A Parallel Evolution Strategy for an Earth Imaging Problem in Geophysics

Y. Diouane* S. Gratton† X. Vasseur‡ L. N. Vicente§ H. Calandra¶

October 13, 2015

Abstract

In this paper we propose a new way to compute a rough approximation solution, to be later used as a warm starting point in a more refined optimization process, for a challenging global optimization problem related to Earth imaging in geophysics. The warm start consists of a velocity model that approximately solves a full-waveform inverse problem at low frequency. Our motivation arises from the availability of massively parallel computing platforms and the natural parallelization of evolution strategies as global optimization methods for continuous variables.

Our first contribution consists of developing a new and efficient parametrization of the velocity models to significantly reduce the dimension of the original optimization space. Our second contribution is to adapt a class of evolution strategies to the specificity of the physical problem at hands where the objective function evaluation is known to be the most expensive computational part. A third contribution is the development of a parallel evolution strategy solver, taking advantage of a recently proposed modification of these class of evolutionary methods that ensures convergence and promotes better performance under moderate budgets.

The numerical results presented demonstrate the effectiveness of the algorithm on a realistic 3D full-waveform inverse problem in geophysics. The developed numerical approach allows us to successfully solve an acoustic full-waveform inversion problem at low frequencies on a reasonable number of cores of a distributed memory computer.

Keywords: Evolution strategy, global convergence, Earth imaging, inverse problem, high performance computing (HPC), search space reduction, full-waveform inversion.

*Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), Universit de Toulouse, 31055 Toulouse Cedex 4, France (youssef.diouane@isae.fr).

†ENSEEIHT, INPT, rue Charles Camichel, B.P. 7122 31071 Toulouse Cedex 7, France (gratton@enseeiht.fr).

‡CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France (vasseur@cerfacs.fr).

§CMUC, Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal (lnv@mat.uc.pt).
Support for this author was provided by FCT under grants PTDC/MAT/116736/2010 and PEst-C/MAT/UI0324/2011 and by the Réseau Thématique de Recherche Avancée, Fondation de Coopération Sciences et Technologies pour l'Aéronautique et l'Espace, under the grant ADTAO.

¶TOTAL E&P Research and Technology USA, llc 1201 Lousiana, Suite 1800, Houston, TX 77002-4397 (calandra@total.com).

1 Introduction

Vibrations generated by earthquakes, explosions or similar phenomena and propagated within the Earth or along its surface can yield information about the Earth and its subsurface structure. Such a knowledge, called Earth imaging, is of major interest for economy, environment, and science. Geologists have developed several methods for Earth imaging using seismic wave information. Acoustic full-waveform inversion (FWI) is one of such procedures and it attempts to derive high-resolution quantitative models of the subsurface using the full information of acoustic waves [52]. Following [50], a description of the problem can be given as follows. During the propagation, waves interfere with the environment and the total wavefield is recorded through a certain number of receivers (called hydrophones or geophones). Since the waves are affected by the physical properties of the subsurface, they are carrying information about the environment that can be retrieved by an inversion process. The propagation waves are generated by sources situated in the domain of consideration (see Figure 1 for a simple illustration).

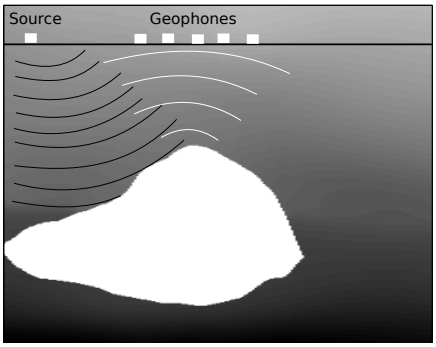


Figure 1: Acoustic waves propagated by a source are reflected by a reflective layer (in white) and are detected by the geophones. The reflective layer represents a salt dome in this example.

For many years acoustic full-waveform inversion has been almost exclusively employed by academic researchers. Only recently it has been adopted by practitioners in industry. Nevertheless, its computational cost is still large compared to other existing methods in seismic exploration. The attractiveness of the approach is the promise of deriving high-fidelity Earth models for seismic imaging. Since our ability to both understand and manage complex nonlinear inversions has improved and since the available computing power has grown at the same time, full-waveform inversion has become more and more practical.

It is known that the acoustic full-waveform inversion, formulated as a nonlinear optimization least-squares minimization problem, can be efficiently solved if the starting propagation velocity model is accurate enough (in the sense of explaining the data at a low frequency and still being a smooth version of the true velocity model, see [10, 52]). Otherwise the inversion procedure suffers from stalled convergence to spurious local minima due to the oscillatory nature of the data [28]. Thus, a crucial step related to full-waveform inversion in seismic imaging consists of finding a good starting model (or point, in an optimization context) without the need of sophisticated a priori knowledge on the velocity model. In industry, first- arrival travel-time tomography [30] is the most popular method to generate an accurate initial propagation velocity model. More recent methods such as stereo tomography [27] and inversion in the Laplace domain [41] are being investigated in academia. It must be also mentioned here that the use of multiscale strategies

can mitigate the nonlinearity and reduce the dependence on the starting velocity model for FWI [34, 45].

In this paper we propose a novel approach to find an initial smooth velocity model for the full-waveform inversion problem *without* any a priori physical knowledge. We are motivated by the recent availability of massively parallel computing platforms (see [36]). First we introduce a new parametrization of the problem to reduce the number of parameters needed to describe a velocity model and therefore the objective function of our optimization problem. Then, we show how to adapt an Evolution Strategy (ES) to take advantage of such a model or space reduction. ES's are a class of evolutionary algorithms designed for searching the global minimum of a function in a continuous space. We are motivated by the modifications in ES's recently proposed in [16] to ensure some form of rigorous convergence and a better computational performance under moderate budgets of function evaluations. Thirdly, based on one of the modified ES's given in [16], we propose a highly parallel ES adapted to the full-waveform inversion setting. By combining model reduction and ES's in a parallel environment, we aim at solving realistic instances of the problem. In fact the numerical results obtained along this direction will show the appropriateness and promise of our approach.

We note that global optimization heuristics have already been already employed to solve related inverse problems. A first attempt to invert the ocean bottom properties [14] has been made through simulated annealing. Later Gerstoft [19] has applied genetic algorithms to invert seismoacoustic data. Training a neural network to compute a reliable estimate of a one-dimensional velocity model has been also proposed in [40]. In all these applications, the heuristics were either applied to problems where the objective function was cheap to evaluate or where a very simple parametrization of the velocity model was used.

The paper is organized as follows. We start by describing in Section 2 the large-scale Earth imaging optimization problem of interest to us. In Section 3, we detail our proposed methodology to reduce the number of unknowns, while representing the full search space as faithfully as possible. In Section 4 we describe a parallel ES adapted to the specificity of the given application. Numerical results for a realistic large-scale public domain inversion problem are then presented and discussed in Section 5. Finally, we draw some conclusions and describe future lines of research in Section 6.

2 Full-waveform inversion

Estimating the subsurface velocity from seismic recordings is the main goal of the full-waveform inversion procedure. One uses the recorded wavefields to guess the physical properties of the medium through which the wavefield has propagated. Two formulations (either time-domain or frequency-domain based) are traditionally used for finding the solution of this inverse problem. Relevant details on both approaches can be found in, e.g., [10, 34, 45, 52]. Since the frequency-domain approach is regarded as more advantageous when solving the full-waveform inversion in the multiple frequency case [52], we will exclusively consider this approach in our paper. Below we briefly detail the full seismic wavefield problem (forward problem) and the associated problem used to recover the velocity model (inverse problem). At the end of the section, we will also introduce the three-dimensional public domain velocity model used in all our numerical illustrations and experiments.

2.1 The forward problem

Given the medium properties (e.g., the subsurface velocity), the forward problem consists of modeling the full seismic wavefield in a three-dimensional parallelipedic domain $\Omega \subset \mathbb{R}^3$ at a given frequency. The wave propagation is usually controlled by a partial differential equation, whose formulation depends on the characteristics of the propagation model [13, 33]. In the frequency domain, the acoustic propagation of a pressure field $u(x, y, z)$ at the position $(x, y, z) \in \Omega$ in a heterogeneous medium is governed by the Helmholtz equation defined as:

$$-\Delta u(x, y, z) - k^2(x, y, z)u(x, y, z) \;=\; s(x, y, z), \tag{1}$$

where $k(x, y, z) = 2\pi f/m(x, y, z)$ is the wavenumber, $f \in \mathbb{R}^+$ is the frequency in Hz, and $m(x, y, z)$ is the variable acoustic-wave velocity model in m/s. In equation (1), Δ represents the standard Laplace operator and $s(x, y, z)$ a Dirac source term. In a more realistic scenario, the source excitation is estimated by solving an inverse problem (see [46]). The wavelength is given by $m(x, y, z)/f$. We consider absorbing boundary conditions: First, a popular approach, called Perfectly Matched Layer formulation (PML) [4, 5], is used to obtain a satisfactory near boundary solution, without many artificial reflections; Second, this artificial boundary layer is used to absorb outgoing waves at any incidence angle as shown in [4]. The acoustic full-waveform inversion requires the solution of three-dimensional Helmholtz problems at various locations of the Dirac sources and thus leads to multiple right-hand side problems [48, 49].

In this paper, we consider a standard second-order accurate seven point finite-difference discretization of the Helmholtz equation (1) on an uniform equidistant Cartesian grid of size $N_x \times N_y \times N_z$. For later use, we define $N = N_x \times N_y \times N_z$, h the corresponding mesh grid size, and Ω_h the discrete computational domain. After discretization, the acoustic full-wave inversion leads to the following linear system with p multiple right-hand sides:

$$A\,U \;=\; S \tag{2}$$

where $S \in \mathbb{C}^{N\times p}$ and $A \in \mathbb{C}^{N\times N}$ is a sparse complex matrix (nonhermitian and nonsymmetric due to the PML formulation). We note that the matrix A embeds the properties of the subsurface and depends on the propagation velocity model m that we want to quantify. Since a stability condition has to be satisfied to correctly represent the wave propagation phenomena [13], we consider numerical discretization schemes with 10 points per wavelength. Consequently, at a given frequency f in Hz, we deduce the mesh grid size h in m as

$$h \;=\; \frac{\min\limits_{(x,y,z)\in\Omega_h} m(x, y, z)}{10\, f}. \tag{3}$$

In practice this last relation imposes the solution of very large systems of equations (see Section 5, where $N \approx 10^6$) at the frequencies of interest for the geophysicists. Such a task may be too computationally and memory expensive when solving linear systems by sparse direct methods. Due to their indefiniteness, these systems are known to be very challenging for iterative methods [18]. Based on previous studies [11, 12, 26], we consider a recently proposed block flexible Krylov subspace method (BFGMRES-S [11, Algorithm 3]) for the solution of the linear system with multiple right-hand sides (2). In [11] the authors have shown the relevance of the BFGMRES-S algorithm combined with a variable two-level preconditioner to address the solution of such large-scale acoustic forward problems in a distributed memory parallel environment.

We refer the reader to [12, Algorithm 5] for a complete description of the geometric two-grid preconditioner and to [33] for additional theoretical properties in relation with Krylov subspace methods. Applications to large-scale forward acoustic problems have been considered in detail in [26, Chapter 4].

Figure 2 depicts a graphical representation of the SEG[1]/EAGE[2] salt dome velocity model (described later in Section 2.3) and the real part of the wavefield at frequency 12 Hz obtained after solving the Helmholtz equation in the case of a single source. The numerical solution has been obtained using BFGMRES-S. We note that the wave propagation is affected by the properties of the velocity model. The interference of the waves with the reflected layer generates reflection waves. The latter are recorded at different time steps using geophones to generate the so-called *seismograms*, i.e., the observed data for the associated inverse problem detailed next.

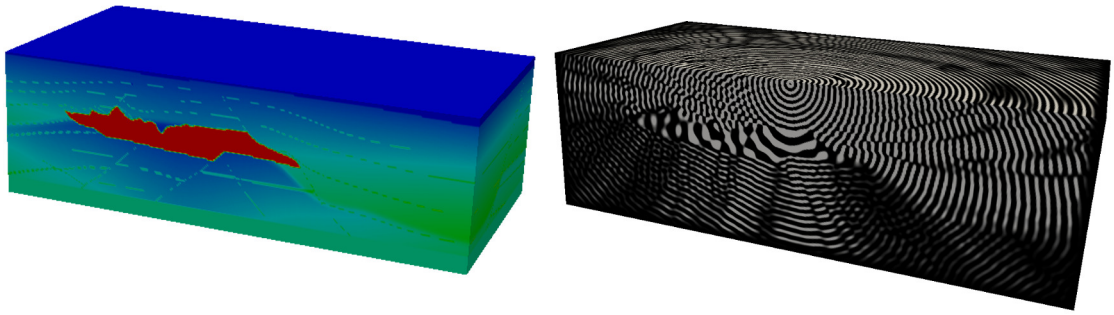


Figure 2: 3D SEG/EAGE salt dome velocity model: problem geometry with velocity distribution (left) and real part of numerical solution at 12 Hz (right). Figures from [26, Chapter 4].

2.2 Full-waveform inversion as a least-squares global optimization problem

The standard formulation of acoustic full-waveform inversion (at a given frequency f) aims at minimizing the following least-squares misfit function [50]:

$$\mathcal{J}(m) \;=\; \frac{1}{2}\sum_{i=1}^{p}(d^i(m)-d^i_{\text{obs}})^\dagger W^i(d^i(m)-d^i_{\text{obs}}), \tag{4}$$

where $\dagger$ denotes the adjoint operator (transpose conjugate). The weighting matrices W^i are in general used to include a priori data information. The misfit vector $d^i(m)-d^i_{\text{obs}} \in \mathbb{R}^n$, related to the i-th source, is computed as the difference at the receiver positions between the

[1]The Society of Exploration Geophysicists.

[2]European Association of Geoscientists and Engineers.

recorded seismic data d^i_{obs} (i.e. seismograms) and the modeled seismic one $d^i(m)$. The latter one corresponds to the modeled seismic wavefield u^i (computed as the i-th column of the U solution of (2)), projected using the operator $\mathcal{P}_{\text{data}}$, which extracts the values of the wavefield at the receiver positions for each source, i.e., $d^i(m) = \mathcal{P}_{\text{data}}(u^i)$. The use of this projection operator makes the full-waveform inversion an ill-posed problem, meaning that an infinite number of velocity models matches the data, leading to the same objective function value. Therefore, an additional regularization term is classically added to the inversion problem to make it well posed [50]. In addition to the velocity model, the source excitation is generally unknown and must be included as an unknown of the problem [35]. Provided that a good starting velocity model m_s is available (good in the sense that smoothly represents the structure of the true velocity model), the minimization of the objective function (4) is in practice solved using a Newton type method (see [52] and references therein).

2.3 The SEG/EAGE salt dome velocity model

We have conducted our numerical studies on the acoustic inverse problem using a three-dimensional public domain velocity model known in the geophysics community as the SEG/EAGE salt dome velocity model.

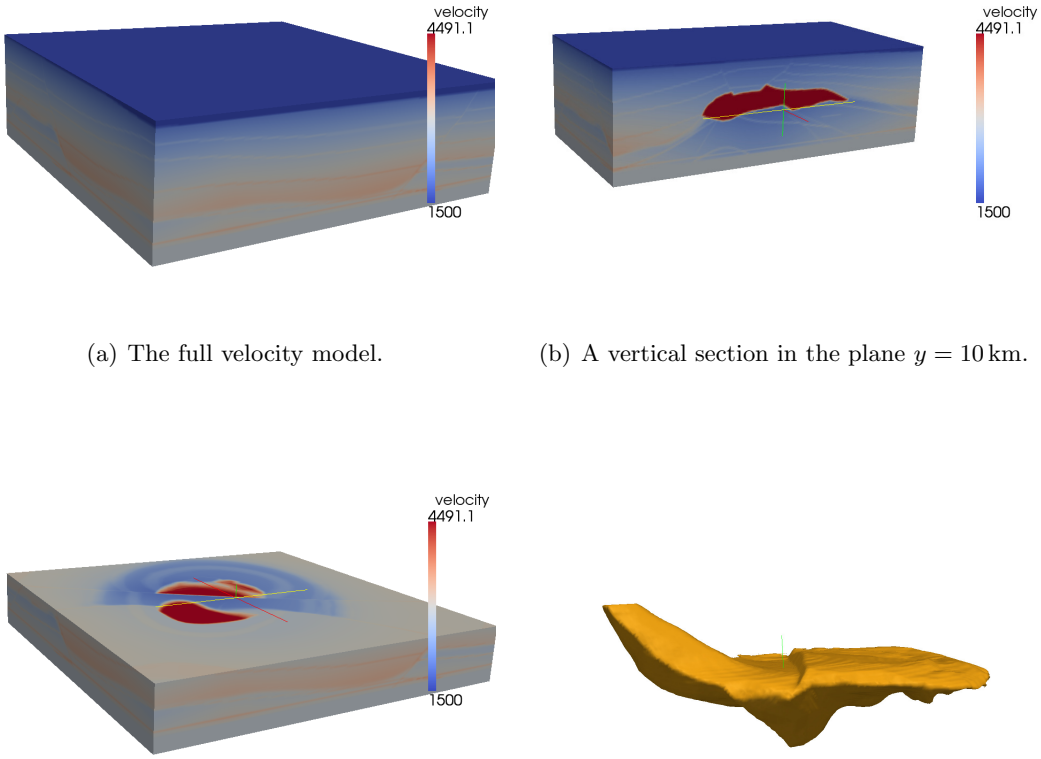
This velocity model (depicted in Figure 3) is based on a typical US Golf coast salt structure, and special care was taken to ensure that it is geologically feasible. Hence it is widely accepted as an adequate benchmark model for seismic imaging in the geophysics community. Next, we will introduce a parametrization procedure to compute an appropriate basis for the velocity models, which will then allows to compute an accurate and smooth representation of the SEG/EAGE salt dome velocity model using a reduced number of parameters.

3 Search space reduction

Evolution Strategies are heuristic methods designed for the solution of global optimization problems (with continuous variables) that have performed well in terms of the quality of the final point computed (see [2, 3, 22, 39]). However, like any other method for global optimization, ES's suffer from the *curse of dimensionality*, meaning that their performance is satisfactory on low dimensional problems, but deteriorates as the dimensionality of the search space increases [29]. For realistic simulations of full-waveform inversion [26, 33], the typical size N of the velocity models exceeds in general 10^6, and thus trying to solve directly the problem using an ES is ruled out.

However, our purpose is not to solve the full-waveform inversion problem but rather to find a good starting velocity model m_s which can be later improved using a local gradient based method. The initial velocity model m_s is only required to represent the general structure of the true model, and such a representation is often smooth and can be expressed using only a few parameters [52]. Given an appropriate and efficient procedure to represent a velocity model using a reduced number of parameters, the ES method will be applied to try to find the values of the parameters that lead to a smooth representation of the unknown velocity model to be inverted.

The problem of search space reduction has been investigated over the past years using subspace approaches [25, 31, 47]. In our context, a velocity model perturbation $\tilde{m} \in \mathbb{R}^N$ can be restricted to lie in an n-dimensional subspace of $\mathbb{R}^N$, spanned by the vectors $\{v_i\}_{i=1,\dots,n}$, with



(a) The full velocity model.

(b) A vertical section in the plane $y = 10\,\text{km}$.

(c) An horizontal section in the plane $z = 2.5\,\text{km}$.

(d) The salt dome.

Figure 3: Visualization of the 3D SEG/EAGE salt dome velocity model using ParaView [24]. The geophysical domain is of size $20 \times 20 \times 5\,\text{km}^3$. The seismic waves propagate in water and in the salt dome at the minimal and maximal velocities of 1500 m/s and 4418 m/s, respectively. The occurrence of a salt dome in the subsurface of Earth abruptly increases the velocity of propagation of the waves.

$n \leq N$ (as with N, n will also later have a 3-D interpretation). The model perturbation can be then written as follows:

$$\tilde{m} \;=\; \sum_{i=1}^{n} g_i v_i \;=\; Vg,$$

where $g \in \mathbb{R}^n$ are the new parameters to invert, and $V = [v_1, \ldots, v_n] \in \mathbb{R}^{N\times n}$ is the so-called *reduction basis*. Subspace approaches lead to an important simplification of the problem [25, 47], but are unfortunately very sensitive to the choice of the reduction basis. In fact, by restricting the search space to directions in a subspace, the neglected ones could be the vectors which are important in finding a local or global minimum of the objective function $\mathcal{J}$ given in (4). Often researchers use a sinusoidal basis as a reduction basis and try to find a vector g of parameters which produces an acceptable agreement to the observation [31]. The existing subspace methods,

previously cited, use gradient information of the objective function, thus the reduction process is problem dependent. In our case, no information about the objective function will be used, thus our subspace technique can be applied regardless of the problem.

Inspired by the methodology used in image compressing, we propose in this work a new procedure to construct this basis using a combination of sinusoidal and rectangular basis functions, more specifically a Discrete Cosine Transform (DCT) [9] and a step function transform. In fact, the step function used to magnify the vector parameter $g \in \mathbb{R}^n$, so that it fits the original space $\mathbb{R}^N$, usually leads to a pixelization effect. Thus a DCT is then applied to produce a smooth velocity model, reducing such pixelization. To simplify the exposition, we will first explain our approach in the one-dimensional case, and then give a generalization to cover the realistic 3D geometry of interest. We refer the reader to [15, Chapter 7] for a complete description of these procedures.

3.1 One-dimensional space reduction

There are three main procedures in our space reduction scheme: a reduction, a duplication, and a magnification.

In the **reduction procedure**, given a vector $m \in \mathbb{R}^N$ discretizing a velocity vector, n subdivisions are first created. Then, it is taken the mean value the parameters included in each subdivision. In our context, the reduction operation will be only used to estimate how efficient is our magnification procedure.

The **duplication procedure** consists of building a vector $m \in \mathbb{R}^N$ using a small-size vector $g \in \mathbb{R}^n$ with $n < N$. We first construct an empty vector m of size N, considering n subdivisions of indices $[x_i, x_{i+1}]$. Each subdivision contains around $\delta = [N/n]$ parameters and thus $\delta = x_{i+1} - x_i$ and $x_i = (i-1)\delta + 1$. The n parameters of the velocity vector g are distributed over the n subdivisions. The value associated to each subdivision is then duplicated over the δ parameters of m assigned for that subdivision. The duplication procedure, as presented here, introduces a pixelization effect in the constructed vector m. A DCT can then be applied to improve the quality of the duplication and to remove the subdivision discontinuities.

A **magnification procedure** aims in general at removing noise or producing a less pixelated image. The most used smoothing algorithms are based on Gaussian smoothing [1], bilateral filters [51], and sinusoidal based approaches [9]. As a smoothing procedure, we choose to work with a sinusoidal basis since it is one of the most popular subspace approaches for FWI to generate a smooth approximation vector using few coefficients [31]. We will smooth the pixelization effect in the magnified vector using a DCT [9].

As we have seen before, when duplicating a velocity vector from a vector of smaller size, the value in each subdivision is constant and thus it can be seen as a mean of all the subdivision values. Such a property will be imposed as well on the magnified velocity vector m in the sense that it corresponds to a model $m(\cdot)$ such that

$$\frac{1}{x_i - x_{i+1}} \int_{x_{i+1}}^{x_i} m(x)dx \;=\; g(i), \quad i = 1, \ldots, n. \tag{5}$$

In turn, this velocity model $m(\cdot)$ is expressed using a discrete cosine basis in the following way:

$$m(x) \;=\; \sum_{j=1}^{n} a_j \cos\left(\frac{(j-1)\pi}{N}(x-1)\right), \tag{6}$$

where $a = (a_j)_{1 \le j \le n} \in \mathbb{R}^n$. By incorporating the expression (6) into the equations (5), we can obtain the vector a by solving a linear system of the form

$$Ca \;=\; g, \tag{7}$$

where $C \in \mathbb{R}^{n \times n}$ is a matrix with the following coefficients:

$$C_{ij} \;=\; \begin{cases} 1 & \text{if } j = 1, \\ \frac{2N}{(j-1)\pi\delta} \cos\left(\frac{\pi}{N}(j-1)(i-\frac{1}{2})\delta\right) \sin\left(\frac{\delta\pi}{2N}(j-1)\right) & \text{otherwise.} \end{cases}$$

The matrix C is nonsingular (see the proof in the appendix). The one-dimensional smoothed vector m is then built by evaluating (8) for all $i \in \{1, \ldots, N\}$, which amounts to

$$m_i \;=\; \sum_{j=1}^{n} a_j \cos\left(\frac{(j-1)(i-1)\pi}{N}\right),$$

or, equivalently, to

$$m \;=\; Mg, \tag{8}$$

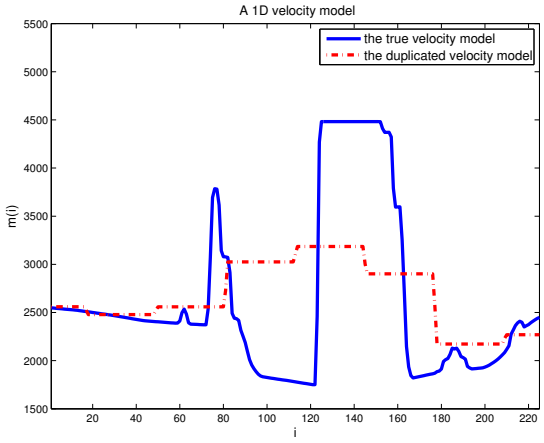
where $M = KC^{-1} \in \mathbb{R}^{N \times n}$ and $K \in \mathbb{R}^{N \times n}$ is the matrix defined by $K_{ij} = \cos(\frac{(j-1)(i-1)\pi}{N})$. Equation (8) shows that the magnification procedure corresponds to the application of a linear operator. The magnification cost is negligible compared to an objective function evaluation. In fact, the magnification is accomplished by first computing $C^{-1}g \in \mathbb{R}^n$ and then multiplying it by $K \in \mathbb{R}^{N \times n}$, resulting in $K(C^{-1}g) \in \mathbb{R}^N$, while an evaluation of the objective function requires the solution of a larger linear system of size $N \times N$ with p multiple right-hand sides (see (2)).

3.2 Three-dimensional space reduction

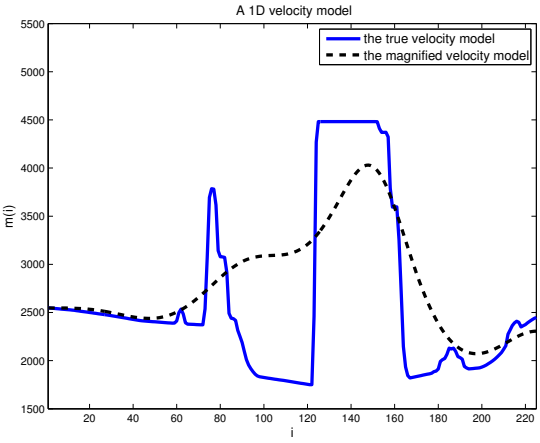
A multi-dimensional transform (known as Fast Direct multi-dimensional DCT) can be carried out by using a composition of the one-dimensional magnification procedure along each dimension [9]. Equation (8) can then be immediately extended to 2D or 3D velocity models. A detailed description of the extension of equation (8) to higher dimensions is given in [9]. In the case of three-dimensional data, given a 3D velocity model G of $n = n_x \times n_y \times n_z$ parameters, we ought to build a magnified 3D velocity model m of size $N = N_x \times N_y \times N_z \gg n$ parameters. The magnification procedure is obtained by applying (8) first to the x axis, then to y, and finally to z as follows (using Matlab notation):

$$\begin{aligned} V(:,:,k) &= M_x G(:,:,k), \quad k = 1, \ldots, n_z, \\ T(:,:,k) &= V(:,:,k) M_y^\top, \quad k = 1, \ldots, n_z, \\ m(i,:,:) &= T(i,:,:) M_z^\top, \quad i = 1, \ldots, N_x, \end{aligned}$$

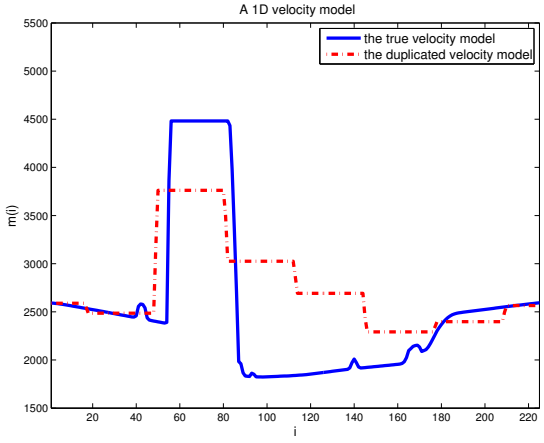
where $M_x \in \mathbb{R}^{N_x \times n_x}$, $M_y \in \mathbb{R}^{N_y \times n_y}$, and $M_z \in \mathbb{R}^{N_z \times n_z}$ are the one-dimensional smoothing matrices defined in (8) along the axes x, y, and z, respectively.



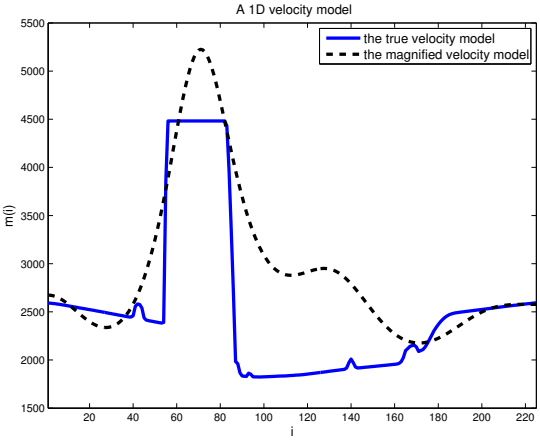
(a) The duplication procedure following x axis.



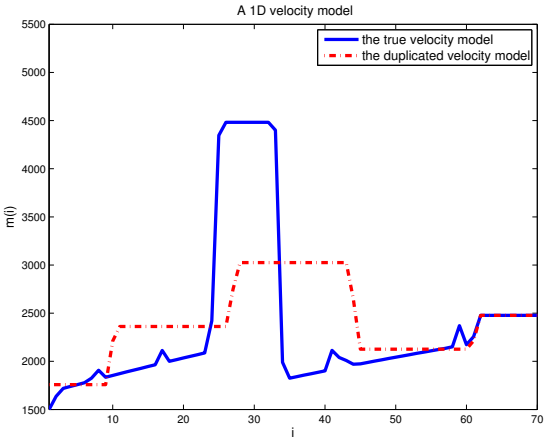
(b) The magnification procedure following x axis.



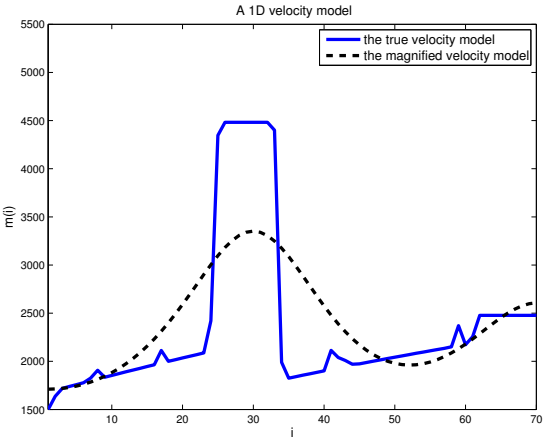
(c) The duplication procedure following y axis.



(d) The magnification procedure following y axis.



(e) The duplication procedure following z axis.



(f) The magnification procedure following z axis.

Figure 4: Illustration using $1D$ SEG/EAGE salt dome velocity profiles.

3.3 Application to the SEG/EAGE salt dome velocity model

To illustrate numerically the performance of the three-dimensional approximation procedure, we have used the SEG/EAGE salt dome velocity model introduced in Section 2.3.

Figure 4 outlines three one-dimensional SEG/EAGE salt dome velocity profiles (with respect to the x, y, and z axes, respectively). Reduction procedures were then applied to these velocity profiles to create vectors for duplication and magnification. Following the x axis (resp. y axis), the velocity profile has been selected at the position $y = 8.9\,\text{km}$ and $z = 2.5\,\text{km}$ (resp. $x = 8.9\,\text{km}$ and $z = 2.5\,\text{km}$). In both cases the reduced velocity vector is built using $n = 8$ parameters only, while $N = 225$ parameters are required for the true velocity vector. Following the z axis, the velocity profile (selected at the position $x = 8.9\,\text{km}$ and $y = 8.9\,\text{km}$) is reduced using $n = 5$ parameters compared to the true $N = 70$ ones. From the results obtained, we observe that the smoothing effect of the DCT transform improves the quality of the duplicated velocity profiles as a representation of the original ones, and that the space reduction approach can work relatively well with a reduced number of parameters.

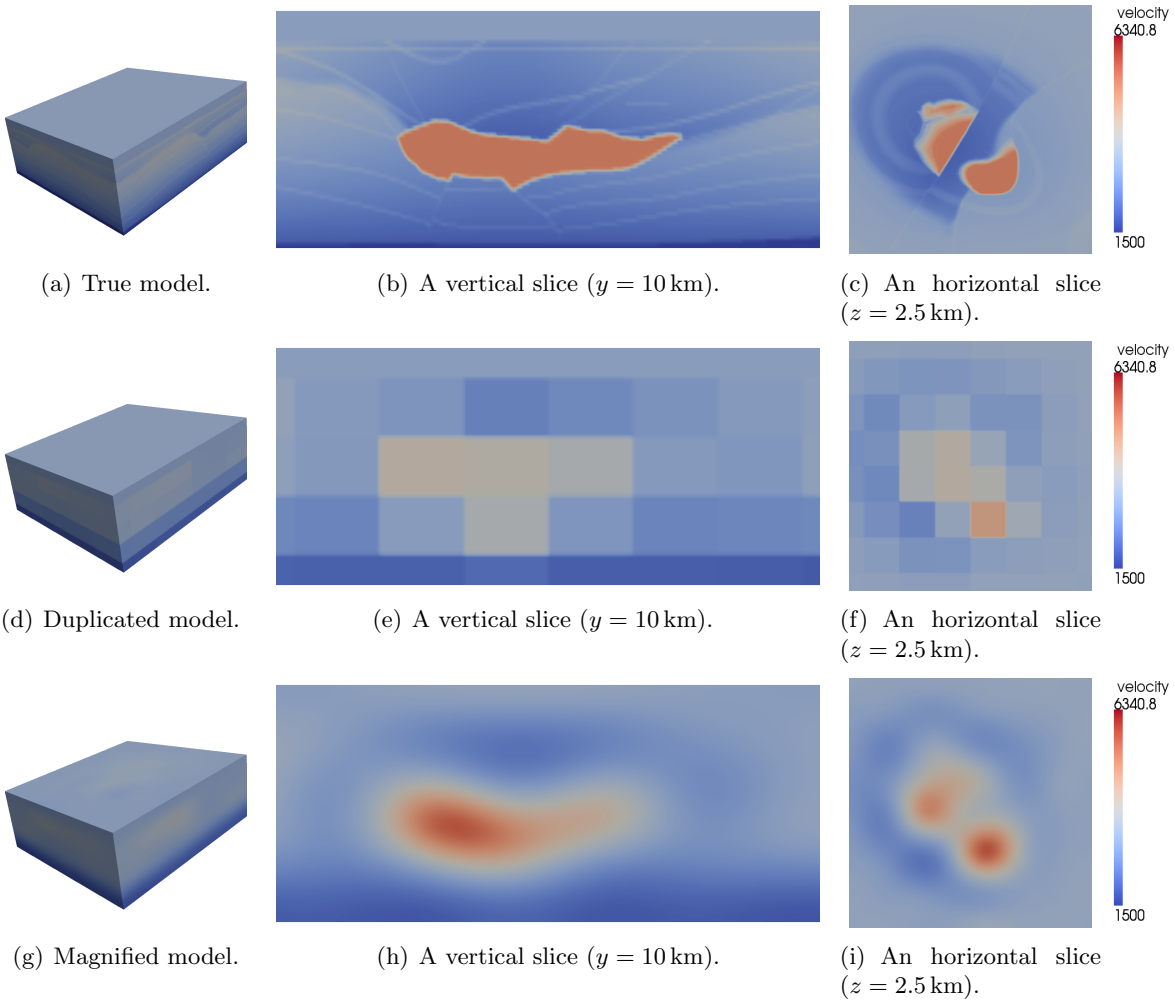
For the three dimensional case, we have found that the true velocity model can be relatively well approximated using $n = 8 \times 8 \times 5 = 320$ parameters instead of the original $N = 225 \times 225 \times 70 = 3543750$ ones, in the sense of still representing the main structure (i.e., the salt dome) of the true velocity model. Figure 5 outlines an illustration of the obtained results. As expected, the magnification procedure using DCT (see Figures 5(g)–5(i)) gives better results compared to the duplication procedure which is based on the step function transform (see Figures 5(d)–5(f)). Although we use only a few parameters to represent the velocity model, our smooth magnification procedure preserves the main specificity of the true model, in particular the salt dome.

4 A parallel ES for acoustic full-waveform inversion

In this section we start by briefly reviewing the existing methods to compute a satisfactory initial velocity model for seismic inversion. Then we explain how to apply ES's for this purpose when using the space reduction introduced before. A parallel implementation of the resulting ES's is also proposed.

4.1 Existing methods

The acoustic full-waveform inversion problem introduced in Section 2.2 is nonconvex, and thus its solution by optimization algorithms crucially depends on the starting velocity model m_s. In fact, it is known that the inversion procedure converges to satisfactory results only if the starting velocity model is situated not far from a global minimizer [52]. Hence, before applying the full-waveform inversion, a starting model is generally built. To do this, the most common techniques are first-arrival travel-time tomography (FATT) [30], stereotomography [27] or, more recently, inversion in the Laplace domain [41]. For many years, FATT has proven to be stable in generating smooth velocity models of the subsurface. Some examples of application of FWI to real data using a starting model built by FATT are described in [32, 37]. Similarly, the stereotomography is regarded as one of the most promising methods for building a smooth velocity model. It exploits the arrival time of locally coherent events within an automatic procedure to select a seismogram collection [27]. Some applications to synthetic and real data sets are shown in [7, 8]. Finally,



(a) True model. (b) A vertical slice ($y = 10\,\text{km}$). (c) An horizontal slice ($z = 2.5\,\text{km}$).

(d) Duplicated model. (e) A vertical slice ($y = 10\,\text{km}$). (f) An horizontal slice ($z = 2.5\,\text{km}$).

(g) Magnified model. (h) A vertical slice ($y = 10\,\text{km}$). (i) An horizontal slice ($z = 2.5\,\text{km}$).

Figure 5: 3D duplicated and magnified models of the SEG/EAGE salt dome velocity model. The velocity models are built using $n = 8 \times 8 \times 5 = 320$ parameters, while $N = 225 \times 225 \times 70 = 3543750$ parameters are required for the true velocity model.

the inversion in the Laplace domain can be viewed as a frequency domain inversion using a pure imaginary complex valued frequency that controls the time damping of the seismic wavefield. Applications of Laplace domain FWI to synthetic and real data are proposed in [41, 42, 43].

4.2 Evolution strategies and CMA-ES

Evolutions strategies are a class of evolutionary algorithms designed for the optimization of a possibly nonconvex function in a continuous domain without using derivatives. It has been originally developed in [38] for the unconstrained optimization of a function, $\min_{v \in \mathbb{R}^n} f(v)$, and has been extensively investigated and tested (see, e.g., [6, 21] and the references therein). We are interested in a large class of ES's denoted by (μ, λ)–ES, with μ and λ integers such that $1 < \mu \leq \lambda$, where a certain number λ of points are randomly generated in each iteration, among which μ of them are selected as the best in terms of the objective function f.

CMA-ES [23] (where CMA stands for Covariance Matrix Adaptation) is regarded as one of the best in the class (μ, λ)–ES in terms of numerical performance [2, 3, 22, 39]. More precisely, CMA-ES belongs to the ES family denoted by $(\mu/\mu_W, \lambda)$–ES, where the subscript 'W' indicates the use of 'recombination' via weights. Broadly speaking, at iteration k, a candidate minimizer $\bar{x}_k$ is used to produce a generation of λ 'offspring', each consisting of adding to $\bar{x}_k$ a random direction multiplied by a parameter controlling the length or size of the steps; the best μ of these are retained as 'parents' in a 'selection', and $\bar{x}_{k+1}$ is taken as a weighted combination ('recombination') of these parents. The CMA component considers a Gaussian distribution of mean zero for the random generation of the directions and provides a scheme for updating the corresponding covariance matrix as well as the step length.

4.3 A modified CMA-ES

CMA-ES has exhibited robust performance for difficult ill-conditioned, non-separable, and highly multi-modal problems [2, 3, 22, 39]. Its main drawback is, however, that a large budget is required to provide outstanding results. Recently, the authors in [16] have proposed modifications to the class of algorithms in $(\mu/\mu_W, \lambda)$–ES to make them enjoying a favorable convergence property and performing better for smaller budgets. The modifications have been essentially the imposition of a sufficient decrease on the objective function values to accept new iterates and the reduction of the step size when such a condition is not satisfied. Under such modifications, these ES's can converge globally (meaning independently of the starting point) for a first-order stationary point. Algorithm 1 shows an adaptation of the globally convergent ES's proposed in [16] to the context of full-waveform inversion.

The authors in [16] have proposed three different globally convergent ES versions named mean/mean, max/max, and max/mean. On a large data set of problems, the mean/mean version has performed numerically the best. However, the incorporation of the mean/mean sufficient decrease condition requires an extra objective function evaluation $\mathcal{J}(m_{k+1}^{trial})$ at each iteration, where m_{k+1}^{trial} is the trial mean parent computed as the mean of the best μ generated velocity models. The mean/mean version would therefore corrupt the parallel nature of ES's. In fact, if one supposes that the evaluation of $\mathcal{J}$ at the offspring is performed at the same time using synchronized parallel clusters, the mean parent evaluation $\mathcal{J}(m_{k+1}^{trial})$ will force all these clusters to wait until the end of such an evaluation to be able to restart a new offspring generation. Alternatively, the max/max version has shown good performance (not as good as the mean/mean version) without the need of any extra objective function evaluation to impose the corresponding sufficient decrease condition. Consequently, we have adopted the max/max version in Algorithm 1.

We also note that the update of the weights to enforce the sufficient decrease condition, originally proposed in [16], has not been activated in our setting since we aim at the least amount of changes in the original ES's and since such an update did not seem to have a real impact on the results for the max/max version (see [16]).

Moreover, in the evaluation procedure of the objective function, one needs to satisfy the relation (3). Hence we have imposed a lower bound on the velocity equal to a known minimum value $m_{min} = 1500m/s$ (the velocity model value of the water). A maximum value on the velocity model of $m_{max} = 4500m/s$ has been also imposed to avoid propagation by meaningless velocity models. Both requirements were guaranteed by projecting the offspring models onto the feasible domain defined by the bounds, an approach that has been shown to be globally

convergent in [17]. Since there are no other constraints rather than simple bounds on the variables one can use the simple orthogonal ℓ_2-projection.

Algorithm 1: An adaptation of the ES algorithm to FWI setting.

Initialization: Choose positive integers λ and μ such that $\lambda \geq \mu$. Select an initial $\bar{x}_0 \in \mathbb{R}^n$, generate a velocity model $m_0 \in \mathbb{R}^N$ (using the magnification procedure) and evaluate $\mathcal{J}(m_0)$. Choose initial step lengths $\sigma_0, \sigma_0^{\mathrm{ES}} > 0$ and initial weights $(\omega_0^1, \ldots, \omega_0^\mu)$ in a simplex S. Let $\mathcal{C}_0$ be a normal distribution with the identity matrix as covariance matrix. Choose the constants $\beta_1, \beta_2, d_{\min}, d_{\max}$ such that $0 < \beta_1 \leq \beta_2 < 1$ and $0 < d_{\min} < d_{\max}$. Select a forcing function $\rho(\cdot)$. Set $k = 0$.

Until some stopping criterion is satisfied:

1. **Generation of velocity models:** Generate λ velocity models $M_{k+1} = \{m_{k+1}^1, \ldots, m_{k+1}^\lambda\}$ using the magnification procedure (imposing the bounds m_{min} and m_{max} by projection) based on the sample points $Y_{k+1} = \{\bar{y}_{k+1}^1, \ldots, \bar{y}_{k+1}^\lambda\}$ such that

$$\bar{y}_{k+1}^i \;=\; \bar{x}_k + \sigma_k d_k^i,$$

where $d_k^i \in \mathbb{R}^n$ is drawn from the distribution $\mathcal{C}_k$ and obeys $d_{\min} \leq \|d_k^i\|_2 \leq d_{\max}$, $i = 1, \ldots, \lambda$.

2. **Parent selection:** Evaluate $\mathcal{J}(m_{k+1}^i)$, $i = 1, \ldots, \lambda$, and reorder the offspring points in $Y_{k+1} = \{\tilde{y}_{k+1}^1, \ldots, \tilde{y}_{k+1}^\lambda\}$ by increasing order: $\mathcal{J}(\tilde{m}_{k+1}^1) \leq \cdots \leq \mathcal{J}(\tilde{m}_{k+1}^\lambda)$.
Select the new parents as the best μ offspring sample points $\{\tilde{y}_{k+1}^1, \ldots, \tilde{y}_{k+1}^\mu\}$.

3. **Imposing sufficient decrease:**

If $\mathcal{J}(\tilde{m}_{k+1}^\mu) \leq \mathcal{J}(m_k^\mu) - \rho(\sigma_k)$, then consider the iteration successful. Compute the weighted mean of the new parents

$$\bar{x}_{k+1}^{trial} \;=\; \sum_{i=1}^{\mu} \omega_k^i \tilde{y}_{k+1}^i.$$

Magnify $\bar{x}_{k+1}^{trial}$ to obtain the velocity model m_{k+1}^{trial}. Set $\bar{x}_{k+1} = \bar{x}_{k+1}^{trial}$, $m_{k+1} = m_{k+1}^{trial}$, and $\sigma_{k+1} \geq \sigma_k$ (for example $\sigma_{k+1} = \max\{\sigma_k, \sigma_k^{\mathrm{ES}}\}$). Set also $m_{k+1}^\mu = \tilde{m}_{k+1}^\mu$.

Otherwise, consider the iteration unsuccessful, set $\bar{x}_{k+1} = \bar{x}_k$, $m_{k+1} = m_k$ and $\sigma_{k+1} = \bar{\beta}_k \sigma_k$, with $\bar{\beta}_k \in (\beta_1, \beta_2)$. Set $m_{k+1}^\mu = m_k^\mu$.

4. **ES updates:** Update the ES step length $\sigma_{k+1}^{\mathrm{ES}}$, the distribution $\mathcal{C}_k$, and the weights $(\omega_{k+1}^1, \ldots, \omega_{k+1}^\mu) \in S$. Increment k and return to Step 1.

4.4 A parallel implementation

The proposed ES implementation consists of a synchronized parallel optimizer composed of λ clusters (typically, the population size). Each cluster is composed of a group of processors, which is assigned to evaluate the objective function (4). At a given iteration k, the clusters are synchronized and not activated until the new mean parent m_{k+1} is defined, which in turn depends on the iteration state (successful or not). Figure 6 sketches in detail a given iteration of our proposed parallel implementation of Algorithm 1.

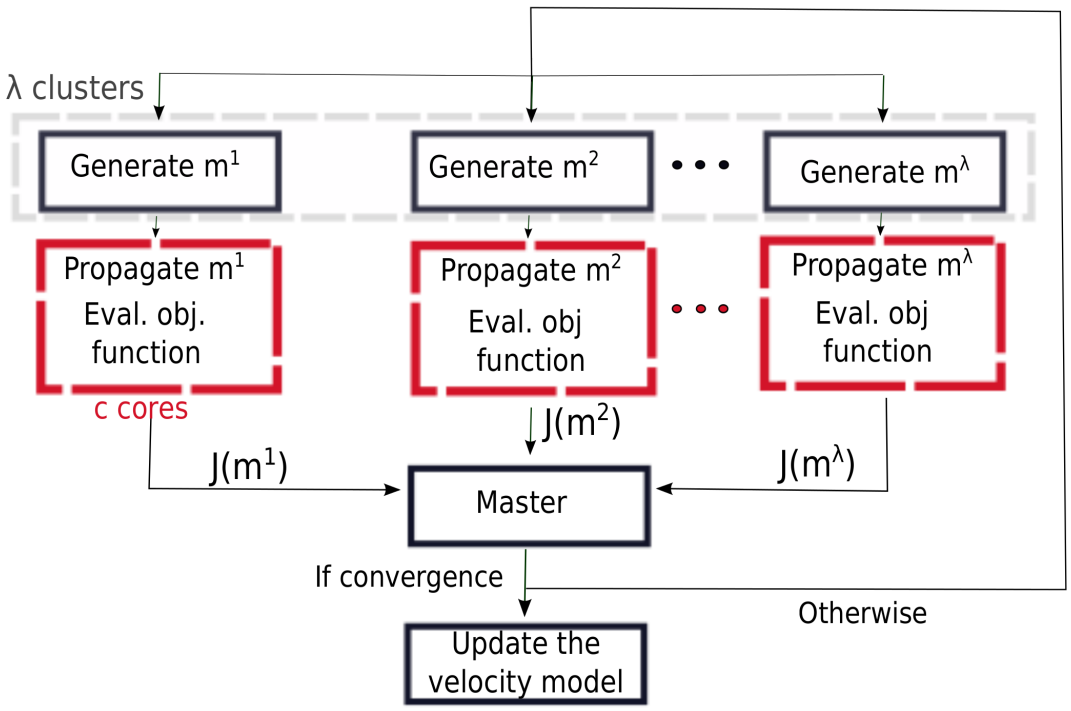


Figure 6: A graphical illustration for a given iteration of the parallel evolution strategy.

The proposed parallel implementation can be described as follows. In addition, at each iteration, λ clusters, represented in Figure 6 by the components [`Generate` m^i], are launched in a synchronous manner. Each of these clusters generates a reduced velocity model based on the ES parameters and strategies. Once the velocity model is generated, the related cluster evokes the component [`Propagate` m^i].

The wave propagation simulation related to each velocity model considers the p source terms (i.e., p right-hand sides) at once. The [`Propagate` m^i] component is based on MPI (Message Passing Interface) and is responsible for discretizing and generating the linear system related to the forward problem (see Section 2.1) and to provide the information needed to evaluate the objective function $\mathcal{J}$ given in (4). The last component will just return the value of the objective

function to the unique master processor [`Master`]. Once the master processor has received results from the λ clusters, it will choose the best, decide whether the iteration is successful or not, and update the mean parent accordingly. The [`Master`] component updates also the ES parameters (e.g., the distribution and the step length) and repeats the loop until a convergence criterion is achieved.

The propagation in itself behaves as a black box process, hiding from the ES the complexity of the discretization and of the forward solver. Changing the discretization and/or forward solver parameters will not incur in any rewriting of the ES implementation. MPI-2 has been used with the `MPI_COMM_SPAWN` primitive which allows an MPI process to spawn a number of clusters. Each newly spawned cluster has a specific `MPI_COMM_WORLD` intracommunicator that allows us to launch easily the propagation simulations over a number of CPU cores. We note that the proposed ES implementation is portable and that the propagator itself can be a standalone client. When the available cluster number is less than λ, one can launch many propagation simulations on the same cluster until we obtain the needed function evaluations.

5 Numerical experiments

In this section we first describe the validation scenario and detail the parameters of the global optimization algorithm. Then we analyze the numerical results obtained for the inversion procedure at two frequencies. The numerical simulations have been performed on CURIE [3], a parallel cluster located at TGCC, France (two eight-cores Intel Sandy Bridge EP (E5-2680) at 2.7 GHz and 64 GB RAM per computing node with InfiniBand QDR Full Fat Tree interconnect) using a Fortran 2003 implementation with MPI in single precision arithmetic. The code has been compiled by the Intel ifort compiler suite with standard compiling options and linked with the MKL library. The numerical experiments have been performed on a fixed number of cores (2048, i.e., 128 computing nodes) with a maximal allowed elapsed computing time of 24 hours.

5.1 Problem and algorithm specifications

We have used a simple scenario where the p source terms are supposed to be Dirac functions and where the observed data d^i_{obs} (i.e. seismograms [50]) are generated from the propagating velocity model that we are trying to invert (see Section 2.3). The p sources (with $p = 16$) are uniformly distributed in a survey plan located at 500 meters of depth (10% of the exploration depth).

In this scenario we consider the acoustic full waveform inversion procedure at two different frequencies (1 Hz and 2 Hz respectively) with only 320 parameters for the initial velocity model (see Section 3). Table 1 reports the dimensions $N_x \times N_y \times N_z$ of the forward problem in agreement with the stability condition (3), the number of clusters, as well as the population size λ versus the frequency. As the frequency increases, the number of cores dedicated to the objective function evaluation becomes larger. In fact, the forward problem gets more complicated to solve as far as the frequency f increases. The initial iterate $\bar{x}_0$, for the parallel ES algorithm in the case of 1 Hz, is built using the magnification procedure based on two given velocity values ($3000m/s$ and $1500m/s$, distributed as follows $2 = 1 \times 1 \times 2$). Both the nonlinearity and the ill-posedness of the FWI problem are in practice tackled in the frequency domain using a multi-scale approach

[3] http://www-hpc.cea.fr/fr/complexe/tgcc-curie.htm

where one starts the inversion procedure in a low frequency range to mitigate the nonlinearity of the inversion, and then incorporate progressively higher frequencies [44, 52]. Hence the velocity model solution obtained in the case of 1 Hz will serve as a warm-starting point for the 2 Hz case.

In the context of CMA-ES, the choice $\lambda = \text{floor}(4 + 3\log(n))$ (where floor$(\cdot)$ rounds to the nearest integer no larger than the number given) has been shown to provide a good compromise between quality of the solution found and effort in determining it (see [20]). However, given the strong need for global exploration at 1 Hz, we have chosen an even larger value, equal to $\lambda = 512$, to better take advantage of the number of cores and cores per cluster, as we explain next. Since doubling the frequency doubles the number of discretization points in each direction, the number of cores per cluster is multiplied by at least a factor of eight when considering the case of 2 Hz. A factor of 16 has been retained in these numerical experiments. Given that the total number of cores is fixed to 2048 in both experiments, we note that the number of clusters at 2 Hz is reduced from $2048/8 = 256$ to $2048/128 = 16$. Since the population size λ is chosen as twice the number of available clusters, we had to run two evaluations on the same cluster to get the entire offspring population (for both frequencies).

Frequency	$N_x \times N_y \times N_z$	Number of clusters	Population size λ
1 Hz	$136 \times 136 \times 34$	256 (8 cores/cluster)	512 (2 evaluations/cluster)
2 Hz	$272 \times 272 \times 68$	16 (128 cores/cluster)	32 (2 evaluations/cluster)

Table 1: The distribution of the clusters and the population size versus frequency.

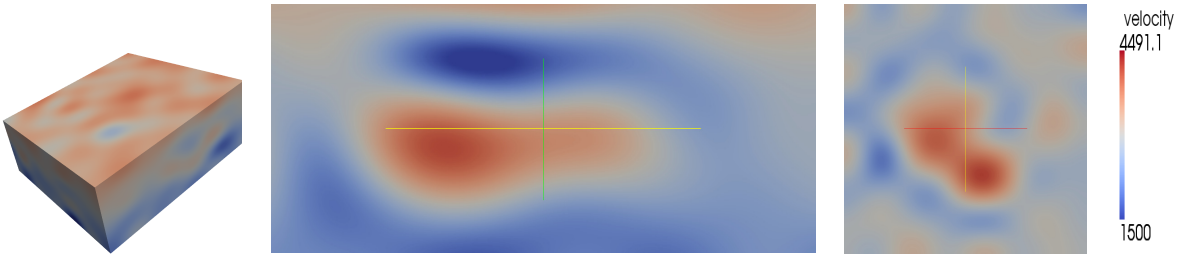
The parameters of the optimization algorithm are chosen similarly to those of CMA-ES for unconstrained optimization (see [20]): $\mu = \text{floor}(\lambda/2)$ and $\omega_0^i = a_i/(a_1 + \cdots + a_\mu)$ with $a_i = \log(\lambda/2 + 1/2) - \log(i)$, $i = 1, \ldots, \mu$. The choices of the distribution $\mathcal{C}_k$ and of the update of σ_k^{ES} also followed CMA-ES for unconstrained optimization (see [20]). The selected forcing function was $\rho(\sigma) = 10^{-4}\sigma^2$. To reduce the step length in unsuccessful iterations we have used $\sigma_{k+1} = 0.5\sigma_k$ which corresponds to set $\beta_1 = \beta_2 = 0.5$. Finally, the initial step size σ_0 is set to half of the difference between the velocity value on the bottom and the one on the top, given (in SEG/EAGE salt dome velocity model) by 3000 m/s and 1500 m/s, respectively.

5.2 Numerical results for full-waveform inversion

Figure 7 reports a graphical representation of the inverted velocity model at 1 Hz. These numerical results are found to be satisfactory since a smooth version of the true velocity model is obtained. We can indeed consider that we are able to invert the general structure of the regarded velocity model, since in particular the salt dome structure is recovered. We also remark that such a model can be considered as a good starting point when using gradient based methods [52].

Figure 9(a) depicts the values of the objective function at the best population point at each iteration for the case 1 Hz. The variation of the objective function is more significant only at the early stages of the inversion process. Such a behavior is due to the sufficient decrease condition which monitors the quality of the sampling procedure. We note that the objective function value decreases from 2042.113 to 575.8082, and that after 278 iterations the inversion procedure is stopped due to the maximal elapsed computational time.

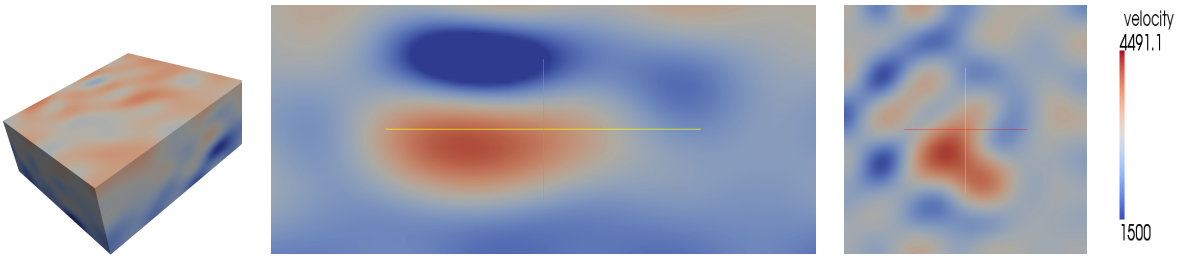
Figure 8 shows the results obtained for $f = 2$ Hz. Even though the optimization procedure is started from the inverted velocity model obtained at 1 Hz, we note that the inversion result



(a) The inverted model. (b) A vertical slice ($y = 10\,\text{km}$). (c) An horizontal slice ($z = 2.5\,\text{km}$).

Figure 7: Inversion results for the salt dome velocity model using $n = 320$ parameters. The 1 Hz case.

is getting less accurate and rather far from being a good approximation of the targeted velocity model.



(a) The inverted model. (b) A vertical slice ($y = 10\,\text{km}$). (c) An horizontal slice ($z = 2.5\,\text{km}$).

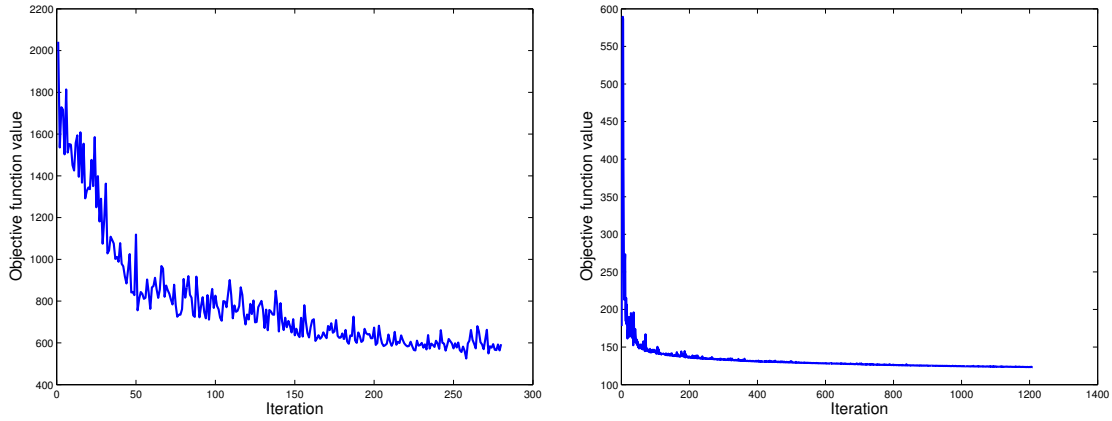
Figure 8: Inversion results for the salt dome velocity model using $n = 320$ parameters. The 2 Hz case.

Figure 9(b) shows that the optimization procedure is converging to a fixed velocity model with an objective function value of 123.2841 after 1220 iterations. Unlike the 1 Hz case, this plot at 2 Hz indicates that the inversion process is getting stacked at a local minimum. The explanation of such results is that the objective function becomes more and more noisy and multi-modal as far as the frequency increases [44]. A possible way to overcome such a difficulty is to increase the population size of the ES in order to encourage the global exploration.

6 Conclusions

In this paper we have presented a numerical method based on evolution strategies for the solution of the acoustic full-waveform inversion problem arising in Earth imaging in geophysics. Considering that each parameter in the velocity model to be inverted is an additional variable to optimize, we have proposed a new parametrization of the problem, reducing the number of parameters needed to faithfully representing the velocity models. We have illustrated the efficiency of our parametrization on the public domain SEG/EAGE salt dome velocity model, where we were able to reconstruct a rather satisfactory approximation using very few parameters.

We have then adapted the chosen ES's to a reduction of the original space, in particular



(a) 1 Hz. (b) 2 Hz.

Figure 9: Objective function values at the best population point during the application of Algorithm 1 for FWI.

to the full-waveform inversion setting when using reduced parameterized velocity models. The main purpose of the incorporation of ES's in the inversion procedure is to find a good starting velocity model without the need of sophisticated a priori knowledge on the background velocity model.

Given the high cost of the problem function evaluations, a highly parallel scheme for such ES's has been derived and validated on parallel distributed memory platforms. The parallel implementation has been tested in combination with the new model parametrization. The obtained numerical results have shown that a great improvement can be obtained in the automation of the inversion procedure.

The testing scenario used for the acoustic full-waveform inversion problem is a simple one where the source excitations are known and the observed data (i.e. seismograms) is generated using the propagating velocity model which we are trying to invert. We note that such a scenario is not truly realistic for the following two reasons: (a) the source excitation is generally unknown and must be rather included as an unknown of the problem, and (b) the observed data is generally given by geophones located at the surface of the exploration domain. Thus, a more realistic test case has to be investigated in the future to fully understand the potential of the proposed approach.

Finally we remark that our model reduction procedure can certainly be generalized to cover other fields of applications. Similarly, the developed parallel approach can also be applied to other optimization problems in geosciences (e.g. well placement in reservoir modeling or analysis of permeability of rocks in petrophysics).

Acknowledgments The authors would like to acknowledge GENCI (Grand Equipement National de Calcul Intensif) for letting us used the CURIE computer at CCRT, Bruyères-le-Châtel, France. This work was granted access to the HPC resources of CCRT under allocation 2014065068 made by GENCI.

Appendix

Theorem 6.1 *Given $n \in \mathbb{N}, N \in \mathbb{N}$ with $n < N$ and $\delta = [N/n]$, the coefficient matrix $C \in \mathbb{R}^{n \times n}$ defined as*

$$C_{ij} \;=\; \frac{2N}{(j-1)\pi\delta} \cos\left(\frac{\pi}{N}(j-1)(i-\frac{1}{2})\delta\right) \sin\left(\frac{\delta\pi}{2N}(j-1)\right)$$

is nonsingular. (Note that for $j = 1$ we have used the convention $sin(0)/0 = 1$, i.e., $C_{i1} = 1$.)

Proof.
A direct calculation shows that

$$\det(C) \;=\; \prod_{j=1}^{n} \frac{2N}{(j-1)\pi\delta} \sin\left(\frac{\delta\pi}{2N}(j-1)\right) \det(M),$$

where $M \in \mathbb{R}^{n \times n}$ is defined as $M_{ij} = \cos(\frac{\pi}{N}(j-1)(i-\frac{1}{2})\delta)$. The nonsingularity of M can be deduced from the properties of DCT-III [9].

■

References

[1] G. Aditya, B. Akhilesh, and C. Kuntal. A comprehensive review of image smoothing techniques. *Inter. J. of Advanced Research in Computer Engineering & Technology*, 1:315–319, 2012.

[2] A. Auger, D. Brockhoff, and N. Hansen. Benchmarking the local metamodel CMA-ES on the noiseless BBOB'2013 test bed. In *Proceedings of the 15th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '13 Companion, pages 1225–1232, New York, NY, USA, 2013. ACM.

[3] A. Auger, N. Hansen, Z. J. Perez Zerpa, R. Ros, and M. Schoenauer. Experimental comparisons of derivative free optimization algorithms. In *8th International Symposium on Experimental Algorithms*, volume 5526 of *Lecture Notes in Computer Science*, pages 3–15. Springer Verlag, 2009.

[4] J.-P. Berenger. A perfectly matched layer for absorption of electromagnetic waves. *J. Comp. Phys.*, 114:185–200, 1994.

[5] J.-P. Berenger. Three-dimensional perfectly matched layer for absorption of electromagnetic waves. *J. Comp. Phys.*, 127:363–379, 1996.

[6] H.-G. Beyer and H.-P. Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1:3–52, 2002.

[7] F. Billette, S. Le Bégat, P. Podvin, and G. Lambaré. Practical aspects and applications of 2D stereotomography. *Geophysics*, 68:1008–1021, 2003.

[8] F. Billette and G. Lambaré. Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophys. J. Int.*, 135:671–690, 1998.

[9] V. Britanak, P.C. Yip, and K.R. Rao. *Discrete Cosine and Sine Transforms.* Academic Press, Oxford, 2006.

[10] R. Brossier. *Imagerie Sismique à Deux Dimensions des Milieux Visco-élastiques par Inversion des Formes d'Ondes : Développements Méthodologiques et Applications.* PhD thesis, Université de Nice-Sophia Antipolis, 2009.

[11] H. Calandra, S. Gratton, R. Lago, X. Vasseur, and L. Carvalho. A modified block flexible GMRES method with deflation at each iteration for the solution of non-hermitian linear systems with multiple right-hand sides. *SIAM J. Sci. Comput.*, 35:S345–S367, 2013.

[12] H. Calandra, S. Gratton, J. Langou, X. Pinel, and X. Vasseur. Flexible variants of block restarted GMRES methods with application to geophysics. *SIAM J. Sci. Comput.*, 34:A714–A736, 2012.

[13] G. Cohen. *Higher-Order Numerical Methods for Transient Wave Equations*. Springer-Verlag, Berlin, 2002.

[14] M. D. Collins and W. A. Kuperman. Nonlinear inversion for ocean bottom properties. *The Journal of the Acoustical Society of America*, 92:2770–2782, 1992.

[15] Y. Diouane. *Globally Convergent Evolution Strategies with Application to an Earth Imaging Problem in Geophysics*. PhD thesis, Institut National Polytechnique de Toulouse, 2014.

[16] Y. Diouane, S. Gratton, and L. N. Vicente. Globally convergent evolution strategies. *Math. Program.*, 152:467–490, 2015.

[17] Y. Diouane, S. Gratton, and L. N. Vicente. Globally convergent evolution strategies for constrained optimization. *Comput. Optim. Appl.*, 62:323–346, 2015.

[18] O. Ernst and M. J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods. In O. Lakkis I. Graham, T. Hou and R. Scheichl, editors, *Numerical Analysis of Multiscale Problems*, pages 325–363. Springer, 2012.

[19] P. Gerstoft. Nonlinear inversion for ocean bottom properties. *The Journal of the Acoustical Society of America*, 95:770–782, 1994.

[20] N. Hansen. The CMA Evolution Strategy: A tutorial. Available at `https://www.lri.fr/~hansen/cmatutorial.pdf`, 2011.

[21] N. Hansen, D. V. Arnold, and A. Auger. Evolution strategies. In J. Kacprzyk and W. Pedrycz, editors, *Handbook of Computational Intelligence*. Springer, Berlin, to appear.

[22] N. Hansen, A. Auger, R. Ros Raymond, S. Finck, and P. Pošík. Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009. In *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '10, pages 1689–1696, New York, NY, USA, 2010. ACM.

[23] N. Hansen, A. Ostermeier, and A. Gawelczyk. On the adaptation of arbitrary normal mutation distributions in evolution strategies: The generating set adaptation. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms, Pittsburgh*, pages 57–64, 1995.

[24] A. Henderson. *ParaView Guide*. A Parallel Visualization Application. Kitware Inc, 2007.

[25] B. L. Kennett, M. S. Sambridge, and P. R. Williamson. Subspace methods for large inverse problems with multiple parameter classes. *Geophys. J. Int.*, 94:237–247, 1988.

[26] R. Lago. *A Study on Block Flexible Iterative Solvers with Application to Earth Imaging Problem in Geophysics*. PhD thesis, Institut National Polytechnique de Toulouse, 2013.

[27] G. Lambaré. Stereotomography. *Geophysics*, 73:VE25–VE34, 2008.

[28] W. A. Mulder and R. E. Plessix. Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56:827–841, 2008.

[29] O. M. Nabi and L. Xiaodong. A comparative study of CMA-ES on large scale global optimisation. In Jiuyong Li, editor, *Australasian Conference on Artificial Intelligence*, volume 6464 of *Lecture Notes in Computer Science*, pages 303–312. Springer, 2010.

[30] G. Nolet, editor. *Seismic Tomography: With Applications in Global Seismology and Exploration Geophysics*. D. Reidel publishing Company, Dordrecht, 1987.

[31] D. W. Oldenburg, P. R. McGillivray, and R. G. Ellis. Generalized subspace methods for large-scale inverse problems. *Geophys. J. Int.*, 114:12–20, 1993.

[32] S. Operto, J. Virieux, J. X. Dessa, and G. Pascal. Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern Nankai Trough. *J. Geophys. Res.*, 159:1032–1056, 2006.

[33] X. Pinel. *A Perturbed Two-level Preconditioner for the Solution of Three-Dimensional Heterogeneous Helmholtz Problems with Applications to Geophysics*. PhD thesis, Institut National Polytechnique de Toulouse, 2010.

[34] G. R. Pratt and M. H. Worthington. Inverse theory applied to multi-source cross-hole tomography. Part 1: Acoustic wave-equation method. *Geophysical Prospecting*, 38(3):287–310, 1990.

[35] R. G. Pratt. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics*, 64:888–901, 1999.

[36] T. Rauber and G. Rünger. *Parallel Programming: for Multicore and Cluster Systems*. Springer, Berlin, second edition, 2013.

[37] C. Ravaut, S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: Application to a thrust belt. *Geophys. J. Int.*, 159:1032–1056, 2004.

[38] I. Rechenberg. *Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann-Holzboog, 1973.

[39] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: A review of algorithms and comparison of software implementations. *J. Glob. Optim.*, 56:1247–1293, 2013.

[40] G. Röth. *Application of Neural Networks to Seismic Inverse Problems*. PhD thesis, Institut de Physique du Globe de Paris, 1994.

[41] C. Shin and Y. H. Cha. Waveform inversion in the Laplace domain. *Geophys. J. Int.*, 173:922–931, 2008.

[42] C. Shin and W. Ha. A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains. *Geophysics*, 73:VE119–VE133, 2008.

[43] C. Shin and W. Ha. Laplace-domain full-waveform inversion of seismic data lacking low-frequency information. *Geophysics*, 77:R199–R206, 2012.

[44] L. Sirgue. The importance of low frequency and large offset in waveform inversion. In *Extended Abstracts*, volume A037 of *68th Confererence & Technical Exhibition*. EAGE, 2006.

[45] L. Sirgue and R. G. Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69:231–248, 2004.

[46] B. Sjogreen and N. A. Petersson. Source estimation by full wave form inversion. *Journal of Scientific Computing*, 59:247–276, 2014.

[47] J. Skilling and R. K. Bryan. Maximum entropy image reconstruction - general algorithm. *Monthly Notices of the Royal Astronomical Society*, 211:111–124, 1984.

[48] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J. Y. L' Excellent. FWT2D: A massively parallel program for frequency-domain Full-Waveform Tomography of wide-aperture seismic data – Part 1: Algorithm. *Computer & Geosciences*, 35:487–495, 2009.

[49] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J. Y. L' Excellent. FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data – part 2: numerical examples and scalability analysis. *Computer & Geosciences*, 35:496–514, 2009.

[50] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, 2005.

[51] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 839–846, Washington, DC, USA, 1998. IEEE Computer Society.

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