Seismic imaging of complex onshore structures by
two-dimensional elastic frequency-domain full-waveform
inversion

Romain Brossier*, Stéphane Operto† and Jean Virieux‡

*Géozur, Université Nice-Sophia Antipolis, CNRS, IRD, Observatoire de la Côte d’Azur,
Valbonne, France. E-mail: brossier@geoazur.unice.fr.

†Géozur, Université Nice-Sophia Antipolis, CNRS, IRD, Observatoire de la Côte d’Azur,
Villefranche-sur-mer, France. E-mail: operto@geoazur.obs-vlfr.fr

‡Laboratoire de Géophysique Interne et Tectonophysique, Université Joseph Fourier,
CNRS, IRD, Grenoble, France. E-mail: Jean.Virieux@obs.ujf-grenoble.fr.

(June 5, 2009)

Running head: 2D elastic waveform inversion

ABSTRACT

Quantitative imaging of the elastic properties of the subsurface at depth is essential for civil engineering applications and for oil and gas reservoir characterization. A realistic synthetic example provides for an assessment of the potential and limits of two-dimensional elastic full-waveform inversion of wide-aperture seismic data, for recovering high resolution P-wave and S-wave velocity models of complex onshore structures. Full-waveform inversion of land data is challenging, because of the increased non-linearity introduced by free-surface effects, such as the propagation of surface waves in the heterogeneous near-surface. Moreover, the short wavelengths of the shear wavefield require an accurate S-wave velocity starting model
if low frequencies are not available in the data. Different multiscale strategies with the aim of mitigating these non-linearities are here evaluated. Massively parallel full-waveform inversion is implemented in the frequency domain. The numerical optimization relies on a limited-memory quasi-Newton algorithm that out-performs the more classic preconditioned conjugate-gradient algorithm. The forward problem is based upon a discontinuous Galerkin method on triangular mesh, which allows accurate modeling of free-surface effects. Sequential inversions of increasing frequencies define the most natural level of hierarchy in the multiscale imaging. In the case of land data involving surface waves, the regularization introduced by hierarchical frequency inversions is not enough for adequate convergence of the inversion. A second level of hierarchy implemented with complex-valued frequencies is necessary, and provides convergence of the inversion towards acceptable P-wave and S-wave velocity models. Among the possible strategies for sampling frequencies in the inversion, successive inversions of slightly overlapping frequency groups proves to be the most reliable when compared with the more standard sequential inversion of single frequencies. This suggests that simultaneous inversion of multiple frequencies is critical when considering complex wave phenomena.
INTRODUCTION

Quantitative imaging of the elastic properties of the subsurface is essential for oil and gas reservoir characterization, and for the monitoring of CO$_2$ sequestration with time-lapsed acquisitions. Indeed, fluids and gas have significant effects on the elastic properties of the subsurface in terms of the Poisson-ratio anomalies. This quantitative imaging is also required for near-surface imaging in the framework of civil engineering applications, because the shear properties of the shallow weathered layers have a strong impact on the elastic wavefield. Moreover, at the near-surface scale, short propagation times do not easily allow the separation in time of the body waves and the surface waves. In this case, filtering or muting surface waves is not easy, and both types of waves need to be involved in the imaging, which requires solving the elastic-wave equation. Evolution of acquisition systems towards wide-aperture/wide-azimuth geometries and multi-component recordings is another motivation behind the development of elastic-imaging methods, because the occurrence of P-to-S mode conversion is dominant at wide-aperture angles.

Reservoir characterization is classically performed by amplitude-versus-offset analysis in the prestack domain (e.g., Jin et al., 2000; Buland and Omre, 2003). An alternative is the full-waveform inversion (FWI) of the elastic wavefield recorded by multiple components with the aim of reconstructing the P-wave and S-wave velocity ($V_P$ and $V_S$) models of the subsurface (or other related parameters, such as impedances, if density is involved in the inversion) with a resolution limit of the order of half a wavelength. The misfit between the recorded and the modeled wavefields is minimized through resolution of a numerical optimization problem (Tarantola, 1987; Nocedal and Wright, 1999). The FWI forward problem is based on the complete solution of the full (two-way) wave equation.

One drawback is that elastic FWI is a computationally challenging problem. Recent ad-
vances in high performance computing on large-scale distributed memory platforms allow
two-dimensional (2D) problems of representative sizes to be tackled, while the 3D problems
start to be investigated only in the acoustic case (Sirgue et al., 2007; Ben-Hadj-Ali et al.,
2008; Vigh and Starr, 2008; Warner et al., 2008). Moreover, the complexity of the elastic
wavefield at wide aperture makes the inverse problem highly non-linear and ill-posed. This
is even more dramatic for onshore applications where surface waves with high energy as
well as body waves can be used in the optimization process. FWI is conventionally solved
with local optimization (linearized) approaches, which makes the inversion sensitive to the
limited accuracy of the starting model. As such, it is essential to use realistic synthetic case
studies to investigate under which conditions the non-linearity of the elastic FWI can be
mitigated.

Most of the recent applications of FWI to real data have been performed under acoustic
approximation (Pratt and Shipp, 1999; Hicks and Pratt, 2001; Ravaut et al., 2004; Gao
et al., 2006; Operto et al., 2006; Bleibinhaus et al., 2007). Although reliable results can be
obtained with acoustic approximation if judicious data pre-processing and inversion precon-
ditioning is applied (Brenders and Pratt, 2007b), elastic FWI is desirable for applications
that involve the detection of fluids and gas, and for CO₂ sequestration. Moreover, acoustic
FWI can lead to erroneous models when applied to elastic data when the velocity models
show high velocity contrasts and when no specific FWI pre-processing and tuning is applied
to the data (Barnes and Charara, 2008).

Only a few applications of elastic FWI have been presented in the literature recently. In
the early applications, Crase et al. (1990, 1992) applied elastic FWI to real land and marine
reflection data using limited offsets. In this framework, the FWI was applied as quanti-
tative migration processing for imaging P and S impedances. With the benefit provided
by wide-aperture data to build the large and intermediate wavelengths of the subsurface recognized by Mora (1987, 1988), acoustic and elastic FWI evolved as an attempt to build high resolution velocity models. Shipp and Singh (2002) performed 2D time-domain FWI of a small wide-aperture marine data subset recorded by a long streamer (12-km long). Although the forward problem was solved using the elastic wave equation, only the $V_P$ parameter was reconstructed during FWI, assuming an empirical relationship between $V_P$ and $V_S$ and between $V_P$ and density. They designed a hierarchical multi-step approach based on layer-stripping, and offset and time windowing, where the aim was to mitigate the non-linearity of the inverse problem. Sears et al. (2008) designed a similar multi-step strategy to perform elastic time-domain FWI of multi-component ocean-bottom-cable data. Their strategy, was based on selection of data component, parameter class and arrival type (by time windowing), and it revealed itself especially useful when the amount of P-to-S conversion was small at the sea bottom, which makes the reconstruction of the $V_S$ model particularly ill-posed. Choi and Shin (2008) and Choi et al. (2008) applied elastic frequency-domain FWI to onshore and offshore versions of the synthetic Marmousi2 model (Martin et al., 2006), respectively. They successfully imaged the model using a velocity-gradient starting model and a very low starting frequency (0.16 Hz). Shi et al. (2007) applied elastic time-domain FWI to marine data collected from a gas field in western China. They successfully imaged a zone of Poisson-ratio anomalies associated with gas layers. Accurate starting $V_P$ and $V_S$ models were built from the P-wave and P-SV-wave velocity analysis and from a priori information of several well logs along the profile. Gelis et al. (2007) implemented a 2D elastic frequency-domain FWI using the Born and the Rytov approximations for the linearization of the inverse problem. They highlighted the dramatic footprint of the surface waves on the imaging of small inclusions in homogeneous background models. To mitigate
this footprint, they only involved body waves during the early stages of the inversion, by
only selecting short-offset traces.

The present study presents a 2D massively parallel elastic frequency-domain FWI algorithm
based on a discontinuous Galerkin (DG) forward problem (Brossier et al., 2008), and its
application to a realistic synthetic onshore case study. The aim of this application was
to assess whether surface waves and body waves recorded by wide-aperture acquisition ge-
ometries can be jointly inverted to build high resolution $V_P$ and $V_S$ of complex onshore
structures. We implement FWI in the frequency domain, which presents some distinct ad-
vantages with respect to the time-domain formulation (Pratt and Worthington, 1990; Pratt,
1999; Sirgue and Pratt, 2004). The inverse problem can be solved with a local optimiza-
tion approach using either a conjugate gradient or a quasi-Newton method (Nocedal and
Wright, 1999). The gradient of the objective function is computed with the adjoint-state
technique (Plessix, 2006). Successive inversions of increasing frequency provide a natural
framework for multiscale imaging algorithms. Moreover, by sacrificing the data redundancy
of multifold acquisitions, the inversion of a limited number of frequencies can be enough to
build reliable velocity models, provided the acquisition geometry spans over sufficiently long
offsets. This limited number of frequencies can be efficiently modeled in the 2D case for
multiple shots once the impedance matrix that results from discretization of the frequency-
domain wave equation has been factorized through a LU decomposition (Marfurt, 1984;
Nihei and Li, 2007). Finally, attenuation can be implemented in the forward problem in
a straightforward way, and without extra computational cost, by using complex velocities.

The main drawback of the frequency-domain FWI formulation arises from the difficulty of
time windowing of the modeled data when inverting one or a few sparsely sampled frequen-
cies at a time. Time windowing allows a selection of specific arrivals to be included in the
various stages of the inversion. A last resort is the use of complex-valued frequencies, which damp arrivals starting at a given traveltime (Shin et al., 2002).

In the next section of the present study, we briefly review the theory of frequency-domain elastic full-wavefield modeling and inversion. In the following section, we review several multiscale strategies to mitigate the non-linearity of the elastic inverse problem. These strategies involve two nested levels of hierarchy over frequencies and aperture angles in the inversion algorithm. The effectiveness of these strategies is first illustrated with a simple two-parameter elastic problem with a free surface. Then we have applied the elastic frequency-domain FWI algorithm to a realistic synthetic example, for the reconstruction of a dip section of the SEG/EAGE Overthrust model, assuming a constant Poisson ratio. The results of the different analyses show that simultaneous inversions of multiple frequencies and data preconditioning by time damping are critical to obtain reliable results when surface waves propagating in a heterogeneous near-surface are present in the elastic wavefield. We will also illustrate the improvements provided by quasi-Newton algorithms, compared to more conventional conjugate-gradient approaches.

METHOD AND ALGORITHM

Forward problem

Two-dimensional elastic frequency-domain FWI requires computation of the frequency solution of the elastic P-SV equations in heterogeneous media. We present a short review of the $P_0$ DG method used in this study. The reader is referred to Brossier et al. (2008) for more details.

We consider the first-order hyperbolic system where both particle velocities ($V_x,V_z$) and
stresses \((\sigma_{xx}, \sigma_{zz}, \sigma_{xz})\) are unknown quantities, as described by the system:

\[
\begin{align*}
-i\omega V_x &= \frac{1}{\rho(x)} \left\{ \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z} \right\} + f_x \\
-i\omega V_z &= \frac{1}{\rho(x)} \left\{ \frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z} \right\} + f_z \\
-i\omega \sigma_{xx} &= (\lambda(x) + 2\mu(x)) \frac{\partial V_x}{\partial x} + \lambda(x) \frac{\partial V_z}{\partial z} - i\omega \sigma_{xx0} \\
-i\omega \sigma_{zz} &= \lambda(x) \frac{\partial V_x}{\partial x} + (\lambda(x) + 2\mu(x)) \frac{\partial V_z}{\partial z} - i\omega \sigma_{zz0} \\
-i\omega \sigma_{xz} &= \mu(x) \left\{ \frac{\partial V_x}{\partial z} + \frac{\partial V_z}{\partial x} \right\} - i\omega \sigma_{xz0},
\end{align*}
\]

where the Lamé coefficients that describe the medium are denoted by \(\lambda\) and \(\mu\), the density by \(\rho\), and the angular frequency by \(\omega\). Source terms are either point forces \((f_x, f_z)\) or applied stresses \((\sigma_{xx0}, \sigma_{zz0}, \sigma_{xz0})\). \(i\) is the purely imaginary term defined by \(i = \sqrt{-1}\). Only isotropic media are considered in this study.

Equations 1 are discretized and solved with the DG method (Käser and Dumbser, 2006). We used low-order \(P_0\) interpolation, which corresponds to piecewise constant velocity and stress fields, and physical parameters in each cell. The DG method is applied to the weak formulation of the system 1, and the partial derivatives are computed through numerical fluxes exchanged at interfaces between cells. The perfectly-match layers (PML) method are used for absorbing boundary conditions along the edges of the model (Berenger, 1994).

An explicit free-surface boundary condition for arbitrary complex topographies is implemented on top of the model by canceling fluxes of normal stresses along the boundary that consist of edges of triangles. A sufficient level of accuracy for FWI can be obtained using the DG \(P_0\) method with ten to fifteen cells per minimum shear wavelength in regular equilateral meshes (Brossier et al., 2008). These regular meshes will be used for the simulations presented below. Note that extension of the DG method to higher order interpolations is required for modeling in unstructured meshes and for arbitrarily combining different inter-
polations during one simulation (e.g., Dumbser and Käser, 2006).

Discretization of equations 1 leads to solving the linear system:

$$\mathbf{A} \mathbf{v} = \mathbf{s},$$  \hspace{1cm} (2)

where the coefficients of the impedance matrix $\mathbf{A}$, namely, the forward modeling operator, depends on the modeled frequency and the medium properties. Vector $\mathbf{s}$ represents the source term. Vector $\mathbf{v}$ represents the unknowns for particle velocities and stress in each cell. Note that only the vertical and horizontal particle velocity wavefields will be inverted in this study.

**Inverse Problem**

FWI is an optimization problem which can be recast as a linearized least-squares problem that attempts to minimize the misfit between the recorded and the modeled wavefields (Tarantola, 1987). The inverse problem can be formulated in the frequency domain (Pratt and Worthington, 1990) and the associated objective function to be minimized is defined by:

$$\mathcal{C}^{(k)} = \frac{1}{2} \Delta \mathbf{d}^\dagger \mathbf{W}_d \Delta \mathbf{d} = \frac{1}{2} \Delta \mathbf{d}^\dagger \mathbf{S}_d^\dagger \mathbf{S}_d \Delta \mathbf{d},$$  \hspace{1cm} (3)

where $\Delta \mathbf{d} = \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{calc}}^{(k)}$ is the data misfit vector, the difference between the observed data $\mathbf{d}_{\text{obs}}$ and the modeled data $\mathbf{d}_{\text{calc}}^{(k)}$ computed in the model $\mathbf{m}^{(k)}$. Superscript $\dagger$ indicates the adjoint (transposed conjugate) and $\mathbf{S}_d$ is a diagonal weighting matrix applied to the misfit vector to scale the relative contributions of each of its components. $k$ is the iteration number.
The gradient $G^{(k)}$ of the objective function is given by:

$$G^{(k)} = R \left\{ J^t W_d \Delta d^* \right\}$$  

where $J$ is the Fréchet derivative matrix. Adjoint-state formalism allows efficient computation of the gradient with the back-propagation technique, without explicit computation of $J$ (Plessix, 2006). This leads to the following expression of the gradient with respect to the parameter $m_i$ (Pratt et al., 1998; Gelis et al., 2007):

$$G_{m_i}^{(k)} = R \left\{ v^t \left[ \frac{\partial A^t}{\partial m_i} \right] A^{-1} W_d \Delta d^* \right\},$$  

which shows that the gradient can be recast as a product between the incident wavefields $v$ and the back-propagated wavefields $A^{-1} W_d \Delta d^*$, using residuals at receiver positions as a composite source. Therefore, only two forward problems per shot are required for gradient building. In equation 5, we exploited the reciprocity of the Greens functions to remove the transpose operator in the expression of the back-propagated residuals ($A^{-1} W_d \Delta d^* = A^{-1} W_d \Delta d^*$).

The radiation pattern of the diffraction by the model parameter $m_i$ is denoted by $\partial A / \partial m_i$. The analysis of these radiation patterns suggests that $V_P$ and $V_S$ parameterization is that which is optimal for large diffraction angles (i.e., for wide-angle reflections), while the $I_p$ and $I_s$ impedances should provide a better decoupling between the two classes of parameters for small diffraction angles (i.e., short-angle reflections) (Tarantola, 1986).

A second-order Taylor expansion of the objective function provides the perturbation model $\delta \mathbf{m}$, which minimizes the objective function assumed to be locally parabolic, expressed as:

$$B^{(k)} \delta \mathbf{m} = -G^{(k)},$$  

where $B^{(k)}$ is the Hessian of the objective function. Due to the cost of the computation of $B^{(k)}$, Newton and Gauss-Newton methods are generally not considered for realistic
size problems (Pratt et al., 1998). Steepest-descent or conjugate-gradient methods preconditioned by the diagonal terms of an approximate Hessian are more conventionally used (Pratt et al., 1998; Operto et al., 2006). Shin et al. (2001) used the diagonal part of the pseudo-Hessian, a less computationally demanding approximation of the truncated Hessian. Accounting for the Hessian accelerates the convergence of the inversion and improves the resolution of the imaging by correctly scaling and deconvolving the gradient by geometrical amplitude and limited-bandwidth effects.

In the following, we use a quasi-Newton method for the FWI problem. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method is commonly used in numerical optimization to solve large-scale, non-linear problems (Nocedal, 1980). As indicated, L-BFGS is a limited memory version of the BFGS method, and it appears to be one of the most robust and efficient limited-memory quasi-Newton algorithms (Nocedal and Wright, 1999). The quasi-Newton L-BFGS(n) method estimates curvature information contained in the Hessian matrix from a limited number (n) of gradient difference vectors and model difference vectors associated with the n most recent iterations (n is usually chosen between 3 and 20). The iterative process is preconditioned by an initial guess of the Hessian, which is typically the diagonal terms of an approximate Hessian. Using the Sherman-Morrison-Woodburry formula (Nocedal and Wright, 1999), at each iteration the L-BFGS algorithm computes an improved estimation of the inverse $H^{(k)}$ of the Hessian matrix $B^{(k)}$. Therefore, resolution of the linear system 6 is avoided and the perturbation model is directly inferred from:

$$\delta m = -H^{(k)} \tilde{G}^{(k)}.$$  \hspace{1cm} (7)

The double-loop recursive algorithm designed by Nocedal (1980) does not explicitly build and store $H^{(k)}$, but directly computes the right-hand side of equation 7 with additions, differences and inner products of the vectors.
Finally, the model is updated with the perturbation vector:

$$m^{(k+1)} = m^{(k)} + \alpha \delta m,$$  \hspace{1cm} (8)

where \( \alpha \) denotes the step length, which minimizes the objective function. In this study, we estimate \( \alpha \) by parabola fitting. When the Hessian matrix is taken into account in the inversion, perturbation models associated with each parameter class computed with equation 7 are correctly scaled with respect to each other (Nocedal and Wright, 1999). Therefore, an estimation of only one step length \( \alpha \) is necessary to minimize the objective function for the multi-parameter classes.

A subspace method (Sambridge et al., 1991) has also been tested that led to similar results; however, this was more computationally intensive because it required extra forward problem simulations.

**Parallel implementation**

The massively parallel direct solver MUMPS (Amestoy et al., 2006) is based on a multifrontal method (Duff and Reid, 1983) and it is used to solve the linear system that results from the discretization of the forward problem (Equation 2). Parallel LU factorization of the \( A \) matrix allows the speeding up of the factorization by more than one order of magnitude, compared to sequential execution. Moreover, LU factors are stored in a distributed form over the in-core memory of the processors (Sourbier et al., 2009), making this quite efficient for solving larger problems without intensive I/O resources. After the substitution phase, the multiple solutions are distributed over the processors following the domain decomposition driven by the distribution of the LU factors. Equation 5 shows that the
gradient computation is a weighted product of the forward problem solutions; namely, the incident wavefields and the back-propagated residual wavefields. This product can be performed easily in parallel by assigning one processor to each subdomain. To improve the load balancing over the processors, MUMPS-distributed solutions are re-ordered with message passing interface (MPI) point-to-point communications before gradient computation, using a well-balanced mesh partitioning that is performed with METIS software (Karypis and Kumar, 1999). The gradient and the initial estimation of Hessian are therefore efficiently computed in parallel, before being centralized on the master processor for perturbation-model building.

FULL-WAVEFORM INVERSION DATA PRECONDITIONING AND MULTISCALE STRATEGIES

CPU-efficient frequency-domain FWI is generally carried out by successive inversions of single frequencies, by proceeding from the low frequencies to the higher ones (Pratt and Williamson, 1990; Sirgue and Pratt, 2004). This defines a multiresolution framework that helps mitigation of the non-linearity of the inverse problem associated with high frequency cycle-skipping artifacts. CPU-efficient algorithms can be designed by selecting a few coarsely sampled frequencies such that the wavenumber redundancy that results from dense sampling of frequencies and aperture angles is decimated. This strategy, which is referred to as the sequential inversion approach in the following, has proven to be effective for several applications of acoustic FWI (Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007a). However, it might lack robustness when complex wave phenomena are present. For example, reconstruction of a low velocity layer in the Overthrust model was improved when several frequencies were inverted simultaneously, rather than successively, during acoustic
FWI (Sourbier et al., 2009). More significant wave propagation effects are expected in elastic FWI because of conversions, dispersive surface waves, and frequency-dependent attenuation.

More robust, but more computationally expensive, multiscale FWI schemes should be designed by partially preserving the redundancy of multifold seismic data. The first scheme, referred to as the Bunks approach in the following, is an adaptation in the frequency domain of the time-domain approach of Bunks et al. (1995). It consists of successive inversions of overlapping frequency groups. The first group contains only the starting frequency, and one higher frequency is added from one group to the next. The second approach, which is referred to as the simultaneous inversion approach in the following, consists of successive inversions of slightly overlapping frequency groups. The main difference from the Bunks approach is that several frequencies are simultaneously inverted at each inversion stage, and that overlapping between two next-frequency groups is minimized. Definition of the frequency bandwidth of each group should be driven by the best compromise between the need to avoid high frequency cycle-skipping artifacts, and the need to simultaneously invert as many frequencies as possible to stabilize the inversion.

Non-linearity of FWI can also be efficiently mitigated by selecting a subset of specific arrivals (i.e., early arrivals, reflected phases) in the data by time windowing (e.g., Sheng et al., 2006; Sears et al., 2008). Frequency-domain wave modeling is not as flexible as the time-domain system for the preconditioning of the data by time windowing, since a limited number of frequencies is conventionally processed at a given step of the inversion. Data preconditioning can, however, be applied in the frequency domain by means of complex frequencies \( \omega + i\gamma \), which is equivalent to damp seismograms in the time domain (Shin et al., 2002; Brenders and Pratt, 2007b). The Fourier transform of a signal \( f(t) \) damped in
Time damping applied from the first-arrival time can be viewed as a heuristic way to select aperture angles of P-waves in the data. Arrivals located in time close to the first-arrival times correspond to wide-aperture P-wave events, while the later arriving phase from the first-arrival traveltime corresponds to shorter aperture P-wave events and converted waves. Since aperture angle is an additional parameter to frequency in the control of the spatial resolution of FWI (Miller et al., 1987; Wu and Toksoz, 1987; Sirgue and Pratt, 2004), aperture selection can be exploited to implement a second level of hierarchy in FWI in addition to that naturally introduced by frequency selection. Another benefit expected from complex-valued frequencies is the damping during the early FWI iterations of converted P-S waves, free-surface multiples and surface waves, which introduces additional non-linearities into the inversion. In practice, this second level of hierarchy can be implemented by progressively relaxing the time damping during each frequency-group inversion. Note that when complex-valued frequencies are used in FWI, the amplitude term $\exp^{-\gamma(t-t_0)}$ must be introduced in the weighting matrix $S_d$ (equation 3) to apply the same damping to the partial derivative wavefields and to the data residuals, which are cross-correlated during gradient building.

We designed the elastic frequency-domain FWI algorithm such that each of the above-mentioned strategies can be easily tested. The FWI algorithm implements the two above-mentioned nested hierarchical levels through an outer loop over the frequency groups and an inner loop over the damping terms (i.e., the imaginary part of the complex-valued frequencies). Here, we should note that a frequency group is a set of real frequencies that are
simultaneously inverted. A third loop is over inversion iteration, and a fourth is over the frequencies of the group. An updated model is produced after one iteration of the inversion involving one frequency group and one damping term. The major steps of the frequency-domain FWI algorithm can be summarized as follows:

\[
\text{for } \text{frequency
group} = \text{group}_1 \text{ to } \text{group}_n \text{ do}
\]

\[
\text{for } \text{data
damping} = \text{high}_{\text{damping}} \text{ to } \text{low}_{\text{damping}} \text{ do}
\]

\[
\text{while } (\text{NOT convergence OR } \text{iter} < \text{niter}_{\text{max}}) \text{ do}
\]

\[
\text{for } \text{frequency} = \text{frequency}_1 \text{ to } \text{frequency}_n \text{ do}
\]

Propagate wavefields from sources

Compute Residuals $\Delta d$ and Cost function $C_{mk}$

Backpropagate Residuals from receivers

end for

Build gradient vector $G_{mk}$

if $\text{iter} = 1$ then

Compute diagonal of Pseudo-Hessian

end if

Compute perturbation vector $\delta m$ with Quasi-Newton L-BFGS method

Define optimal step length $\alpha$ by parabola fitting

Update model $m_{k+1} = m_k + \alpha \delta m$

end while

end for

end for
APPLICATION TO A CANONICAL MODEL

FWI is an ill-posed, non-linear problem even for apparently simple models involving few parameters. Mulder and Plessix (2008) illustrated analytically the non-linearity of 3D acoustic FWI with a 1D velocity gradient model defined by two parameters. The objective function showed multiple local minima around the true global minimum. We consider here a similar two-parameter problem for a 2D elastic model with a free surface on top of it.

A 1D $V_P$ gradient model defined by $V_P(z) = V_0 + \eta z$ is considered, where $V_0$ is the P-wave velocity at the surface and $\eta$ is the vertical velocity gradient. The S-wave velocity is inferred from the P-wave velocity by considering a constant Poisson ratio of 0.24. A vertical point-force is located just below the free surface, and 350 receivers record horizontal and vertical particle velocities on the free surface along a 17-km-long profile.

The objective function is plotted as a function of $V_0$ and $\eta$ for the 5.8 Hz frequency in Figure 1a. The global minimum is located at the coordinates of the true model ($V_0 = 4$ km/s, $\eta = 0.35$ s$^{-1}$). Cross sections along the $\eta$ axis for $V_0 = 4$ km/s and along the $V_0$ axis for $\eta = 0.35$ s$^{-1}$ show the non-convex shape of the objective function (Figure 1). Multiple local minima are present, particularly on the $\eta$ section, even for this simple gradient model and the low frequency content in the data.

We repeated the same simulations for damped data using $\gamma = 3.33$ (equation 9). Using this data preconditioning, the objective function is now convex (Figures 1b and 1c). The convex shape of the objective function should ensure the convergence of the inversion towards the global minimum of the objective function for all of the starting models sampled in Figure 1.
SEG/EAGE Overthrust model and experimental setup

We considered a 20 km $\times$ 4.65 km dip section of the SEG/EAGE Overthrust model to assess the potential of 2D elastic frequency-domain FWI for imaging complex onshore structures (Aminzadeh et al., 1997)(Figure 2a). The Overthrust model was a 20 km $\times$ 20 km $\times$ 4.65 km 3D acoustic model that represents an onshore complex thrusted sedimentary succession constructed on top of a basement block. Several faults and channels were present in the model, as well as a complex weathering zone on the surface. For elastic FWI, a $V_S$ model was built from the $V_P$ model using a constant Poisson ratio of 0.24. A uniform density of 1000 kg.m$^{-3}$ was considered and assumed to be known during the inversion. A free surface was set on top of the model.

The onshore, wide aperture survey consisted of 199 explosive sources spaced every 100 m and located 25 m below the free surface. All of the shots were recorded by 198 vertical and horizontal geophones, which were spaced every 100 m on the surface. The vertical and horizontal components of particle velocity were used as the dataset for the elastic FWI. The data were computed with the same algorithm for both observed and computed data in inversion. The source signature was assumed to be known in FWI. An elastic shot gather is shown in Figure 3(a-b) for the horizontal and vertical components of particle velocity. The corresponding shot gather computed when an absorbing boundary condition is set on top of the model is shown for comparison in Figure 3(c-d), to highlight the additional wave complexities introduced by free-surface effects (i.e., surface waves and body-wave reflexion from the free surface). Of note, the high amplitudes of the surface waves dominate the wavefield especially on the vertical component (Figure 3(a-b)). Starting $V_P$ and $V_S$ mod-
els for FWI were computed by smoothing the true velocity models with a 2D Gaussian function of vertical and horizontal correlation ranges of 500 m (Figure 2b). This starting model was proven to be accurate enough to image the Overthrust model by 2D acoustic frequency-domain FWI using a realistic starting frequency of 3.5 Hz (Sourbier et al., 2009). For elastic inversions presented hereafter, we used a lower starting frequency of 1.7 Hz. Using a starting frequency of 3.5 Hz for elastic FWI led to a deficit of long wavelengths in the $V_S$ models, which made the inversion converge towards a local minimum.

The different behavior of acoustic and elastic FWI for the Overthrust case study highlights the increased sensitivity of elastic FWI with respect to the limited accuracy of the starting models. Five discrete frequencies (1.7, 2.5, 3.5, 4.7 and 7.2 Hz) were used for the elastic FWI. This frequency sampling should allow continuous sampling of the wavenumber spectrum according to the criterion of Sirgue and Pratt (2004). In the following, we shall consider the three different strategies to manage frequencies described in the section Full-waveform inversion data preconditioning and multiscale strategies: successive inversion of single frequencies, successive inversion of frequency groups of increasing bandwidth, and successive inversion of slightly overlapping frequency groups. For each frequency group, the inversion is subdivided into two steps. In the first step, no offset-dependent gain was applied to the data. Although we scale the gradient by the diagonal terms of the Hessian, we observed some lack of reconstruction in the deep part of the model, suggesting that the near-offset traces have a dominant contribution in the objective function. This layer-stripping effect may provide additional regularization of the inversion, in addition to that provided by the frequency and aperture angle selections. In the second step, we applied a quadratic gain with offset to the data, to strengthen the contribution of long-offset data in the inversion, and hence to improve the imaging of the deeper part of the model.
During the two-step inversion, the coefficients of the diagonal weighting operator $S_d$ were respectively given by:

$$S_d = \exp^{\gamma t_0}$$

$$S_d = \exp^{\gamma t_0} |\text{offset}|^2,$$

(10)

with the reminder that the coefficients $\exp^{\gamma t_0}$ account for the offset-dependent time damping; equation 9. For all of the tests presented below, except for the first, we used five damping factors per frequency to precondition the data ($\gamma = 1.5, 1.0, 0.5, 0.1, 0.033$). A shot gather computed for the first four damping factors is shown in Figure 4, to illustrate the amount of information preserved in the data. Note how the high damping limits the offset range over which surface waves are seen. Inversion was also regularized by Gaussian smoothing of the perturbation model, the aim of which was to cancel high frequency artifacts in the gradient. The diagonal terms of the pseudo-Hessian matrix (Shin et al., 2001) provided an initial guess of the Hessian for the L-BFGS algorithm without introducing extra computational costs during gradient building. Five differences of gradients and models vectors were used for the L-BFGS algorithm. The model parameters for inversion were $V_P$ and $V_S$, which are suitable for wide aperture acquisition geometries (Tarantola, 1986). The loop over the inversion iteration of one complex-valued frequency group was stopped when a maximum iteration number of 45 was reached or when the convergence criterion was reached (relative decrease of two successive cost functions lower than $5.10^{-5}$). The schedule of the frequencies and damping terms used in the sequential approach, the Bunks approach and the simultaneous approach are outlined in Table 1.

In the following, we shall quantify the data misfit for each test with the normalized misfit
\( \bar{C} \), defined by:

\[
\bar{C} = \frac{\sum_{i=1}^{5} \| \Delta d_i(m_f) \|_2}{\sum_{i=1}^{5} \| \Delta d_i(m_0) \|_2} \tag{11}
\]

where \( \Delta d_i(m_f) \) denotes the data misfit vector for the \( i^{th} \) frequency and for the final FWI model \( m_f \), and \( m_0 \) denotes the starting model shown in Figure 2b.

The FWI model quality will be quantified by:

\[
mq = \frac{1}{N} \| \frac{m_f - m_{true}}{m_{true}} \|_2 \tag{12}
\]

where \( m_{true} \) denotes the exact model either for \( V_P \) or \( V_S \), and \( N \) is the number of grid points in the computational domain. The normalized misfit and the model quality for the different tests presented hereafter are outlined in Table 2.

**Raw data inversion**

A first inversion test was performed without data damping (i.e., without using complex-valued frequencies), which implies that all of the arrivals were involved in the inversion.

The five frequencies (Table 1) were successively inverted with the sequential approach. The FWI \( V_P \) and \( V_S \) models after inversion are shown in Figure 5. The inversion clearly failed to converge towards the true models for both of the \( V_P \) and \( V_S \) parameters, even at low frequencies.

**Successive single-frequency inversions of damped data**

We repeated the previous experiment, except that the five damping terms (\( \gamma = 1.5, 1.0, 0.5, 0.1, 0.033 \)) were used to stabilize the inversion (Table 1). The final FWI \( V_P \) and \( V_S \) models are shown in Figure 6. Contrary to the previous experiment, most of the layers were now successfully reconstructed. Comparison between 1D vertical profiles extracted from the
true model, the starting model and the FWI models shows a reliable estimate of velocity amplitudes despite a low maximum frequency of 7.2 Hz (Figure 7). To take into account the limited bandwidth effect of the source in the FWI model appraisal, we also plotted the vertical profiles of the true models after low-pass filtering at the theoretical resolution of FWI for a maximum frequency of 7.2 Hz: the true models were converted from depth to time using the velocities of the starting model, and low-pass filtered with a cut-off frequency of 7 Hz, before conversion back to the depth domain. We noted that the $V_S$ model is more affected by spurious artifacts than the $V_P$ model, especially in the deep part of the model. This may be due to a deficit of small wavenumbers in the $V_S$ models, resulting from the shorter propagated wavelengths, which makes the reconstruction of the $V_S$ parameter more non-linear. Secondly, we saw some inaccuracies in the reconstruction of both the $V_P$ and $V_S$ parameters in the shallowest parts of the models (Figure 7). The resulting residuals of the surface waves and reflections from the free surface may have been erroneously back-projected in the deeper part of the model, leading to the above-mentioned noise (Figure 8). The final normalized $L_2$ misfit computed for the five frequencies was $4.12 \times 10^{-1}$. The $V_P$ and $V_S$ model qualities are $5.54 \times 10^{-2}$ and $6.47 \times 10^{-2}$, respectively.

**Full-waveform inversion without free-surface effects**

To assess the footprint of surface waves and free-surface reflections on elastic FWI, we inverted the data computed in the Overthrust model with an absorbing boundary condition on top of it, instead of a free surface. The same inversion process was used as in the previous section (sequential approach with damped data). The final FWI $V_P$ and $V_S$ models were very close to the low-pass filtered versions of the true models, and they were not affected by any spurious artifacts (Figure 9). The $V_P$ and $V_S$ model qualities are $4.09 \times 10^{-2}$ and
3.89 $10^{-2}$, respectively. Comparison with the previous results (Figure 6) illustrates the substantial increase of non-linearity introduced by surface waves and free-surface reflections in elastic FWI.

**Simultaneous multi-frequency inversion of damped data**

We now investigated the influence of simultaneous multiple-frequency inversion strategies for FWI improvement. We first considered the Bunks approach for FWI. The different frequency groups and damping terms are outlined in Table 1. Each mono-frequency dataset belonging to a frequency group was computed with a unit Dirac source wavelet, which implies that each frequency of a group has a similar weight in the inversion. This is equivalent to inversion of deconvolved data (i.e., data with a flat amplitude spectrum).

The final FWI $V_P$ and $V_S$ models shown in Figure 10 are slightly improved compared to those of the sequential approach (Figure 6). The improvements are more obvious on the vertical profiles of the final FWI models (Figure 11). Near-surface instabilities in the $V_P$ and $V_S$ models were mitigated, although not fully removed, and the estimations of the velocity amplitudes were improved in most parts of the model (compare Figures 7 and 11). Mitigation of the near-surface instabilities translated into a significant misfit reduction for the surface waves and free-surface reflections (Figure 12). The final normalized $L_2$ misfit decreased in this case to $1.54 \times 10^{-1}$. The $V_P$ and $V_S$ qualities are $5.22 \times 10^{-2}$ and $5.33 \times 10^{-2}$, respectively.

In a second step, we considered the simultaneous approach implemented by successive inversions of two overlapping frequency groups (Table 1). The frequency bandwidth of each group was chosen such that cycle-skipping artifacts were avoided. For example, we tried to simultaneously invert the five frequencies listed in Table 1. In this case, the inversion failed
to converge towards acceptable models. Note that the computational cost of the simultaneous approach is similar to that of the sequential one if the same convergence rate for the two approaches is assumed. The total number of iterations in the simultaneous approach is less important, because the iterations are performed over fewer frequency groups at the expense of more factorization and substitution phases per frequency group. The extra cost of the simultaneous approach is only due to the overlap between the frequency groups.

The final FWI $V_P$ and $V_S$ velocity models shown in Figure 13 were improved with respect to the velocity models produced by the sequential approach (Figure 6) and the Bunks approach (Figure 10), especially for the $V_S$ velocity model in the thrust zone. The vertical profiles extracted from the final FWI models do not show near-surface instabilities anymore (Figure 14), which allowed a significant data misfit reduction for the surface waves and free-surface reflections (Figure 15). A significant misfit reduction of the wide aperture arrivals recorded at large offsets was also seen. The final $L2$ misfit is $1.46 \times 10^{-1}$. The $V_P$ and $V_S$ model qualities are $5.03 \times 10^{-2}$ and $5.39 \times 10^{-2}$, respectively. We note, however, slightly underestimated velocity amplitudes in the deep part of the $V_P$ and $V_S$ models at the thrust location (see below 3 km depth in Figure 14a, c). We attribute this amplitude deficit to a slower convergence of the simultaneous approach when compared to that of the sequential one, which results from the fact that more information is simultaneously inverted in the simultaneous approach. The imaging was further improved by decreasing the frequency interval by a factor of 2 within each frequency group (five frequencies instead of three frequencies per group). This resampling contributes to the strengthening of the spectral redundancy in the imaging. Close-ups of the $V_P$ and $V_S$ models centered on the thrust zone show how the resolution and the signal-to-noise ratio of the velocity models were still improved by involving more frequencies in one inversion iteration (Figure 16). Note that using five frequencies
instead of three in each group leads to a factor of 5/3 in computational costs.

**L-BFGS versus preconditioned conjugate-gradient optimizations**

The sequential approach was applied using a preconditioned conjugate-gradient (PCG) algorithm for numerical optimization. Preconditioning is performed by scaling the gradient by the diagonal terms of the pseudo-Hessian matrix. The same frequencies and damping terms were used as for the L-BFGS run based on the sequential approach (Table 1). The velocity models recovered with the PCG algorithm are shown in Figure 17 and can be compared with the corresponding L-BFGS ones (Figure 6). Amplitude estimations and focusing of structures were improved by the L-BFGS algorithm (Figure 18), leading to sharper models. The improvements in the model resolution and the quantitative estimate of the model parameters can be attributed to the approximate estimation of the off-diagonal terms of the Hessian performed by the L-BFGS algorithm. These off-diagonal elements help to deconvolve the models from limited-bandwidth effects resulting from the limited source bandwidth and the limited extent of the acquisition geometry (e.g., Pratt et al., 1998).

Figure 19 shows the objective functions as a function of iteration number for the L-BFGS and PCG algorithms. L-BFGS provides accelerated and improved convergence when compared to PCG. The PCG convergence level (i.e., the minimum value of the objective function reached during optimization) cannot reach that of L-BFGS because the off-diagonal information of Hessian estimated by L-BFGS cannot be retrieved by more iterations of PCG. Note that the final $L^2$ misfit of PCG is $6.71 \times 10^{-1}$, whereas that of L-BFGS is $4.12 \times 10^{-1}$. These two issues open promising applications of L-BFGS for computationally challenging problems, such as for 3D FWI.
Computational aspect

All of the simulations were performed on the cluster of the SIGAMM computer center, which is composed of a 48-node cluster, with each node comprising 2 dual-core 2.4-GHz Opteron processor, providing 19.2 Gflops peak performance per node. This computer has a distributed memory architecture, where each node has 8 GBytes of RAM. The interconnection network between processors is Infiniband 4X. Twenty-four processors were used for each simulation, leading to the best compromise between execution time and numerical resources used. A single regular equilateral mesh composed of 265675 cells was designed for the simulations. Although the mesh could have been adapted to the inverted frequency, we did not consider this strategy here, and the mesh was kept constant whatever the inverted frequency. Table 3 outlines the memory requirements and computational time of the major tasks performed by the parallel FWI algorithm. Of note, most of the memory and computational time were dedicated to the LU factorization and substitution phases performed during the multi-source forward problem. Computation of the gradient had a negligible computational cost, due to the domain-decomposition parallelism. The L-BFGS algorithm required a negligible extra amount of memory and computational time compared to a steepest-descent or PCG algorithm, suggesting that this optimization scheme can be efficiently used for realistic 2D and 3D FWI applications.

DISCUSSION

Application of elastic FWI to the Overthrust model has highlighted the strong non-linearity of the inversion resulting from free-surface effects. The impact of these effects on FWI can be assessed by comparing the FWI results inferred from the data including or without the
free-surface effects (compare Figures 6 and 9). As the best models were obtained when free-surface effects are not considered, this shows that for this case study, inversion of the surface waves was not useful towards an improvement of the reconstruction of the near-surface structure.

We interpret the failure of the raw-data inversion as the footprint of surface waves, the amplitudes of which dominate the wavefield and carry no information of the deep part of the model (Figure 5). Similar effects of surface waves on elastic FWI were also seen on a smaller scale by Gelis et al. (2007). Comparison between the sequential FWI results obtained with the raw data and the preconditioned data illustrates how the time damping helps to mitigate the non-linearities of FWI by injecting progressively more complex wave phenomena in the inversion (compare Figures 5 and 6).

The further improvements obtained by simultaneous inversion of multiple frequencies combined with hierarchical inversions of damped data show that preserving some wavenumber redundancy in elastic FWI is critical to mitigate the non-linearity of the inversion associated with the propagation of surface waves in weathered near-surface layers and free-surface reflections. The more stable results obtained with the simultaneous approach compared to the Bunks approach, especially in the near-surface, suggest that several frequencies must be simultaneously inverted from the early stage of the inversion (compare Figures 10 and 13). Strengthening the wavenumber redundancy by decreasing the frequency interval in each frequency group further improved the imaging (Figure 16).

Another factor that increased the non-linearity of elastic FWI was the short S wavelengths that may require more accurate starting models or lower frequencies to converge towards an acceptable model. The maximum frequency of the starting frequency group must be chosen such that it prevents cycle-skipping artifacts that can result from the limited accuracy of
the starting S-wave velocity model. Laplace-domain waveform inversion, which has been recently proposed as a reliable approach to build smooth initial elastic models of the subsurface (Pyun et al., 2008; Shin and Cha, 2008), may represent an approach to tackle the issue of building the starting model. An alternative approach is PP-PS stereotomography (Alerini et al., 2002), including the joint inversion of refraction and reflection traveltimes of wide-aperture data.

As mentioned above, with this case study, we were not able to illustrate the usefulness of the surface waves for the reconstruction of the near-surface structure, since the most accurate FWI models were inferred without involving free-surface effects in the data (Figure 9). Instead, we have shown how to manage the non-linearities introduced by the surface waves, by means of judicious data preconditioning and FWI tuning. Alternatively, the surface waves in the recorded and modeled data can be filtered out or muted. We did not investigate this approach at this stage because efficient filtering of the modeled surface waves in the frequency domain is not straightforward. Note that the surface waves must be filtered out not only at the receiver positions, but also at each position in the computational domain where they have significant amplitudes, in order to remove their footprint from the gradient of the objective function. This investigation still requires further work.

More realistic applications of elastic FWI in more complex models still need to be investigated. Areas of complex topography, such as foot-hills, will lead to conversions from surface waves to body waves and vice versa, which may carry additional information on the near-surface. The robustness of elastic FWI for imaging models with heterogeneous Poisson ratios is a second field of investigation, especially in areas of soft seabed where the P-S-converted wavefield may have a limited signature in the data (Sears et al., 2008). In the present study, we inverted data computed with a constant density that was assumed to
be known. For real data inversion, estimation of the density is required for a more reliable amplitude match. Reliable estimation of the density by FWI is difficult, because the P-wave velocity and density have similar radiation patterns at short apertures (Forgues and Lambaré, 1997). The benefit provided by wide apertures to uncouple these two parameters needs to be investigated. Other extensions of isotropic elastic FWI may relate to reconstruction of attenuation factors and some anisotropic parameters. Vertical transversely isotropic elastic FWI should be easily implemented from isotropic elastic FWI, since only the expression of the coefficients of the P-SV elastodynamic system need to be modified compared to the isotropic case (e.g., Carcione et al., 1988).

Application of 2D elastic FWI to real data will require additional data preprocessing that was not addressed in this study, such as source estimation (Pratt, 1999) and 3D to 2D amplitude corrections (Bleistein, 1986; Williamson and Pratt, 1995). The sensitivity of the elastic FWI to the approximations and errors underlying this processing will need to be determined.

CONCLUSION

This study presents a new massively parallel 2D elastic frequency-domain FWI algorithm, with an application to a dip section of the SEG/EAGE onshore Overthrust model. Strong non-linearities of elastic FWI arise both from the presence of converted and surface waves, and from the limited accuracy of the $V_S$ starting model. These two factors prevent convergence of FWI on the global minimum of the objective function if no specific preconditioning is applied to the data and no low starting frequency is available. Data preconditioning performed by time damping is necessary to converge towards acceptable velocity models, whatever the frequency sampling strategy is. Secondly, successive inversions of overlapping
frequency groups out-perform successive inversions of single frequencies for the removal of instabilities in the near-surface of the FWI models. The bandwidth of the frequency groups must be chosen such that cycle-skipping artifacts are avoided, while injecting a maximum amount of redundant information into the frequency groups. The quasi-Newton algorithm of L-BFGS outperforms the most popular preconditioned conjugate-gradient algorithm in terms of convergence rate and convergence level, without significant extra computational costs, and hence is shown to be very useful for this application.

ACKNOWLEDGMENTS

This study was funded by the SEISCOPE consortium http://seiscope.unice.fr, sponsored by BP, CGG-VERITAS, EXXON-MOBIL, SHELL and TOTAL, and by Agence Nationale de la Recherche (ANR) under project ANR-05-NT05-2-42427. The linear systems were solved with MUMPS package, available on http://graal.ens-lyon.fr/MUMPS/index.html. The mesh generation was performed with help of TRIANGLE, available on http://www.cs.cmu.edu/~quake/triangle.html. Access to the high-performance computing facilities of the SIGAMM (Observatoire de la Côte d’Azur) and IDRIS computer center (project 082280) provided the required computer resources, and we gratefully acknowledge both of these facilities and the support of their staff. Many thanks go to Serge Gratton (CERFACS and CNES, Toulouse, FRANCE) for interesting and stimulating discussions on quasi-Newton optimization methods. We would like to thank Isabelle Lecomte, Christophe Barnes, Hansruedi Maurer and Changsoo Shin for their fruitful comments on the manuscript.
REFERENCES


marmousi: The Leading Edge, 25, 156–166.


Shi, Y., W. Zhao, and H. Cao, 2007, Nonlinear process control of wave-equation inversion
and its application in the detection of gas: Geophysics, 72, no. 1, R9–R18.


Vigh, D. and E. W. Starr, 2008, 3d prestack plane-wave, full-waveform inversion: Geo-
physics, 73, no. 5, VE135–VE144.


TABLE HEADINGS
Table 1: Inversion parameters for the sequential, Bunks, and simultaneous approaches. FG: frequency group number; F: frequencies within a frequency group; $\gamma$: damping factors (imaginary part of frequency).

Table 2: Final $L^2$ misfit and model quality $mq$ for the different reconstructed models. Seq: Sequential approach; Bunks: Bunks approach; Sim: Simultaneous approach; without FS: Sequential approach without free-surface effects; PCG: Sequential approach computed with PCG optimization.

Table 3: Computational cost of the main tasks performed by FWI. Time estimations were averaged over several iterations.
FIGURE CAPTIONS
Figure 1: Objective function as a function of the two model parameters for a) full data set and b) damped data set. c) Cross-sections of maps shown above for $V_0=4$ km/s. The solid line and the dashed line correspond to the full data-set and the damped data-set, respectively. d) Cross-sections of maps shown above for $\eta=0.35$ s$^{-1}$. The solid line and the dashed line correspond to the full data-set and the damped data set respectively.

Figure 2: a) Dip section of the synthetic SEG/EAGE Overthrust model. P-wave velocity is depicted. b) Starting model used for elastic FWI.

Figure 3: Seismograms computed in Overthrust model for (a) horizontal and (b) vertical components of particle velocity. The shot is located at a horizontal distance of 3 km. A free-surface was set on top of the model. (c-d) As for Figure 3(a-b), except that an absorbing boundary condition was implemented on top of the model.

Figure 4: Seismograms for vertical component of particle velocity computed in the dip section of the Overthrust model using four values of imaginary frequency. a) $\gamma = 1.5$ s$^{-1}$, b) $\gamma = 1.0$ s$^{-1}$, c) $\gamma = 0.5$ s$^{-1}$, d) $\gamma = 0.1$ s$^{-1}$. Time-damping was applied from the first-arrival traveltime to preserve long-offset information.
Figure 5: Sequential inversion of raw data - (a) $V_P$ and (b) $V_S$ models after frequency of 7.2 Hz.

Figure 6: Sequential inversion of damped data - (a) $V_P$ and (b) $V_S$ models after inversion. The L-BFGS algorithm was used for optimization. Five frequency components were inverted successively. Five damping coefficients were successively used for data preconditioning during each mono-frequency inversion.

Figure 7: Sequential inversion of damped data - Vertical profiles for the $V_P$ (a-b) and $V_S$ (c-d) parameters. Profiles (a)-(c) and (b)-(d) are at horizontal distances of 7.5 km and 14 km respectively. Profiles of the starting and the true models are plotted with dashed gray lines and solid black lines, respectively. A low-pass filtered version of the true model at the theoretical resolution of FWI is plotted with a dashed black line for comparison with the FWI results. The profiles of the FWI models of Figure 6 are plotted with solid gray lines.

Figure 8: Sequential inversion of damped data - Seismograms computed in the FWI models of Figure 6 for shot located at a horizontal distance of 3 km. a) Horizontal component. b) Vertical component. (c-d) Residuals between seismograms computed in the true $V_P$ and $V_S$ models (Figure 3a) and in the FWI models of Figure 6. c) Vertical component. d) Horizontal component.
Figure 9: Sequential inversion without free-surface effects - (a) $V_P$ and (b) $V_S$ models. The models can be compared with that of Figures 6 to assess the footprint of free-surface effects on elastic FWI.

Figure 10: Bunks inversion - (a) $V_P$ and (b) $V_S$ models obtained with the frequency-domain adaptation of the multiscale approach of Bunks et al. (1995).

Figure 11: Same as Figure 7, but for the profiles extracted from the models recovered by the Bunks approach (Figure 10).

Figure 12: Same as Figure 8, but for seismograms computed in the FWI models recovered by the Bunks approach (Figure 10).

Figure 13: Simultaneous inversion - Final models obtained by successive inversions of two overlapping frequency groups composed of three frequencies each. (a) $V_P$ model. (b) $V_S$ model.

Figure 14: Same as Figure 7, but for the profiles extracted from the models recovered by the simultaneous approach (Figure 13).
Figure 15: Same as Figure 8, but for seismograms computed in the FWI models recovered by successive inversion of two overlapping frequency groups (Figure 13).

Figure 16: Comparison between FWI models obtained by successive inversion of two overlapping groups of frequencies (simultaneous approach) when three and five frequencies per group are used in the inversion, respectively. The frequency bandwidth is the same for each experiment but the frequency interval differs. a) Close-up of the true $V_P$ model after low-pass filtering at the theoretical resolution of FWI. (c)-(e) Close-up of the FWI $V_P$ model when five frequencies (c) and three frequencies (e) per group are used respectively. (b)-(d)-(f) Same as (a)-(c)-(e) for the $V_S$ model.

Figure 17: FWI velocity models obtained with the PCG algorithm. a) $V_P$ parameter. b) $V_S$ parameter. The sequential approach with five damping terms was used. The velocity models can be compared with those recovered by the L-BFGS algorithm (Figure 6).

Figure 18: Comparison between velocity profiles extracted from the FWI models recovered by the L-BFGS (solid gray line) and the PCG algorithm (dashed black line). a) $V_P$ models. b) $V_S$ models. Starting and true models are depicted with dashed gray line and solid black line, respectively.
Figure 19: L-BFGS and PCG objective functions plotted as a function of iteration number for the inversion of the complex frequencies \((1.7 + i1.5)\) Hz and \((4.7 + i1.5)\) Hz. The curves associated with one frequency are normalized by the PCG objective functions at the first iteration. Other frequencies and damping factors show similar trends.
Table 1: Inversion parameters for the sequential, Bunks, and simultaneous approaches. FG: frequency group number; F (Hz): frequencies within a frequency group; \( \gamma \) (1/s): damping factors (imaginary part of frequency).

<table>
<thead>
<tr>
<th>FG</th>
<th>Sequential - F(Hz)</th>
<th>Bunks - F(Hz)</th>
<th>Simultaneous - F(Hz)</th>
<th>( \gamma ) (1/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7, 2.5, 3.5</td>
<td>1.5, 1, 0.5, 0.1, 0.033</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>1.7, 2.5</td>
<td>3.5, 4.7, 7.2</td>
<td>1.5, 1, 0.5, 0.1, 0.033</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>1.7, 2.5, 3.5</td>
<td></td>
<td>1.5, 1, 0.5, 0.1, 0.033</td>
</tr>
<tr>
<td>4</td>
<td>4.7</td>
<td>1.7, 2.5, 3.5</td>
<td>4.7</td>
<td>1.5, 1, 0.5, 0.1, 0.033</td>
</tr>
<tr>
<td>5</td>
<td>7.2</td>
<td>1.7, 2.5, 3.5</td>
<td>4.7, 7.2</td>
<td>1.5, 1, 0.5, 0.1, 0.033</td>
</tr>
</tbody>
</table>
Table 2: Final $L^2$ misfit and model quality $mq$ for the different reconstructed models. Seq: Sequential approach; Bunks: Bunks approach; Sim: Simultaneous approach; without FS: Sequential approach without free-surface effects; PCG: Sequential approach computed with PCG optimization.

<table>
<thead>
<tr>
<th>Test</th>
<th>Data misfit $\bar{C}$</th>
<th>$V_P \ mq$</th>
<th>$V_S \ mq$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq</td>
<td>$4.12 \times 10^{-1}$</td>
<td>$5.54 \times 10^{-2}$</td>
<td>$6.47 \times 10^{-2}$</td>
</tr>
<tr>
<td>Bunks</td>
<td>$1.54 \times 10^{-1}$</td>
<td>$5.22 \times 10^{-2}$</td>
<td>$5.33 \times 10^{-2}$</td>
</tr>
<tr>
<td>Sim</td>
<td>$1.46 \times 10^{-1}$</td>
<td>$5.03 \times 10^{-2}$</td>
<td>$5.39 \times 10^{-2}$</td>
</tr>
<tr>
<td>Without FS</td>
<td>$9.03 \times 10^{-2}$</td>
<td>$4.09 \times 10^{-2}$</td>
<td>$3.89 \times 10^{-2}$</td>
</tr>
<tr>
<td>PCG</td>
<td>$6.71 \times 10^{-1}$</td>
<td>$5.56 \times 10^{-2}$</td>
<td>$6.89 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
Table 3: Computational cost of the main tasks performed by FWI. Time estimations were averaged over several iterations for 24 processors.

<table>
<thead>
<tr>
<th>Task Description</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time for factorization (s)</td>
<td>21.3</td>
</tr>
<tr>
<td>MUMPS total memory for factorization (Gbytes)</td>
<td>4.8</td>
</tr>
<tr>
<td>Time for substitutions (199 shots) (s)</td>
<td>22.9</td>
</tr>
<tr>
<td>MUMPS total memory for substitution (Gbytes)</td>
<td>20.2</td>
</tr>
<tr>
<td>Time for gradient build up (s)</td>
<td>13.8</td>
</tr>
<tr>
<td>Time for L-BFGS perturbation computation (s)</td>
<td>1.2</td>
</tr>
<tr>
<td>Memory for L-BFGS(5) history (Mbytes)</td>
<td>21.2</td>
</tr>
</tbody>
</table>
FIGURES
Figure 1: Objective function as a function of the two model parameters for a) full data set and b) damped data set. c) Cross-sections of maps shown above for $V_0=4$ km/s. The solid line and the dashed line correspond to the full data-set and the damped data-set, respectively. d) Cross-sections of maps shown above for $\eta=0.35$ s$^{-1}$. The solid line and the dashed line correspond to the full data-set and the damped data set respectively.
Figure 2: a) Dip section of the synthetic SEG/EAGE Overthrust model. P-wave velocity is depicted. b) Starting model used for elastic FWI.
Figure 3: Seismograms computed in Overthrust model for (a) horizontal and (b) vertical components of particle velocity. The shot is located at a horizontal distance of 3 km. A free-surface was set on top of the model. (c-d) As for Figure 3(a-b), except that an absorbing boundary condition was implemented on top of the model.
Figure 4: Seismograms for vertical component of particle velocity computed in the dip section of the Overthrust model using four values of imaginary frequency. a) $\gamma = 1.5 \ s^{-1}$, b) $\gamma = 1.0 \ s^{-1}$, c) $\gamma = 0.5 \ s^{-1}$, d) $\gamma = 0.1 \ s^{-1}$. Time-damping was applied from the first-arrival traveltime to preserve long-offset information.
Figure 5: Sequential inversion of raw data - (a) $V_P$ and (b) $V_S$ models after frequency of 7.2 Hz.
Figure 6: Sequential inversion of damped data - (a) $V_P$ and (b) $V_S$ models after inversion. The L-BFGS algorithm was used for optimization. Five frequency components were inverted successively. Five damping coefficients were successively used for data preconditioning during each mono-frequency inversion.
Figure 7: Sequential inversion of damped data - Vertical profiles for the $V_P$ (a-b) and $V_S$ (c-d) parameters. Profiles (a)-(c) and (b)-(d) are at horizontal distances of 7.5 km and 14 km respectively. Profiles of the starting and the true models are plotted with dashed gray lines and solid black lines, respectively. A low-pass filtered version of the true model at the theoretical resolution of FWI is plotted with a dashed black line for comparison with the FWI results. The profiles of the FWI models of Figure 6 are plotted with solid gray lines.
Figure 8: Sequential inversion of damped data - Seismograms computed in the FWI models of Figure 6 for shot located at a horizontal distance of 3 km. a) Horizontal component. b) Vertical component. (c-d) Residuals between seismograms computed in the true $V_P$ and $V_S$ models (Figure 3a) and in the FWI models of Figure 6. c) Vertical component. d) Horizontal component.
Figure 9: Sequential inversion without free-surface effects - (a) $V_P$ and (b) $V_S$ models. The models can be compared with that of Figures 6 to assess the footprint of free-surface effects on elastic FWI.
Figure 10: Bunks inversion - (a) $V_P$ and (b) $V_S$ models obtained with the frequency-domain adaptation of the multiscale approach of Bunks et al. (1995).
Figure 11: Same as Figure 7, but for the profiles extracted from the models recovered by the Bunks approach (Figure 10).
Figure 12: Same as Figure 8, but for seismograms computed in the FWI models recovered by the Bunks approach (Figure 10).
Figure 13: Simultaneous inversion - Final models obtained by successive inversion of two overlapping frequency groups composed of three frequencies each. (a) $V_P$ model. (b) $V_S$ model.
Figure 14: Same as Figure 7, but for the profiles extracted from the models recovered by the simultaneous approach (Figure 13).
Figure 15: Same as Figure 8, but for seismograms computed in the FWI models recovered by successive inversion of 2 overlapping frequency groups (Figure 13).
Figure 16: Comparison between FWI models obtained by successive inversion of two overlapping groups of frequencies (simultaneous approach) when three and five frequencies per group are used in the inversion, respectively. The frequency bandwidth is the same for each experiment but the frequency interval differs. a) Close-up of the true $V_P$ model after low-pass filtering at the theoretical resolution of FWI. (c)-(e) Close-up of the FWI $V_P$ model when five frequencies (c) and three frequencies (e) per group are used respectively. (b)-(d)-(f) Same as (a)-(c)-(e) for the $V_S$ model.
Figure 17: FWI velocity models obtained with the PCG algorithm. a) $V_P$ parameter. b) $V_S$ parameter. The sequential approach with five damping terms was used. The velocity models can be compared with those recovered by the L-BFGS algorithm (Figure 6).
Figure 18: Comparison between velocity profiles extracted from the FWI models recovered by the L-BFGS (solid gray line) and the PCG algorithm (dashed black line). a) $V_p$ models. b) $V_S$ models. Starting and true models are depicted with dashed gray line and solid black line, respectively.
Figure 19: L-BFGS and PCG objective functions plotted as a function of iteration number for the inversion of the complex frequencies $(1.7 + i1.5)$ Hz and $(4.7 + i1.5)$ Hz. The curves associated with one frequency are normalized by the PCG objective functions at the first iteration. Other frequencies and damping factors show similar trends.