Eigenvector models for solving the seismic inverse problem for the Helmholtz equation

Florian Faucher¹, ¹, ² Otmar Scherzer¹, ³ and Hélène Barucq²

¹Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria. E-mail: florian.faucher@univie.ac.at
²Inria Project-Team Magique 3D, E2S UPPA, CNRS, Pau, France
³Johann Radon Institute for Computational and Applied Mathematics (RICAM), Linz, Austria

SUMMARY
We study the seismic inverse problem for the recovery of subsurface properties in acoustic media. In order to reduce the ill-posedness of the problem, the heterogeneous wave speed parameter is represented using a limited number of coefficients associated with a basis of eigenvectors of a diffusion equation, following the regularization by discretization approach. We compare several choices for the diffusion coefficient in the partial differential equations, which are extracted from the field of image processing. We first investigate their efficiency for image decomposition (accuracy of the representation with respect to the number of variables). Next, we implement the method in the quantitative reconstruction procedure for seismic imaging, following the full waveform inversion method, where the difficulty resides in that the basis is defined from an initial model where none of the actual structures is known. In particular, we demonstrate that the method may be relevant for the reconstruction of media with salt-domes. We use the method in 2-D and 3-D experiments, and show that the eigenvector representation compensates for the lack of low-frequency information, it eventually serves us to extract guidelines for the implementation of the method.

Key words: Image processing; Numerical modelling; Waveform inversion; Seismic tomography; Wave propagation.

1 INTRODUCTION
We consider the inverse problem associated with the propagation of time-harmonic waves which occurs, for example, in seismic applications, where the mechanical waves are used to probe the Earth. Following the non-intrusive geophysical setup for exploration, we work with measured seismograms that record the waves at the surface (i.e. partial boundary measurements) and one-side illumination (backscattered/reflection data). In the last decades, this problem has encountered a growing interest with the increase in numerical capability and the use of supercomputers. However, the accurate recovery of the deep subsurface structures remains a challenge, due to the non-linearity and ill-posedness of the problem, the availability of partial reflection data only, and the large scale domains of investigation.

In the context of seismic, the quantitative reconstruction of physical properties using an iterative minimization of a cost function originally follows the work of Bamberger et al. (1979), Lailly (1983) and Tarantola (1984, 1987) in the time-domain, Pratt et al. (1990, 1996, 1998) for the frequency approach. The method is commonly referred to as full waveform inversion (FWI), which takes the complete observed seismograms for data. One key of FWI is that the gradient of the misfit functional is computed using the adjoint-state method (Lions & Mitter 1971; Chavent 1974), to avoid the formation of the (large) Jacobian matrix; we refer to Plessix (2006) for a review in geophysical applications. Then, Newton-type algorithms represent the traditional framework to perform the iterative minimization. Due to the large computational scale of the domain investigated, seismic experiments may have difficulties to incorporate second order (Hessian) information in the algorithm, and alternative techniques have been proposed, for example, in the works of Pratt et al. (1998), Akcelik et al. (2002), Choi et al. (2008), Métivier et al. (2013) and Jun et al. (2015). The quantitative (as opposed to qualitative) reconstruction methods based upon iterative minimization are naturally not restricted to seismic and we refer, among others, to Ammari et al. (2015); Barucq et al. (2018) and the references therein for additional applications using similar techniques.

The main difficulty of FWI (in both exploration and global seismology) lies in the high non-linearity of the problem and the presence of local minima in the misfit functional, which are due to the time shifts and cycle-skipping effect (Bunks et al. 1995), in particular when...
the background velocity (the low-frequency profile) is not correctly anticipated (Gauthier et al. 1986; Luo & Schuster 1991; Fichtner et al. 2008; Barucq et al. 2019b). For this reason, the phase information is included in the traveltime inversion by Luo & Schuster (1991) by using a cross-correlation function between measurements and simulations, where the relative phase shift is given by the maximum of the correlation. The method is further generalized by Gee & Jordan (1992), while Van Leeuwen & Mulder (2010) propose to select the phase shift using a weighted norm. The choice of misfit has further encountered a growing interest in the past decade: in application to global-scale seismology, Fichtner et al. (2008) compare the phase, correlation-based, and envelope misfit functionals, the latter being also studied by Bozdağ et al. (2011). In exploration seismic, comparisons of phase and amplitude inversion are performed by Bednar et al. (2007); Pyun et al. (2007). The L1 norm is studied by Brossier et al. (2010) while approaches based upon optimal transport are considered by Métivier et al. (2016); Yang et al. (2018). In the context where different fields are measured, Alessandrini et al. (2019) and Faucher et al. (2019) advocate for a reciprocity-based functional, which further connects to the correlation-based formulas (Faucher et al. 2019). In the case of accurate knowledge of the background velocity, the inverse problem is close to linear or quasi-linear as the Born approximation holds and then, alternative methods of linear inverse problem can be applied, such as the Backus–Gilbert method (Backus & Gilbert 1967, 1968). The difficulty to recover the background velocity variation has also motivated alternative parametrization of the inverse problem: for instance the MBTT (migration based traveltime) reformulation of FWI (Clément et al. 2001; Barucq et al. 2019b), which represents the wave speed from a background and a data-space reflectivity models.

In order to diminish the ill-posedness of the inverse problem, a regularization criterion can be incorporated. It introduces an additional constraint (in addition to the fidelity between observations and simulations), which, however, may be complicated to select a priori and problem dependent (with ‘tuning’ parameters). For instance, we refer to the body of work of Kirsch (1996), Isakov (2006), Kern (2016), Kaltenbacher (2018) and the references therein. In the regularization by discretization approach, the model representation plays the role of regularizing the functional, by controlling (and limiting) the number of unknowns in the problem, and possibly defining (i.e. constraining) the shape of the unknown (e.g. to force smoothness). Controlling the number of unknowns influences the resolution of the outcome, but also the stability and convergence of the procedure. The use of piecewise constant coefficients appears natural for numerical applications, and is also motivated by stability results (Alessandrini & Vessella 2005; Beretta et al. 2016). However, such a decomposition can lead to an artificial ‘block’ representation (cf. Beretta et al. 2016; Faucher 2017) which would not be appropriate in terms of resolution. For this reason, a piecewise linear model representation is explored by Alessandrini et al. (2018, 2019), still motivated by the stability properties. We also mention the wavelet-based model reductions, that offer a flexible framework and are used for the purpose of regularization in seismic tomography by Loris et al. (2007, 2010). In the work of Yuan et al. (2014, 2015), FWI is carried out in the time-domain with a model represented from a wavelet-based decomposition.

In our work, we will use a model decomposition based upon the eigenvectors of a chosen diffusion operator, as introduced by De Buhan & Kray (2013), Grote et al. (2017) and Grote & Nahum (2019). Note that this decomposition is shown (with the right choice of operator) to be related with the more standard total variation (TV) or Tikhonov regularizations. The main difference in our work is that we study several alternatives for the choice of the operator following image processing techniques, which traditionally also relies on such diffusion PDEs (e.g. Weickert 1998). We first investigate the performance of the decomposition depending on the choice of PDE, and, next, the performance of such a model decomposition as parameterization of the reconstruction procedure in seismic FWI. It shows that the efficient choice of PDE should change depending on the situation. In addition, we provide a series of experiment to extract the robust guidelines for the implementation of the method in seismic.

We specifically target the reconstruction of subsurface salt domes (i.e. media with high contrasting objects), which is particularly challenging, because (in addition to the usual restrictive data) of the change of the kinematics involved and the lack of low frequency data (Farmer et al. 1996; Chironi et al. 2006; Virieux & Operto 2009; Faucher et al. 2019a). In such cases, the use of the total-variation regularization (Rudin et al. 1992) with FWI is becoming more and more prominent, and consists in incorporating an additional constraint on the model in the minimization problem. Its efficiency is shown in the context of acoustic media with salt-dome contrasts by, for example Brandsberg-Dahl et al. (2017), Esser et al. (2018), Kalita et al. (2019) and Aghamiry et al. (2019). In our work, we study several alternatives and demonstrate that the model representation with the criterion extracted from Geman & Reynolds (1992) appears the most appropriate for the eigenvector decomposition method in the presence of salt-domes. We also show the limitations of the method, in particular, it appears that the decomposition fails to represent models which are composed of several thin structures.

In Section 2, we define the inverse problem associated with the Helmholtz equation and introduce the iterative method for the reconstruction of the wave speed. In Section 3, we review several possibilities for the model decomposition using the eigenvectors of the diffusion operators. The numerical implementation of the method and subsequent time cost are given in Appendix A. The process of model (image) decomposition is illustrated in Section 4. Then, in Section 5, we carry out the iterative reconstruction with FWI experiments in two and three dimensions. Here, the model decomposition is based upon the initial model, which does not contain a priori information on the target, hence increasing the differences of performance depending on the selection of the basis. It allows us to identify the best candidate for the recovery of salt dome, and to extract some guidelines for applications in quantitative reconstruction.
2 INVERSE TIME-HARMONIC WAVE PROBLEM

2.1 Forward problem

We focus on acoustic media where, for simplicity, the density is taken as a constant, leading us to the identification of a single heterogeneous parameter: the wave speed $c$. We consider a domain $\Omega$ in two or three dimensions, with $\Omega \subset \mathbb{R}^2$ or $\Omega \subset \mathbb{R}^3$. The boundary of $\Omega$, $\Gamma$ is separated into two parts: the upper (top) boundary, $\Gamma_1$, represents the interface between the Earth and the air, with a free surface boundary condition imposing the pressure to zero. The other part of the boundary, $\Gamma_2$, corresponds to the numerical restriction of the Earth for the computations and we use absorbing boundary conditions (ABC, cf. Engquist & Majda 1977) to ensure that the waves are not reflected back to the computational domain. For a source $f$ and an angular frequency $\omega$, the forward problem writes in terms of the pressure field $p$ solution to the Helmholtz equation,

\[
\begin{cases}
-\Delta - \omega^2 c^2(x) p(x) = f(x), & \text{in } \Omega, \\
p(x) = 0, & \text{on } \Gamma_1, \\
\partial_\nu p(x) - i \omega c^{-1}(x) p(x) = 0, & \text{on } \Gamma_2,
\end{cases}
\]

where $\partial_\nu$ is the normal derivative, $c$ denotes the wave speed.

The subsurface imaging works with partial surface measurements which correspond to observation of the wave propagation at the (discrete) device/receiver locations. We refer to this set of positions as $\Sigma$ and we define the forward map at frequency $\omega$, $\mathcal{F}_\omega$, (which links the model to the data), such that,

\[ \mathcal{F}_\omega : m \rightarrow p(x)|_\Sigma. \]

We have introduced $m(x) := c^{-2}(x)$, which is our choice of parameter for the reconstruction. In seismic, the data are further generated from several point sources (excited one by one) and all devices (sources and receivers) remain near the surface.

2.2 Quantitative reconstruction using iterative minimization

The inverse problem aims the reconstruction of the unknown medium squared slowness $m$ (i.e. the wave speed) from data $y_\omega$ that connects to the forward map $\mathcal{F}_\omega(m)$ for a reference (target) model $m_t$ with

\[ y_\omega = \mathcal{F}_\omega(m_t) + \mathcal{E}_\omega, \]

where $\mathcal{E}_\omega$ represents the noise in the measurements (from the inaccuracy of the devices, model error, etc), and is possibly frequency dependent. Our choice to invert the squared slowness (instead of the velocity or slowness) is motivated by the Helmholtz eq. (1) and is further discussed by Tarantola (1986), Brossier (2011), Kohn et al. (2012) and (Faucher 2017, section 5.4).

For the reconstruction, we follow the FWI method (Tarantola 1984; Pratt et al. 1998), which relies on an iterative minimization of a misfit functional defined as the difference between the observed data and the computational simulations:

\[ \min_m \mathcal{J}(m) = \frac{1}{2} \sum_\omega \| \mathcal{F}_\omega(m) - y_\omega \|^2, \]

where we use the standard least-squares minimization but alternatives have also been studied, as indicated in the introduction.

A sequence of increasing frequencies is usually taken for the minimization Problem 4 (Bunks et al. 1995; Pratt & Worthington 1990; Sirgue & Pratt 2004; Brossier et al. 2009; Barucq et al. 2018; Faucher 2017) and we use a sequential progression in our experiments (instead of band/windows of frequencies), as advocated by Faucher et al. (2019a) to increase the convergence radius. Then, an iterative minimization algorithm is used for the resolution of Problem 4 in the framework of the Newton methods. Starting with an initial guess $m^{(0)}$, the model is updated at each iteration $k$, using a search direction $s^{(k)}$, such that

\[ m^{(k+1)} = m^{(k)} + \mu s^{(k)}, \quad k > 0. \]

Several possibilities exist for the search direction (e.g., Newton, Gauss–Newton, BFGS, gradient descent, etc.) and we refer to Nocedal & Wright (2006) for an extensive review of the methods. The scalar coefficient $\mu$ is approximated using a line search algorithm (Nocedal & Wright 2006). In our implementation, we rely on a gradient-based optimization, with the non-linear conjugate gradient method, and a backtracking line search (Nocedal & Wright 2006). A review of the performance of first order-based minimization algorithms and the influence of line search step selection is further investigated by Barucq et al. (2018) in the context of inverse scattering. The computation of the gradient of the misfit functional is carried out using the adjoint-state method (Chavent 1974; Plessix 2006), which specific steps for complex-valued fields can be found in (Barucq et al. 2019b, appendix A)
3 REGULARIZATION BY DISCRETIZATION: MODEL DECOMPOSITION

In this section, we introduce a representation of the unknown, that is, the model decomposition, based upon the eigenvectors of a diffusion equation. The objective is to reduce the dimension of the unknown to mitigate the ill-posedness of the inverse problem. We provide several possibilities for the choice of the eigenvectors, following the literature in image processing.

3.1 Regularization and diffusion operators

The resolution of the inverse problem using a quantitative method introduces an optimization problem (4) where the misfit functional $\mathcal{J}$ accounts for a fidelity term. The resolution of such a problem is also common in the context of image processing (e.g. for denoising or edge enhancement) where the fidelity term corresponds to the matching between the original and processed images. It is relatively common (for both the quantitative reconstruction methods and in image processing) to incorporate an additional term in the minimization, for the purpose of regularization. The primary function of this additional term is to reduce the ill-posedness of the problem, by adding a constraint. It has been the topic of several studies, we refer to, for example, Kirsch (1996), Isakov (2006), Charbonnier et al. (1994), Robert & Deriche (1996), Rudin et al. (1992), Vogel & Oman (1996), Lobet et al. (1997), Kern (2016), Qiu et al. (2016) and Kaltenbacher (2018). The regularized minimization problem writes as

$$\min_{m} \mathcal{J}_r(m) = \frac{1}{2} \sum_{a} \| \mathcal{F}_a(m) - y_a \|^2 + \mathcal{I}(m),$$

where $\mathcal{I}$ stands for the regularization term.

In many applications such as image processing, $\mathcal{I}$ is usually defined to only depend on the gradient of the variable (image), such that

$$\mathcal{I}(m) = \int_{\Omega} \phi(|\nabla m|) \, d\Omega,$$

where $\phi \in L^2(\Omega)$. In particular, the minimum of $\mathcal{I}$ with respect to $m$ verifies the Euler–Lagrange equations (Evans 2010; Dubrovin et al. 1992). In one dimension, it is given by (Dubrovin et al. 1992, theorem 31.1.2) and is extended for higher dimensions with (Dubrovin et al. 1992, theorem 37.1.2) (further simplified in our case because $\mathcal{I}$ only depends on the gradient). It states that the minimizer of $\mathcal{I}$ is the solution of the diffusion equation:

$$\nabla \cdot \left( \frac{\phi(|\nabla m|)}{|\nabla m|} \nabla m \right) = 0 \quad \text{in } \Omega.$$  

The minimization of $\mathcal{J}_r$ can be performed using Newton-type algorithms or, when rewriting with the Euler–Lagrange formulation, be recast as a time dependent evolution problem, e.g., in image processing, Weickert (1998); Catté et al. (1992); Rudin et al. (1992); Alvarez et al. (1992).

For the sake of clarity, we introduce the following notation:

$$A(y, \eta) := -\nabla \cdot (\eta(y)\nabla), \quad \text{with} \quad \eta(y) = \frac{\phi'(|\nabla y|)}{|\nabla y|}.$$  

3.2 Diffusion coefficients from image processing

There exist several possibilities for the choice of diffusion coefficient $\eta$ (also referred to as the weighting function) in (9), inherited from image processing theory and applications. In the following, we investigate the most common formulations, see Table 1, for which we have mainly followed the ones that are reviewed by (Blanc-Féraud et al. 1995, table 1) and Robert & Deriche (1996). Furthermore, we incorporate a scaling coefficient $\beta > 0$ for the diffusion coefficient, which impacts on the magnitude. For consistency in the different models, the norms that we used are scaled with the maximal values so that they remain between 0 and 1. We define,

$$g_1(m, x) = \frac{\nabla m(x)}{\gamma_1}, \quad g_2(m, x) = \frac{\nabla m(x)^2}{\gamma_2}, \quad \text{with}$$  

$$\gamma_1 = \max \left( |\nabla m(x)| \right), \quad \gamma_2 = \max \left( |\nabla m(x)|^2 \right), \quad |\nabla m(x)| = \sqrt{\sum_{k=1}^{d} \left( \frac{\partial m}{\partial x_k} \right)^2},$$

where $d$ is the space dimension ($d = 2$ or $d = 3$ in our experiments) and $x$ the space coordinates: $x = \{x_i\}_{i=1}^d$. In order to simplify the formulas, we will omit the space dependency in the following. Note that in the numerical experiments, we calculate the eigenvalues and eigenvectors from the linear differential operator $A(m, \eta)$ defined in (9), where the diffusion coefficient $\eta$ is taken from the non-linear PDE model.

Remark 1. Regarding the diffusion coefficients that are introduced in Table 1, we note that the PDE (8) using the Tikhonov diffusion coefficient $\eta_t$ coincides with the Laplace equation while the formulation with $\eta_b$ corresponds to the total variation (TV) regularization (Rudin...
we remain with Problem (4), and use a specific representation for the model (unknown). We follow the work of De Buhan & Osses (2010), De Buhan & Kray (2013), Grote & Nahum (2019) (Lorentzian), Grote & Nahum (2019) (Gaussian) and Grote & Nahum (2019) with the ‘Adaptive Inversion’ or ‘Adaptive Eigenspace Inversion’ method. Namely, the unknown is represented via a decomposition into the basis of eigenvectors computed from a diffusion PDE. The purpose is to control the number of unknowns in the representation, and consequently reduce the ill-posedness of the inverse problem. The decomposition uses the steps given by Grote et al. (2017), depicted in Algorithm 1. The numerical implementation is given in Appendix A.

### Table 1. List of formula for the coefficient of the diffusion operator which is used to decompose the image.

| Reference (name) | Definition | $\beta \to 0$ | $\beta \to \infty$ |
|------------------|------------|---------------|-------------------|
| Perona & Malik (1988, 1990) | $\eta_1(m, \beta) = \frac{\beta}{\beta + g_2(m)}$ | 0 | 1 |
| Perona & Malik (1988, 1990) | $\eta_2(m, \beta) = \exp \left( - \frac{g_2(m)}{\beta} \right)$ | 0 | 1 |
| Geman & Reynolds (1992) | $\eta_3(m, \beta) = \frac{2\beta}{(\beta + g_2(m))^2}$ | 0 | 0 |
| Green (1990) | $\eta_4(m, \beta) = \tanh \left( \frac{g_1(m)}{\beta} \right)$ | $+\infty$ | 0 |
| Charbonnier et al. (1994) | $\eta_5(m, \beta) = \frac{1}{\beta} \left( \frac{\beta + g_2(m)}{\beta} \right)^{-1/2}$ | $+\infty$ | 0 |
| Grote & Nahum (2019) (Lorentzian) | $\eta_6(m, \beta) = \frac{\beta}{1 + \beta g_2(m)^2}$ | 0 | 0 |
| Grote & Nahum (2019) (Gaussian) | $\eta_7(m, \beta) = \left( \beta \exp \left( \frac{g_2(m)}{\beta} \right) \right)^{-1}$ | 0 | 0 |
| Rudin et al. (1992) (Total Variation, TV) | $\eta_8(m) = \frac{1}{g_1(m)}$ | n/a | n/a |
| Tikhonov | $\eta_9 = 1$ | n/a | n/a |

et al. (1992; Vogel & Oman 1996). For the formulation of $\eta_4$ and $\eta_3$, the coefficient is not defined where the gradient is zero and we impose that $\eta_4 = \eta_8 = 1$ for the positions $x_i$ where $|\nabla m(x_i)| < 10^{-12}$.

### 3.3 Eigenvector model decomposition in FWI

In our work, we use the **regularization by discretization** approach: instead of adding the regularization term $I$ in the minimization problem, we remain with Problem (4), and use a specific representation for the model (unknown). We follow the work of De Buhan & Osses (2010), De Buhan & Kray (2013), Grote et al. (2017) and Grote & Nahum (2019) with the ‘Adaptive Inversion’ or ‘Adaptive Eigenspace Inversion’ method. Namely, the unknown is represented via a decomposition into the basis of eigenvectors computed from a diffusion PDE. The purpose is to control the number of unknowns in the representation, and consequently reduce the ill-posedness of the inverse problem. The decomposition uses the steps given by Grote et al. (2017), depicted in Algorithm 1. The numerical implementation is given in Appendix A.

**Eigenvector decomposition**: given an initial model $m(x)$, a selected integer value $N > 0$, and the selected diffusion coefficient $\eta_i$.

1. Compute $m_0$, the solution of the linear PDE
   \[
   \begin{align*}
   A(m, \eta) m_0 &= 0, & \text{in } \Omega, \\
   m_0 &= m, & \text{on } \Gamma.
   \end{align*}
   \] (11)

2. Compute the subset of $N$ eigenfunctions $\{\psi_k\}_{k=1,...,N}$ which are associated to the $N$ smallest eigenvalues $\{\lambda_k\}_{k=1,...,N}$ such that, for all $k$,
   \[
   \begin{align*}
   A(m, \eta) \psi_k &= \lambda_k \psi_k, & \text{in } \Omega, \\
   \psi_k &= 0, & \text{on } \Gamma.
   \end{align*}
   \] (12)

3. Compute the model decomposition using $N$ eigenvectors:
   \[
   m = m_0 + \sum_{k=1}^{N} \alpha_k \psi_k,
   \] (13)

   where $\alpha_k$ is a scalar and $\psi_k$ a vector. Here, the set of $\alpha$ is chosen to minimize $\|m - m\|_2^2$; and the $\psi_k$, $k = 1, \ldots, N$ are the eigenvectors associated with the $N$ smallest eigenvalues $\lambda_k$, computed in Step (ii).

**Algorithm 1**: The model decomposition using the eigenvectors of the diffusion equation associated with the smallest eigenvalues. We refer to the model decomposition as $m(m, N, \eta, \Psi)$ where $\Psi(m, \eta, N)$ is the set of eigenvectors, see (14) and (15). The computational details for implementation are given in Appendix A.

Following Algorithm 1, we introduce the notation,

$\Psi(m, \eta, N) = \{\psi_k\}_{k=1,...,N}$ the set of $N$ eigenvectors associated with the model $m$ and th diffusion coefficient $\eta_i$ computed from (11) and (12);

\[
(14)
\]
and
\[ m(m, N, \psi) \] is the decomposition of the model \( m \) using \( N \) vectors from \( \psi \), (13).

Therefore, the model is represented via \( N \) coefficients \( \alpha \) in (13) in the basis given by the diffusion operator. The reconstruction procedure follows an iterative minimization of (4), and performs successive update of the coefficients \( \alpha \). The key is that \( N \) is much smaller than the dimension of the original representation of \( m \), but allows an accurate resolution, as we illustrate in Sections 4 and 5. Algorithm 2 details the procedure.

Inputs: the measurements \( \psi \); the initial model \( m(0) \); the selected number of iterations \( n_{\text{iter}} \); the list of frequencies \( \omega_i, i = 1, \ldots, n_{\omega} \); the decomposition dimension associated with frequency: \( N_i \), for \( i = 1, \ldots, n_{\omega} \).

Using Algorithm 1,
(i) compute the eigenvector basis for \( m(0) \) using the selected \( \eta \) and the highest integer \( N_{\text{max}} = \max\left\{ \{N_i\}_{i=1}^{n_{\omega}} \right\} \), \( \psi = \psi(m, \eta, N_{\text{max}}) \).
(ii) Decompose the initial model using the initial decomposition dimension:
\[ m^{(0)} = m(m^{(0)}, N_1, \psi) = m_0 + \sum_{l=1}^{N_1} \alpha_l^{(0)} \psi_l. \]

Frequency loop for \( i \in \{1, \ldots, n_{\omega}\} \) do
| Optimization loop for \( j \in \{1, \ldots, n_{\text{iter}}\} \) do |
| --- |
| Set \( k := (i - 1)n_{\text{iter}} + j - 1. \) |
| Solve the Helmholtz eq. (1) at frequency \( \omega_i \) with model \( m^{(k)} \). |
| Compute the misfit functional \( J \) in (4). |
| Compute the gradient of the misfit functional \( \nabla_{\alpha} J \) using the adjoint-state method. |
| Compute the search direction \( s^{(k)} \). |
| Compute the descent step \( \mu \) using the line search algorithm. |
| Update the coefficient \( \alpha \) with \( \alpha^{(k+1)} = \alpha^{(k)} - \mu s^{(k)}. \) |
| Update the model: \( m^{(k+1)} = m_0 + \sum_{l=1}^{N} \alpha_l^{(k+1)} \psi_l. \) |

\[ m^{(0)} = m(m^{(0)}, N_1, \psi) = m_0 + \sum_{l=1}^{N_1} \alpha_l^{(0)} \psi_l. \]

Algorithm 2: The iterative minimization algorithm (FWI) using the model decomposition to control the number of variables. The model is represented in the eigenvector basis, for which the weights are updated along with the iterations.

3.4 Gradient computation

In our implementation, the gradient is first computed with respect to the original (nodal) representation and we use the chain rule to retrieve the gradient with respect to the decomposition coefficients \( \alpha \):
\[ \frac{\partial J}{\partial \alpha} = \frac{\partial J}{\partial m} \frac{\partial m}{\partial \alpha}. \]

It is straightforward, from (13), that the derivation for a chosen coefficient \( \alpha_i \) gives \( \frac{\partial \alpha_i}{\partial m} = \psi_i \). Therefore, it is computationally easy to introduce the formulation with respect to the eigenvector decomposition from an already existing ‘classical’ (i.e. when the derivative with respect to the model is performed) formulation: it only necessitates one additional step with the eigenvectors.

Remark 2. In Algorithm 2, the basis of eigenvectors remains the same for the complete set of iterations, and is extracted from the initial model. Only the number of vectors taken for the representation, \( N_i \), changes. Namely, from \( N_1 \) to \( N_2 > N_1 \), the decomposition using \( N_2 \) still has the same \( N_1 \) first eigenvectors in its representation (with different weights \( \alpha \)), and additional \( (N_2 - N_1) \) eigenvectors.
Figure 1. Seismic velocity models used to illustrate the eigenvector decomposition. They are both represented with $921 \times 301$ nodal coefficients.

Table 2. Minimal relative error obtained and associated scaling coefficient: $\mathcal{E} (\beta)$, for the decomposition of the Marmousi model Fig. 1(a). The definition of $\eta$ is given Table 1.

| Coeff. | $N = 10$ | $N = 20$ | $N = 50$ | $N = 100$ | $N = 250$ | $N = 500$ |
|--------|---------|---------|---------|---------|---------|---------|
| $\eta_1$ | $6\% \times 10^{-6}$ | $5\% \times 10^{-6}$ | $4\% \times 10^{-6}$ | $4\% \times 10^{-6}$ | $3\% \times 10^{-6}$ | $3\% \times 10^{-7}$ |
| $\eta_2$ | $14\% \times 5.10^{-5}$ | $13\% \times 5.10^{-5}$ | $12\% \times 5.10^{-5}$ | $9\% \times 10^{-5}$ | $7\% \times 10^{-5}$ | $5\% \times 10^{-5}$ |
| $\eta_3$ | $8\% \times 10^{-5}$ | $7\% \times 10^{-5}$ | $6\% \times 10^{-5}$ | $5\% \times 10^{-5}$ | $5\% \times 10^{-5}$ | $5\% \times 10^{-5}$ |
| $\eta_4$ | $14\% \times 5.10^{-4}$ | $14\% \times 10^{-4}$ | $13\% \times 10^{-4}$ | $13\% \times 10^{-4}$ | $12\% \times 5.10^{-4}$ | $10\% \times 5.10^{-4}$ |
| $\eta_5$ | $13\% \times 10^{-3}$ | $12\% \times 10^{-3}$ | $12\% \times 10^{-3}$ | $10\% \times 10^{-3}$ | $10\% \times 10^{-3}$ | $6\% \times 10^{-3}$ |
| $\eta_6$ | $8\% \times 10^{-3}$ | $7\% \times 10^{-3}$ | $6\% \times 10^{-3}$ | $5\% \times 10^{-3}$ | $5\% \times 10^{-3}$ | $4\% \times 10^{-3}$ |
| $\eta_7$ | $14\% \times 5.10^{-2}$ | $13\% \times 5.10^{-2}$ | $12\% \times 5.10^{-2}$ | $11\% \times 5.10^{-2}$ | $9\% \times 5.10^{-2}$ | $7\% \times 5.10^{-2}$ |
| $\eta_8$ | $15\% \times n/a$ | $14\% \times n/a$ | $13\% \times n/a$ | $12\% \times n/a$ | $10\% \times n/a$ | $9\% \times n/a$ |
| $\eta_9$ | $14\% \times n/a$ | $14\% \times n/a$ | $14\% \times n/a$ | $13\% \times n/a$ | $12\% \times n/a$ | $11\% \times n/a$ |

Table 3. Minimal relative error obtained and associated scaling coefficient: $\mathcal{E} (\beta)$, for the decomposition of the SEAM model Fig. 1(b). The definition of $\eta$ is given Table 1.

| Coeff. | $N = 10$ | $N = 20$ | $N = 50$ | $N = 100$ | $N = 250$ | $N = 500$ |
|--------|---------|---------|---------|---------|---------|---------|
| $\eta_1$ | $4\% \times 10^{-6}$ | $3\% \times 10^{-6}$ | $3\% \times 10^{-6}$ | $2\% \times 10^{-6}$ | $2\% \times 10^{-6}$ | $2\% \times 10^{-7}$ |
| $\eta_2$ | $8\% \times 5.10^{-5}$ | $6\% \times 5.10^{-5}$ | $5\% \times 10^{-5}$ | $3\% \times 10^{-5}$ | $2\% \times 10^{-5}$ | $2\% \times 10^{-5}$ |
| $\eta_3$ | $6\% \times 10^{-5}$ | $5\% \times 10^{-5}$ | $4\% \times 10^{-5}$ | $3\% \times 10^{-5}$ | $2\% \times 10^{-5}$ | $2\% \times 10^{-5}$ |
| $\eta_4$ | $10\% \times 10^{-4}$ | $9\% \times 10^{-4}$ | $7\% \times 10^{-4}$ | $6\% \times 10^{-4}$ | $4\% \times 10^{-4}$ | $4\% \times 10^{-4}$ |
| $\eta_5$ | $7\% \times 10^{-4}$ | $6\% \times 10^{-4}$ | $4\% \times 10^{-4}$ | $3\% \times 10^{-4}$ | $3\% \times 10^{-4}$ | $2\% \times 10^{-4}$ |
| $\eta_6$ | $6\% \times 10^{-4}$ | $5\% \times 10^{-4}$ | $4\% \times 10^{-4}$ | $3\% \times 10^{-4}$ | $2\% \times 10^{-4}$ | $2\% \times 10^{-4}$ |
| $\eta_7$ | $8\% \times 5.10^{-3}$ | $6\% \times 5.10^{-3}$ | $5\% \times 10^{-3}$ | $4\% \times 5.10^{-3}$ | $3\% \times 5.10^{-3}$ | $2\% \times 5.10^{-3}$ |
| $\eta_8$ | $18\% \times n/a$ | $17\% \times n/a$ | $14\% \times n/a$ | $12\% \times n/a$ | $8\% \times n/a$ | $7\% \times n/a$ |
| $\eta_9$ | $12\% \times n/a$ | $10\% \times n/a$ | $7\% \times n/a$ | $7\% \times n/a$ | $5\% \times n/a$ | $4\% \times n/a$ |

4 ILLUSTRATION OF MODEL DECOMPOSITION

First, we illustrate the eigenvector model decomposition with two geophysical media in two dimensions:

(i) the Marmousi velocity model, which consists in structures and faults, of size $9.2 \times 3$ km, see Fig. 1(a);

(ii) a velocity model extracted from the SEAM (SEG Advanced Modeling Program) Phase I benchmark, which consists in a salt dome (an object of high contrast velocity) and subsalt layer structures, of size $17.5 \times 3.75$ km, see Fig. 1(b).

In order to produce consistent comparisons, both models are represented on a structured nodal grid of size $n_x \times n_z = 921 \times 301$, that is with 277 221 coefficients. We perform the decomposition of the models by application of Algorithm 1, and steps (11), (12) and (13) (and we recall that we use the linear PDE problem). We study the main parameters of the decomposition: the number of eigenvectors $N$, the choice of $\eta$ and $\beta$ (see Table 1). The accuracy of the decomposition is estimated using the $L^2$ norm of the relative difference between the decomposition and the original representation such that

$$\mathcal{E} = 100 \frac{\|m - \mathbf{m}\|}{\|m\|}, \quad \text{Relative Error (per cent)},$$

where $m$ is the original model (Fig. 1) and $\mathbf{m}$ the decomposition using the basis of eigenvectors from Algorithm 1.

We decompose the SEAM and Marmousi models using the nine possibilities for $\eta$, that are given in Table 1. For the choice of scaling coefficient $\beta$ (which does not affect $\eta_k$ and $\eta_9$), we roughly cover an interval from $10^{-7}$ up to $10^6$, namely: $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 5 \times 10^{-2}, 10^{-1}, 5 \times 10^{-1}, 1, 5, 10, 10^2, 10^3, 10^4, 10^5, 10^6\}$. In Tables 2 and 3, we show the best relative error (i.e. minimal value) obtained for the Marmousi model of Fig. 1(a) and the SEAM model of Fig. 1(b). We test all choices of $\eta$ and values of $N$ between 10 (coarse) and 500 (refined). The corresponding values of the scaling parameter $\beta$ which gives the best (i.e. the minimal) error are also given in parenthesis.

1see https://wiki.seg.org/wiki/Open_data.
Figure 2. Decomposition of the Marmousi and SEAM velocity models of Figs 1(a) and (b) using $N = 50$ eigenvectors, following Algorithm 1. The relative error is computed from (19) for two selected formulations of $\eta$ (see Table 1) and different scaling parameters $\beta$.

Figure 3. Decomposition of the Marmousi and SEAM model of Fig. 1 using the basis $\psi(m^\dagger, \eta, N)$. As expected, we observe that the more eigenvectors are chosen (higher $N$), the better will be the decomposition. When using 500 eigenvectors, which represents about 2 per cent of the original number of coefficients ($921 \times 301$ in Fig. 1), the error is of a few per cent only. This can be explained by the redundancy of information provided by the original fine grid where the model is represented. Comparing the methods and models, we see that

(i) the Marmousi model (Table 2) is harder to retrieve than the SEAM model (Table 3) as it gives higher errors. In particular for low $N$, the SEAM model can be acutely decomposed (possibly 4 per cent error with $N = 10$).

(ii) For both models, it appears that four methods stand out: $\eta_1$ (Perona–Malik), $\eta_3$ (Geman & Reynolds 1992), $\eta_5$ (Charbonnier et al. 1994) and $\eta_6$ (Lorentzian), with a slight advantage towards $\eta_1$.

(iii) The scaling coefficient that minimizes the error may vary with respect to $N$, but remains with a similar amplitude. When changing the model, it may also require the modification of $\beta$.

To investigate further the last point, we show the evolution of relative error $E$ with respect to the scaling coefficient $\beta$ for the decomposition of the Marmousi and SEAM models in Fig. 2, where we compare two selected formulations for $\eta$. We observe some flexibility in the choice of $\beta$ that gives an accurate decomposition but it still appears relatively model-dependent.

We then picture the resulting images obtained after the decomposition of both models. For the sake of clarity, we only show two formulations in Fig. 3. The pictures illustrate that the decomposition of the SEAM model necessitates less eigenvectors than the Marmousi one which, in addition, is not accurately represented depending on the choice of $\eta$ while the upper boundary of the SEAM salt dome is always well captured.

5 EXPERIMENTS OF RECONSTRUCTION WITH FWI

In this section, we perform seismic imaging following the FWI Algorithm 2 for the identification of the subsurface physical parameters. We focus on media encompassing salt domes, as we have shown that the decomposition is more appropriate, and it also gives a more challenging configuration in seismic applications. We follow a seismic context, where the forward problem is given by (1) and the data are restricted to be acquired near the surface. The challenges of our experiments is threefold: (i) the recovery of salt domes is recognized to be difficult
Figure 4. Data associated with a single centered source. The data are first generated in the time-domain, then we incorporate white noise and proceed to the Fourier transform. In this experiment, the complete seismic acquisition is composed of 91 independent sources and 183 receivers for each source.

Figure 5. Target model and starting model for FWI. The models are of size 9.2 km × 3 km. The initial model corresponds to a 1-D variation of low velocity due to the change of kinematics (Farmer et al. 1996; Virieux & Operto 2009; Faucher et al. 2019a); (ii) we consider an initial guess that has no information on the subsurface structures and where the background velocity amplitude is incorrect and (iii) we avoid the use of the (unrealistically) low frequencies (below 2 Hz in exploration seismic).

5.1 Reconstruction of 2-D salt model
We first consider a 2-D salt model of size 9.2 × 3 km, which consists in three domes, see Fig. 5(a). We generate the data using 91 sources and 183 receivers (i.e. data points) per source. Both devices are located near the surface: the sources are positioned on a line at 20 m depth and the receivers at 80 m depth. In order to establish a realistic situation despite having a synthetic experiment, the data are generated in the time-domain and we incorporate white noise in the measurements. The level of the signal to noise ratio in the seismic trace is of 10 dB, the noise is generated independently for all receivers record associated with every source. Then we proceed to the discrete Fourier transform to obtain the signals to be used in the reconstruction algorithm. In Fig. 4, we show the time-domain data with noise and the corresponding frequency data for one source, located at 20 m depth, in the middle of the x-axis.

For the reconstruction of the salt dome model, the starting and true model are given in Fig. 5. We do not assume any a priori knowledge of the contrasting object in the subsurface, and start with a 1-D variation, which has a drastically lower amplitude (i.e. the background velocity is unknown).

5.1.1 Fixed decomposition, single frequency reconstruction
We first only use 2 Hz frequency data, and perform 180 iterations for the minimization. In Fig. 6, we show the reconstruction where the decomposition has not been used, that is the model representation follows the original piecewise constant decomposition of the model (one value per node). In Fig. 7, we compare the reconstruction using Algorithm 2 for the different formulations of η given in Table 1, using a fixed \( N = 50 \). We use \( N = 100 \) for Fig. 8. For the sake of clarity, we focus on \( \eta_3 \) (the most effective formulation), and \( \eta_8 \) (which relates to the Total Variation regularization).

We observe that

(i) the traditional FWI algorithm (without decomposition), see Fig. 6, fails to recover any dome. It only shows some thin layers of increasing velocity, with amplitudes much lower than the original ones.
Figure 6. Reconstruction of the salt velocity model from the starting medium Fig. 5(b) using 2 Hz frequency data. The reconstruction does not apply eigenvector decomposition. The model is parametrized following the domain discretization, using piecewise constant representation with one value per node on a 921 × 301 grid.

Figure 7. Reconstruction of the salt velocity model from the starting medium Fig. 5(b) using 2 Hz frequency data. The eigenvector decomposition uses $N = 50$ and the formulations of $\eta$ from Table 1. It uses the same color scale as Fig. 5.

Figure 8. Reconstruction of the salt velocity model from the starting medium Fig. 5(b) using 2 Hz frequency data. The eigenvector decomposition uses $N = 100$.

(ii) The decomposition using $N = 50$ is able to discover the largest object with formulation $\eta_3$, which recovers the three domes; while $\eta_1$, $\eta_5$ and $\eta_8$ show artifacts in the lower right corner. The other decompositions fail. We note that, due to the lack of velocity background information, the positions of the domes are slightly above the correct ones to compensate for the low travel times.

(iii) In this experiment, the method behaves much better with a restrictive number of eigenvectors. With $N = 100$ (Fig. 8), the iterative reconstruction only results in artifacts. The restrictive number of eigenvectors provides a regularization of the problem by reducing the number of parameters, which is crucial. For instance, the stability is known to deteriorate exponentially with the number of parameters in the piecewise constant case (Beretta et al. 2016).

Opposite to the decomposition of images (Section 4), the quantitative reconstruction using a model represented with a basis of eigenvectors from a diffusion PDE shows drastic differences between the formulations, where the procedure can fail depending on the choice of $\eta$. In addition, the number of eigenvectors for the representation has to be carefully selected, see Subsection 5.3.
Table 4. List of experiments for the reconstruction of the 2-D salt dome model (Fig. 5b). For each combination of frequency and associated number of eigenvectors $N$ in the decomposition, 30 iterations are performed ($n_{\text{iter}}$ in Algorithm 2).

| Experiment  | Frequency List | List of $N$ | Total Iterations | Reference |
|-------------|----------------|-------------|------------------|-----------|
| Experiment 1 | 2 Hz           | 50          | 180              | Fig. 7    |
| Experiment 2 | 2 Hz           | $\{50, 60, 70, 80, 90, 100\}$ | 180              | Fig. 9    |
| Experiment 3 | $\{2, 3, 4, 5\}$ Hz | 50          | 120              | Fig. 9    |
| Experiment 4 | $\{2, 3, 4, 5\}$ Hz | $\{50, 60, 70, 80\}$ | 120              | Fig. 9    |

Figure 9. Results of Experiments 2, 3 and 4 of Table 4 for the reconstruction of the salt velocity model Fig. 5.

5.1.2 Experiments with increasing $N$ and multiple frequencies

We investigate the performance of the eigenvector decomposition for multiple frequency data, and with progressive evolution of the number of eigenvectors in the representation $N$. We have a total of four different experiments, which are summarized in Table 4. The reconstructions, for $\eta_3$ and $\eta_8$, are shown in Fig. 9.

From these experiments using multiple $N$ and/or frequency contents, we observe the performance of the method. The best results are obtained using a single 2 Hz frequency with either progression of increasing $N$ or constant $N$: Experiments 2 and 1. The progression of $N$, Experiment 2, appears the most robust.

In this experiment, using multiple frequencies does not improve the results. It is probably due to the lack of knowledge of the velocity background which prevents us from recovering finer scale (i.e. kinematic error, Bunks et al. 1995). In particular, local minima in the misfit functional become more and more prominent in the high-frequency regime (Bunks et al. 1995; Barucq et al. 2019b). We illustrate the performance of the reconstruction with the results in the data-space: In Fig. 10, we show the time-domain seismograms for the true, starting and reconstructed velocity models. We observe that when we filter out frequencies above 2 Hz (first line of Fig. 10), the trace from the reconstructed model is indeed very similar to the measured one. However, when encompassing all frequency contents (bottom line of Fig. 10),
Eigenvector models for inverse problem

Figure 10. Comparison of the time-domain seismic traces for a central shot using different velocity models of the salt medium reconstruction experiment.

Figure 11. Comparisons of the frequency-domain data (pressure field) at the receivers location for a central source. We compare the data obtained from the target wave speed with salt of Fig. 5(a), the starting wave speed of Fig. 5(b) and the reconstruction of Fig. 9(a). Those three models are respectively denoted $c_1^t$, $c_{\text{start}}$ and $c_{\text{rec}}$.

important differences arise, in particular, one can see the travel time of the first reflection which is earlier with the recovered model. This indicates that the location of the salt in the reconstructed velocity is above its ‘true’ position.

Then, while we incorporate the higher frequency in the minimization procedure, the FWI is not amenable to improve the results (see Fig. 9) and it is most likely due to the missing velocity background which is not improved during the first iterations, and still missing. In Fig. 11, we show the frequency-domain data at 2 and 4 Hz: the observed data at 2 Hz are accurately obtained with the model reconstructed with the decomposition in eigenvectors, which confirms the pertinence of the method. Interestingly, at 4 Hz, while the frequency is not even used in the inversion scheme (we only use 2 Hz for Fig. 9a), we already have a good correspondence near the source and only the parts further away show a shift. In order to overcome the issue of recovering the background velocity, one would need lower frequency content, or one could try alternative strategies, such as the MBTTT (migration based travel time) method, which parametrizes the inverse problem with the background velocity model and the reflectivity (Clement et al. 2001). Here, the decomposition in eigenvectors appropriately recover the reflectivity part (better than traditional FWI), but the background model remains missing.
5.1.3 On the choice of the number of eigenvectors

We have shown in Figs 7 and 8 that one should take an initial \( N \) relatively low for the reconstruction algorithm to succeed. It remains to verify if the appropriate \( N \) can be selected ‘a priori’, or based upon minimal experiments. In Fig. 12(a), we show the evolution of the misfit functional with 30 iterations, for different values of \( N \), from 10 to 250. We compare, in Fig. 12(b), with the progression of \( N \), which follows Experiment 2 of Table 4.

From Fig. 12(a), we see that all of the choices of \( N \) have the same pattern: first the decrease of the functional and then its stagnation. We notice that the good choice for \( N \) is not reflected by the misfit functional. Indeed, it shows lower error for larger \( N \), while they are shown to result in erroneous reconstructions (Fig. 8 compared to Fig. 7). It is most likely that using larger \( N \) leads to local minima and/or deteriorates the stability (see, for the piecewise constant case, Beretta et al. (2016)). It results in the false impression (from the misfit functional point of view) that it would improve the efficiency of the method. Using a progression of \( N \), Fig. 12(b), eventually gives the same misfit functional value than the large \( N \), but it needs more iterations. This increase of iterations and ‘slow’ convergence is actually required, because it leads to an appropriate reconstruction (see Fig. 9).

Therefore, we cannot anticipate a good choice for \( N \) a priori (with a few evaluation of the misfit functional). the guideline we propose, as a safe and robust approach, is the progression of increasing \( N \), from low to high: it costs more in terms of iterations, but it converges properly.

5.2 Reconstruction of the SEAM Phase I model

We now consider the recovery of the SEAM Phase I model, which is expected to be more challenging as it contains both a salt-dome and subsalt layers. The starting model for the reconstruction is shown in Fig. 13, where, for the sake of clarity, we also picture the reference model which was previously used for the decomposition. This medium is of size 17.5 \( \times \) 3.75 km and the starting model we use is a smooth version of the reference one, where the contrasting objects and layers are missing.

We follow the same configuration as in the previous experiment: the data are generated in the time-domain, and noise is incorporated to the seismograms using a signal-to-noise ratio of 10 dB. Next, we proceed with the discrete Fourier-transform to use the iterative procedure with time-harmonic waves. In this experiment, the smallest available frequency is 2 Hz.

For the sake of conciseness, we only present the reconstruction results with the representation using \( \eta_3 \) (which was the most efficient). We follow a slow increase of \( N \) in a fixed basis (analogous to Experiment 2 of Table 4) which was the more stable approach. The reconstruction using 2 Hz is shown in Fig. 14.

We observe that the standard FWI algorithm gives artifacts over the medium, with oscillatory patterns, in particular on the sides. On the other hand, the reconstruction using the representation based upon the eigenvector decomposition is stable, and is able to capture accurately the upper boundary of the salt dome. Because of the frequency used, only the long wavelengths are recovered at this stage. In Fig. 15, we...
Figure 14. Reconstruction of the SEAM phase I model from an initially smooth model, see Fig. 13, using 2 Hz frequency data.

Figure 15. Comparisons of the frequency-domain data (pressure field) at the receivers location for a centrally located source at 2 Hz frequency for the SEAM Phase 1 wave speed of Fig. 13(a), the starting wave speed of Fig. 13(b) and the reconstruction using the eigenvector decomposition, Fig. 14(b). Those three models are respectively denoted $c^\dagger$, $c_{\text{start}}$ and $c_{\text{rec}}$.

Figure 16. Reconstruction of the SEAM phase I model starting from the reconstruction obtained with eigenvector decomposition of Fig. 14(b), using data of frequency from 2 to 10 Hz, with 30 iterations per (sequential) frequency.

Further illustrate the recovery by showing the frequency-domain data at 2 Hz frequency using the different wave speed models. We can see that the data from the starting model are out of phase as soon as we move away from the source, with cycle-skipping effects. However, the reconstruction using the eigenvector decomposition is able to retrieve this information and accurately capture the oscillations of the signal, and only the amplitude is inaccurate.

We now continue the procedure using increasing frequencies. Because of the smoothing effect of the decomposition, we use the algorithm without the eigenvector representation (alternatively, one could use the decomposition but with large $N$). Therefore, the velocity model obtained from the FWI with eigenvector decomposition of Fig. 14(b) is used as an initial model for restarted FWI with multiple frequencies, from 2 to 10 Hz [we use the sequential progression advocated by Faucher et al. (2019a)]. Eventually, the reconstruction after 10 Hz iterations is pictured in Fig. 16.

The reconstruction is able to capture the finer details of the velocity model: the salt dome is clearly defined, and the subsalt layer starts to appear. Therefore, the eigenvector decomposition method can also serve to build initial models for the FWI algorithm, where it appears as an
interesting alternative to overcome the lack of low frequency. To illustrate the different steps of the reconstruction, we show the time-domain seismograms at the different stages of the reconstruction in Fig. 17 (while the 2 Hz frequency-domain data are given in Fig. 15). We compare the traces resulting from the initial and true wave speed models, from the partial reconstruction obtained with the eigenvector decomposition (Fig. 14b), and from the final reconstruction (Fig. 16).

The trace that uses the starting model mainly shows the first arrivals, with some minor reflections coming from the smoothing of the salt dome. The one using the partial reconstruction with eigenvector decomposition and only the 2 Hz data accurately resolves the main multiple reflections between the salt upper boundary and the surface (the thick line in the center of the red ellipses in Fig. 17) but it misses the events of smaller importance. Eventually, the final reconstruction, that uses up to 10 Hz data contents is able to reproduce some of the events of smaller amplitudes. We also note that the amplitude of the trace for the final reconstruction is slightly high, which indicates that the contrasts are even too strong.

5.3 Implementation of the method

The eigenvector decomposition for the model representation depends on different choices of parameters, and it is not trivial to efficiently implement the method in the iterative reconstruction procedure. From the experiments we have carried out, we propose the following strategy.

(i) Regarding the choice of weighting coefficient, \( \eta_3 \) from Geman & Reynolds (1992) is the most efficient in these applications, and supersedes the standard Total Variation approach (i.e. \( \eta_8 \)).

Secondly, the difficulty resides in the number of eigenvectors to take for the decomposition: \( N \). More important, it appears that the misfit functional does not provide us with a good indication (see Fig. 12).

(i) The number of eigenvector \( N \) for the decomposition should initially takes a low value, and progressively evolves to higher values with iterations. It may not be the fastest convergence, but it is the most robust approach.

Finally, the reconstruction can serve as an initial model for multifrequency data:

(i) the (partial) reconstruction with eigenvector decomposition is used as a starting model for multifrequency algorithm. It allows the recovery of the finer details, which depend on the smaller wavelengths and where the smoothing effect is misleading.

Namely, the decomposition is particularly efficient to overcome the lack of low-frequency in the data.
Figure 18. 3-D model incorporating contrasting objects. The domain is of size $2.46 \times 1.56 \times 1.2$ km. We highlight a horizontal section at 550 m depth and vertical section at $y = 670$ m.

Figure 19. Initial model taken for the reconstruction of the 3-D medium with vertical section at $y = 670$ m. It consists in a 1-D variation in depth.

Figure 20. Time-domain data and corresponding Fourier transform at 5 Hz frequency. The 3-D trace corresponds with the evolution receivers recordings (positioned on a 2D map in the $x$-$y$ plane) with time. There are 1000 data points per time step (i.e. 1000 receivers on the domain) and we highlight sections at fixed time (0.5 s) and for a line of receivers (positioned at $y = 710$ m).

5.4 3-D experiment

The method extends readily for 3-D model reconstruction, simply incurring a larger computational cost (as larger matrices are involved for the eigenvector decomposition and the forward problem discretization). We proceed with a 3-D experiment, where we consider a subsurface medium of size $2.46 \times 1.56 \times 1.2$ km, encompassing several salt domes, illustrated in Fig. 18. The seismic acquisition consists in 96 sources, positioned on a 2-D plane at 10 m depth; one thousand receivers are positioned at 100 m depth. Similar to the previous experiments, the data are first generated in the time-domain and we incorporate noise before we proceed to the Fourier transform. Fig. 20 shows the time-domain data associated with a centrally located source, and the corresponding Fourier transform at 5 Hz frequency. For the reconstruction, we start with a 1-D variation, in depth only, where none of the objects is intuited, see Fig. 19, and the velocity background is incorrect.

In Fig. 21, we show the reconstruction without using the eigenvector decomposition, where the wave speed has a piecewise constant representation on a $124 \times 79 \times 61$ nodal grid. We only use 5 Hz frequency data, and 30 iterations. Next, we use the eigenvector model
Figure 21. Velocity model reconstruction with vertical section at \( y = 670 \) m, from the starting medium Fig. 19 using 5 Hz frequency data. The reconstruction does not apply the eigenvector decomposition. The model is parametrized following the domain discretization, using piecewise constant representation with one value per node on a \( 124 \times 79 \times 61 \) grid.

Figure 22. Velocity model reconstruction with vertical section at \( y = 670 \) m and horizontal section at 550 m depth, from the starting medium Fig. 19 and using 5 Hz frequency data. The reconstruction applies the eigenvector decomposition with \( \eta_3 \) and progression of \( N = \{20, 30, 50, 75, 100\} \), with 30 iterations per \( N \), see Algorithm 2.

For visualization, we focus on the 2-D vertical and horizontal sections which illustrate the positions and shapes of the objects. This experiment is consistent with the 2-D results: the classical FWI reconstruction fails to discover the subsurface objects, with only a narrow layer, misplaced (see Fig. 21). However, the reconstruction with the eigenvector model representation is able to accurately capture the largest subsurface salt domes (see Fig. 22).

6 CONCLUSION

In this paper, we have investigated the use of a representation based upon a basis of eigenvectors for image decomposition and quantitative reconstruction in the seismic inverse problem, using 2-D and 3-D experiments. We have implemented several diffusion coefficients, and compared their performance, depending on the target medium.

(i) In the context of image decomposition, the case of contrasting objects (salt domes) is clearly more appropriate than the layered media (such as Marmousi). All of the diffusion coefficients behave well and provide satisfactory results for salt domes image decomposition, even in the presence of noise.

(ii) For the decomposition of images with layered patterns, only a few formulations perform well \((\eta_1, \eta_3, \eta_6)\). It would be interesting to investigate further the performance of anisotropic or directional diffusion coefficients, mentioned by Weickert (1998) and Grote & Nahum (2019).

Next, we have considered the quantitative reconstruction procedure in seismic, where only partial, backscattered (i.e. reflection) data are available, from one side illumination. We have probed the performance of the method by considering initial guesses with minimal \textit{a priori} information, and by avoiding the low frequency data, which are not accessible in seismic applications. In this context, the FWI algorithm based upon the eigenvector model representation has shown promising results of subsurface 3-D salt dome media. The method only requires the preliminary computation of the basis of eigenvectors associated with the diffusion operator, and a trivial modification of the gradient computation. Namely, the overall additional cost of the method remains marginal compared to the cost of FWI. Our findings are the following.

(i) For reconstruction, the result depends on the choice of diffusion coefficient. We recommend \( \eta_3 \), from Geman & Reynolds (1992), which was the most robust in our applications, even with a fixed \( N \) and a few iterations.
(ii) We have shown that the choice of $N$ is not trivial, and one cannot rely on the misfit functional evaluation. Therefore, we have proposed a progression of increasing $N$, which appears to stabilize the reconstruction.

(iii) Because the method has a smoothing effect, it focuses on the long wavelength structures. Thus, the reconstruction using the decomposition can serve as an initial model to iterate with higher frequency contents, in order to improve the resolution.

Following these analyses, some difficulties remain regarding the optimal choice of parameters. For instance, it is possible that $\eta$ has to be selected differently depending on the model (as illustrated with the Marmousi decomposition). Similarly, the scaling coefficient $\beta$ would affect the performance but the acceptable range appears, hopefully, quite large.

For future work, it seems that the lack of background velocity information would not be overcome by the decomposition, possibly resulting in artifacts. Therefore, we envision the use of multiple basis to parametrize the velocity (e.g. using the background/reflectivity decomposition idea of Clément et al. (2001) and Barucq et al. (2019b), with a dedicated smooth eigenvector basis to represent the background, and another to represent the reflectors).

Eventually, the method can readily extend to multiparameter inversion (e.g. for elastic medium), upon taking a separate basis per parameter (e.g. one for each of the Lamé parameters in linear elasticity). However, a more appropriate approach would be the use of joint-basis by, for example, considering a system of PDE instead of the scalar diffusion operator. We have in mind strategies such that joint-sparsity and tensor decomposition.

ACKNOWLEDGEMENTS

The authors would like to thank Prof Marcus Grote for thoughtful comments and discussions. The authors also thank the editor and the reviewers which have contributed to improve the quality of the paper. FF is funded by the Austrian Science Fund (FWF) under the Lise Meitner fellowship M 2791-N. The research of FF is also supported by the Inria–TOTAL strategic action DIP. OS is supported by the FWF, with SFB F68, project F6807-N36 (Tomography with Uncertainties). OS also acknowledges support from the FWF via the project I3661-N27 (Novel Error Measures and Source Conditions of Regularization Methods for Inverse Problems). HB acknowledges funding from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement Number 777778 (Rise action Mathrocks) and with the E2S–UPPA CHICKPEA project.

REFERENCES

Aghamiry, H.S., Gholami, A. & Operto, S., 2019. Implementing bound constraints and total-variation regularization in extended full-waveform inversion with the alternating direction method of multiplier: application to large contrast media, Geophys. J. Int., 218(2), 855–872.
Akcelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in Proceedings of the Supercomputing, ACM/IEEE 2002 Conference.
Alessandrini, G. & Vessella, S., 2005. Lipschitz stability for the inverse conductivity problem, Adv. Appl. Math., 35(2), 207–241.
Alessandrini, G., de Hoop, M.V., Gaburro, R. & Sincich, E., 2018. Lipschitz stability for a piecewise linear Schrödinger potential from local cauchy data, Asympt. Anal., 108(3), 115–149.
Alessandrini, G., De Hoop, M.V., Faucher, G., Gaburro, R. & Sincich, E., 2019. Inverse problem for the Helmholtz equation with Cauchy data: reconstruction with conditional well-posedness driven iterative regularization, ESAIM: M2AN, 53(3), 1005–1030.
Alvarez, L., Lions, P-L. & Morel, J.-M., 1992. Image selective smoothing and edge detection by nonlinear diffusion, II, SIAM J. Numer. Anal., 29(3), 845–866.
Amestoy, P.R., Duff, I.S., L’Excellent, J.-Y. & Koster, J., 2001. A fully asynchronous multifrontal solver using distributed dynamic scheduling, SIAM J. Matrix Anal. Appl., 23(1), 15–41.
Amestoy, P.R., Guermouche, A., L’Excellent, J.-Y. & Pralet, S., 2006. Hybrid scheduling for the parallel solution of linear systems, Parallel Comput., 32(2), 136–156.
Ammari, H., Seo, J.K. & Zhou, L., 2015. Viscoelastic modulus reconstruction using time harmonic vibrations, Math. Model. Anal., 20(6), 836–851.
Anderson, E. et al., 1999. LAPACK Users’ Guide, 3rd edn, Society for Industrial and Applied Mathematics.
Backus, G. & Gilbert, F., 1968. The resolving power of gross earth data, Geophys. J. Int., 16(2), 169–205.
Backus, G.E. & Gilbert, J., 1967. Numerical applications of a formalism for geophysical inverse problems, Geophys. J. Int., 13(1-3), 247–276.
Bamberger, A., Chavent, G. & Lailly, P., 1979. About the stability of the inverse problem in the 1-D wave equation, J. appl. Math. Opt., 5, 1–47.
Barucq, H., Faucher, F. & Pham, H., 2018. Localization of small obstacles from back-scattered data at limited incident angles with full-waveform inversion, J Comput. Phys., 370, 1–24.
Faucher, F., Chavent, G., Barucq, H. & Calandra, H., 2019a. A priori estimates of attraction basins for velocity model reconstruction by time-harmonic full-waveform inversion and data space reflectivity formulation, to appear in Geophysics, Research Report RR-9253, Magique 3D; Inria Bordeaux Sud-Ouest, Université de Pau et des Pays de l’Adour.
Barucq, H., Chavent, G. & Faucher, F., 2019b. A priori estimates of attraction basins for nonlinear least squares, with application to Helmholtz seismic inverse problem, Inverse Probl., 35, 115004, 30 pp.
Bednar, J.B., Shin, C. & Pyun, S., 2007. Comparison of waveform inversion, part 2: phase approach, Geophys. Prospect., 55(4), 465–475.
Beretta, E., de Hoop, M.V., Faucher, F. & Scherzer, O., 2016. Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates, SIAM J. Math. Anal., 48(6), 3962–3983.
Blanc-Féraud, L., Charbonnier, P., Aubert, G. & Barlaud, M., 1995. Nonlinear image processing: modeling and fast algorithm for regularization with edge detection, in Image Processing, 1995. Proceedings., International Conference on, Vol. 1, pp. 474–477, IEEE.
Bozdag, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, Geophys. J. Int., 185(2), 845–870.
Brandsberg-Dahl, S., Chemingui, N., Valenciano, A., Ramos-Martinez, J. & Qiu, L., 2017. Fwi for model updates in large-contrast media, Leading Edge, 36(1), 81–87.
Brossier, R., 2011. Two-dimensional frequency-domain visco-elastic full waveform inversion: parallel algorithms, optimization and performance, Comput. Geosci., 37(4), 444–455.
Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex on-shore structures by 2D elastic frequency-domain full-waveform inversion, Geophys. 74(6), WCC105–WCC118.
Brossier, R., Operto, S. & Virieux, J., 2010. Which data residual norm for seismic imaging of complex on-shore structures by 2D elastic frequency-domain full-waveform inversion, Geophys. Prospect., 58(4), WCC115–WCC120.
Brossier, R., Operto, S. & Virieux, J., 2019. Inverse problem for the Helmholtz equation with Cauchy data, ESAIM: M2AN, 53(4), 1085–1105.
Brossier, R., Operto, S. & Virieux, J., 2019b. A priori estimates of attraction basins for nonlinear least squares, with application to Helmholtz seismic inverse problem, Inverse Probl., 35, 115004, 30 pp.
Briggs, W.L., Henson, V.E. & McCormick, S.F., 2000. A multigrid tutorial, 2nd edn, Society for Industrial and Applied Mathematics.
Bruneau, D., Chavent, G. & Barucq, H., 2019. Inverse problem for the Helmholtz equation with Cauchy data, ESAIM: M2AN, 53(4), 1085–1105.
Burrus, C.S., Gopinath, R. & Guo, H., 1998. Introduction to wavelet and wavelets, Prentice Hall.
Burrus, C.S., Gopinath, R. & Guo, H., 2008. Introduction to wavelet and wavelets, 2nd edn, Prentice Hall.
Calandra, H., Amestoy, P.R., Duff, I.S.,高兴栋, M.V. and Koster, J., 2001. An asynchronous method for solving linear systems, in Proceedings of the 16th International Conference on Parallel Processing, 2001.
Calandra, H., Amestoy, P.R., Duff, I.S.,高兴栋, M.V. and Koster, J., 2001. An asynchronous method for solving linear systems, in Proceedings of the 16th International Conference on Parallel Processing, 2001.
Calandra, H., Amestoy, P.R., Duff, I.S.,高兴栋, M.V. and Koster, J., 2001. An asynchronous method for solving linear systems, in Proceedings of the 16th International Conference on Parallel Processing, 2001.
Robert, L. & Deriche, R., 1996. Dense depth map reconstruction: a minimization and regularization approach which preserves discontinuities, in *European Conference on Computer Vision*, pp. 439–451, Springer.

Rudin, L.I., Osher, S. & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Phys. D.* 60(1-4), 259–268.

Sigure, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, 69(1), 231–248.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49, 1259–1266.

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, 51(10), 1893–1903.

Tarantola, A., 1984. Sh-wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, 49(11), 1933–1942.

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

Vogel, C.R. & Oman, M.E., 1996. Iterative methods for total variation denoising, *SIAM J. Scient. Comput.*, 17(1), 227–238.

Wang, S., de Hoop, M.V. & Xia, J., 2011. On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver, *Geophys. Prospect.*, 59(5), 857–873.

Weickert, J., 1998. *Anisotropic Diffusion in Image Processing*, Vol. I, Teubner Stuttgart.

Yang, Y., Engquist, B., Sun, J. & Hamfeldt, B.F., 2018. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, *Geophysics*, 83(1), R43–R62.

Yuan, Y.O. & Simons, F.J., 2014. Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics*, 79(3), WA79–WA95.

Yuan, Y.O., Simons, F.J. & Bozdağ, E., 2015. Multiscale adjoint waveform tomography for surface and body waves, *Geophysics*, 80(5), R281–R302.

APPENDIX: NUMERICAL IMPLEMENTATION OF THE EIGENVECTOR DECOMPOSITION

In this appendix, we give the numerical details for the implementation of the decomposition in eigenvectors. Our code is developed in *Fortran90*, it uses both *mpi* and *OpenMP* parallelism and run on cluster for efficiency. The forward wave operator is discretized using a Finite Differences scheme, for example *Virieux* (1984), *Operto* et al. (2009) and *Wang* et al. (2011). The discretization of the Helmholtz operator generates a large sparse matrix, for which we use the direct solver *MUMPS* (Amestoy et al. 2010). A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophys. J. Int.*, 182(3), 1383–1394.

Virieux, J., 1984. Sh-wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, 49(11), 1933–1942.

Virieux, J., 1984. Sh-wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, 49(11), 1933–1942.

APPENDIX: NUMERICAL IMPLEMENTATION OF THE EIGENVECTOR DECOMPOSITION

In this appendix, we give the numerical details for the implementation of the decomposition in eigenvectors. Our code is developed in *Fortran90*, it uses both *mpi* and *OpenMP* parallelism and run on cluster for efficiency. The forward wave operator is discretized using a Finite Differences scheme, for example *Virieux* (1984), *Operto* et al. (2009) and *Wang* et al. (2011). The discretization of the Helmholtz operator generates a large sparse matrix, for which we use the direct solver *MUMPS* (Amestoy et al. 2010). A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophys. J. Int.*, 182(3), 1383–1394.

Virieux, J., 1984. Sh-wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, 49(11), 1933–1942.

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

Vogel, C.R. & Oman, M.E., 1996. Iterative methods for total variation denoising, *SIAM J. Scient. Comput.*, 17(1), 227–238.

Wang, S., de Hoop, M.V. & Xia, J., 2011. On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver, *Geophys. Prospect.*, 59(5), 857–873.

Weickert, J., 1998. *Anisotropic Diffusion in Image Processing*, Vol. I, Teubner Stuttgart.

Yang, Y., Engquist, B., Sun, J. & Hamfeldt, B.F., 2018. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, *Geophysics*, 83(1), R43–R62.

Yuan, Y.O. & Simons, F.J., 2014. Multiscale adjoint waveform-difference tomography using wavelets, *Geophysics*, 79(3), WA79–WA95.

Yuan, Y.O., Simons, F.J. & Bozdağ, E., 2015. Multiscale adjoint waveform tomography for surface and body waves, *Geophysics*, 80(5), R281–R302.

A1 Eigenvectors associated with the lowest eigenvalues

The Lanczos and Arnoldi methods are particularly efficient to compute the largest eigenvalues and associated eigenvectors of the matrix, and only require matrix vector multiplication. However, we are interested in the lowest eigenvalues for our decomposition. The idea is simply to use that the lowest eigenvalues of the discretized diffusion matrix, say $A$, are simply the largest eigenvalues of the matrix $A^{-1}$. Then, the matrix–vector multiplication, say $Av$ for a vector $v$, becomes a resolution of a linear system $A^{-1}v$. It may appear computationally expensive but it is not thanks to the use of the direct solver MUMPS (see above), which, once the factorization is obtained, is very efficient for the resolution procedure. Hence, the computation follows the steps:

(i) compute the (sparse) matrix discretization of the selected diffusion operator: $A$;

(ii) compute the factorization of the matrix $A$ using MUMPS,

(iii) use the package *ARPACK* to compute the largest eigenvalues of $A^{-1}$, by replacing the matrix-vector multiplication step in the iterations by the resolution of a linear system using MUMPS.

We have observed important reduction of time cost when allowing some flexibility in the accuracy with this threshold criterion. However, in the computational experiments, we do not use this option, as the numerical efficiency is not the primary objective of our study.

$ARPACK$ has the possibility to compute the smallest eigenvalues using matrix–vector multiplication, however, we have observed a drastic increase of the computational time compared to using the inverse matrix and resolution of linear system.

$2$The experiments have been performed on the cluster PlaFRIM (Plateforme Fédérative pour la Recherche en Informatique et Mathématiques, https://www.pl afrim.fr/fr, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d’Aquitaine).

$3$www.caam.rice.edu/software/ARPACK/, *ARPACK* uses sequential computation, hence, contrary to the rest of our code, this part does not use parallelism.

Future developments include the implementation of the parallel version of the package: *PARPACK*.

We have observed important reduction of time cost when allowing some flexibility in the accuracy with this threshold criterion. However, in the computational experiments, we do not use this option, as the numerical efficiency is not the primary objective of our study.

*ARPACK* has the possibility to compute the smallest eigenvalues using matrix–vector multiplication, however, we have observed a drastic increase of the computational time compared to using the inverse matrix and resolution of linear system.
Figure A1. Comparison of the computational time for the eigenvector decomposition for different $\eta$. The computation of the eigenvectors uses the discretized matrix of a diffusion operator of size $277,221 \times 277,221$. It corresponds with the decomposition of the model Fig. 5(a). Other formulations of $\eta$ from Table 1 are not shown for clarity, but follow the same pattern.

A2 Computation of the coefficients $\alpha$

The last step is to retrieve the appropriate coefficients $\alpha_k$ in (13) for the decomposition. It basically consists in the resolution of a dense linear system (from least squares method). We use LAPACK, Anderson et al. (1999) (contrary to MUMPS, LAPACK is adapted to dense linear system). Note that, because we usually consider a few hundreds of coefficients for the decomposition, this operation remains relatively cheap compared to the eigenvectors computation.

A3 Numerical illustration

We compare the computational time for the eigenvectors computation and model decomposition in Fig. A1 for different values of $N$ and $\eta$. We first note that the choice of $\eta$ does not really modify the computational time. Then, we see that the two operations are mostly linear in $N$, and that the time to solve the least squares problem with LAPACK is much smaller than the time to compute the eigenvectors with ARPACK, namely, hundred time smaller. For the largest case: $N = 1000$ for a squared matrix of size $277,221$, it takes about 30min to retrieve the eigenvectors, and $10$ s to compute the $\alpha$ (in our applications we usually take $N < 100$).