FWT2D: a massively parallel program for frequency-domain Full-Waveform Tomography of wide-aperture seismic data - Part 1: algorithm ⋆

Florent Sourbier [a], Stéphane Operto [b,*], Jean Virieux [c],

Patrick Amestoy [d], Jean-Yves L'Excellent [e],

[a] *Géosciences Azur - CNRS - IRD - UNSA - UPMC, Sophia-Antipolis, FRANCE*

[b] *Géosciences Azur - CNRS - IRD - UNSA - UPMC, Villefranche/mer, FRANCE*

[c] *Laboratoire Géophysique Interne et Tectonophysique, BP 53, 38041 Grenoble CEDEX 9, FRANCE*

[d] *ENSEEIHT-IRIT, BP 7122 - F31071, Toulouse Cedex 7*

[e] *INRIA, Université de Lyon, Laboratoire de l'Informatique du Parallélisme (CNRS - ENS Lyon - INRIA - UCBL)*

Abstract

This is the first paper in a two-part series that describes a massively parallel code that performs 2D frequency-domain full-waveform inversion of wide-aperture seismic data for imaging complex structures. Full waveform inversion methods, namely quantitative seismic imaging methods based on the resolution of the full wave equation, are computationally expensive. Therefore, designing efficient algorithms which

⋆ Code available at http://seiscope.unice.fr/opendownload.php

take advantage of parallel computing facilities is critical for the appraisal of these approaches when applied to representative case studies and for further improvements.

Full waveform modelling requires the resolution of a large sparse system of linear equations which is performed with the massively parallel direct solver MUMPS for efficient multiple shot simulations. Efficiency of the multiple-shot solution phase (forward/backward substitutions) is improved by using the BLAS3 library. The inverse problem relies on a classic local optimization approach implemented with a gradient method. The direct solver returns the multiple-shot wavefield solutions distributed over the processors according to a domain decomposition driven by the distribution of the LU factors. The domain decomposition of the wavefield solutions is used to compute in parallel the gradient of the objective function and the diagonal Hessian, this latter providing a suitable scaling of the gradient. The algorithm allows one to test different strategies for multiscale frequency inversion ranging from successive mono-frequency inversion to simultaneous multi-frequency inversion. These different inversion strategies will be illustrated in the following companion paper. The parallel efficiency and the scalability of the code will also be quantified.

Key words: seismic imaging, full-waveform inversion, parallel computation

1 Introduction

Seismic imaging has widespread applications including civil engineering, seismic hazards, hydrocarbon exploration and crustal-scale imaging for more fundamental applications. Among seismic imaging methods, full-waveform inver-

* Corresponding author. S. Operto. Tel: (33) 4 93 76 37 52; Fax: (33) 4 93 76 37 66; email: operto@geoazur.obs-vlfr.fr

Email addresses: sourbier@geoazur.unice.fr (Florent Sourbier), operto@geoazur.obs-vlfr.fr (Stéphane Operto), viri@geoazur.unice.fr (Jean Virieux), amestoy@enseeiht.fr (Patrick Amestoy), Jean-Yves.L.Excellent@ens-lyon.fr (Jean-Yves L'Excellent).

sion (e.g., Tarantola (1984)) is that with the greatest potential in terms of resolution power and quantification of the material properties since the whole information contained in the recorded wavefields is theoretically exploited thanks to the complete resolution of the full wave equation embedded in the optimization process. However, the number of convincing applications of full-waveform inversion to real data case studies has remained extremely limited due to the computational cost of these methods and their sensitivity to several source of errors and limitations such as the inaccuracy of the starting model in the frame of local optimization approaches, incomplete modelling of the wave propagation physics, noise, lack of low frequencies source and limited-aperture acquisition systems. In order to improve full-waveform inversion schemes and to appraise their relevance on representative case studies, it is critical to design optimized algorithms which take advantage of parallel computing facilities. In this framework, we present a massively parallel algorithm which performs 2D frequency-domain full-waveform inversion of wide-aperture seismic data.

The frequency-domain formulation of full-waveform inversion or tomography (FWT) has prompted renewed interest during last decade to build accurate velocity models of complex structures from dense global offset (or wide-aperture) acquisition geometries (Pratt, 2004). Global offset acquisition means geometries providing sufficiently long source-receiver offset so that both refracted and reflected waves can be recorded. These acquisitions are generally carried out with dense network of multi-component stations both on land or at sea. One of the main interests of FWT is the extensive use of the full aperture range spanned by global offset geometries for the reconstruction of broad and continuous range of wavelengths in the medium including large to medium wavelengths. In these high-resolution velocity models, theoretical resolution

limit is half the minimum propagated wavelength. Resolution analysis of FWT reveals that both temporal frequency and aperture angle control the spatial resolution of the imaging (Wu and Töksoz, 1987; Sirgue and Pratt, 2004). Therefore, the more the acquisition geometry illuminates a broad range of aperture angles, the more the seismic imaging can resolve a broad and continuous spectrum of wavenumbers. When applied to wide-aperture data, the frequency-domain approach of FWT has been shown to be efficient for three main reasons: first, only few discrete frequencies are necessary to develop a reliable image of the medium by decimating the wave number redundancy provided by multi-aperture geometries (Sirgue and Pratt, 2004; Pratt and Worthington, 1990; Pratt, 1999). Second, proceeding sequentially from the low to the high frequencies defines a multiscale imaging framework which helps to mitigate the non linearity of the inverse problem. Indeed, the low frequencies are less sensitive to cycle-skipping artefacts than the higher ones for a given starting model. A third criteria is related to the forward modelling problem. Frequency-domain modelling reduces to the resolution of a large sparse system of linear equations per frequency whose right-hand side (RHS) term is the source and the solution is the monochromatic wavefield. For 2D acoustic problems, the few frequencies involved in the inverse problem can be efficiently modelled in the frequency domain for a large number of sources if a direct solver is used for solving the linear system. Indeed, the matrix factorization is independent of the RHS terms and the solution for multiple shots can be obtained efficiently by substitutions once the matrix has been factorized once (Marfurt, 1984). Furthermore, attenuation effects using complex velocity (Toksöz and Johnston, 1981) can be easily implemented in the frequency domain as well as unsplit Perfectly-Matched Layers (PML) (Berenger, 1994; Hustedt et al., 2004) and 45^o paraxial (Clayton and Engquist, 1977) absorbing

boundary conditions. For all these reasons, the frequency-domain approach of FWT is more computationally efficient than the time-domain counterpart to tackle 2D acoustic problems. Moreover, the frequency domain provides a more natural framework than the time domain to design a multiscale approach by successive inversion of increasing frequencies which mitigates the risk of convergence towards a local minimum of the objective function (see Bunks et al. (1995) for a multiscale FWT in the time domain).

The memory and time complexity of direct solvers as well as their limited scalability will probably limit the use of direct solver to solve very large 2D elastic or 3D high-frequency acoustic problems. Especially, the memory and time complexity of direct solver dramatically increase from the 2D case to the 3D one, i.e., from $\mathcal{O}(n^2\, Log_2\, n)$ and $\mathcal{O}(n^3)$ in 2D to $\mathcal{O}(n^4)$ and $\mathcal{O}(n^6)$ in 3D for the memory and time complexities respectively (George and Liu, 1981; Ashcraft and Liu, 1998). This prompted several authors to perform wave modelling in the frequency domain using iterative solvers (Riyanti et al., 2007; Plessix, 2007; Warner et al., 2007; Stekl et al., 2007) or in the time domain from which the frequency response is extracted for frequency-domain FWT (Nihei and Li, 2007; Sirgue et al., 2007). The main advantage of the iterative or the time-domain approach is their lower memory requirement which allows to perform simulation in large 3D models and their good scalability for both 2D and 3D problems. The main drawback may be the time requirement to perform multiple-shot simulations in the frame of waveform tomography since the time complexity of these approaches linearly increases with the number of sources.

The time and memory complexity of 3D acoustic full-waveform modelling and inversion based on the MUMPS direct solver are quantified in Operto et al.

(2007); **?** who showed that representative problems can be addressed at low frequencies. At the resolution scale provided by frequencies smaller than 7 Hz the velocity models developed by 3D FWT can be conceived as macromodel for prestack depth migration. The feasibility of 3D FWT will not be addressed further in this paper which is focused on the implementation of a parallel algorithm to perform 2D acoustic FWT. However, the numerical strategies presented in this paper can be applied equally well in 3D (**?**Ben-Hadj-Ali et al., 2007).

Although the frequency-domain formulation of FWT has shown to be very attractive, its computational cost remains high. Few applications of sequential frequency-domain FWT to 2D real data were recently presented (Bleibinhaus et al., 2007; Jaiswal et al., 2007). A crustal-scale application was presented by Operto et al. (2006) for which only the LU factorization was performed in parallel using 12 processes. For a computational grid of 4.4 millions of nodes, the FWT took 20 days to perform 20 iterations of 13 frequencies inversions. Therefore, it is critical that FWT algorithms take advantage of recent advances in high-performance computing as that provided by large Beowulf clusters.

This is the first paper in a two-part series that describes a massively parallel code that performs 2D full-waveform inversion of wide-aperture seismic data for imaging complex structures. In this first paper, we focus on the description of the parallel algorithm. In the companion paper, we shall validate the FWT program with synthetic examples of increasing complexity and present a scalability analysis of the algorithm thanks to a real data case study. Our code is written in Fortran 90 and uses the Message Passing Interface (MPI) for parallelism. Although the models are parametrized by heterogeneous P-wave velocity, density and attenuation for wave propagation modelling, only the P-wave

velocity is currently reconstructed in the inversion. The inverse problem is solved by an iterative local optimization approach based on a steepest-descent (or gradient) method (Tarantola, 1987; Pratt et al., 1998). The gradient of the objective function is computed using the adjoint method which allows to avoid the explicit computation of the sensitivity matrix. The inversion is iterated non linearly, which means that the final model of the current iteration is used as the starting model for the subsequent iteration. The algorithm has been originally designed so that discrete frequencies are inverted successively by proceeding from the low to the high frequencies. However, we also implemented the possibility to perform simultaneous inversion of multiple frequencies. Full-waveform modelling is performed by a finite-difference (FD) frequency-domain method (Jo et al., 1996; Hustedt et al., 2004). For solving the discrete linear system of equations resulting from the finite-difference discretization of the forward problem, we use the MUlti frontal Massively Parallel direct Solver (MUMPS) for distributed-memory computers (Amestoy et al., 2006, June 2007). The MUMPS solver is a key component of the FWT program as the parallelism of the FWT algorithm is strongly driven by several functionalities of MUMPS. One of this functionality performs parallel multiple-shot substitutions after LU factorization taking advantage of high-performance BLAS3 (Basic Linear Algebra Subprograms; http://www.netlib.org/blas) library. After resolution, the wavefield solutions for all sources are left in core and spatially distributed on the processors allowing straightforward parallelization of subsequent tasks in the FWT program such as gradient computation.

In the first part of the paper, we briefly review the theory of frequency-domain full-waveform modelling and inversion and we introduce the main equations to be implemented. In the second part, we describe the parallel implementation

of the FWT program. We first review the main functionalities of the MUMPS solver. Second, we describe the parallel FWT program.

2 Method

2.1 Finite-difference frequency-domain full-waveform modelling

The 2-D visco-acoustic wave equation is written in the frequency domain as

$$\begin{aligned} &\tfrac{\omega^2}{\kappa(x,z)} p\left(x,z,\omega\right) + \tfrac{\partial}{\partial x}\left(\tfrac{1}{\rho(x,z)}\tfrac{\partial p(x,z,\omega)}{\partial x}\right) + \tfrac{\partial}{\partial z}\left(\tfrac{1}{\rho(x,z)}\tfrac{\partial p(x,z,\omega)}{\partial z}\right) \\ &= -s\left(x,z,\omega\right), \end{aligned} \tag{1}$$

where $\rho\left(x,z\right)$ is the density, $\kappa\left(x,z\right)$ the complex bulk modulus, ω the frequency, $p\left(x,z,\omega\right)$ the pressure field and $s\left(x,z,\omega\right)$ is the source. Attenuation effects can be straightforwardly implemented thanks to complex P-wave velocities in the expression of the bulk modulus (Toksöz and Johnston, 1981).

Since the relation between the pressure wave field and the source is linear, the visco-acoustic wave equation, Eq. 1, can be recast in a matrix form as

$$\mathbf{A}\,\mathbf{p} = \mathbf{s} \tag{2}$$

where the complex-valued impedance matrix $\mathbf{A}$ depends on the frequency and medium properties. The 2-D pressure $\mathbf{p}$ and source $\mathbf{s}$ fields at one frequency ω are stored as vectors of dimension $nx \times nz$ where nx and nz are the dimensions of the regular FD grid, with a grid interval h. The pressure field is obtained by solving the system of linear equations, Eq. 2. If a direct solver can be

used to solve Eq. 2, solutions for multiple sources (i.e., multiple RHS) can be efficiently obtained by forward and backward substitutions once the matrix $\mathbf{A}$ was factorized using a LU decomposition scheme.

We discretized Eq. 1 with the so-called parsimonious mixed-grid method (Hustedt et al., 2004). The aim of the mixed-grid method is to design both accurate and spatially-compact FD stencil for frequency-domain modelling based on a direct solver. The use of a spatially-compact stencil is critical if a direct solver is used to solve the system 2 to limit the numerical bandwidth of the matrix and hence its filling during LU factorization. Spatially-compact stencils were implemented with second-order accurate differencing operators. Accuracy in terms of both numerical anisotropy and dispersion is achieved using the following recipes: first, the differential operators are discretized on different rotated coordinate systems and are combined linearly following the so-called mixed-grid strategy (Jo et al., 1996; Stekl and Pratt, 1998; Hustedt et al., 2004). Second, the mass term at the collocation point is replaced by its weighted average over the grid points involved in the stencil (Marfurt, 1984). Combining these two tricks allows one to use a discretization rule 4 grid points per wavelength although 2^{nd}-order accurate stencils are used. This discretization of the forward problem is optimal for FWT whose resolution limit is half a wavelength. Concerning the discretization of the differential operators, the paper of Hustedt et al. (2004) has clarified the relation between the original mixed-grid approach of Jo et al. (1996) and the staggered-grid methods applied to the first-order hyperbolic velocity-stress formulation of the wave equation (Virieux, 1984, 1986; Saenger et al., 2000) through a parsimonious strategy, originally developed for the time-domain wave equation (Luo and Schuster, 1990).

The numerical bandwidth of the matrix is $\mathcal{O}(n_1)$ where n_1 is the smallest dimension of the 2D computational grid and the number of non-zero coefficients per row is 9. For this kind of sparse matrix, the memory and time complexity of the LU factorization using a nested-dissection ordering is $\mathcal{O}(n^2\, Log_2\, n)$ and n^3 respectively for a square computational grid of dimension $n \times n$ (George and Liu, 1981).

Absorbing boundary conditions are implemented with a combination of 45^o paraxial condition (Clayton and Engquist, 1977) along the outer edges of the model and Perfectly-Matched Layer (PML) conditions (Berenger, 1994). Attenuation is implemented with complex velocities which can depend on frequencies.

The solution of the system, Eq. 2, was computed through a distributed-memory parallel multifrontal scheme developed by Amestoy et al. (2006, June 2007). A brief review of MUMPS is provided latter in the paper.

2.2 Frequency-domain full-waveform tomography

Theory of frequency-domain full-waveform tomography is developed in Pratt et al. (1998). The specific inversion formula including regularization and scaling that is implemented in our FWT program is provided in Operto et al. (2006). Only a brief review is given here.

The weighted least-squares objective function is given by

$$\mathcal{C}(\mathbf{m}) = \Delta\mathbf{d}^{\dagger}\mathbf{W}_d\Delta\mathbf{d} \tag{3}$$

where $\Delta\mathbf{d}$ is the misfit vector, the difference between the observed data and the data computed in model $\mathbf{m}$. The superscript $\dagger$ indicates the adjoint, i.e.,

the transpose conjugate. $\mathbf{W}_d$ is a weighting operator applied to the data that scales the relative contribution of each component of the vector $\Delta\mathbf{d}$ in the inversion.

In the steepest-descent method, a perturbation model $\delta\mathbf{m}$ is searched in the vicinity of a starting model $\mathbf{m}_0$ in the opposite direction of the gradient of the objective function.

One can obtain for the velocity perturbation located at a diffractor point i

$$\Delta m_i = \alpha \sum_S Re \left\{ \mathbf{p}^t \left[\frac{\partial \mathbf{A}}{\partial m_i} \right]^t \mathbf{A}^{-1} \mathbf{W}_d \Delta \mathbf{d}^*. \right\} \tag{4}$$

where α is the step length and S labels sources. Re denotes the real part of a complex number.

This equation indicates that the perturbation model is computed by multiplication of the forward wavefield $\mathbf{p}$ with the backpropagated residuals $\mathbf{A}^{-1}\mathbf{W}_d\Delta\mathbf{d}^*$. Therefore, two forward problems (two forward and backward substitutions once the matrix $\mathbf{A}$ was LU-factorized) per shot are necessary to compute the gradient. The term $\frac{\partial \mathbf{A}}{\partial m_i}$ is a sparse operator obtained by computing the spatial derivative of the coefficients of the matrix $\mathbf{A}$ with respect to the model parameter m_i. In the frame of diffraction tomography, this operator is the radiation pattern of the diffraction by the ficticious heterogeneity δm_i. For the P-wave velocity parameter, the operator $\frac{\partial \mathbf{A}}{\partial m_i}$ reduces to one scalar located on the diagonal of the i^{th} row of matrix $\mathbf{A}$. Equation 4 corresponds to the inversion of one frequency. Simultaneous inversion of multiple frequencies can be easily implemented by summation over frequencies in Eq. 4. Eq. 4 can be derived by explicitely deriving the partial derivative wavefields in a first step and exploiting the reciprocity of Green functions in a second step (Pratt et al., 1998) or

by using the adjoint-state method (Plessix, 2006). The first derivation draws clear connection between FWT and generalized diffraction tomography (Wu and Töksoz, 1987).

In order to obtain reliable perturbation models, we applied some scaling and regularization to the gradient, Eq. 4. We implemented the following inversion formula

$$\Delta m_i = -\alpha \left(diag \mathbf{H}_a + \epsilon I \right)^{-1} \mathcal{G}_{\mathbf{m}} Re \left\{ \mathbf{p}^t \left[\frac{\partial \mathbf{A}^t}{\partial m_i} \right] \mathbf{A}^{-1} \mathbf{W}_d \Delta \mathbf{d}^*. \right\} \tag{5}$$

where $diag\ \mathbf{H_a} = diag Re\{\ \mathbf{J}^t\ \mathbf{W}_d\ \mathbf{J}^*\}$ denotes the diagonal elements of the weighted approximate Hessian $\mathbf{H_a}$ and $\mathcal{G}_{\mathbf{m}}$ is a spatial smoothing operator. $\mathbf{J}$ is the sensitivity matrix whose explicit estimation is required to compute the diagonal Hessian. This estimation requires to compute $N_{shot} + N_{rec}$ solutions where N_{shot} and N_{rec} are the number of non redundant shot and receiver positions respectively. Since the diagonal approximate Hessian requires the explicit computation of the sensitivity matrix and therefore can be expensive to compute, we implement the possibility to compute it only at the first iteration of one frequency inversion and for a decimated acquisition (Operto et al., 2006). The diagonal of the approximate Hessian provides a preconditioner of the gradient that properly scales the perturbation model (Shin et al., 2001).

The smoothing operator $\mathcal{G}_{\mathbf{m}}$ is a 2-D Gaussian spatial filter whose correlation lengths are adapted to the inverted frequency (Ravaut et al., 2004).

The operator $\mathbf{W}_d$ is an amplitude gain with offset applied to each seismic

trace.

$$W_d(o_{sr}) = |o_{sr}|^g \tag{6}$$

where the scalar g controls the amplitude of the gain with respect to the source-receiver offset o_{sr}. This weighting allows to strengthen the relative contribution of large offsets at the partial expense of smaller ones.

We used two criteria to stop iterations. The first one is that the iteration number must not exceed a maximum number of iterations. The second one requires that the objective function reduction must be superior to a given percentage of the objective function computed in the starting model:

$$(C(i-1) - C(i)) \geq thresc \times C(i-1) \tag{7}$$

The step length in Eq. 4 is estimated by parabolic fitting. It requires to compute at least 2 factorizations and $2 \times N_{shot}$ solutions to define 3 pairs $(0, C(\mathbf{m_0}))$, $(\alpha_1, C(\mathbf{m_1}(\alpha_1))$ and $(\alpha_2, C(\mathbf{m_2}(\alpha_2))$ through which a parabola is fitted. The minimum of the parabola provides the optimal step length.

The FWT program can either be applied to hydrophone or geophone data generated by explosive sources. In case of geophone data, we exploit the reciprocity of Green functions to match geophone data with pressure synthetics by replacing explosive sources by vertical forces in the forward problem (Operto et al., 2006).

3 Numerical implementation

3.1 The massively parallel MUMPS solver

MUMPS, a multifrontal massively parallel sparse direct solver package for distributed-memory platforms, developed by CERFACS, ENSEEIHT-IRIT and INRIA, is used here to solve the system of linear equations, Eq. 2 (Amestoy et al., June 2007, 2006). Like other direct methods, the multifrontal technique used in MUMPS is based on an elimination tree, which is a transitive reduction of the graph of L, where L is the Cholesky factor of the matrix $(\mathbf{A}+\mathbf{A}^{\mathbf{T}})$, and is the smallest data structure representing dependencies between operations (Duff and Reid, 1983; Liu, 1992). The multifrontal algorithm is an approach to organize right-looking sparse matrix factorization in which the factorization of the matrix is done by performing a succession of partial factorizations of small dense matrices called frontal matrices. The factorization process is given by the assembly tree where a frontal matrix is associated to each node. Reordering (i.e. renumbering the entries of a sparse linear system) is a well known technique to reduce fill in the final factors, and this has a big impact on the memory size. MUMPS offers several ordering algorithms, particularly the METIS package from University of Minnesota which relies on an hybrid method combining a multilevel nested dissection with a multiple minimum degree algorithm (Karypis and Kumar, 1998) and an AMD ordering (Amestoy et al., 1996). An analysis of the impact of reordering on the memory of the multifrontal method is presented in Guermouche et al. (2003).

In the multifrontal method, the resolution of the linear system is generally subdivided into 3 main tasks. The first one is an analysis phase or sym-

bolic factorization. This analysis is currently sequential in MUMPS and is carried out by the master processor. Reordering of the matrix coefficients is first performed in order to minimize fill-ins, namely, the additional non zero coefficients introduced during the elimination process. Then, the dependency graph is estimated as well as the memory required to perform the subsequent numerical factorization. If several simulations need to be performed in slightly different velocity models during iterative FWT, the analysis phase needs to be performed only once per frequency. The sparse impedance matrix is generally assembled on the master processor before the analysis. For very large 2D problems (i.e., typically for problems involving few tens of millions of unknowns), the sparse impedance matrix may not fit the memory assigned to the master processor. In that case, the impedance matrix can be assembled in distributed format over the processors involved during the subsequent parallel factorization. It is also possible with MUMPS to avoid involving the master processor in the LU factorization in the case of large centralized matrix to avoid unbalanced workload. Note also than for large 2D problems it may be necessary to compile MUMPS with an option making default integer variables 8 bytes long in order to avoid overflow during the sequential analysis. The second task is the numerical factorization performed with the multifrontal method. At the end of the factorization, the LU factors are distributed over the processors. The third task is the solution phase performed by forward and backward substitutions. During the solution phase, multiple RHS terms can be provided in sparse format on the master processor. They are broadcast from the host to the processors over which the LU factors are distributed. The multiple-shot solutions can be computed simultaneously from the LU factors taking advantage of BLAS3 library and are either assembled on the host or kept distributed on the processors for subsequent parallel computations. To

maximize the efficiency of the solve phase for multiple RHS, MUMPS allocates workspace memory of size proportional to the blocking factor for multiple RHS (MUMPS parameter KEEP(84)). This parameter needs to be tuned to obtain the minimum computational time for the multi-RHS solution phase.

We performed the factorization and the solution phases in single precision. To reduce the condition number of the matrix, a row and column scaling is applied in MUMPS before factorization. The sparsity of the matrix and suitable equilibration have made single precision factorization accurate enough so far for the 2D and 3D problems we tackled (see Hustedt et al. (2004) for examples of 2D simulations in very contrasted media).

3.2 Flowchart of the FWT algorithm

The processing flow performed by the FWT program is summarized in Fig. 1. It contains mainly two sequential loops over groups of frequencies and inversion iterations (by sequential loop is meant that each processor does all the iterations of the do loop). These loops are necessarily sequential due to the non linearity of the iterative inversion, i.e., the final model of one iteration of a frequency group inversion is used as a starting model for the next iteration. A group of frequencies defines several frequencies that are simultaneously inverted. Defining only one frequency per group is equivalent to perform successive mono-frequency inversion following the classic multiscale approach promoted by Pratt (1999). On the opposite, defining only one group is equivalent to perform simultaneous multi-frequency inversion. A third approach performs of successive inversions of overlapping frequency groups of increasing frequency content. This means that a given frequency group contains the frequencies

from the previous groups plus a higher frequency. This multiscale approach can be viewed as the equivalent in the frequency domain of the time-domain approach of Bunks et al. (1995).

The first task within the loop over iterations reads the starting model and initializes the gradient and the diagonal Hessian. The gradient and diagonal Hessian are computed by summation over the frequencies of a group. In the following, the contribution of one frequency to the diagonal Hessian and gradient will be referred to as *partial* diagonal Hessian and *partial* gradient respectively. Within the loop over iterations, a sequential loop over frequencies of one group is opened. Within this loop, the impedance matrix $\mathbf{A}$ whose coefficients depend on frequency and velocity model properties is assembled in sparse format on the master processor and is subsequently LU factorized in parallel with MUMPS. At the end of the factorization, the LU factors are kept distributed on the processors. The program proceeds with the computation of the partial diagonal Hessian which first requires to solve system 2 by forward/backward substitutions for each non redundant shot and receiver positions. At the end of the solve, each processors stores in core a spatial subdomain of all the wavefield solutions. This domain decomposition which is driven by the distribution of the LU factors is illustrated for one wavefield and 12 processes in Fig. 2. The partial diagonal Hessian is basically computed by weighted summation over the source and receiver wavefields. We take advantage of the domain decomposition of the wavefields to perform in parallel this weighted summation i.e., each processor computes the spatial subdomain of the partial diagonal Hessian corresponding to the subdomain of the wavefields stored on this processor. Once the partial diagonal Hessian has been computed in distributed format on each processor, all the subdomains of the partial di-

agonal Hessian are gathered on the master processor and the diagonal Hessian is updated by addition of the newly-computed partial diagonal Hessian. Note that the update of the diagonal Hessian is performed on the master processor and not in distributed format over the processors because the domain decomposition can slightly change from one frequency to the next due to pivoting during the LU factorization. The next task is the computation of the gradient. It first requires N_{shot} solutions of system 2 to compute the incident wavefields. Again these incident wavefields are left in core and distributed over the processors following the same domain decomposition as for the diagonal Hessian. Once these solutions are computed, the incident wavefields are extracted at the receiver positions on each subdomain containing receivers. At this step, the program can perform a source estimation for the current frequency by solving a linear inverse problem (see Eq. 17 in Pratt (1999)). After the source estimation, partial data residuals are computed on each subdomain and are assembled on the master processor to build the composite residual sources in sparse format. A second series of N_{shot} resolutions is performed to obtain the backpropagated residual wavefields. The partial gradient is computed in distributed format over the processors and gathered in the end on the master processor following the same strategy than for the diagonal Hessian. The total gradient is updated by addition of the newly-computed partial gradient on the master processor. Once all the frequencies of the group are processed, the gradient is scaled by the diagonal Hessian and smoothed by the 2D Gaussian filter. Note that we have assumed so far that all the wavefield solutions and the LU factors can remain in core over the processors. If not enough memory is available, we split the RHS terms in several partitions and process sequentially each partition during the solve phases and the computation of the diagonal Hessian and gradient without hard disk access. A second remark is

that no point-to-point communication is required during the diagonal Hessian and gradient computations when the P-wave velocity is considered as model parameter. This is due to the fact that the matrix $\frac{\partial \mathbf{A}}{\partial m_i}$, Eq. 4, reduces to a scalar located on the i^{th} row of the diagonal of $\mathbf{A}$. In this case, estimation of the gradient at a point m_i depends only on the wavefield solutions at this same point. This comment would apply for the attenuation parameter as well. On the contrary, if one consider parameter such as density, estimation of the gradient at the point m_i will require the knowledge of the incident and back-propagated residual wavefields at the next positions. If these next positions belong to another domain than the point m_i some bidirectional point-to-point communications will be necessary during the gradient computation. Involvement of other classes of parameters in the inversion will be the aim of a future release of our FWT program. A third remark is that the loop over frequencies of one group is currently sequential. Since simultaneously accounting for multiple frequencies is a linear process, this loop may be distributed over M sets of processors where M is the number of frequencies in the group. Indeed, this strategy implies that each group of processors have enough distributed memory to solve the system 2 for a significant number of RHSs. This strategy was not investigated yet.

The program proceeds with the step length estimation which requires at least two factorizations and two N_{shot} resolutions for each frequency as mentioned before.

4 Conclusion and perspectives

We described a massively parallel algorithm to perform frequency-domain full-waveform modelling and inversion for imaging 2D acoustic media using a steepest-descent algorithm. Frequency-domain full-waveform modelling reduces to the resolution of a large and sparse system of linear equations. We have used a massively parallel direct solver for efficient resolution of this system for a large number of sources, a critical issue in tomography applications. The efficiency of the parallelism of the FWT algorithm is strongly driven by the functionalities implemented in the parallel direct solver. In addition to the high-performance LU factorization performed with a multifrontal method, three of these functionalities which are useful for tomographic applications are the sparse storage of the right-hand side (source) terms, the multi-RHS solutions in parallel and the storage of the multi-RHS solutions in distributed format thanks to a domain decomposition. This latter functionality provides the optimal framework to compute in parallel the gradient of the objective function.

Our FWT program allows to test in a flexible way different strategies for selecting and clustering frequencies in the inversion ranging from successive mono-frequency inversion to simultaneous multi-frequency inversion. In the companion paper (this issue), the main functionalities of the code are illustrated with numerical examples and a detailed scalability analysis is performed with a large crustal-scale case study.

5 Acknowledgements

We would like to thank the two anonymous reviewers for their suggestions to improve the manuscript. This work was funded by the SEISCOPE consortium sponsored by BP, CGG-VERITAS, EXXON-MOBIL, SHELL, TOTAL and by the Agence Nationale de la Recherche (project ANR-05-NT05-4247). Partial support is acknowledged for the use of computer facilities at the MESOCENTRE SIGAMM computer center. We are grateful to A. Miniussi for his assistance during the installation of the modelling and MUMPS programs on the MESOCENTRE SIGAMM. This work has been carried out partially with the contribution of the National Institute of Geophysics and Volcanology and of the Department of Civil Protection.

References

Amestoy, P., Davis, T. A., Duff, I. S., 1996. An approximate minimum degree ordering algorithm. SIAM J. Matrix. Anal. and Applics. 17, 886–905.

Amestoy, P. R., Duff, I. S., L'Excellent, J. Y., Koster, J., June 2007. Multifrontal massively parallel solver (MUMPS version 4.7): Users'guide. http://enseeiht.fr/apo/MUMPS/ and http://graal.ens-lyon.fr/MUMPS/.

Amestoy, P. R., Guermouche, A., L'Excellent, J. Y., Pralet, S., 2006. Hybrid scheduling for the parallel solution of linear systems. Parallel computing 32, 136–156.

Ashcraft, C., Liu, J. W. H., 1998. Robust ordering of sparse matrices using multisection. SIAM Journal on Matrix Analysis and Applications 19 (3), 816–832.

URL citeseer.ist.psu.edu/ashcraft96robust.html

Ben-Hadj-Ali, H., Operto, S., Virieux, J., Sourbier, F., 2007. 3D acoustic frequency-domain full-waveform inversion. In: Extended Abstracts. Society of Exploration Geophysicists, pp. 1730–1734.

Berenger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves. Journal of Computational Physics 114, 185–200.

Bleibinhaus, F., Hole, J. A., Ryberg, T., Fuis, G. S., 2007. Structure of the california coast ranges and san andreas fault at SAFOD from seismic waveform inversion and reflection imaging. Journal of Geophysical Research 112 (B06315), doi:10.1029/2006JB004611.

Bunks, C., abd S. Zaleski, F. M. S., Chavent, G., 1995. Multiscale seismic waveform inversion. Geophysics 60 (5), 1457–1473.

Clayton, R., Engquist, B., 1977. Absorbing boundary conditions for acoustic and elastic wave equation. Bull. Seismol. Soc. Am. 67, 1529–1540.

Duff, I. S., Reid, J. K., 1983. The multifrontal solution of indefinite sparse symmetric linear systems. ACM Trans. Math. Softw. 9, 302–325.

George, A., Liu, J. W., 1981. Computer solution of large sparse positive definite systems. Prentice-Hall, Inc.

Guermouche, A., L'Excellent, J. Y., Utard, G., 2003. Impact of reordering on the memory of a multifrontal solver. Parallel computing 29, 1191–1218.

Hustedt, B., Operto, S., Virieux, J., 2004. Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. Geophysical Journal International 157, 1269–1296.

Jaiswal, P., Zelt, C. A., Dasgupta, R., 2007. Traveltime and full-waveform inversion for improved seismic imaging in geologically comples areas. In: Expanded Abstracts. EAGE, p. C032.

Jo, C. H., Shin, C., Suh, J. H., 1996. An optimal 9-point, finite-difference,

frequency-space 2D scalar extrapolator. Geophysics 61, 529–537.

Karypis, G., Kumar, V., Sep. 1998. METIS - A software package for partitioning unstructured graphs, partitioning meshes and computing fill-reducing orderings of sparse matrices - Version 4.0. University of Minnesota.

Liu, J. W. H., 1992. The multifrontal method for sparse matrix solution: theory and practice. SIAM review 34 (1), 82–109.

Luo, Y., Schuster, G. T., 1990. Parsimonious staggered grid finite-differencing of the wave equation. Geophysical Research Letters 17 (2), 155–158.

Marfurt, K., 1984. Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation. Geophysics 49, 533–549.

Nihei, K. T., Li, X., 2007. Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (td-psd). Geophysical Journal International 169, 1069–1078.

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J. Y., Giraud, L., Ben-Hadj-Ali, H., 2007. 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. Geophysics 72 (5), SM195–SM211.

Operto, S., Virieux, J., Dessa, J. X., Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern nankai trough. Journal of Geophysical Research 111 (B09306), doi:10.1029/2005JB003835.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophysical Journal International 167, 495–503.

Plessix, R. E., 2007. A helmholtz iterative solver for 3d seismic-imaging problems. Geophysics 72, SM185–SM194.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain, part

517 I : theory and verification in a physic scale model. Geophysics 64, 888–901.
518 Pratt, R. G., 2004. Velocity models from frequency-domain waveform tomog-
519 raphy: past, present and future. In: Expanded Abstracts. Eur. Ass. Expl.
520 Geophys.
521 Pratt, R. G., Shin, C., Hicks, G. J., 1998. Gauss-newton and full newton
522 methods in frequency-space seismic waveform inversion. Geophysical Jour-
523 nal International 133, 341–362.
524 Pratt, R. G., Worthington, M. H., 1990. Inverse theory applied to multi-source
525 cross-hole tomography. Part 1: acoustic wave-equation method. Geophysical
526 Prospecting 38, 287–310.
527 Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., dell'Aversana,
528 P., 2004. Multi-scale imaging of complex structures from multi-fold wide-
529 aperture seismic data by frequency-domain full-wavefield inversions: appli-
530 cation to a thrust belt. Geophysical Journal International 159, 1032–1056.
531 Riyanti, C. D., Kononov, A., Erlangga, Y. A., Vuik, C., Oosterlee, C., Plessix,
532 R. E., Mulder, W. A., 2007. A parallel multigrid-based preconditioner for
533 the 3d heterogeneous high-frequency helmholtz equation. Journal of Com-
534 putational physics 224, 431–448.
535 Saenger, E. H., Gold, N., Shapiro, S. A., 2000. Modeling the propagation of
536 elastic waves using a modified finite-difference grid. Wave motion 31, 77–92.
537 Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., Chung, S.,
538 Shin, S., 2001. Efficient calculation of a partial derivative wavefield using
539 reciprocity for seismic imaging and inversion. Geophysics 66 (6), 1856–1863.
540 Sirgue, L., Etgen, T. J., Albertin, U., Brandsberg-Dahl, S., 2007. Sys-
541 tem and method for 3D frequency-domain waveform inversion based on
542 3D time-domain forward modeling. US Patent Application Publication

US2007/0282535 A1.

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

Stekl, I., Pratt, R. G., 1998. Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators. Geophysics 63, 1779–1794.

Stekl, I., Warner, M. R., Umpleby, A. P., 2007. 3D frequency domain waveform inversion - synthetic shallow channel example. In: Abstract book. EAGE, p. C026.

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

Tarantola, A., 1987. Inverse problem theory: methods for data fitting and model parameter estimation. Elsevier, New York.

Toksöz, M. N., Johnston, D. H., 1981. Geophysics reprint series, No. 2: Seismic wave attenuation. Society of exploration geophysicists, Tulsa, OK.

Virieux, J., 1984. SH wave propagation in heterogeneous media, velocity-stress finite difference method. Geophysics 49, 1259–1266.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media, velocity-stress finite difference method. Geophysics 51, 889–901.

Warner, M., Stekl, I., Umpleby, A., 2007. Full wavefield seismic tomography - iterative forward modelling in 3d. In: Expanded Abstracts,. EAGE, p. C025.

Wu, R.-S., T̈oksoz, M. N., 1987. Diffraction tomography and multisource holography applied to seismic imaging. Geophysics 52, 11–25.

6 Captions of Figures

Fig. 1. Flowchart of the FWT algorithm. The algorithm is structured within 2 main nested loops over frequency groups and iterations. Within these loops, 3 loops over frequencies of one group are performed. The diagonal Hessian and the gradient are computed within the first one while the second and third ones are required to estimated the step length (not detailed in the figure). In the diagonal Hessian and gradient boxes, by *partial* diagonal Hessian and *partial* gradient are meant the specific contributions of the current frequency to the diagonal Hessian and gradient computation. Nomenclature: P_0 labels the master process. $Nshot$: number of shots; $Nrec$: number of receivers. RHS: right-hand side terms.

Fig. 2. Illustration of the domain decomposition of the wavefield solutions during the solution step in MUMPS. Computation was performed using 12 processes. The black areas delineate the 12 subdomains of the wavefield solutions distributed over the processors. The model dimensions are 105 x 25 km and correspond to the real data case study presented in the companion paper (this issue). The gradient of the objective function (i.e., the unscaled perturbation model) is computed in parallel with the same domain decomposition that shown in the figure.

7 Figures

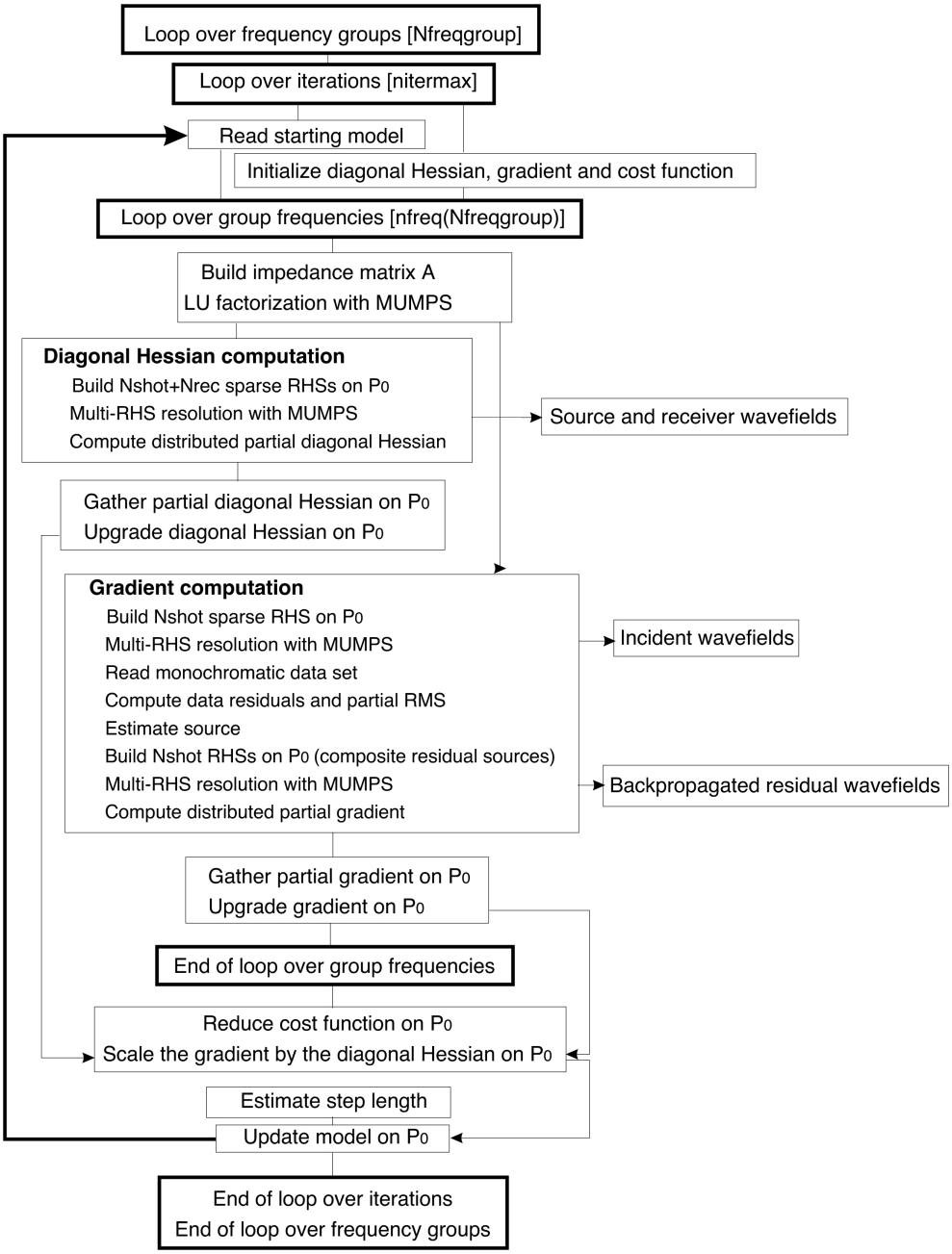

Fig. 1.

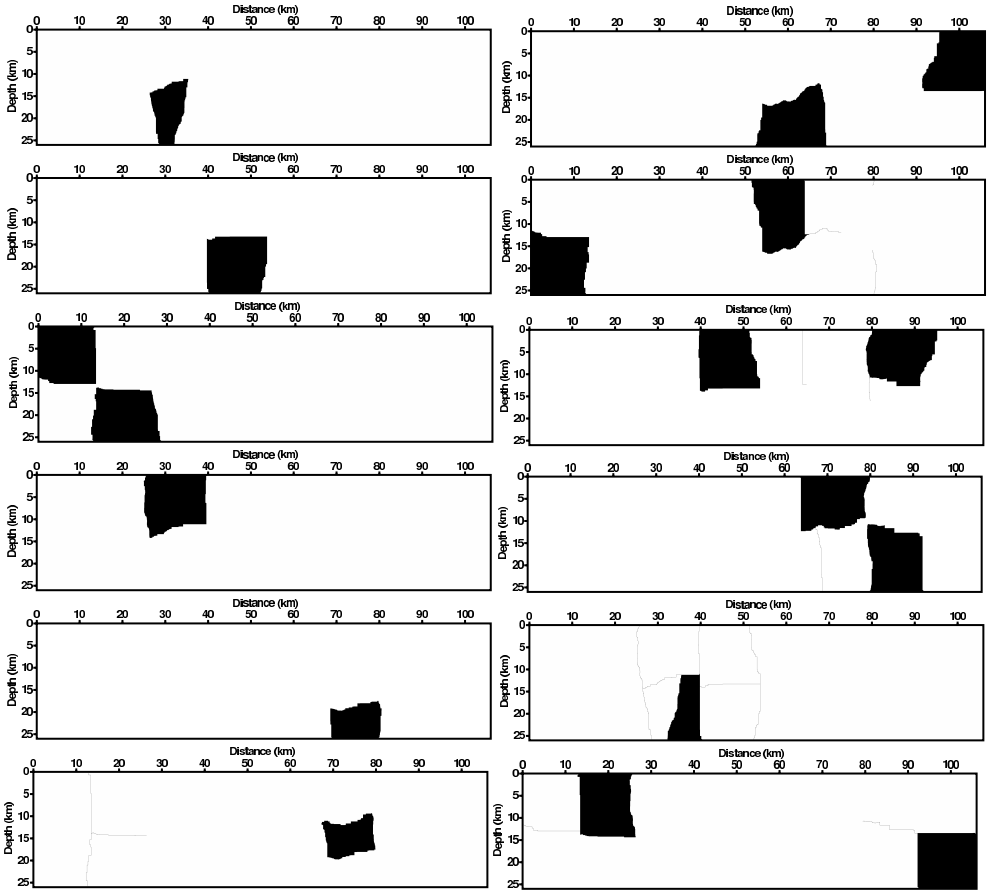

Fig. 2.