization

Abstract

We present the new method Coupled Multiscenario Full Waveform Inversion (CMFWI) for the solution of the seismic inversion problem. As in the case of Full Waveform Inversion (FWI), the proposed method is based on seismic reflection signals and it tries to recover the subsoil velocity profile by minimizing a suitable misfit functional between recorded and computed data. CMFWI suitably combines data generated by shooting one source at a time, but sharing the effect of this signal with the other sources. Moreover, CMFWI differs from FWI employed with the same sources shot together, and we numerically show that it performs better than FWI. In particular, this comparison focuses on different types of boundary conditions, wave sources, initial guesses of the velocity profile, and signal-to-noise ratio.

*This work has been supported by the Projects MIUR-PRIN 2010/2011 "Innovative methods for water resources under hydro-climatic uncertainty scenarios" and "Data-Centric Genomic Computing" (GenData 2020)

1 Introduction and motivation

The discovery of oil and gas fields is the main goal of seismic exploration. Clearly, an accurate identification of the position and extension of the reservoirs becomes mandatory because oil exploration is an expensive, high-risk operation.

Visible surface features such as oil or natural gas seeps and pockmarks (underwater craters caused by escaping gas) provide basic evidence of hydrocarbon generation (be it shallow or deep in the earth). However, these features do not allow to exactly detect the extent of these deposits. This justifies the employment of sophisticated exploration geophysics technologies.
Therefore, several numerical techniques have been developed in order to facilitate locating hydrocarbon reservoirs. A first category of methods is provided by refraction and reflection tomography techniques [1], which are based on travel time kinematics of the seismic data. As an alternative, Full Waveform Inversion (FWI) aims to estimate high-resolution velocity models by minimizing the misfit between observed and modelled seismic waveforms, and uses the additional information provided by the amplitude and phase of the seismic waveform. From a mathematical viewpoint, FWI amounts to solving an inverse problem hinging on the wave equation and on a suitable misfit functional [18, 16, 17, 12, 21]. An overview of FWI can be found in [20].

FWI is actually based on seismic reflection data, which are collected during seismic surveys (see Fig. 1). In practice, FWI takes place either through marine or land exploration. In the first case, air guns are shot to generate waves which propagate in the marine environment, while in the second setting, seismic vibrators are employed to trigger waves in land geometries. In both cases, the outcome of the exploration is a set of signals recorded at certain locations by hydrophones in the marine case, and geophones in the land configuration. The recorded data are post-processed and then compared with the ones predicted by a numerical simulation of the same experiment.
It is well known that the inverse problem associated with FWI is, in general, ill-

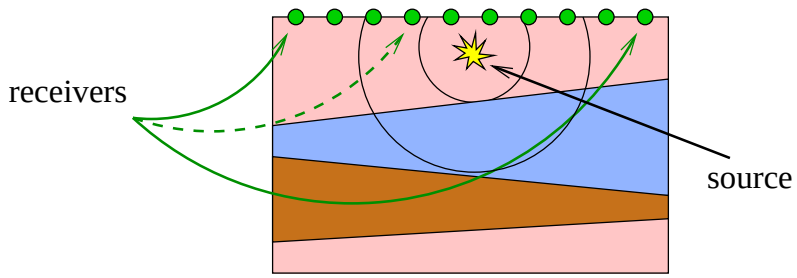


Figure 1: Scheme of seismic survey

posed, because different velocity profiles can lead to (almost) the same recorded data, and since the misfit functional may exhibit many local minima. In order to try and improve this ill-posedness, a Tikhonov regularization is usually added

to the misfit functional. Often, this workaround does not suffice to guarantee an accurate modeling due to the high sensitivity of FWI to the initial velocity guess.

The approach that we propose in this work aims at reducing this sensitivity as well as at providing a regularization less aggressive than Tikhonovs' correction. The main idea is to suitably combine data generated by shooting one source at a time, but sharing the effect of this signal with the other sources. This approach, which we denote by Coupled Multiscenario FWI (CMFWI), differs from the FWI method used with the same sources shot together and we numerically show that it outperforms FWI. As an improvement over Tikhonov regularization, we resort to the Total Variation (TV) correction, more suited to track discontinuty velocity profiles.

2 Full Waveform Inversion

The objective of seismic inversion is to estimate the unknown velocity of the wave, for a given set of measurements, recorded as a seismogram. In particular, Waveform Inversion is a data-fitting procedure, based on wavefield modeling. The data are usually recorded signals in correspondence with some receivers located at the points $\{\mathbf{x}_i\}$, coinciding with the pressure sampled at discrete time levels.

In this section, we provide the classical FWI method [18], completed with a Total Variation (TV) regularization [22, 21] to mitigate the ill conditioning of the problem.

Following [18], we consider the acoustic approximation of the wave equation for isotropic media

$$\begin{cases} \sigma \dfrac{\partial^2 p}{\partial t^2} = \nabla \cdot (\nabla p) + f & \text{in } \Omega \times (0,T) \\ + \text{ b.c.} & \text{on } \partial\Omega \times (0,T) \\ p = 0,\ \dfrac{\partial p}{\partial t} = 0 & \text{on } \Omega \times \{0\} \end{cases} \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is a polygonal computational box identifying the geological domain, T is the final observation time, p is the acoustic pressure, $\sigma(\mathbf{x}) = 1/v^2$ is the squared slowness, with v the wave speed into the medium, and f is a source term. Concerning the boundary conditions (b.c.), we consider homogeneous Dirichlet and absorbing boundary conditions, possibly assigned on different parts of the boundary [8]. Even though the Dirichlet conditions are not thoroughly realistic, they are often adopted in seismic modeling since they can be implemented in a straightforward way.

We introduce the functional

$$J(p,\sigma) = \frac{1}{2} \sum_{\mathbf{x}_i \in X_r} \int_0^T \left(p(\mathbf{x}_i,t) - p_{obs}(\mathbf{x}_i,t)\right)^2 dt, \tag{2}$$

which measures the misfit in time between observed data and computed signals with respect to the L^2-norm. Standard notation is employed to denote Sobolev spaces and their norms [6]. The solution p to (1) depends on σ, p_{obs} is the recorded data, and X_r collects the locations of the receivers.

The goal of FWI is to solve the minimization problem

$$\min_{\sigma} \mathcal{J}(\sigma) = \min_{\sigma} J(p(\sigma), \sigma) = \min_{\sigma} \Big[\frac{1}{2} \sum_{\mathbf{x}_i \in X_r} \int_0^T \left(p(\mathbf{x}_i, t) - p_{obs}(\mathbf{x}_i, t) \right)^2 dt \Big]. \quad (3)$$

In the following, to simplify notation, the dependence on $(\mathbf{x}, t)$ in understood. It is well known that (3) is an ill-posed problem since $\mathcal{J}$ exhibits many local minima. Therefore, some regularization is usually added to $\mathcal{J}$. With this aim, the most typical choice is a Tikhonov regularization, for example the L^2-norm or the H^1-seminorm of σ [19]. As an alternative, we consider a Total Variation (TV) regularization, i.e., the integral of the absolute value of the gradient of σ [2, 22, 21]. The idea is inherited from the signal/image processing framework to damp the excessive and possibly spurious details characterized by a high total variation, while preserving only the important details of the signal [14].

Functional $\mathcal{J}$ in (3) is thus modified in

$$\min_{\sigma} \mathcal{J}^{\mathrm{FWI}}(\sigma) = \min_{\sigma} J^{\mathrm{FWI}}(p(\sigma), \sigma) = \min_{\sigma} \Big\{ \mathcal{J}(\sigma) + \eta \int_{\Omega} \phi \left(|\nabla \sigma| \right) d\mathbf{x} \Big\}, \quad (4)$$

where $\phi(s) = \sqrt{\delta^2 + s^2}$, δ is a regularization parameter, and η is the total variation constant. In practice, $\phi(|\nabla \sigma|) = \sqrt{\delta^2 + |\nabla \sigma|^2}$ is the "lifted" absolute value with δ the (usually small) positive lifting parameter [4].

3 An optimal control approach to FWI

In order to solve (4), we apply the adjoint-state method, as done in [18, 12, 20]. Essentially, this approach amounts to interpreting the unknown field σ as a control variable, which is to be determined by minimizing $\mathcal{J}^{\mathrm{FWI}}$. In order to deal with this minimization, we go back to the functional J^{FWI} subject to the constraint that p satisfies the wave equation. For the sake of simplicity, we use homogeneous Dirichlet boundary conditions for equation (1). However, the derivation of the adjoint problem in the case of absorbing boundary conditions can be performed in an analogous fashion (see the Appendix for the details).

Let us consider the Lagrangian

$$\mathcal{L}(p, \sigma, \lambda) = J^{\mathrm{FWI}}(p, \sigma) - \int_{\Omega} \int_0^T \lambda \left(\sigma \frac{\partial^2 p}{\partial t^2} - \nabla \cdot (\nabla p) - f \right) dt d\mathbf{x}, \quad (5)$$

where p, σ, and λ are considered as independent variables, and J^{FWI} is thought of as a function of p and σ. The variable p is the state variable and solves the

state or forward problem (1), while λ is the Lagrange multiplier and corresponds to the adjoint variable.

According to the classical result of constrained optimization [11, 5], the first-order necessary optimality condition requires that the solution of the optimization problem (4) is obtained as the stationary points of $\mathcal{L}$. In particular, if (p, σ, λ) is a saddle point of $\mathcal{L}$, then σ is a minimum point of $\mathcal{J}^{\mathrm{FWI}}$.

For this purpose, we now compute the Gâteaux derivatives of $\mathcal{L}$ with respect to λ, p, and σ, in order to obtain the gradient of $\mathcal{J}^{\mathrm{FWI}}$ with respect to σ.
By enforcing that the derivatives of $\mathcal{L}$ with respect to λ and p vanish, we obtain the forward (or state), and backward (or adjont) problem, for p and λ, respectively. Finally, the gradient of $\mathcal{J}^{\mathrm{FWI}}$ coincides with the derivative of $\mathcal{L}$ with respect to σ, evaluated at $(p(\sigma), \sigma, \lambda(\sigma))$.

Now we provide the details of these computations.

Forward problem The forward problem can be obtained by imposing that the Fréchet derivative, $\dfrac{D\mathcal{L}}{D\lambda}$, be identically equal to zero.

For this purpose, we compute the Gâteaux (directional) derivative of $\mathcal{L}$ at (p, σ, λ) with respect to λ in the generic direction η. This yields

$$\begin{aligned}\frac{D\mathcal{L}}{D\lambda}[\eta] &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\Big(\mathcal{L}(p, \sigma, \lambda + \varepsilon\eta) - \mathcal{L}(p, \sigma, \lambda)\Big) \\ &= -\int_\Omega \int_0^T \eta \left(\sigma \frac{\partial^2 p}{\partial t^2} - \nabla \cdot (\nabla p) - f\right) dt d\mathbf{x} = \langle \frac{D\mathcal{L}}{D\lambda}, \eta \rangle,\end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $L^2(0, T; L^2(\Omega))$ and itself. Thus $\dfrac{D\mathcal{L}}{D\lambda} = 0$ if we require that p be the solution to the forward problem (1), where the boundary and initial conditions are strongly enforced.

Backward problem Analogously to the previous paragraph, we derive the backward problem by imposing that $\dfrac{D\mathcal{L}}{Dp} = 0$.

The Gâteaux derivative of $\mathcal{L}$ at (p, σ, λ) with respect to p in the generic direction q, which satisfies the same boundary and initial conditions as p, is

$$\frac{D\mathcal{L}}{Dp}[q] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\Big(\mathcal{L}(p + \varepsilon q, \sigma, \lambda) - \mathcal{L}(p, \sigma, \lambda)\Big).$$

We compute each term of $\mathcal{L}(p + \varepsilon q, \sigma, \lambda) - \mathcal{L}(p, \sigma, \lambda)$ separately, and then we sum

them up. We have

$$\begin{aligned}(I) =& J^{\mathrm{FWI}}(p+\varepsilon q,\sigma) - J^{\mathrm{FWI}}(p,\sigma)\\ &= \frac{1}{2}\sum_{\mathbf{x}_i\in X_r}\int_0^T \left((p+\varepsilon q)-p_{obs}\right)^2 dt - \frac{1}{2}\sum_{\mathbf{x}_i\in X_r}\int_0^T \left(p-p_{obs}\right)^2 dt\\ &= \frac{1}{2}\sum_{\mathbf{x}_i\in X_r}\int_0^T\int_\Omega 2\varepsilon q\delta_{\mathbf{x}_i}\big(p-p_{obs}\big)d\mathbf{x}dt + \frac{1}{2}\sum_{\mathbf{x}_i\in X_r}\int_0^T \varepsilon^2 q^2 dt\\ &= \varepsilon\int_0^T\int_\Omega \Big(\sum_{\mathbf{x}_i\in X_r}(p-p_{obs})\delta_{\mathbf{x}_i}\Big)q d\mathbf{x}dt + \frac{\varepsilon^2}{2}\int_0^T\sum_{\mathbf{x}_i\in X_r} q^2 dt,\end{aligned}$$

$\delta_{\mathbf{x}_i}$ being the Dirac distribution associated with the location $\mathbf{x}_i$, and

$$\begin{aligned}(II) =& -\int_\Omega\int_0^T \lambda\left(\sigma\frac{\partial^2(p+\varepsilon q)}{\partial t^2} - \nabla\cdot(\nabla(p+\varepsilon q))\right)dtd\mathbf{x}\\ &+\int_\Omega\int_0^T\lambda\left(\sigma\frac{\partial^2 p}{\partial t^2}-\nabla\cdot(\nabla p)\right)dtd\mathbf{x}\\ =& -\varepsilon\int_\Omega\int_0^T\lambda\left(\sigma\frac{\partial^2 q}{\partial t^2}-\nabla\cdot(\nabla q)\right)dtd\mathbf{x}.\end{aligned}$$

Integrating by parts, we obtain

$$\begin{aligned}(II) =& \varepsilon\left[\int_\Omega\int_0^T\sigma\frac{\partial\lambda}{\partial t}\frac{\partial q}{\partial t}dtd\mathbf{x} - \int_\Omega\left(\sigma\lambda\frac{\partial q}{\partial t}\right)\Big|_0^T d\mathbf{x}\right.\\ &\left.-\int_\Omega\int_0^T\nabla\lambda\cdot\nabla q dtd\mathbf{x} + \int_0^T\int_{\partial\Omega}\lambda\nabla q\cdot\mathbf{n}dtd\boldsymbol{\gamma}\right]\\ =& \varepsilon\left(-\int_\Omega\int_0^T\sigma\frac{\partial^2\lambda}{\partial t^2}q dtd\mathbf{x} + \int_\Omega\left(\sigma\frac{\partial\lambda}{\partial t}q\right)\Big|_0^T d\mathbf{x} - \int_\Omega\left(\sigma\lambda\frac{\partial q}{\partial t}\right)\Big|_0^T d\mathbf{x}\right.\\ &\left.+\int_\Omega\int_0^T\nabla\cdot(\nabla\lambda)q dtd\mathbf{x} - \int_0^T\int_{\partial\Omega}\nabla\lambda\cdot\mathbf{n}q\,dtd\boldsymbol{\gamma} + \int_0^T\int_{\partial\Omega}\lambda\nabla q\cdot\mathbf{n}dtd\boldsymbol{\gamma}\right)\\ =& \varepsilon\left(-\int_\Omega\int_0^T\sigma\frac{\partial^2\lambda}{\partial t^2}q dtd\mathbf{x} + \int_\Omega\left(\sigma\frac{\partial\lambda}{\partial t}q\right)\Big|_{t=T} d\mathbf{x} - \int_\Omega\left(\sigma\lambda\frac{\partial q}{\partial t}\right)\Big|_{t=T} d\mathbf{x}\right.\\ &\left.+\int_\Omega\int_0^T\nabla\cdot(\nabla\lambda)q dtd\mathbf{x} + \int_0^T\int_{\partial\Omega}\lambda\nabla q\cdot\mathbf{n}dtd\boldsymbol{\gamma}\right).\end{aligned}$$

By summing up (I) and (II) and by dividing by ε, we obtain $\dfrac{D\mathcal{L}}{Dp}[q]$ as the limit for $\varepsilon\to 0$. Requiring that $\dfrac{D\mathcal{L}}{Dp}[q]=0$, for any q, we obtain the backward

problem for the adjoint variable λ

$$\begin{cases} \sigma \dfrac{\partial^2 \lambda}{\partial t^2} = \nabla \cdot (\nabla \lambda) + \displaystyle\sum_{\mathbf{x}_i \in X_r} (p - p_{obs}) \delta_{\mathbf{x}_i} & \text{in } \Omega \times (0,T) \\ \lambda = 0 & \text{on } \partial\Omega \times (0,T) \\ \lambda = 0,\ \dfrac{\partial \lambda}{\partial t} = 0 & \text{on } \Omega \times \{T\}. \end{cases} \tag{6}$$

According to (6), the adjoint variable is obtained by propagating the residual backward in time, namely the difference between recorded and computed data, through the wave equation. The backward direction induces the final conditions on λ.

Gradient equation The gradient equation is obtained by setting to zero the gradient of $\mathcal{J}^{\text{FWI}}$ with respect to σ, i.e., by computing $\dfrac{D\mathcal{L}}{D\sigma}$. The Gâteaux derivative of $\mathcal{L}$ at (p, σ, λ) with respect to σ in the generic direction r is

$$\begin{aligned} \frac{D\mathcal{L}}{D\sigma}[r] &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Big(\mathcal{L}(p, \sigma + \varepsilon r, \lambda) - \mathcal{L}(p, \sigma, \lambda) \Big) \\ &= - \int_\Omega \int_0^T \lambda \frac{\partial^2 p}{\partial t^2} r dt d\mathbf{x} + \eta \int_\Omega \frac{\nabla \sigma \cdot \nabla r}{\sqrt{\delta^2 + |\nabla \sigma|^2}} d\mathbf{x} \\ &= - \int_\Omega \int_0^T \lambda \frac{\partial^2 p}{\partial t^2} r dt d\mathbf{x} - \eta \int_\Omega \nabla \cdot \left(\frac{\nabla \sigma}{\sqrt{\delta^2 + |\nabla \sigma|^2}} \right) r d\mathbf{x} + \eta \int_{\partial\Omega} \frac{\nabla \sigma \cdot \mathbf{n}\, r}{\sqrt{\delta^2 + |\nabla \sigma|^2}} d\boldsymbol{\gamma} \end{aligned}$$

which yields the gradient equation

$$\nabla \mathcal{J}^{\text{FWI}}(\sigma) = \frac{D\mathcal{L}}{D\sigma}(p(\sigma), \sigma, \lambda(\sigma)) = - \int_0^T \lambda \frac{\partial^2 p}{\partial t^2} dt - \eta \nabla \cdot \left(\frac{\nabla \sigma}{\sqrt{\delta^2 + |\nabla \sigma|^2}} \right) = 0, \tag{7}$$

after assuming homogeneous Neumann conditions for σ.

3.1 Numerical approximation

The minimization problem (4) is tackled by applying the Barzilai Borwein (BB) method [3, 13]. This is an iterative two-step descent method suitable for the solution of complex and high-dimensional minimization problems. At each step, given the current value of σ, the BB method requires computing the funcional $\mathcal{J}^{\text{FWI}}$ and its gradient.
Generally, the iteration of the BB method applied to the minimization of a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by: let $\mathbf{w}_{-1}, \mathbf{w}_0 \in \mathbb{R}^n$ be two initial guesses; then, for $k \geq 0$

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \beta_k \alpha_k \nabla f(\mathbf{w}_k),$$

where

$$\alpha_k = \frac{(\nabla f(\mathbf{w}_k) - \nabla f(\mathbf{w}_{k-1})) \cdot (\mathbf{w}_k - \mathbf{w}_{k-1})}{\|\nabla f(\mathbf{w}_k) - \nabla f(\mathbf{w}_{k-1})\|^2},$$

with the step-length β_k chosen via a line-search method [13], $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^n$.
When applied to (4), we identify f with $\mathcal{J}^{\mathrm{FWI}}$ and $\mathbf{w}$ with the discrete counterpart of σ. Moreover, each evaluation of the gradient of $\mathcal{J}^{\mathrm{FWI}}$ requires solving both the forward and the backward problems derived in the previous section.

The discretization of the problems (1) and (6), of $\mathcal{J}^{\mathrm{FWI}}$ in (4) and of its gradient in (7) is performed by a finite difference method, based on second-ordered centered schemes for both the space and time derivatives. For this purpose, we set $\Omega = (a, b) \times (c, d)$ and introduce a Cartesian grid characterized by the uniform steps h_x, h_z, along the x and z direction, respectively. We thus have $N_x = (b - a)/h_x$ and $N_z = (d - c)/h_z$ subintervals identifying the spatial nodes

$$\mathbf{x}_{ij} = (x_i, z_j) = (x_0 + ih_x, z_0 + jh_z) \quad i = 0, ..., N_x + 1,\ j = 0, ..., N_z + 1,$$

with $x_0 = a$, $x_{N_x+1} = b$, $z_0 = c$, $z_{N_z+1} = d$. As for the time discretization, we use a constant time step Δt to partition the time window $[0, T]$ into $N_t = T/\Delta t$ intervals, with end points $t^k = k\Delta t$, $k = 0, ..., N_t + 1$.

We use the standard notation $v^k_{i,j}$ to denote the space-time values $v(\mathbf{x}_{ij}, t^k)$ of the mesh function v at $(\mathbf{x}_{ij}, t^k)$. To evaluate the integrals involved in the definition of $\mathcal{J}^{\mathrm{FWI}}$ and of its gradient, we adopt the composite trapezoidal quadrature rule.

We detail the discrete scheme associated with (1) in the case of both homogeneous Dirichlet and absorbing boundary conditions. The discretization of the backward problem (6) can be obtained in a similar way, just reversing the time direction.

Let us first consider the case of homogeneous Dirichlet boundary conditions. The overall scheme is: given $p^0_{i,j} = p^1_{i,j} = 0$ for $i = 0, ..., N_x + 1$, $j = 0, ..., N_z + 1$, compute $p^{k+1}_{i,j}$, for $k = 1, ..., N_t$, satisfying

$$p^{k+1}_{i,j} = 2p^k_{i,j} - p^{k-1}_{ij} + \frac{\Delta t^2}{\sigma_{i,j}} \left[\frac{p^k_{i+1,j} - 2p^k_{i,j} + p^k_{i-1,j}}{h_x^2} + \frac{p^k_{i,j+1} - 2p^k_{i,j} + p^k_{i,j-1}}{h_z^2} + f^k_{i,j} \right],$$

for $i = 1, \ldots, N_x$, $j = 1, \ldots, N_z$, and

$$p^{k+1}_{i,j} = 0 \quad \text{for } i \in \{0, N_x + 1\}, j \in \{0, N_z + 1\}.$$

Let us now consider the assignment of absorbing boundary conditions (ABC), following the approximation proposed in [8], and to fix ideas, we focus on the ABC on the bottom side of a rectangular domain Ω. Applying the Fourier

transform in space ($\partial/\partial x \leftrightarrow -\iota\kappa_x$, $\partial/\partial z \leftrightarrow -\iota\kappa_z$, with $\iota = \sqrt{-1}$) and time ($\partial/\partial t \leftrightarrow \iota\omega$), to the homogeneous wave equation with a constant slowness

$$\sigma \frac{\partial^2 p}{\partial t^2} - \Delta p = 0,$$

we obtain the dispersion relation of a wave travelling in a homogeneous medium:

$$\sigma\omega^2 - \kappa^2 = 0, \tag{8}$$

where ω is the angular frequency, and $\kappa = \sqrt{\kappa_x^2 + \kappa_z^2}$ is the absolute value of the wave vector of components κ_x and κ_z. From this relation, the vertical wave number, κ_z, can be expressed as a function of $\kappa_x/(\sigma^{1/2}\omega)$, i.e.,

$$\kappa_z = \pm\, \kappa_z \left(\frac{\kappa_x}{\sigma^{1/2}\omega} \right). \tag{9}$$

The quantity $h = \kappa_x/(\sigma^{1/2}\omega)$ depends on the incidence angle of the wave. For quasi-vertical plane waves (quasi-horizontal wave front), this ratio is small. Equation (9) has two solutions, corresponding to waves that move vertically, downward (+), and upward (−). In order to control the reflections occurring at the boundary, we neglect the negative solution in (9). Thus, the positive solution represents the exact absorbing boundary condition.

The dispersion relation (9) can be rewritten as

$$\kappa_z = \sigma^{1/2}\omega\sqrt{1-h^2} \quad \text{with } |h| \ll 1. \tag{10}$$

From this relation, we obtain the ABC conditions via a Taylor expansion:

$$\kappa_z = \kappa_z(0) + \kappa_z'(0)h + \frac{\kappa_z''(0)}{2}h^2 + \dots . \tag{11}$$

In particular, the zero-th order Taylor truncation yields

$$\kappa_z = \sigma^{1/2}\omega.$$

By Fourier antitransforming this relation, we obtain

$$\frac{\partial p}{\partial z} = -\sigma^{1/2}\frac{\partial p}{\partial t}, \tag{12}$$

which represents the ABC condition adopted in this work. With reference to the stencil in Fig. 2, this equation can be discretized as derived in [8]:

$$D_z^- \left(p_{i,N_z+1}^k + p_{i,N_z+1}^{k+1} \right) + \sigma_{i,N_z+1}^{1/2} D_t^+ \left(p_{i,N_z+1}^k + p_{i,N_z}^k \right) = 0, \tag{13}$$

for $i = 1, \dots, N_x$, where D_z^- is the backward difference with respect to z and D_t^+ is the forward difference with respect to t, i.e.,

$$D_z^- v_{i,j}^k = (v_{i,j}^k - v_{i,j-1}^k)/h_z, \quad D_t^+ v_{i,j}^k = (v_{i,j}^{k+1} - v_{i,j}^k)/\Delta t.$$

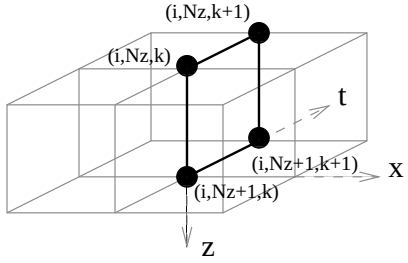


Figure 2: Stencil for the discretization of the ABC condition (12) on the bottom boundary.

Expanding these discrete operators, (13) can be rewritten as

$$p_{i,N_z+1}^{k+1} = \frac{\sigma_{i,N_z+1}^{1/2} \dfrac{p_{i,N_z+1}^{k} - p_{i,N_z}^{k+1} + p_{i,N_z}^{k}}{\Delta t} - \dfrac{p_{i,N_z+1}^{k} - p_{i,N_z}^{k+1} - p_{i,N_z}^{k}}{h_z}}{\dfrac{1}{h_z} + \sigma_{i,N_z+1}^{1/2} \dfrac{1}{\Delta t}}.$$

In principle, higher-order ABC could be obtained by keeping successive terms in (11). However, stable higher-order (≥ 2) absorbing boundary conditions can be derived only via a Padé approximation.

For a general boundary, the lowest order ABC is

$$\sigma^{1/2} \frac{\partial p}{\partial t} + \nabla p \cdot \mathbf{n} = 0,$$

where $\mathbf{n}$ is the unit outward normal vector to the boundary.

Approximation of the TV term

We focus on the numerical approximation of the term related to the TV regularization, appearing in the expression of the gradient of $\mathcal{J}^{\mathrm{FWI}}$ in (7), i.e.,

$$\nabla \cdot \left(\frac{\nabla \sigma}{\sqrt{\delta^2 + |\nabla \sigma|^2}} \right). \tag{14}$$

We introduce D_x and D_z as the second-order finite difference operators with respect to x and z, respectively, given by

$$D_x \sigma_{i,j} = \begin{cases} \left(-3\sigma_{0,j} + 4\sigma_{1,j} - \sigma_{2,j}\right)/(2h_x) & i = 0 \\ \left(\sigma_{i+1,j} - \sigma_{i-1,j}\right)/(2h_x) & i = 1, \cdots, N_x \\ \left(\sigma_{N_x-1,j} - 4\sigma_{N_x,j} + 3\sigma_{N_x+1,j}\right)/(2h_x) & i = N_x + 1 \end{cases}$$

and analogously for D_z. Notice that we have omitted the dependence on the super-index k since σ does not depend on time. We can therefore approximate

(14) by

$$(\nabla\sigma)_{i,j} \simeq \begin{bmatrix} D_x\sigma_{i,j} \\ D_z\sigma_{i,j} \end{bmatrix}, \quad |\nabla\sigma|^2_{i,j} \simeq (D_x\sigma_{i,j})^2+(D_z\sigma_{i,j})^2, \quad (\nabla\cdot\mathbf{v})_{i,j} \simeq D_x v^x_{i,j}+D_z v^z_{i,j},$$

where $\mathbf{v} = (v^x, v^z)^T$, with v^x, v^z generic mesh functions. Combining these approximations, the discretization of (14) is

$$\nabla \cdot \left(\frac{\nabla\sigma}{\sqrt{\delta^2 + |\nabla\sigma|^2}} \right)_{i,j} \simeq$$
$$D_x \left[\frac{D_x\sigma_{i,j}}{\sqrt{\delta^2 + (D_x\sigma_{i,j})^2 + (D_z\sigma_{i,j})^2}} \right]_{i,j} + D_z \left[\frac{D_z\sigma_{i,j}}{\sqrt{\delta^2 + (D_x\sigma_{i,j})^2 + (D_z\sigma_{i,j})^2}} \right]_{i,j}.$$

Remark 3.1 (Checkpointing) *In order to deal with he backward problem in a computationally efficient way, we have implemented the checkpointing strategy [15, 9, 10]. The main idea of this technique is to store the forward solution at some times instead of performing a full backup, and to recompute missing information for the backward problem when missing. Uniform checkpointings have been used, despite the optimal distribution in [9] could be employed.*

4 Coupled Multiscenario Full Waveform Inversion (CMFWI)

In this section, we formalize the new approach, named Coupled Multiscenario Full Waveform Inversion (CMFWI), proposed to enhance the standard FWI in order to strengthen its behavior with respect to the choice of the initial guess. The general idea of this method is to introduce a multiscenario configuration, where each scenario is associated with a particular slowness, depending on a specific placement of the shots. The final outcome is unique and combines the different configurations by a least-squares penalization. As shown in Section 5, this multiscenario model makes FWI robuster, due to the more richness of information compared with the standard FWI.

For the sake of simplicity, we detail the method on problem (1) supplemented with homogeneous Dirichlet boundary conditions, and we assume that each scenario is driven by a single source term.

We introduce $\mathcal{S}$ scenarios, and for $s = 1, ..., \mathcal{S}$, we denote by $\mathbf{x}_s$ the position of the corresponding source term. The problem for the pressure p_s behind the s-th scenario is

$$\begin{cases} \sigma_s \dfrac{\partial^2 p_s}{\partial t^2} = \nabla \cdot (\nabla p_s) + f_s & \text{in } \Omega \times (0,T) \\ p_s = 0 & \text{on } \partial\Omega \times (0,T) \\ p_s = 0, \ \dfrac{\partial p_s}{\partial t} = 0 & \text{on } \Omega \times \{0\}, \end{cases} \tag{15}$$

where σ_s is the squared slowness associated with the s-th source f_s. We collect all of the slowness fields in the vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_{\mathcal{S}})^T$.
The idea of CMFWI is to apply FWI to the separate $\mathcal{S}$ scenarios, each one independent of the others, and to enforce the different slowness fields to be somewhat the same (since there is only one Earth). More precisely, for $s = 1, \ldots, \mathcal{S}$, we define the s-th functional

$$\begin{aligned} \mathcal{J}_s^{\mathrm{CMFWI}}(\boldsymbol{\sigma}) &= \frac{1}{2} \sum_{\mathbf{x}_i \in X_r} \int_0^T \left(p_s(\mathbf{x}_i, t) - p_{s,obs}(\mathbf{x}_i, t)\right)^2 dt + \frac{\mu}{4} \sum_{\substack{j=1 \\ j \neq s}}^{\mathcal{S}} \|\sigma_s - \sigma_j\|^2_{L^2(\Omega)} \\ &+ \eta \int_\Omega \sqrt{\delta^2 + |\nabla \sigma_s|^2}\, d\mathbf{x}, \end{aligned} \tag{16}$$

where the first term accounts for the misfit between modelled and measured data, the second quantity penalizes the discrepancy among the s-th slowness field and all fields associated with the other scenarios, with respect to the $L^2(\Omega)$-norm, and we have added a TV regularization term, but only to σ_s. Notice that, in this case, we have $\mathcal{S}$ sets of observations, each set being associated with one source only.
The global functional is obtained by summing all the single-source functionals as

$$\mathcal{J}^{\mathrm{CMFWI}}(\boldsymbol{\sigma}) = \sum_{s=1}^{\mathcal{S}} \mathcal{J}_s^{\mathrm{CMFWI}}(\boldsymbol{\sigma}). \tag{17}$$

This functional represents the quantity to be actually minimized. At each step of the overall minimization algorithm, we let the different $\mathcal{J}_s^{\mathrm{CMFWI}}$ be relatively free to change, under the condition that $\mathcal{J}^{\mathrm{CMFWI}}$ decreases.
By proceeding in the same way as in classical FWI, we thus minimize (17) under the $\mathcal{S}$ contraints (15), by applying the adjoint-state method. Hence, we have $\mathcal{S}$ adjoint variables, where λ_s, $s = 1, \ldots, \mathcal{S}$, solves the backward problem

$$\begin{cases} \sigma_s \dfrac{\partial^2 \lambda_s}{\partial t^2} = \nabla \cdot (\nabla \lambda_s) + \displaystyle\sum_{\mathbf{x}_i \in X_r} (p_s - p_{s,obs}) \delta_{\mathbf{x}_i} & \text{in } \Omega \times (0,T) \\ \lambda_s = 0 & \text{on } \partial\Omega \times (0,T) \\ \lambda_s = 0,\ \dfrac{\partial \lambda_s}{\partial t} = 0 & \text{on } \Omega \times \{T\}. \end{cases} \tag{18}$$

The computation of the gradient of $\mathcal{J}^{\mathrm{CMFWI}}$ with respect to $\boldsymbol{\sigma}$ is analogous to the one of classical FWI. In particular, the second term appearing in (16) acts as a regularization term with respect to the $L^2(\Omega)$-norm, and accounts for the misfit of σ_s with respect to a given reference value, say σ_0, usually used in seismic inversion, that is, $\|\sigma_s - \sigma_0\|^2_{L^2(\Omega)}$ [18]. Therefore, it is straightforward to check that the s-th component of the gradient of $\mathcal{J}^{\mathrm{CMFWI}}$ can be written as

$$\frac{\partial \mathcal{J}^{\mathrm{CMFWI}}}{\partial \sigma_s}(\boldsymbol{\sigma}) = -\int_0^T \lambda_s \frac{\partial^2 p_s}{\partial t^2} dt + \mu \sum_{\substack{j=1 \\ j \neq s}}^{\mathcal{S}} (\sigma_s - \sigma_j) - \eta \nabla \cdot \left(\frac{\nabla \sigma_s}{\sqrt{\delta^2 + |\nabla \sigma_s|^2}} \right). \tag{19}$$

5 Numerical results

In this section, we compare FWI and CMFWI for different source terms and boundary conditions. The constant characterizing the TV-regularization is picked as $\delta = 10^{-2}$.
We consider the wave equation (1), where the domain is the square $\Omega = (0, 0.5) \times (0, 0.5)$Km, and the source term, f, is chosen as a sine or a Ricker wavelet in time, i.e.,

$$f(\mathbf{x}, t) = \sum_{\mathbf{x}_i \in X_s} \delta_{\mathbf{x}_i}(\mathbf{x})\, g(t),$$

where X_s is the set of the source locations, while $g(\cdot)$ is the sine wavelet of frequency f_p, or the classical Ricker wavelet

$$g(t) = A(2(\pi f_p(t - t_0))^2 - 1)e^{-(\pi f_p(t-t_0))^2},$$

where A is the maximum amplitude, f_p is the peak frequency, and t_0 is the time when the maximum amplitude is reached.
Both sources and receivers are located just beneath the surface, but they can be located at different x-positions.

For both FWI and CMFWI, we employ the Barzilai Borwein (BB) method to deal with the corresponding minimization problems. We enforce a lower bound for σ (implying an upper bound for the speed) at each iteration in order to guarantee the CFL condition. The parameter involved in the coupling of the different σ_s in CMFWI is set to $\mu = 10^{-4}$, after a trial-and-error procedure. We show the results obtained after 100 iterations of the BB method, and by using a peak frequency $f_p = 8$ Hz for the source. In all test cases, the observations are obtained by running the forward problem after feeding in the exact slowness profile, with the same space-time discretization parameters and boundary conditions used in the actual runs.

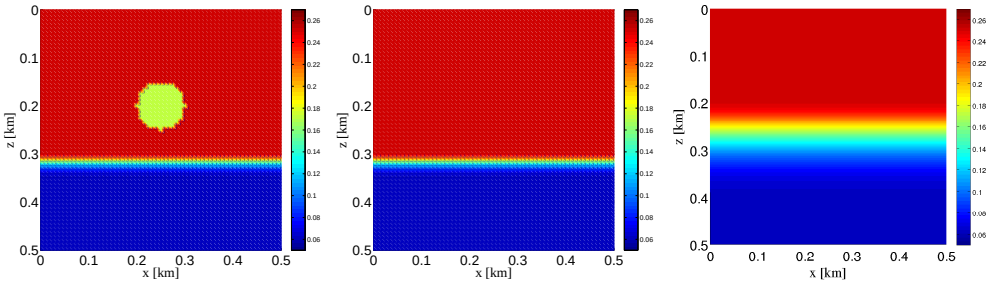
5.1 Test Case I

We set $T = 0.5$s, $h_x = 8.3$m, and $\Delta t = 0.001$s. In Fig. 3(a), we plot the exact profile in $\mathrm{s}^2/\mathrm{Km}^2$

$$\sigma = \begin{cases} 0.25 & \{(x, z) : z \leq 0.3\} \setminus \mathcal{C} \\ [3 - (\cos(\pi(z - 0.3)/0.05)]^{-2} & 0.3 < z < 0.35 \\ 0.0625 & z \geq 0.35 \\ 0.1736 & \text{in } \mathcal{C}, \end{cases}$$

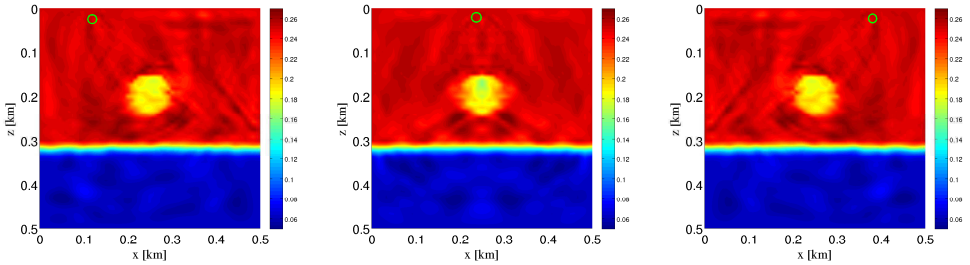
with $\mathcal{C} = \{(x, z) : (x - 0.25)^2 + (z - 0.2)^2 < 0.0025\}$. This consists of a steep variation along the z direction, plus an embedded circular discontinuity.

We fist assess the two methods in the case of Dirichlet boundary conditions. The initial guess σ_0 for σ is shown in Fig. 3(b), and coincides with the exact



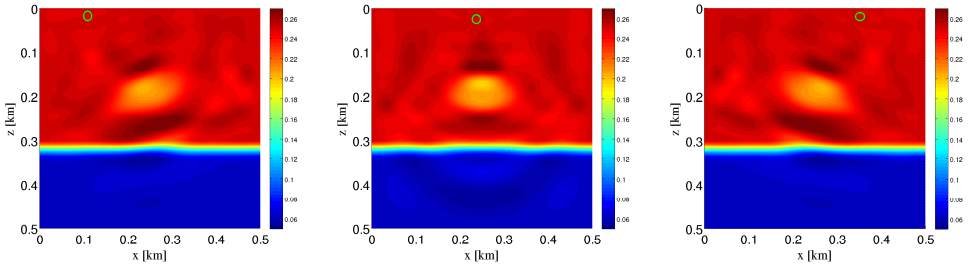
(a) True σ (b) Initial guess σ_0 (c) Initial guess $\widetilde{\sigma}_0$

Figure 3: Test case I: exact solution and initial guesses



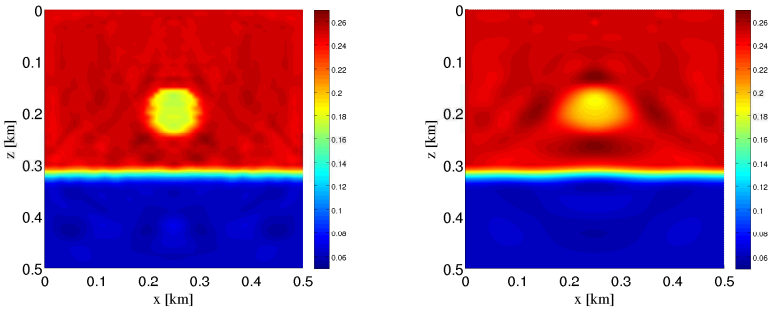
(a) $x_s = L_x/4$ (b) $x_s = 2L_x/4$ (c) $x_s = 3L_x/4$

Figure 4: Test case I: solutions for different source locations (circular marks) using FWI with a sine wavelet g and initial guess σ_0



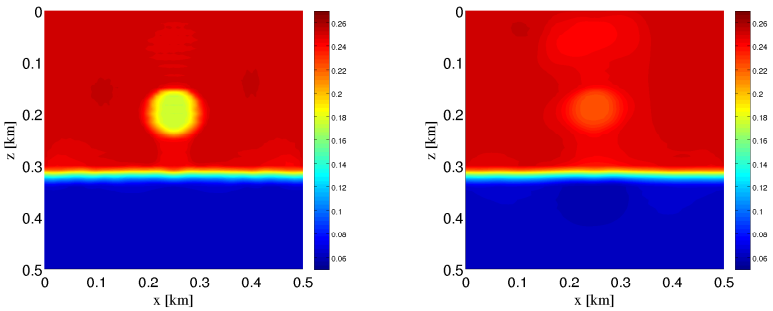
(a) $x_s = L_x/4$ (b) $x_s = 2L_x/4$ (c) $x_s = 3L_x/4$

Figure 5: Test case I: solutions for different source locations (circular marks) using FWI with a Ricker wavelet g and initial guess σ_0



(a) Sine function (b) Ricker wavelet

Figure 6: Test case I: solutions using CMFWI with three evenly spaced sources, and for different wavelets g and initial guess σ_0



(a) $\eta = 10^{-7}$ (b) $\eta = 10^{-6}$

Figure 7: Test case I: solutions using CMFWI with a sine wavelet g, for different values of the TV constant η and initial guess σ_0

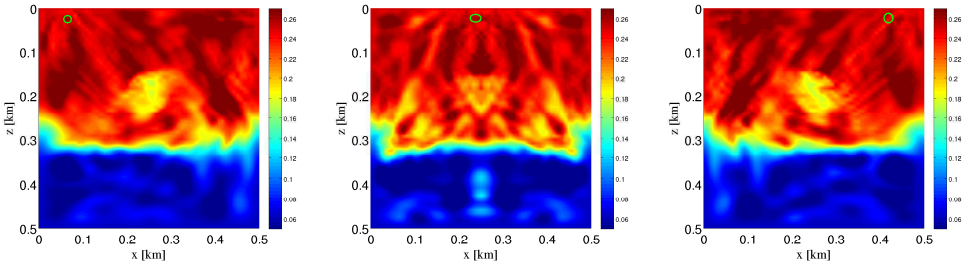
σ, except for the absence of the circular pattern. The sources are located at $x_i = i\,L_x/4$, for $i = 1, 2, 3$, with $L_x = 0.5$Km. FWI and CMFWI are assessed for both types of sources. We observe (see Figures 4-6) that both methods behave better when we use the sine source. Moreover, the solution obtained by applying the FWI method always depends on the source location. Instead, CMFWI takes into account all the scenarios at the same time, and it thus yields a more global imaging of the subsoil. Moreover, all the $\mathcal{S}$ profiles σ_s are very similar, and we always show the prediction associated with the leftmost source, σ_1. In some sense, the similarity among all the predictions supports the uniqueness of the final scenario.

Finally, in Fig. 7, we show that by tuning the TV constant η, it is possible to obtain a sharper and more realistic velocity profile than without any regularization term (compare Fig. 6(a) with Fig. 7). It appears that an excessive regularization tends to blur the predicted scenario.

We now compare the choice of the boundary conditions, focusing on both Dirichlet and absorbing boundary conditions. For this purpose, we consider a more challenging initial guess, $\widetilde{\sigma}_0$ displayed in Fig. 3(c), which is less close to the true σ due the smoother transition between the two constant values. Moreover, we consider seven sources, located at $x_i = i\,L_x/8$, for $i = 1, \ldots, 7$. We show the results obtained with the FWI method, using either one single source at a time, or the seven sources together, and with CMFWI. We apply FWI with seven sources together to understand whether the better performance obtained with CMFWI depends only on the richness of information characterizing this last approach, or by the intrinsic superiority of CMFWI itself.

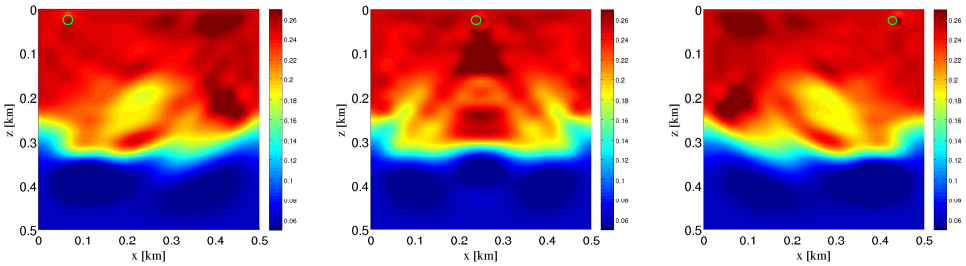
First, we compare the results computed without any regularization term, and with Dirichlet boudary conditions. Figures 8-11 gather the corresponding results without TV regularization. In particular, Figures 8 and 9 show that FWI with a single source provides for this smoothed initial guess a very poor profile. The sine source yields a more noisy slowness profile compared with Fig. 9, whereas the Ricker wavelet leads to a blurred approximation. In general, the FWI method does not capture the discontinuity of the squared slowness in the background profile, and this causes the failure of the method in identifying the desired profile. On the other hand, by increasing the number of sources, the performance of FWI actually improves, as shown in Fig. 10, with an analogous behavior with respect to the type of wavelet. The quality of the predicted profile still improves in the case of the new proposed method, as shown in Fig. 11. Therefore, we may claim that CMFWI is indeed a richer method than the standard FWI.

Finally, in Figures 12 and 13, we compare the results obtained with the two methods, both with seven sources and with the two types of wavelets, using absorbing boundary conditions. As expected, CMFWI behaves better than FWI. Moreover, the Dirichlet boundary conditions provide sharper slowness profiles for both FWI and CMFWI as shown on comparing Fig. 10 with 12, and Fig. 11 with 13.



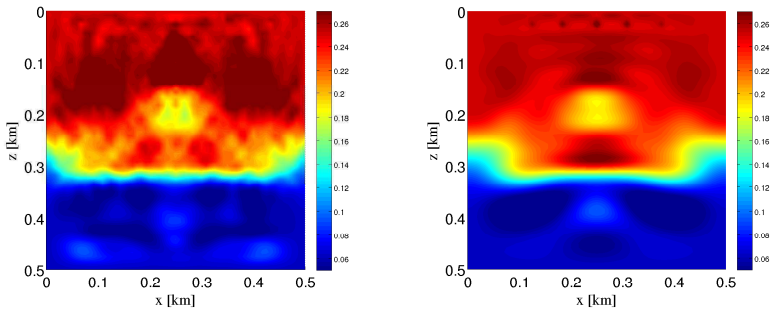
(a) $x_s = L_x/8$ (b) $x_s = 4L_x/8$ (c) $x_s = 7L_x/8$

Figure 8: Test case I: solutions for different source locations using FWI with a sine wavelet and Dirichlet boundary conditions and initial guess $\widetilde{\sigma}_0$



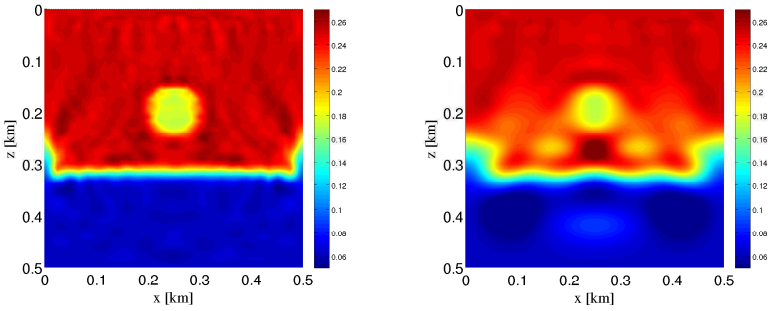
(a) $x_s = L_x/8$ (b) $x_s = 4L_x/8$ (c) $x_s = 7L_x/8$

Figure 9: Test case I: solutions for different source locations using FWI with a Ricker wavelet and Dirichlet boundary conditions and initial guess $\widetilde{\sigma}_0$



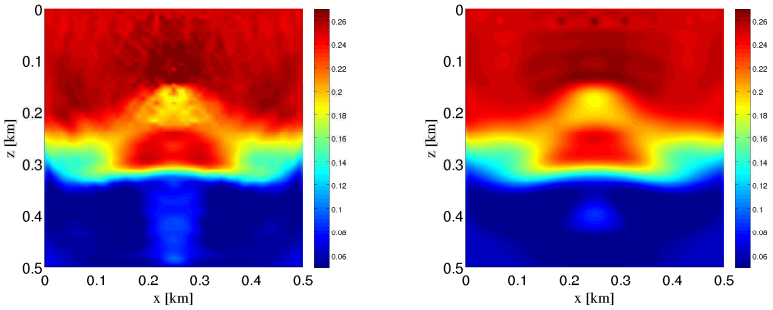
(a) Sine wavelet (b) Ricker wavelet

Figure 10: Test case I: solutions with seven sources using FWI with Dirichlet boundary conditions and initial guess $\widetilde{\sigma}_0$



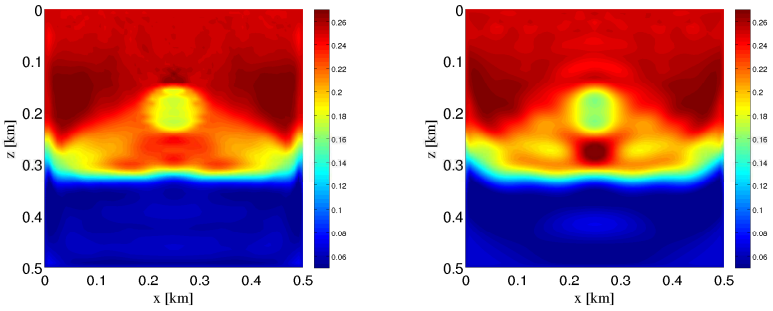
(a) Sine wavelet (b) Ricker wavelet

Figure 11: Test case I: solutions with seven sources using CMFWI with Dirichlet boundary conditions and initial guess $\widetilde{\sigma}_0$



(a) Sine wavelet (b) Ricker wavelet

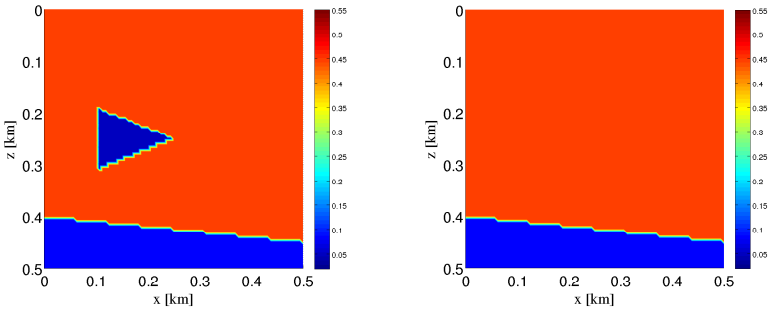
Figure 12: Test case I: solutions using FWI with seven sources with absorbing boundary conditions and initial guess $\widetilde{\sigma}_0$



(a) Sine wavelet (b) Ricker wavelet

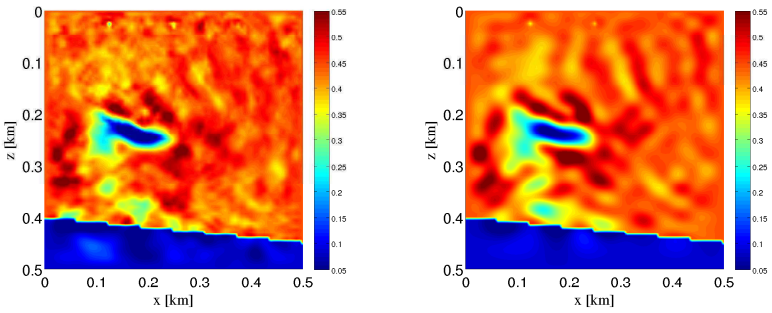
Figure 13: Test case I: solutions with seven sources using CMFWI with absorbing boundary conditions and initial guess $\widetilde{\sigma}_0$

5.2 Test Case II



(a) True σ (b) Initial guess

Figure 14: Test case II: exact solution and initial guess



(a) Sine wavelet (b) Ricker wavelet

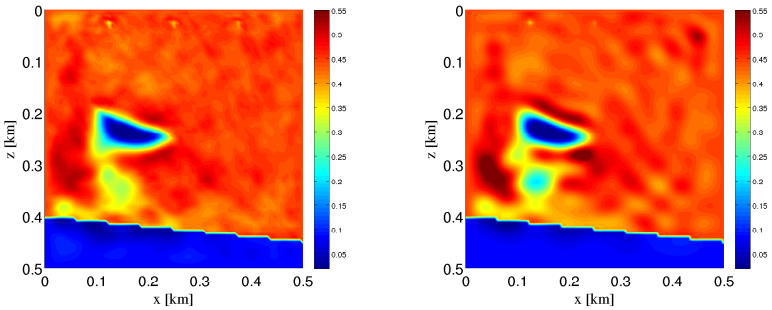
Figure 15: Test case II: solutions using FWI with three evenly spaced sources

In this test case, we consider a variant of a test case considered in [7]. We select the profile in Fig. 14(a), given by

$$\sigma = \begin{cases} 0.4444 & z \geq \mathcal{R} \setminus \mathcal{T} \\ 0.0816 & z < \mathcal{R} \\ 0.0494 & \text{in } \mathcal{T}, \end{cases}$$

with $\mathcal{R} = \{(x, z) : z = 0.1x + 0.4\}$ and $\mathcal{T}$ the (open) triangle defined by the vertices: $(0.1, 0.3)$, $(0.1, 0.2)$, and $(0.25, 0.25)$. The geometric configuration is more complex than the one in the previous test case, due to both the wedge-shaped body and the dipping reflector. In particular, we expect the bottom of the body to be the most challenging part to be detected by a reflection-based seismic inversion procedure.

We consider the FWI and CMFWI methods with the three sources placed at $x_i = i\,Lx/4$, with $i = 1, 2, 3$. We set $T = 0.9$s, $h_x = 8.3$m, and $\Delta t = 0.001$s. First we neglect the regularization term, and we use only Dirichlet boundary

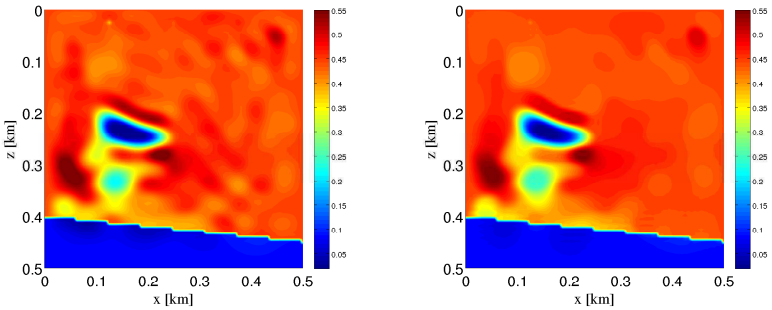


(a) Sine wavelet (b) Ricker wavelet

Figure 16: Test case II: solutions using CMFWI with three evenly spaced sources

conditions.

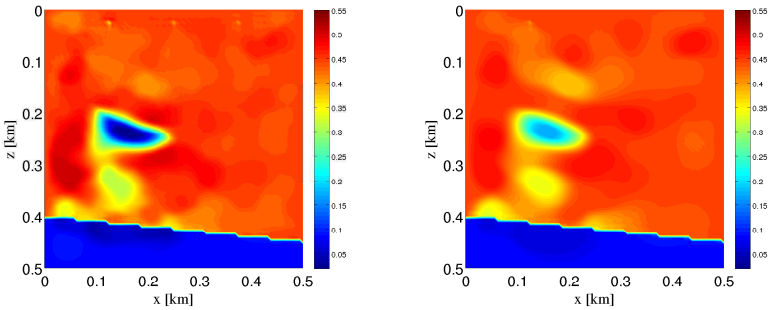
On comparing the results in Fig. 15 and in Fig. 16, we still remark the superiority of CMFWI, for the same number of sources. Moreover, a comparison across wavelets points out that the Ricker signal leads to less reliable scenarios. For both wavelets and methods, some artifacts appear beneath the triangular body, likely due to the complex interaction with the reflector, which seem to be less noticeable with the sine, though.



(a) $\eta = 10^{-7}$ (b) $\eta = 10^{-6}$

Figure 17: Test case II: solutions using CMFWI with three evenly spaced sources and a Ricker wavelet

As a last check, we include the TV regularization in the CMFWI approach. Figures 17 and 18 provide the slowness field for two choices of the parameter η. The effect of η is different for the two wavelets. Indeed, whereas $\eta = 10^{-6}$ does not suffice to detect a clear-cut image with the Ricker wavelet, it turns out to be excessive for the sine wavelet, yielding too blurred a prediction. The artifacts are stronger when using a Ricker signal.



(a) $\eta = 10^{-7}$ (b) $\eta = 10^{-6}$

Figure 18: Test case II: solutions using CMFWI with three evenly spaced sources and a sine wavelet

5.3 Sensitivity analysis to noise

The recorded data are usually affected by noise due to error in measurements or to the instrument precision. To simulate more realistic configurations, we consider FWI and CMFWI when a Gaussian white noise is added to the observed data, which are associated with the exact (noiseless) slowness profiles.

Let us consider a discrete signal $\{X^k\}_{k=0}^{N_t+1}$. The noisy signal $\{Y^k\}_{k=0}^{N_t+1}$ is obtained by adding to X^k a Gaussian white noise in the following way:

$$Y^k = X^k + wn \quad \text{with} \quad wn \sim \mathcal{N}\left(0, \frac{\|X\|^2/(N_t+2)}{\text{SNR}}\right),$$

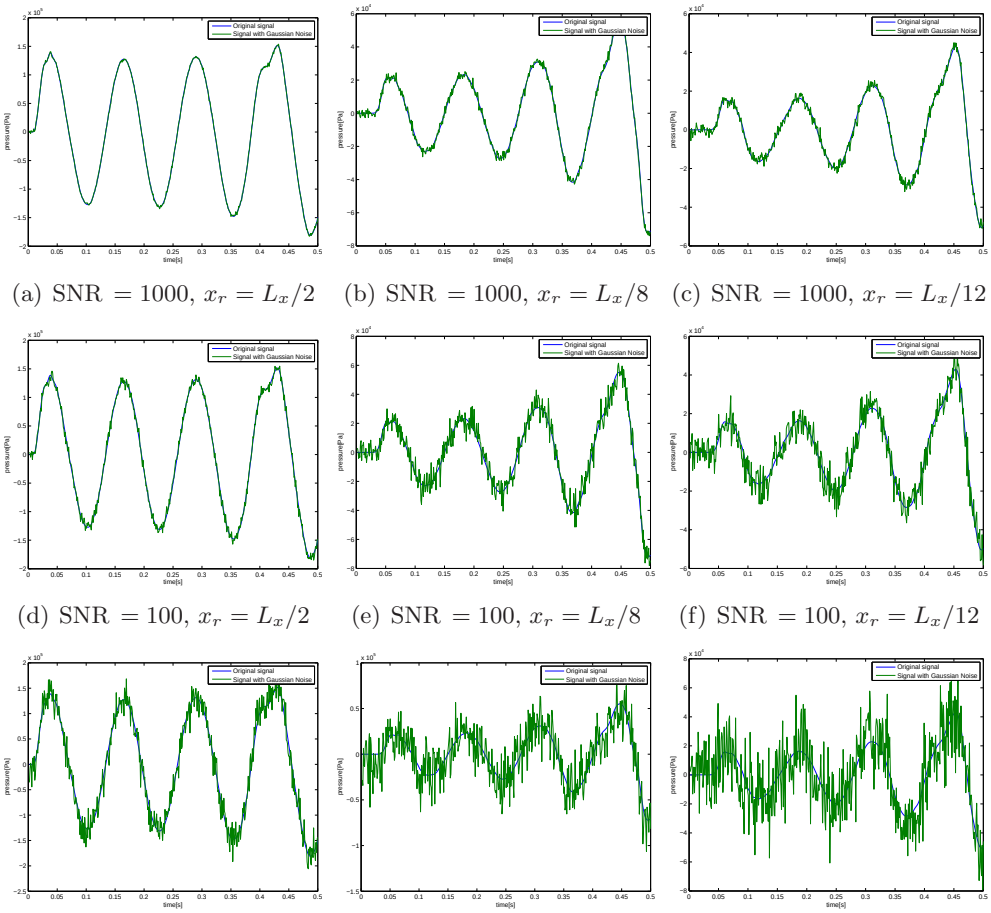
where $\mathcal{N}(m, \sigma)$ identifies a normal distribution with mean m and variance σ, and SNR is the signal-to-noise ratio. The noise wn is independent of the time index k since it is white. No TV regularization is considered.

We consider Test case I and compare FWI with CMFWI by perturbing the observed data p_{obs} in (2) for different values of the parameter SNR , and by shooting three sources simultaneously (see Fig. 19).
Figures 20 and 21 show that the CMFWI method is more robust with respect to noise. In particular, FWI is reliable for SNR $\simeq 1000$, whereas CMFWI furnishes qualitatively reasonale results also for SNR $\simeq 100$.

Increasing the numer of sources, with CMFWI one can even reduce SNR up to 10 without missing important details (see Fig. 23). This is not the case of the FWI appraoch which, even in the presence of seven sources, performs badly (see Fig. 22). The slowness field associated with SNR $= 1000$ is slighly better than the corresponding one in Fig. 20, but it is still very noisy.

6 Conclusions and future developments

We have proposed the Coupled Multiscenario Full Waveform Inversion (CMFWI) method for solving the seismic inversion problem. Moreover, we included a TV



(a) SNR $= 1000$, $x_r = L_x/2$ (b) SNR $= 1000$, $x_r = L_x/8$ (c) SNR $= 1000$, $x_r = L_x/12$

(d) SNR $= 100$, $x_r = L_x/2$ (e) SNR $= 100$, $x_r = L_x/8$ (f) SNR $= 100$, $x_r = L_x/12$

(g) SNR $= 10$, $x_r = L_x/2$ (h) SNR $= 10$, $x_r = L_x/8$ (i) SNR $= 10$, $x_r = L_x/12$

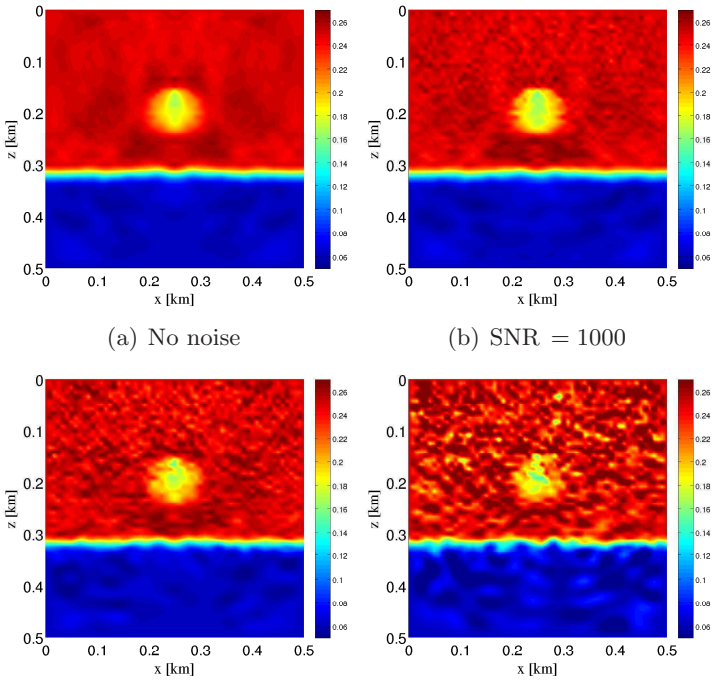
Figure 19: Test case I: seismograms with noise for different SNR for three sources, and for a sine wavelet

regularization term, instead of the more classical Tikhonov regularization, in order to take into account possible abrupt velocity profiles, thus reducing in a more effective way spurious small-scale oscillations.

We have implemented this new method in Matlab® and we have compared it with the classical FWI on two test cases, considering both Dirichlet and absorbing boundary conditions, as well as sine and Ricker wavelet sources. To make a fair comparison, we use FWI with multiple sources, all shooting simultaneously.

The results show that CMFWI is better than classical FWI and robuster with respect to noisy measurements. Moreover, while increasing the number of sources improves the quality of the approximated slowness profile for CMFWI, this is not always the case for FWI, especially with noise.

The current implementation of CMFWI has been developed in serial environment. To improve the computational performance, among the future develop-



(a) No noise (b) SNR = 1000

(c) SNR = 100 (d) SNR = 10

Figure 20: Test case I: FWI solutions with three simultaneous sources and a sine waveket, for different level of noise

ments, we include the parallelization of the code which can substantially exploit the uncoupling of the forward problems, as well as of the backward problems.

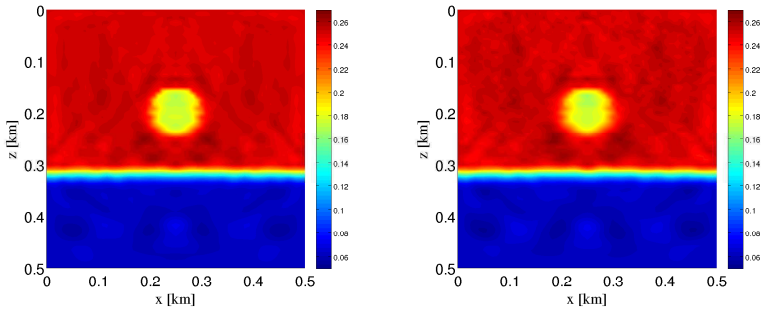
The Dirichlet boundary conditions turn out to be the best choice with a view to a reliable approximation. Also the absorbing boundary conditions do not degrade the solution provided by the CMFWI method.

The comparison between sine and Ricker wavelets seem to suggest that, in general, the first signal yields more accurate results.

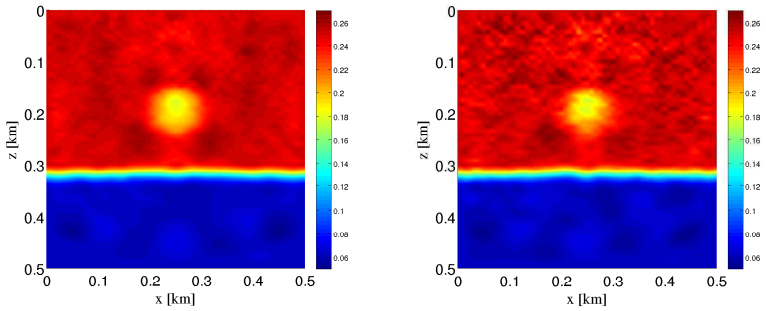
Finally, a verification on more realistic cases is ongoing, to assess the effectiveness of the proposed method.

Appendix

We outline the changes to the backward problem (6) when considering ABC conditions on a portion, say Γ_{ABC}, of the boundary of the domain. The forward



(a) without noise (b) SNR $= 1000$



(c) SNR $= 100$ (d) SNR $= 10$

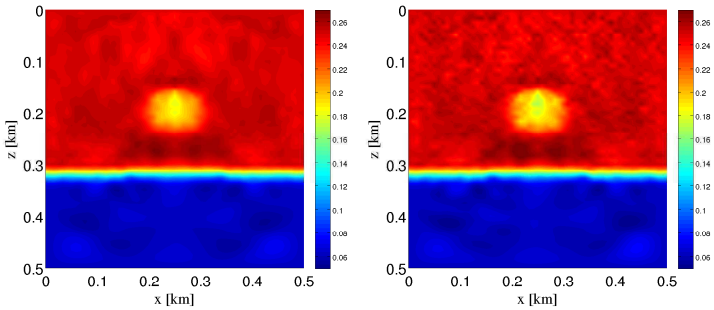
Figure 21: Test case I: CMFWI solutions with three simultaneous sources and a sine waveket, for different level of noise

problem becomes

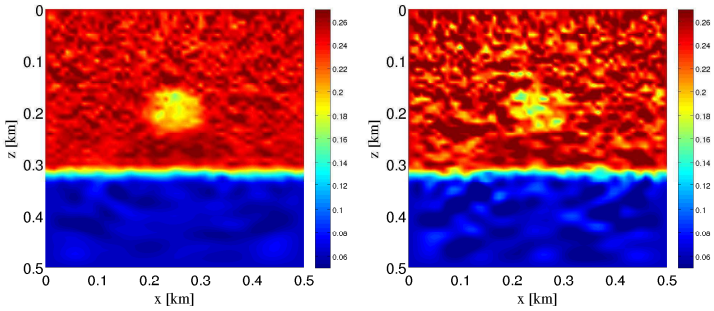
$$\begin{cases} \sigma \dfrac{\partial^2 p}{\partial t^2} = \nabla \cdot (\nabla p) + f & \text{in } \Omega \times (0,T) \\ p = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0,T) \\ \sqrt{\sigma} \dfrac{\partial p}{\partial t} + \nabla p \cdot \mathbf{n} = 0 & \text{on } \Gamma_{\mathrm{ABC}} \times (0,T) \\ p = 0,\ \dfrac{\partial p}{\partial t} = 0 & \text{on } \Omega \times \{0\}, \end{cases} \tag{20}$$

where Γ_{D} is the portion of the boundary where Dirichlet boundary conditions hold. In this case, the Lagrangian does not change, and the only point of the analysis where we should care about boundary conditions is in the derivation of the backward problem. Thus, we focus only on this last issue.

Backward problem with ABC As in the case of homogeneous Dirichlet boundary conditions, the backward problem is obtained by requiring $\dfrac{D\mathcal{L}}{Dp} = 0$, where $\mathcal{L}$ is defined in (5), and we proceed exactly in the same way.
We compute the Gâteaux derivative of $\mathcal{L}$ at (p, σ, λ) with respect to p in the generic direction q, satisfyng the same initial and boundary conditions as p. The



(a) No noise (b) SNR = 1000

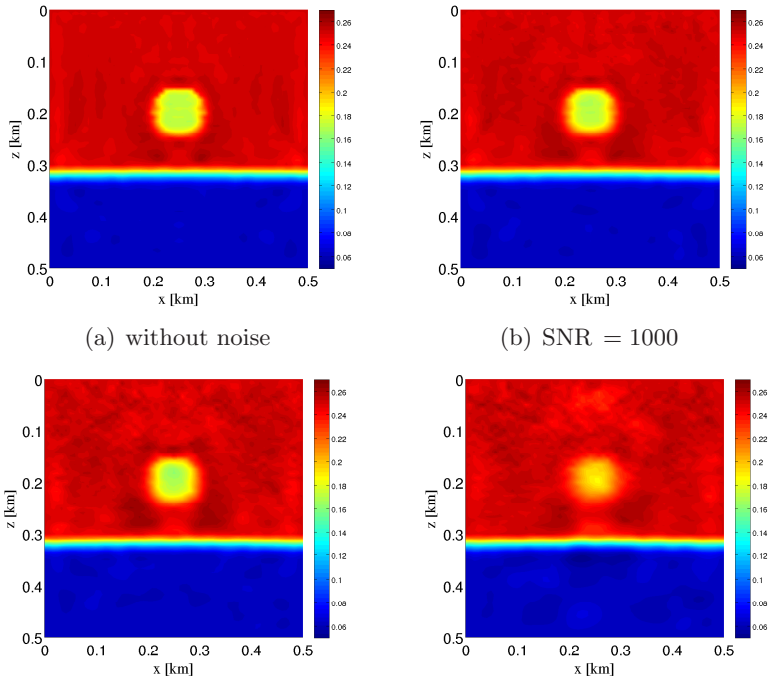


(c) SNR = 100 (d) SNR = 10

Figure 22: Test case I: FWI solutions with seven simultaneous sources and a sine waveket, for different level of noise

contribution (I) associated with J^{FWI} is the same as in the Dirichlet case, and is not developed here. In the term (II), associated with the Lagrange multiplier, only the boundary terms change.

$$
\begin{aligned}
&-\int_0^T \int_{\partial\Omega} \nabla\lambda \cdot \mathbf{n}\, q\, dt d\boldsymbol{\gamma} + \int_0^T \int_{\partial\Omega} \lambda \nabla q \cdot \mathbf{n}\, dt\, d\boldsymbol{\gamma} \\
&= -\int_0^T \int_{\Gamma_{\mathrm{D}}} \nabla\lambda \cdot \mathbf{n}\, q\, dt\, d\boldsymbol{\gamma} - \int_0^T \int_{\Gamma_{\mathrm{ABC}}} \nabla\lambda \cdot \mathbf{n}\, q\, dt\, d\boldsymbol{\gamma} \\
&\quad + \int_0^T \int_{\Gamma_{\mathrm{D}}} \lambda \nabla q \cdot \mathbf{n}\, dt\, d\boldsymbol{\gamma} - \int_0^T \int_{\Gamma_{\mathrm{ABC}}} \lambda \sqrt{\sigma} \frac{\partial q}{\partial t}\, dt\, d\boldsymbol{\gamma} \\
&= -\int_0^T \int_{\Gamma_{\mathrm{ABC}}} \nabla\lambda \cdot \mathbf{n}\, q\, dt\, d\boldsymbol{\gamma} + \int_0^T \int_{\Gamma_{\mathrm{D}}} \lambda \nabla q \cdot \mathbf{n}\, dt\, d\boldsymbol{\gamma} \\
&\quad + \int_0^T \int_{\Gamma_{\mathrm{ABC}}} \sqrt{\sigma} \frac{\partial \lambda}{\partial t}\, q\, dt\, d\boldsymbol{\gamma} - \int_{\Gamma_{\mathrm{ABC}}} \left(\sqrt{\sigma}\lambda q\right)\big|_0^T\, d\boldsymbol{\gamma}.
\end{aligned}
$$



(a) without noise (b) SNR $= 1000$

(c) SNR $= 100$ (d) SNR $= 10$

Figure 23: Test case I: CMFWI solutions with seven simultaneous sources and a sine waveket, for different level of noise

We therefore obtain that the second term (II) is

$$
\begin{aligned}
(II) = \varepsilon \Bigg(& - \int_\Omega \int_0^T \sigma \frac{\partial^2 \lambda}{\partial t^2}\, q \, dt \, d\mathbf{x} + \int_\Omega \left(\sigma \frac{\partial \lambda}{\partial t} q \right) \Big|_{t=T} d\mathbf{x} - \int_\Omega \left(\sigma \lambda \frac{\partial q}{\partial t} \right) \Big|_{t=T} d\mathbf{x} \\
& + \int_\Omega \int_0^T \nabla \cdot (\nabla \lambda)\, q \, dt \, d\mathbf{x} - \int_0^T \int_{\Gamma_{\mathrm{ABC}}} \nabla \lambda \cdot \mathbf{n}\, q \, dt \, d\boldsymbol{\gamma} + \int_0^T \int_{\Gamma_{\mathrm{D}}} \lambda \nabla q \cdot \mathbf{n} \, dt \, d\boldsymbol{\gamma} \\
& + \int_0^T \int_{\Gamma_{\mathrm{ABC}}} \sqrt{\sigma} \frac{\partial \lambda}{\partial t}\, q \, dt \, d\boldsymbol{\gamma} - \int_{\Gamma_{\mathrm{ABC}}} \left(\sqrt{\sigma} \lambda q \right) \big|_{t=T} d\boldsymbol{\gamma} \Bigg) .
\end{aligned}
$$

The overall contributions to $\dfrac{D\mathcal{L}}{Dp}[q]$ are obtained from $(I) + (II)$, after dividing by ε, and passing to the limit for $\varepsilon \to 0$. Requiring that $\dfrac{D\mathcal{L}}{Dp}[q] = 0$, for any q,

the backward problem for λ follows:

$$\begin{cases} \sigma \dfrac{\partial^2 \lambda}{\partial t^2} = \nabla \cdot (\nabla \lambda) + \displaystyle\sum_{\mathbf{x}_i \in X_r} (p - p_{obs}) \delta_{\mathbf{x}_i} & \text{in } \Omega \times (0,T) \\ \lambda = 0 & \text{on } \Gamma_{\mathrm{D}} \times (0,T) \\ -\sqrt{\sigma} \dfrac{\partial \lambda}{\partial t} + \nabla \lambda \cdot \mathbf{n} = 0 & \text{on } \Gamma_{\mathrm{ABC}} \times (0,T) \\ \lambda = 0,\ \dfrac{\partial \lambda}{\partial t} = 0 & \text{on } \Omega \times \{T\}. \end{cases} \tag{21}$$

According to (21), the adjoint variable is still obtained by back propagating the residual, namely the difference between recorded and computed data, and satisfies homogeneous boundary conditions on Γ_{D}. On Γ_{ABC}, the ABC condition is the same as in the forward problem, after a time-reversal, i.e., $t \mapsto \tau = T - t$. Actually, if we rewrite the ABC in (20) with respect to τ, we obtain

$$\sqrt{\sigma} \frac{\partial \lambda}{\partial \tau} + \nabla \lambda \cdot \mathbf{n} = 0 \quad \text{on } \Gamma_{\mathrm{ABC}} \times (0,T).$$

References

[1] U. Albertin, J. Kapoor, R. Randall, M. Smith, G. Brown, C. Soufleris, P. Whitfield, F. Dewey, J. Farnsworth, G. Grubitz, and M. Kemme. The time for depth imaging. *Oilfield Review*, 2002.

[2] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*. Springer-Verlag, New York, 2002.

[3] J. Barzilai and J.M. Borwein. Two-point step size gradient method. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[4] T.F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.

[5] Ph.G. Ciarlet. *Introduction to Numerical Linear Algebra and Optimization*. Cambridge University Press, New York, 1985.

[6] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems I*, volume 5. Springer-Verlag, Berlin, 1992.

[7] E. Duveneck and P.M. Bakker. Stable P-wave modeling for reverse-time migration in tilted TI media. *Geophysics*, 76(2):S65–S75, 2011.

[8] B. Engquist and D. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Math. Comp.*, 31(139):629–651, 1977.

[9] A. Griewank and A. Walther. Algorithm 799: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Soft.*, (26):19–45, 2009.

[10] V. Heuveline and A. Walther. Online Checkpointing for Parallel Adjoint Computation in PDEs: Application to Goal-Oriented Adaptivity and Flow Control. In W.E. Nagel et al., editor, *Euro-Par 2006 Parallel Processing*, volume 4128 of *Lecture Notes in Computer Science*, pages 689–699. Springer Berlin Heidelberg, 2006.

[11] J.L. Lions. *Nonhomogeneous Boundary Value Problems and Applications*. Springer-Verlag, Berlin, 1972.

[12] R.E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys. J. Int.*, 167:495–503, 2006.

[13] L. Qi, K. Teo, and X. Yang. *Optimization and Control with Applications*. Springer Science+Business Media, Inc, 2005.

[14] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[15] W.W. Symes. Reverse time migration with optimal checkpointing. *Geophysics*, pages 213–221, 2007.

[16] W.W. Symes. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56:765–790, 2008.

[17] W.W. Symes. The seismic reflection inverse problem. *IOPscience*, 25, 2009.

[18] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.

[19] A.N. Tychonoff and V.Y. Arsenin. *Solution of Ill-posed Problems*. Winston & Sons, Washington, 1977.

[20] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

[21] Y. Wang, LG. Dong, and YZ. Liu. Improved hybrid iteative optimization method for seismic full waveform inversion. *Appl. Geophys.*, 10(3):265–277, 2013.

[22] Y. Wang, C. Yang, and C. Cui. Hybrid regularization methods for seismic reflectivity inversion. *Int. J. Geomath.*, 2:87–112, 2011.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

58/2014 Menafoglio, A.; Secchi, P.; Guadagnini, A.
A Class-Kriging predictor for Functional Compositions with Application to Particle-Size Curves in Heterogeneous Aquifers

59/2014 Manzoni, A.; Pagani, S.; Lassila, T.
Accurate solution of Bayesian inverse uncertainty quantification problems using model and error reduction methods

60/2014 Signorini, M.; Micheletti, S.; Perotto, S.
CMFWI: Coupled Multiscenario Full Waveform Inversion for seismic inversion

57/2014 Giverso, C.; Verani, M.; Ciarletta P.;
Branching instability in expanding bacterial colonies

56/2014 Antonietti, P. F.; Sarti, M.; Verani, M.; Zikatanov, L. T.
A uniform additive Schwarz preconditioner for the hp-version of Discontinuous Galerkin approximations of elliptic problems

55/2014 Antonietti, P. F.; Houston P.; Sarti, M.; Verani, M.
Multigrid algorithms for hp-version Interior Penalty Discontinuous Galerkin methods on polygonal and polyhedral meshes

54/2014 Ferrario, E.; Pini, A.
Uncertainties in renewable energy generation systems: functional data analysis, monte carlo simulation, and fuzzy interval analysis

53/2014 Ieva, F.; Paganoni, A.M., Pietrabissa, T.
Dynamic clustering of hazard functions: an application to disease progression in chronic heart failure

52/2014 Dede , L..; Quarteroni, A.; S. Zhu, S.
Isogeometric analysis and proper orthogonal decomposition for parabolic problems

51/2014 Dassi, F.; Perotto, S.; Formaggia, L.
A priori anisotropic mesh adaptation on implicitly defined surfaces