Journal of Computational Physics 229 (2010) 8445–8462

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Acoustic inverse scattering via Helmholtz operator factorization and optimization ☆

Shen Wang, Maarten V. de Hoop, Jianlin Xia *

Center for Computational and Applied Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 14 July 2009
Received in revised form 3 July 2010
Accepted 21 July 2010
Available online 2 August 2010

Keywords:
Imaging
Inverse scattering
Wave-equation tomography
Optimization
Helmholtz solver
Structured multifrontal method

ABSTRACT

We present a joint acoustic/seismic inverse scattering and finite-frequency (reflection) tomography program, formulated as a coupled set of optimization problems, in terms of inhomogeneous Helmholtz equations. We use a higher order finite difference scheme for these Helmholtz equations to guarantee sufficient accuracy. We adapt a structured *approximate* direct solver for the relevant systems of algebraic equations, which addresses storage requirements through compression, to yield a complexity for computing the gradients or images in the optimization problems that consists of two parts, viz., the cost for all the matrix factorizations which is roughly $\mathcal{O}(rN)$ (for example $\mathcal{O}(rN\log N)$ when $d=2$) times the number of frequencies, and the cost for all solutions by substitution which is roughly $\mathcal{O}(N)$ (for example $\mathcal{O}(N\log(r\log N))$ when $d=2$) times the number of frequencies times the number of sources (events), where $N=n^d$ if n is the number of grid samples in any direction, and r is a parameter depending on the preset accuracy and the problem at hand. With this complexity, the multi-frequency approach to inverse scattering and finite-frequency tomography becomes computationally feasible with large data sets, in dimensions $d=2$ and 3.

1. Introduction

Computational "full-wave" seismic inverse scattering and reflection tomography are driven by solving the wave-equation in large subdomains of $\mathbb{R}^d$, $d=2,3$, representing Earth's subsurface, while accounting for the broad bandwidth of the available data. Following a multi-frequency approach, the wave-equation is transformed to the Helmholtz equation. One can obtain similar equations, with the Laplace operator replaced by a pseudodifferential operator, for polarized elastic waves in anisotropic media. Subject to certain approximations, one can cast these into the form of coupled Helmholtz equations. Computationally, upon discretization, solving the Helmholtz equation amounts to solving a large linear system of algebraic equations. This has to be repeated for a significant number of frequencies (in the mentioned bandwidth) as well as for a large number of different right-hand sides corresponding with the large number of sources generating a seismic data set. Current seismic data sets consist of observations of the wave fields in large arrays of receivers at Earth's surface generated by a large number of sources indeed. The multi-frequency approach, in the context of seismic imaging, was already discussed by Marfurt [1].

The nature of the multi-frequency formulation of the inverse scattering and reflection tomography problem leads one to explore the use of the LU decomposition or factorization (once per frequency) of the coefficient matrix defining the above

☆ This research was supported in part by the members, BP, ConocoPhillips, ExxonMobil, StatoilHydro and Total, of the Geo-Mathematical Imaging Group.
* Corresponding author.
E-mail addresses: wang273@purdue.edu (S. Wang), mdehoop@math.purdue.edu (M.V. de Hoop), xiaj@math.purdue.edu (J. Xia).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.07.027

mentioned linear system of algebraic equations and a direct solver. However, the system is so large in applications (in particular, in dimension $d = 3$) that the associated memory requirements become unsurmountable. Here, we introduce an alternative approach based on the structured *approximate* direct solver in [2], which addresses the memory requirements through compression.

For different frequencies, the coefficient matrices representing the linear systems share the same nonzero pattern. Thus, an efficient sparse direct solver remains a natural choice for the inverse problem at hand. Sparse direct solvers commonly involve four stages: mesh ordering, symbolic factorization, numerical factorization, and system solution. For all the systems that need to be solved, the mesh ordering and symbolic factorization stages need to be carried out only once. For each frequency, one numerical factorization is needed. However, sparse direct solvers are often considered expensive due to the problem of "fill-in". That is, even if the original matrix is sparse, the factorization generally introduces many new nonzero entries into the factors. Moreover, the accuracy of traditional direct solvers is generally not explicitly controllable. Also, methods used as preconditioners such as incomplete factorizations often suffer from the problem of "breakdown".

The structured approximate direct method in [2] is a black-box solver based on structured approximations of dense fill-in in the direct factorization of the coefficient matrix. The accuracy is controllable. A robustness strategy is used to avoid breakdown. The full integration of graph-based sparse matrix techniques and structured matrix computations makes the method efficient. The cost for the factorization is only $\mathcal{O}(rN\log N)$ for $d = 2$ (roughly $\mathcal{O}(rN)$ for $d = 3$), where N is the matrix size and r is a parameter depending on the preset accuracy and the problem at hand. The storage is $\mathcal{O}(N\log(r\log N))$ for $d = 2$ (roughly $\mathcal{O}(N)$ for $d = 3$).

The solver uses a structured supernodal multifrontal method with nested dissection ordering of the mesh points. It has been observed that in the factorization of some discretized problems, the off-diagonal blocks of these dense matrices have small numerical ranks because the Green's functions of these problems are smooth away from the "diagonal" singularity. For problems such as the one defined by a discrete Helmholtz equation, the off-diagonal numerical ranks may not be small enough. However, [3] and [2] show that the method in [2] can still be satisfactory if only modest accuracy is needed. Dense intermediate matrices are factorized into data-sparse semiseparable matrices. Operations of compact semiseparable matrices are usually fast, such as linear complexity matrix multiplications and system solutions. Fast operations for semiseparable matrices lead to the efficiency of the structured sparse solver exploited here.

Various alternative strategies to solving the discrete Helmholtz equation have been introduced in the past few years. We mention the Krylov-based iterative solvers with preconditioners (Elman et al. [4], Plessix and Mulder [5], Erlangga et al. [6,7] and Riyanti et al. [8]). Operto et al. [9] discuss a (massively parallel) direct solver based on domain decomposition (Larsson [10]).

The space-frequency domain approach has many advantages over the space-time one. In [11] [12], a frequency selection strategy is proposed for seismic inversion based on data redundancy in the frequency domain. It is demonstrated in [8] that in 2D, the frequency domain approach is an order of magnitude faster than the time domain one. However in 3D, the frequency domain approach with LU factorization yields a storage problem. We address this problem in this paper. Moreover, the frequency domain approach plays a key role in FWI (Full Waveform Inversion [13] [14]), in as much as yielding control over convergence of the relevant iteration. In this paper, we consider linearized inversion and demonstrate the accuracy of our approach.

We integrate our discrete Helmholtz equation solver with the optimization formulations of "full-wave" seismic inverse scattering and reflection tomography through the adjoint state method (Section 2). The gradient computation coincides with imaging. In the case of inverse scattering, such a formulation, based on a least-squares misfit, can be found in Lailly [15] and Tarantola [16], while a computational realization using the Helmholtz equation can be found in Mulder and Plessix [12]. We further mention the work of Kühl and Sacchi [17] on least-squares wave-equation migration. Multi-frequency waveform tomography has been developed, also, using a least-squares misfit (see Pratt [18] and Operto et al. [19]). Here, we follow an approach to reflection tomography reminiscent of the one based on annihilators of the data (see Stolk and De Hoop [20]). One class of such annihilators can be constructed from the so-called angle transform in the "downward continuation" approach (derived using a wavefield splitting technique) to seismic imaging; the corresponding adjoint states equations can be found in De Hoop et al. [21]. Under certain conditions, another class of annihilators can be constructed reflecting data redundancy in the source coordinates, following an approach known as "reverse time migration". As in Xie and Yang [22], we use this redundancy in combination with the notion of image continuation [23] to arrive at a misfit criterion and optimization scheme for wave-equation reflection tomography.

In Section 3 we give the discretization of the Helmholtz operator. In Section 4 we discuss the structured approximate direct solver tailored to the discrete Helmholtz equation. We present some comparisons with the direct LU, multilevel Krylov-multigrid and multifrontal methods in terms of computational complexity and storage requirements. The overall complexity for solving all the instances of the discrete Helmholtz equations in our problem has two parts, viz., the cost for all the matrix factorizations which is roughly $\mathcal{O}(rN)$ (for example $\mathcal{O}(rN\log N)$ when $d = 2$) times the number of frequencies, and the cost for all solutions by substitution which is roughly $\mathcal{O}(N)$ (for example $\mathcal{O}(N\log(r\log N))$ when $d = 2$) times the number of frequencies times the number of sources. In Section 5 we present the computational performance of our method through a numerical example ($d = 2$) with a model containing a smoothed, low wavespeed lens leading to the formation of caustics and a flat discontinuity. We give some conclusions in Section 6. On how to connect the inverse scattering program developed here to a multi-scale approach, see Wang et al. [24], and Brytik et al. [25] concerning the tomographic aspects.

2. Inverse scattering program

The causal acoustic-wave Green's function $G(x,t;x_s)$ for a point source at $x = x_s$ is the solution of

$$\left[c(x)^{-2}\partial_t^2 - \Delta\right] G(x,t;x_s) = \delta(t)\delta(x - x_s), \tag{1}$$

with $G = 0$, $t < 0$. Here, $c = c(x)$ is the wavespeed of the medium. The partial differential equation is considered on a set $X \subset \mathbb{R}^d$ ($d \geqslant 2$). Upon applying the Fourier transform, with

$$\widehat{G}(x,\omega;x_s) = \int G(x,t;x_s)\exp(-\mathrm{i}\omega t)\mathrm{d}t,$$

we obtain the corresponding Helmholtz equation,

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\widehat{G}(x,\omega;x_s) = -\delta(x - x_s), \tag{2}$$

with the Sommerfeld radiation condition

$$R\left[\frac{\partial\widehat{G}}{\partial R} + \frac{\mathrm{i}\omega}{c}\widehat{G}\right] \to 0 \quad \text{as } R \to \infty,\ R = |x|.$$

In the following subsections, the inverse scattering scheme is described in terms of the associated Helmholtz operator.

2.1. Linearized modeling

In the development of seismic scattering, imaging and inverse scattering theory, it is common practice to invoke the Born or single scattering approximation. We introduce $r(x) = \delta c(x)/c(x)$, which is a relative perturbation of the wavespeed and can be thought of as a reflectivity. Linearization of the wave-equation yields for the corresponding perturbation of the Green's function,

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\delta\widehat{G}(x,\omega;x_s) = 2\omega^2 c(x)^{-2} r(x)\widehat{G}(x,\omega;x_s), \tag{3}$$

using the solution of (2) in the right-hand side. It is, here, assumed that $c = c(x)$ is a smooth function, while $r = r(x)$ is of low regularity. Essentially, the justification of the Born approximation is based on a separation of scales distinguishing a background from a contrast [26].

Frequency-domain seismic data, $\hat{d}(x_s,x_r,\omega)$ say, are modeled as $\delta\widehat{G}(x,\omega;x_s)$, by solving (3) after solving (2), upon restricting x to points $x_r \in \Sigma_r$ and x_s to Σ_s with Σ_s, $\Sigma_r \subset \partial X$. Moreover, the data are measured in a finite-frequency bandwidth, $\omega \in \Omega = [\omega_1,\omega_2]$ with $0 < \omega_1 < \omega_2$, only. This then defines a so-called single-scattering operator, $\widehat{F} = \widehat{F}[c]:\ r(x) \to \delta\widehat{G}(x_r,\omega;x_s)$, mapping subsurface reflectors to surface reflections. In the further development, we will assume that X is a rectangular domain, and that Σ_s, Σ_r are subsets of its "top" side, representing part of the earth's surface; also, we will assume that the dimension of Σ_r is $d-1$. The results presented here are straightforwardly extended to more generally shaped domains and boundaries upon introducing curvilinear coordinates, as in Stolk et al. [27].

In the seismic inverse problem, the objective is to reconstruct c (referred to as the background) and r (referred to as the reflectivity) from the data, d. The imaging and reconstruction of r is referred to as the inverse scattering problem, while the imaging and reconstruction of c is referred to as the (wave-equation) reflection tomography problem. In the next subsections, we summarize the optimization formulations of these problems in terms of adjoint states equations.

2.2. Imaging of the (relative) contrast: Optimization and gradient

Inverse scattering has been formulated as an optimization problem, that is,

$$\langle r\rangle = \arg\min \frac{1}{2}\left\|\hat{d} - \widehat{F}[c]r\right\|^2_{\Sigma_s\times\Sigma_r\times\Omega},$$

where $\langle r\rangle$ is the estimate of r. The gradient of the underlying mismatch criterion is the image, $\mathcal{I} = \widehat{F}[c]^*\hat{d}$, given by

$$\mathcal{I}(x) = \int_{\Sigma_s}\frac{1}{2\pi}\int\overline{\hat{w}_s(x,\omega;x_s)}\hat{u}^*(x,\omega;x_s)\mathrm{d}\omega\,\mathrm{d}x_s = \int_{\Sigma_s}\mathrm{Re}\frac{1}{\pi}\int_{\Omega}\overline{\hat{w}_s(x,\omega;x_s)}\hat{u}^*(x,\omega;x_s)\mathrm{d}\omega\,\mathrm{d}x_s, \tag{4}$$

and obtained by solving

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\hat{w}_s(x,\omega;x_s) = -\delta(x - x_s), \tag{5}$$

referred to as "pde", and

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\hat{u}^*(x,\omega;x_s) = \omega^2 \int_{\Sigma_r} \hat{d}(x_s,x_r,\omega)\delta(x-x_r)\mathrm{d}x_r, \tag{6}$$

referred to as "pde*" in the diagram,

$$\begin{array}{ccc} & \text{pde} & \\ \text{data} \searrow & & \searrow \\ & & \nearrow \rightarrow \text{image} \\ & \text{pde}^* & \end{array} \tag{7}$$

summarizing the data-field flow in the gradient computation for each source. The integration over frequency in (4) signifies the imaging condition, that is, a cross correlation in time, see Tarantola [16] and others.

We introduce the kernel, $K = K(x;x_s,x_r,t)$, of the imaging operator, $F[c]^*$, that is,

$$\mathcal{I}(x) = \int_{\Sigma_s}\int_{\Sigma_r}\int K(x;x_s,x_r,t)d(x_s,x_r,t)\mathrm{d}t\mathrm{d}x_r\mathrm{d}x_s = \int_{\Sigma_s}\int_{\Sigma_r}\frac{1}{2\pi}\int \overline{\hat{K}(x;x_s,x_r,\omega)}\hat{d}(x_s,x_r,\omega)\mathrm{d}\omega\mathrm{d}x_r\mathrm{d}x_s. \tag{8}$$

This kernel integral representation expresses that each data sample, at (x_{s0},x_{r0},t_0) say, is being "smeared" over the support of $K(x;x_{s0},x_{r0},t_0)$ viewed as a function of x. Accounting for the data frequency bandwidth, this function is referred to as the *finite-frequency isochron* corresponding with (x_{s0},x_{r0},t_0). It is computed by solving

$$\begin{aligned}\left[\Delta + \omega^2 c(x)^{-2}\right]\hat{w}_s(x,\omega;x_{s0}) &= -\delta(x-x_{s0}),\\ \left[\Delta + \omega^2 c(x)^{-2}\right]\hat{u}^*(x,\omega;x_{s0}) &= \omega^2\exp(-\mathrm{i}\omega t_0)\delta(x-x_{r0}),\end{aligned}$$

followed by the cross correlation, $\mathrm{Re}\frac{1}{\pi}\int_\Omega \overline{\hat{w}_s(x,\omega;x_{s0})}\hat{u}^*(x,\omega;x_{s0})\,\mathrm{d}\omega$, cf. (4).

We briefly mention that the action of the normal operator $F[c]^*F[c]$ on r, is obtained by substituting the restricted solution $\delta\widehat{G}(x_r,\omega;x_s)$ of (3) for $\hat{d}(x_s,x_r,\omega)$ in (6). Essentially, the optimization results in compensating the image for this normal operator, while requiring that the so-called Bolker condition is satisfied [20, p.**].

The imaging described by (4)–(6) is a *multi-source*, multi-frequency operation. Upon invoking standard quadratures, and an FFT for the (inverse) Fourier transform, we obtain a discretization with n_s sources, n_ω frequencies and n_r receivers. In the next section, we will introduce the discretization of the Helmholtz operator, on a box of $\mathcal{O}(n^d)$ samples; we have $n_r \approx n^{d-1}$. For seismic data sets, n_s, n_r and n_ω are large. To generate an image, or inverse scattering gradient, one has to solve a couple of Helmholtz equations $n_s{\cdot}n_\omega$ times. This motivates the factorization of the (discretized) Helmholtz operator common to all sources, thus leading to fast solvers once the factorization is completed. We will develop and analyze the performance of such a solution method for frequency-domain imaging and inverse scattering in dimension $d = 2$, and compare it with the performance of iterative methods with preconditioners (such as the one developed by Erlangga et al. [7]), in Sections 4 and 5 of this paper.

2.3. Single-source inverse scattering transform

If the incident field (w_s in (5)) does not form caustics, it is possible to generate an (artifact-free) image using a single source (see Nolan and Symes [28]), with $\Sigma_s = \{x_{s0}\}$ say. By incorporating judiciously chosen operators in the right-hand side of the adjoint equation (cf. (6)) and the imaging integral (cf. (4)), the imaging operator ($F[c]^*$) can be turned into an inverse scattering transform thus bypassing the optimization. We summarize this transform for dimension $d = 2$, see Levy and Esmersoy [29]: Let n denote the outward normal to ∂X. The adjoint equation becomes

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\hat{u}^*(x,\omega;x_{s0}) = -\int_{\Sigma_r}\hat{d}(x_{s0},x_r,\omega)\ 2n(x_r)\cdot\nabla\delta(x-x_r)\mathrm{d}x_r. \tag{9}$$

The imaging integral is adjusted according to

$$\begin{aligned}\mathcal{I}_{s0}(x) = \frac{1}{|\Omega|^{-1}\int_\Omega|\hat{w}_s(x,\omega';x_{s0})|^2\mathrm{d}\omega'}\ \mathrm{Re}\frac{1}{\pi}\int_\Omega\frac{2\mathrm{i}}{\omega^3}\Big[&\omega^2 c(x)^{-2}\overline{\hat{w}_s(x,\omega;x_{s0})}\hat{u}^*(x,\omega;x_{s0})\\ &+\left(\nabla\overline{\hat{w}_s(x,\omega;x_{s0})}\right)\cdot(\nabla\hat{u}^*(x,\omega;x_{s0}))\Big]\mathrm{d}\omega,\end{aligned} \tag{10}$$

which becomes, locally, a direct estimate, $\langle r(x)\rangle$, of $r(x)$, that is, where there is illumination. The general development and analysis of the "common-source" class of inverse scattering transforms, and their relation to interferometry, can be found in Stolk et al. [30].

2.4. Imaging of the background: Optimization and gradient

We consider the imaging of common-source data and assume that the incident field does not form caustics as before:

$$\mathcal{I}_{s0}(x) = \frac{1}{2\pi}\int \overline{\hat{w}_s(x,\omega;x_{s0})}\hat{u}^*(x,\omega;x_{s0})\mathrm{d}\omega,$$

cf. (4). We develop a method for wave-equation reflection from the continuation [23] of this image at a point x_0 under a smooth perturbation, δc, of c, which is described by a complex phase multiplier $\exp[\mathrm{i}\,\delta\Phi(x_0,\omega;x_{s0})]$ applied to $\overline{\hat{w}_s(x_0,\omega;x_{s0})}\hat{u}^*(x_0,\omega;x_{s0})$ evaluated in the original background. Let

$\delta\phi(x_0,\omega;x_{s0}) = \mathrm{Re}\,\delta\Phi(x_0,\omega;x_{s0})$. We follow Xie and Yang [22] and apply the Rytov approximation, when

$$\delta\phi(x_0,\omega;x_{s0}) = \mathrm{Im}\left\{\frac{\overline{\delta\hat{w}_s(x_0,\omega;x_{s0})}}{\overline{\hat{w}_s(x_0,\omega;x_{s0})}} + \frac{\delta\hat{u}^*(x_0,\omega;x_{s0})}{\hat{u}^*(x_0,\omega;x_{s0})}\right\}. \tag{11}$$

Here, $\delta\hat{w}_s(x_0,\omega;x_{s0})$ satisfies

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\delta\hat{w}_s(x,\omega;x_{s0}) = 2\omega^2 c(x)^{-2}\frac{\delta c(x)}{c(x)}\hat{w}_s(x,\omega;x_{s0}),$$

cf. (3) and (5), while $\delta\hat{u}^*(x_0,\omega;x_{s0})$ satisfies

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\delta\hat{u}^*(x,\omega;x_{s0}) = 2\omega^2 c(x)^{-2}\frac{\delta c(x)}{c(x)}\hat{u}^*(x,\omega;x_{s0}),$$

cf. (3) and (6). We identify $\widehat{w_s}(x,\omega;x_s)$ with $\widehat{G}(x,\omega;x_s)$ and note that the oscillatory integral representations of $\hat{u}^*(x,\omega;x_{s0})$ and $\delta\widehat{G}(x,\omega;x_{s0})$ (non-smooth perturbation r) share the same phase function [30].

Upon dividing the phase perturbation by frequency, we identify the source-field and scattered-field finite-frequency traveltime functions:

$$\int\frac{1}{\omega}\delta\phi(x_0,\omega;x_{s0})\mathrm{d}\omega = \delta t_{w_s}(x_0;x_{s0}) + \delta t_{u^*}(x_0;x_{s0}), \tag{12}$$

where

$$\delta t_{w_s}(x_0;x_{s0}) = \int\frac{1}{\omega}\mathrm{Im}\left\{\frac{\overline{\delta\hat{w}_s(x_0,\omega;x_{s0})}}{\overline{\hat{w}_s(x_0,\omega;x_{s0})}}\right\}\mathrm{d}\omega = \int_X K_s(x_0,x';x_{s0})\frac{\delta c(x')}{c(x')}\mathrm{d}x' \tag{13}$$

and

$$\delta t_{u^*}(x_0;x_{s0}) = \int\frac{1}{\omega}\mathrm{Im}\left\{\frac{\delta\hat{u}^*(x_0,\omega;x_{s0})}{\hat{u}^*(x_0,\omega;x_{s0})}\right\}\mathrm{d}\omega = \int_X K_*(x_0,x';x_{s0})\frac{\delta c(x')}{c(x')}\mathrm{d}x', \tag{14}$$

in which

$$K_s(x_0,x';x_{s0}) = \int\frac{1}{\omega}\mathrm{Im}\left\{2\omega^2 c(x')^{-2}\frac{\overline{\widehat{G}(x_0,\omega;x')}\,\overline{\hat{w}_s(x',\omega;x_{s0})}}{\overline{\hat{w}_s(x_0,\omega;x_{s0})}}\right\}\mathrm{d}\omega \tag{15}$$

and

$$K_*(x_0,x';x_{s0}) = \int\frac{1}{\omega}\mathrm{Im}\left\{2\omega^2 c(x')^{-2}\frac{\overline{\widehat{G}(x_0,\omega;x')}\,\hat{u}^*(x',\omega;x_{s0})}{\hat{u}^*(x_0,\omega;x_{s0})}\right\}\mathrm{d}\omega. \tag{16}$$

We let the kernel K_s define the integral operator H_s and the kernel K^* define the integral operator H^*.

The natural formulation of the optimization problem for wave-equation reflection tomography follows to be

$$\langle\delta c\rangle = \arg\min\frac{1}{2}\left\|\int\frac{1}{\omega}\delta\phi(.,\omega;x_{s0})\mathrm{d}\omega - \left(H\frac{\delta c}{c}\right)(.;x_{s0})\right\|^2_{X_0}, \quad H = H_s + H_*.$$

Here, $X_0 \subset X$ denotes the part of X illuminated by the data. The quantity $\frac{1}{\omega}\delta\phi(.,\omega;x_{s0})$ ("data") is to be measured by intercepting the imaging process for inverse scattering. The phase perturbation criterion used here also appears in wave-equation transmission tomography, see Zhao et al. [31]; for an analysis of the relevant integral operator kernel, see De Hoop and Van der Hilst [32]. The gradient of the mismatch criterion is the composite tomographic image, $\mathcal{I} = \mathcal{I}_s + \mathcal{I}_*$, with $\mathcal{I}_s = H_s^*\int\frac{1}{\omega}\delta\phi(.,\omega;x_{s0})\mathrm{d}\omega$ and $\mathcal{I}_* = H_*^*\int\frac{1}{\omega}\delta\phi(.,\omega;x_{s0})\mathrm{d}\omega$. We have

$$\mathcal{I}_s(x) = \int \mathrm{Im}2\omega c(x)^{-2}\overline{\hat{w}_s(x,\omega;x_{s0})}\widehat{U}^*_s(x,\omega;x_{s0})\mathrm{d}\omega, \tag{17}$$

obtained by solving (5), referred to as "pde" as before, and

$$\left[\Delta + \omega^2 c(x)^{-2}\right]\widehat{U}^*_s(x,\omega;x_{s0}) = -\frac{1}{\overline{\hat{w}_s(x,\omega;x_{s0})}}\int\frac{1}{\omega'}\delta\phi(x,\omega';x_{s0})\mathrm{d}\omega', \tag{18}$$

referred to as "pde*" in the diagram,

$$\begin{array}{ccccc} & & \text{pde} & & \\ \text{“data”} & \searrow & \downarrow \text{source} & \nearrow\!\!\searrow & \rightarrow \text{image} \\ & & \text{pde}^* & & \end{array} \tag{19}$$

We note the difference in data-field flow between this diagram for reflection tomography and the diagram in (7) for inverse scattering. Similar adjoint states equations appear in the work of Liu and Tromp [33]. Likewise,

$$\mathcal{I}_*(x) = \int \mathrm{Im} 2\omega c(x)^{-2} \hat{u}^*(x,\omega;x_{s0}) \widehat{U}_*^*(x,\omega;x_{s0}) \mathrm{d}\omega, \tag{20}$$

obtained by solving (6) and

$$\left[\Delta + \omega^2 c(x)^{-2}\right] \widehat{U}_*^*(x,\omega;x_{s0}) = -\frac{1}{\hat{u}^*(x,\omega;x_{s0})} \int \frac{1}{\omega'} \delta\phi(x,\omega';x_{s0}) \mathrm{d}\omega'. \tag{21}$$

The kernel, $K_s(x_0,x;x_{s0}) + K^*(x_0,x;x_{s0})$, viewed as a function of x for x_0 fixed (expressing how a unit "data" sample at x_0 contributes to the tomographic image) is computed following this scheme, upon replacing (21) by

$$\left[\Delta + \omega^2 c(x)^{-2}\right] \widehat{U}_*^*(x,\omega;x_{s0}) = -\frac{1}{\hat{u}^*(x_0,\omega;x_{s0})} \delta(x - x_0).$$

3. Discretization of the Helmholtz equation

Here, we discuss an approximation of the Helmholtz equation based on a centered compact finite-difference discretization. We treat its different appearances in the previous section jointly by considering a general right-hand side, that is,

$$\left[-\Delta - \omega^2 c(x)^{-2}\right] \hat{u}(x,\omega) = \hat{s}(x,\omega), \tag{22}$$

while restricting the computations to the finite domain, X, introduced before. Thus, the original unbounded domain has been truncated, whence the outgoing Sommerfeld radiation condition has to be approximated at ∂X. Though the perfect matched layer (PML) approach (see, for example, [34,35]) could be applied, we choose, here, to use an absorbing boundary condition (ABC, see [36]) on ∂X. We take the second-order one:

$$n \cdot \nabla \hat{u} = -\mathrm{i}k\hat{u} - \frac{\mathrm{i}}{2k} \Delta_\tau \hat{u} \quad \text{on} \partial X, \tag{23}$$

with $k(\omega) = \omega c(x)^{-1}$ the wavenumber, n the outward normal direction to the boundary, and τ representing the boundary components of boundary normal coordinates. For the corners of the domain, we implement a specific corner boundary condition:

$$\sum_{i=1}^{d} \frac{\partial \hat{u}}{\partial n_i} = -\mathrm{i}k\hat{u}\left(d - \frac{1}{2}\right), \quad d = 2,3. \tag{24}$$

A centered compact finite-difference discretization with a $2J$ order scheme gives:

$$\sum_{j=1}^{J} \left[\frac{c_{1j}^{dJ}}{h^2}(\hat{u}_{i_1-j,i_2,\ldots,i_d} + \hat{u}_{i_1+j,i_2,\ldots,i_d}) + \cdots + \frac{c_{dj}^{dJ}}{h^2}(\hat{u}_{i_1,i_2,\ldots,i_d-j} + \hat{u}_{i_1,i_2,\ldots,i_d+j})\right] + \left(\frac{c^J}{h^2} - \frac{\omega^2}{c^2_{i_1,i_2,\ldots,i_d}}\right)\hat{u}_{i_1,i_2,\ldots,i_d} = \hat{s}_{i_1,i_2,\ldots,i_d}, \tag{25}$$

where h is the grid stepsize, $2J$ is the discretization order, and $c_{kj}^{dJ} (k = 1,2,\ldots,d)$ and c^J are the coefficients of the scheme, which is displayed in Fig. 1 in terms of stencils. To guarantee sufficient accuracy required in seismic imaging with typical wavefield sampling, we use a nine point, fourth ($J = 2$) order, finite difference scheme ([37,38]) to discretize (22)–(24). Fig. 1(a) shows the nine point stencil that we use to discretize the Laplace operator $(-\Delta)$. The Helmholtz operator adds another, variable, wavenumber term to the Laplacian $(-\Delta - \omega^2 c(x)^{-2})$. (Since this term is variable, the use of the classical, for $d = 2$, nine point stencil the grid points of which are distributed over a 3×3 square, does not provide the desired fourth order accuracy.)

The fourth order finite difference scheme described above, leads to the linear system of equations,

$$\mathbf{A}(\omega)\mathbf{u} = \mathbf{s}, \tag{26}$$

where $\mathbf{A}(\omega)$ is a sparse complex matrix depending on the frequency, $\mathbf{s}$ denotes the vector of samples $\hat{u}_{i_1,i_2,\ldots,i_d}$ of the source $\hat{s}$, and $\mathbf{u}$ denotes the vector of samples $\hat{u}_{i_1,i_2,\ldots,i_d}$ of $\hat{u}$. Although the original Helmholtz operator is self adjoint, in view of the absorbing boundary conditions, the $\mathbf{A}(\omega)$ is neither Hermitian nor positive definite.

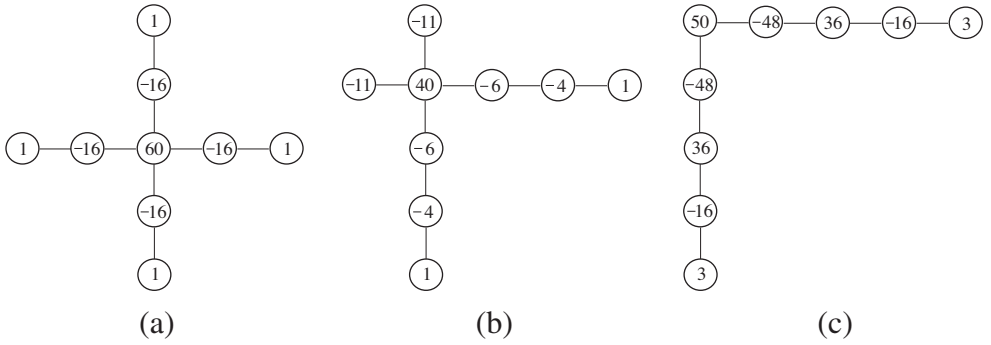

Fig. 1. Fourth-order nine point stencils for $d = 2$; (a) inner grid point stencil, (b) a near-boundary point stencil, and (c) a corner point stencil.

For each fixed ω, the system needs to be solved for multiple right-hand sides, **s**. For different ω, the matrix $\mathbf{A}(\omega)$ maintains the same nonzero pattern. Thus, a robust and efficient direct factorization of $\mathbf{A}(\omega)$ can be attractive. A direct method typically involves four stages [39,40]: Node ordering, symbolic factorization, numerical factorization, and solution. The first and second stages only need to be done once for all frequencies, sources, and receivers. The third stage needs to be done once for each frequency. However, direct solvers are often considered expensive because of the problems of fill-in or loss of sparsity.

Here, we use a recently developed robust approximate structured factorization method [2,3,41]. The main idea of the solver is to fully integrate graph-based sparse matrix techniques, structured matrix compressions, and robust enhancements. The dense fill-in will be approximated by structured matrices and are thus data sparse. Dense off-diagonal blocks are compressed in the direct factorization of the matrix. These compressions improve both the efficiency and the robustness. The approximate factorization can be carried out to any specified accuracy. Moreover, the factorization appears to be relatively insensitive to frequency, or wavelength, in various test problems using the elasticity, Helmholtz, and Maxwell equations [2,3].

Since the current robust structured solver in [2] is designed for symmetric positive definite matrices, we consider the following normal system for (26):

$$\mathbf{M}(\omega)\mathbf{u} = \mathbf{b}, \tag{27}$$

where $\mathbf{M}(\omega) = \mathbf{A}(\omega)^H\mathbf{A}(\omega)$ and $\mathbf{b} = \mathbf{A}(\omega)^H\mathbf{s}$. Matrix $\mathbf{A}(\omega)$ is sparse, and it is not expensive to form (27). It is well known that, in Helmholtz equation problems, the condition number of this $\mathbf{M}(\omega)$ becomes very large for high frequencies. However, we will demonstrate in the next section that our new Helmholtz solver remains numerically stable, which is rarely the case in iterative methods. In addition, $\mathbf{M}(\omega)$ will be denser than $\mathbf{A}(\omega)$. Both $\mathbf{A}(\omega)$ and $\mathbf{M}(\omega)$ have order $N = n^d$, and their half bandwidths are Jn^{d-1} and $2Jn^{d-1}$, respectively, with $J = 2$. On the other hand, our structured solver remains efficient.

4. Robust and efficient solution of the discretized system

In this section, we briefly review the structured solver in [2] specifically for solving (27).

4.1. Nested dissection ordering

During the factorization of **M**, new nonzero entries will be introduced into the factors which causes fill-in. The ordering of the mesh points can highly affect the fill-in and thus the performance of the factorization. For 2D problems, the nested dissection ordering [42] and its generalizations enable $O(N^{3/2})$ complexity factorizations with $O(N\log N)$ storage which are optimal in exact arithmetic, in general [43].

Without loss of generality, assume the mesh is regular and with size $n \times n$, where $n = \sqrt{N}$ (**M** is thus $N \times N$). Nested dissection recursively divides the mesh into subregions with separators. At the first level, a separator divides the entire grid into three pieces, including two subregions and the separator itself. The two subregions are further divided recursively. See Fig. 2. Lower level separators and mesh points are ordered before upper level ones. After reordering, the matrix looks like the one in Fig. 3. Later, we assume **M** is already reordered by nested dissection. During the factorization of **M**, the elimination of a mesh point mutually connects points which are previously connected to it. This creates fill-in. Thus, in standard factorization of **M**, its factor is generally much denser.

4.2. Supernodal multifrontal method

The factorization of **M** is organized by the multifrontal method [44,45]. The basic idea of the method is to reorganize the overall factorization of a large sparse matrix into partial updates and factorizations of small dense matrices. The method

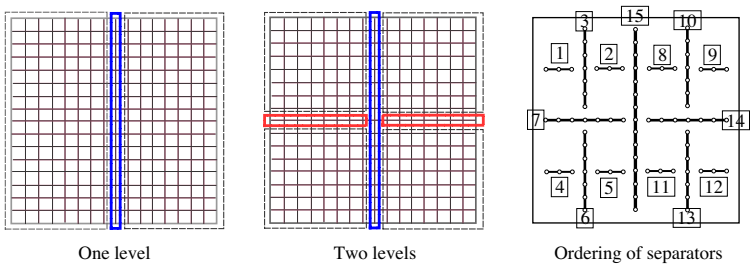

Fig. 2. Nested dissection ordering of mesh points.

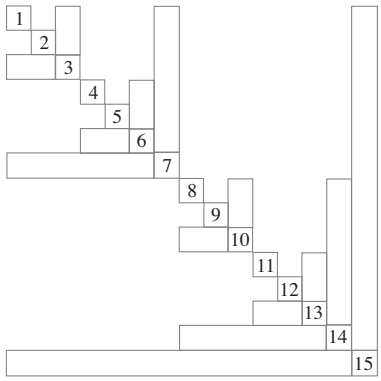

Fig. 3. Block form of **M** corresponding to the third graph of Fig. 2 after nested dissection ordering of the rows and columns.

eliminates variables and accumulates updates according to an assembly tree [46–49] where each variable (or matrix row or column) corresponds to a node in the elimination tree.

Here we use a supernodal version, that is, each separator in nested dissection is treated as a supernode in the assembly tree. See Fig. 4 for an example. Variables are eliminated along the postordering of the assembly tree. Information from a child tree node is passed to its parent only. Two types of matrices are considered in the multifrontal method: *update matrices*, corresponding to the information contributed to parents, and *frontal matrices*, corresponding to the collection of information of nodes themselves and related upper level nodes. Update matrices from children are assembled into frontal matrices with an operation called *extend-add*, which aligns indices, expands matrices, and adds entries.

The multifrontal method is a powerful tool for solving large sparse linear systems [50,51], and has been widely used in numerical methods for partial differential equations, optimization, fluid dynamics, and other areas. It takes good advantage of dense block matrix operations in factorizing sparse problems, and is also good for parallelization.

4.3. Rank property and robust structured factorization

It has been shown that during the factorization of the discretized matrices arising from some problems such as elliptic partial differential equations, the off-diagonal of certain Schur complements have bounded *numerical ranks* independent of mesh sizes [52–54]. (A numerical rank is the number of singular values greater than a given absolute tolerance or a relative tolerance times the largest singular value.) The Green's functions for these problems are smooth away from the diagonal singularity. In the factorization of **M**, we notice that the frontal and update matrices in the supernodal multifrontal

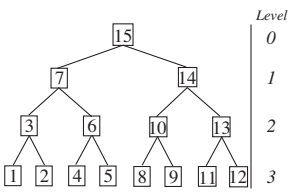

Fig. 4. Elimination tree for the third graph in Fig. 2.

method also have relatively small off-diagonal numerical ranks. This means, the dense frontal and update matrices are compressible. Thus, the method in [2] forms each dense frontal matrix first and then partially factorize it. The leading block is directly factorized into triangular structured factors. The Schur complement remains dense and is used to form the upper level frontal matrix. The solver can still perform well even if the off-diagonal numerical ranks are not extremely small. See [2,3] for some numerical examples.

The solver in [2] integrates a robustness technique using an implicit Schur compensation. During this factorization, off-diagonal blocks are compressed to increase efficiency. In the meantime, the Schur complements are automatically compensated with the information dropped in the compression. No extra cost or stabilization step is needed.

The structured matrices used in the solver are called hierarchically semiseparable (HSS) matrices [55–57]. Fig. 5 shows an example. An off-diagonal block of an HSS structure is defined to be a block row (column) excluding the diagonal block. Off-diagonal blocks are defined recursively for all levels of the partition of the matrix. These off-diagonal blocks are compressed and the compressed representations actually appear in the HSS representation. If the maximum numerical rank of the off-diagonal blocks at all levels is small as compared with the matrix size, we say the matrix has a low-rank property. For HSS matrices with the low-rank property, standard matrix operations such as matrix multiplication, system solution, etc. can be done efficiently in nearly linear time [55–57].

As an example, for the discretized matrix $\mathbf{M}$ in our problem, we consider a frontal matrix H corresponding to the top level separator in nested dissection. For different frequencies and different tolerances in the approximation, the maximum off-diagonal numerical ranks for H are given in Table 1.

After applying the structured factorization algorithm in [2] to $\mathbf{M}$, we obtain an approximate triangular factor $\mathbf{L}$ which is structured. Fig. 6 shows an example with 3 blocks. A traditional factorization generates a factor which is much denser than the original matrix. On the other hand, the structured factorization in [2] gives a data-sparse factor in compressed form.

4.4. Efficiency, robustness, and frequency insensitivity

The algorithm in [2] works as a black-box solver. For $d = 2$, it costs $O(rN\log N)$ flops to approximately factorize $\mathbf{M}$, where r is the maximum numerical rank in all off-diagonal block compressions. The storage requirement is $O(N\log(r\log N))$. The cost

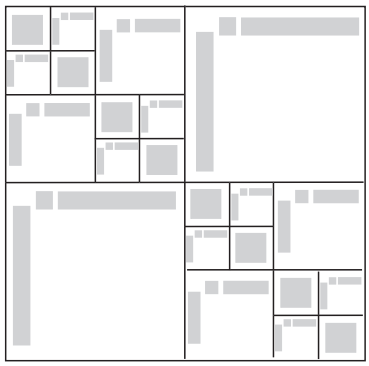

Fig. 5. A Pictorial representation of an HSS matrix showing its partition and the compressed form of the off-diagonal blocks.

Table 1
Maximum off-diagonal numerical ranks of an order 2204 frontal matrix in the supernodal multifrontal method for our problem with different relative tolerances τ.

τ		10^{-1}	10^{-2}	10^{-4}	10^{-6}	10^{-8}	10^{-10}	10^{-12}
$\frac{\omega}{2\pi}$	10	29	45	64	90	136	164	182
	20	29	46	67	139	196	233	254
	30	29	46	79	199	272	313	329
	40	29	46	122	269	368	389	404

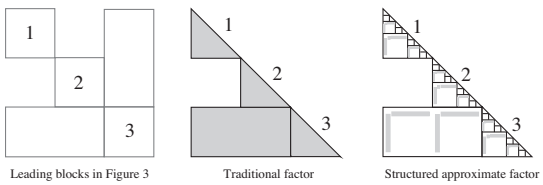

Fig. 6. A matrix after nested dissection ordering, its traditional factor, and its approximate structured factor.

Table 2
Statistics of solving (27) on a 1001 × 551 mesh with different frequencies (from the lens model example described in Sub Section 5.4) which shows the frequency insensitivity of the structured solver in [2].

$\frac{\omega}{2\pi}$	5	10	20	30	40
Factorization cost ($\times 10^{11}$ flops)	2.931	2.957	3.305	3.146	3.290
Solution ($\times 10^9$ flops)	1.468	1.511	1.625	1.766	1.922
Storage ($\times 10^9$)	2.583	2.719	3.081	3.528	4.038

for solving (27) using the approximate factor is $O(N\log(r\log N))$ flops. The parameter r depends on the tolerance and is relatively small as compared to N especially when only modest accuracy is desired. Thus, the storage and solution cost are nearly linear in N.

Due to the robustness technique, the approximate factorization $\mathbf{L}\mathbf{L}^T$ is guaranteed to exist for any tolerance. $\mathbf{L}\mathbf{L}^T$ has enhanced positive definiteness and no breakdown will occur like in many other approximate or incomplete factorizations.

Unlike methods such as multigrid where coarse mesh information is used to approximate fine meshes, this structured factorization uses a reasonable amount of information at all mesh points to approximate the exact matrix. Our numerical tests indicate that the factorization is relatively insensitive to a large range of frequencies, which is different from many other algorithms that are not numerically stable in high frequencies. Table 2 shows the complexity and storage for solving (27) with different ω for a fixed mesh.

5. Numerical tests

5.1. Algorithm summary

We use the approximate factorization algorithm in [2] to directly solve (27) for different frequencies, sources, and receivers. The overall procedure is as follows, in contrast with the standard sparse direct solution process in [39,40].

1. *Node ordering*. Use nested dissection to order mesh points. Organize mesh points into separators. This is done once for all ω.
2. *Symbolic factorization*. Approximately predict fill-in and storage for HSS matrices based on numerical rank estimation with given tolerances. Decide optimal total factorization levels and structured factorization levels. When only modest accuracy is desired, this also only needs to be done once for all ω.
3. *Numerical factorization*. For each ω, factorize **M** once. Separators are eliminated with the supernodal multifrontal method. Each current separator is eliminated, or dense frontal matrices are partially factorized into HSS factors. Schur complements or update matrices are formed. Structured factors are generated following the assembly tree.
4. *System solution*. Use the structured factors to solve the system for each **s**. Traverse the assembly tree to solve the overall system. Solve intermediate structured systems with triangular HSS solvers.

Since the complexity for the solution step is almost linear in N, we can also include few steps of iterative refinements or CG iterations to further improve the solution.

Based on the cost of the structured solver in the previous section, we have the complexity for solving all the systems as shown in Table 3. For comparison, we also included the cost for the classical multifrontal direct factorization with nested dissection whose cost is optimal for exact factorizations for 2D problems. The structured approximate solver with modest accuracy is generally much faster than both the classical direct factorization and the multilevel Krylov-multigrid method (MKMG)[58]. The structured solver for $d = 3$ in Table 3 is still in progress, and its complexity is derived based on a preliminary version without robustness enhancement. We expect to obtain a similar complexity count when robustness enhancement is integrated.

5.2. Accuracy tests

In this subsection, we verify that our discretization scheme does achieve a fourth-order accuracy. We conducted our test in a 1000 m × 1000 m homogeneous medium with wavespeed 1500 m/s at a fixed relatively high frequency, $\frac{\omega}{2\pi} = 30$ Hz. We take

Table 3
Complexity comparison of the multilevel Krylov-multigrid method (MKMG)[58], classical direct factorization with nested dissection ordering, and the new structured approximate factorization with nested dissection, where n_f is the number of frequencies, n_s is the number of sources, n_{it} is the number of iterations for convergence (usually quite large in high frequencies), and $n^d(=N)$ is the matrix size.

Algorithm	Complexity	
	$d = 2$	$d = 3$
Multilevel Krylov-multigrid method	$n_f n_s n_{it} O(n^2)$	$n_f n_s n_{it} O(n^3)$
Classical direct factorization	$n_f[O(n^3) + n_s O(n^2\log n)]$	$n_f[O(n^6) + n_s O(n^4)]$
Structured approximate factorization	$n_f[O(rn^2\log n) + n_s\, O(n^2\log(r\log n))]$	$n_f[O(rn^3) + n_s O(n^3)]$

step sizes h according to h = 10 m, 5 m, 2.5 m, 1.25 m, 0.625 m. We use the solution computed in the finest mesh (h = 0.625 m) as our reference and calculate relative residuals for the four coarser meshes. The results are listed in Table 4 and illustrated in Fig. 7. We observe, for example, an accuracy of 3 digits at a sampling rate of 10 points per (minimum) wavelength.

5.3. Performance tests for linear system solutions

Here, we demonstrate the complexity and storage of the structured approximate solver for problems with dimension d = 2. (Numerical results for d = 3 are expected to appear in future work.) For **M** with different sizes, the costs and storage are listed in Table 5. Every time the mesh dimension n doubles, the matrix size quadruples, and the cost and storage of the method increase. The ratios of increase in the solution cost and storage reduce for large n and approach 4, which indicates nearly linear solution cost and storage. The ratio of increase in the factorization cost reduces slower, but is also expected to be close to 4 when n is sufficiently large.

We also compare this structured factorization with the classical multifrontal direct factorization. See Table 6 and Figs. 8–10. When N quadruples, the factorization cost increases with a factor close to 8, especially when n is large. Clearly, although the classical factorization has optimal complexity for 2D problems, it is still not practical for large problems.

Table 4
Statistics of fourth order accuracy test on a 1000 m × 1000 m model with wavespeed 1500 m/s at fixed frequency 30 Hz from coarser to finer meshes.

Mesh step size $h(m)$	10	5	2.5	1.25
Grid points per wavelength	5	10	20	40
Relative error ($\times 10^{-4}$)	141.45	8.67	0.51	0.03

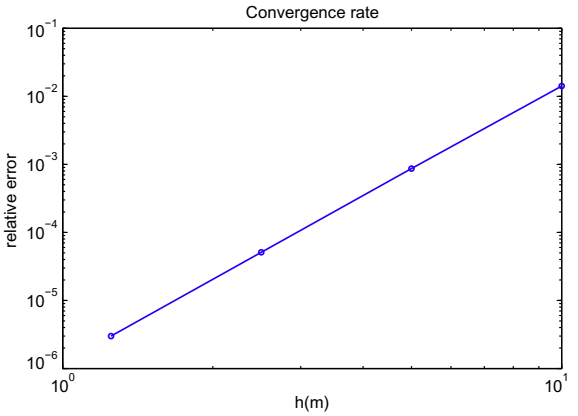

Fig. 7. Computational test of fourth order numerical accuracy in step size.

Table 5
Computational costs and storage for the approximate structured direct solver with relative tolerance $\tau = 10^{-6}$ and $\frac{\omega}{2\pi} = 30$ Hz.

Mesh dimension n	128	256	512	1024	4096
Factorization cost (flops)	2.25*E*9	1.47*E*10	9.77*E*10	6.33*E*11	3.77*E*12
Solution cost (flops)	1.92*E*7	8.95*E*7	4.12*E*8	1.78*E*9	7.40*E*9
Storage	8.69*E*6	3.41*E*7	1.43*E*8	5.94*E*8	2.33*E*9

Table 6
Computational costs and storage for the classical multifrontal direct factorization with $\frac{\omega}{2\pi} = 30$ Hz.

Mesh dimension n	128	256	512	1024	4096
Factorization cost (flops)	2.22*E*9	1.81*E*10	1.44*E*11	1.21*E*12	8.98*E*12
Solution cost (flops)	1.97*E*7	1.03*E*8	5.06*E*8	2.46*E*9	1.10*E*10
Storage	9.21*E*6	4.58*E*7	2.14*E*8	1.00*E*9	4.27*E*9

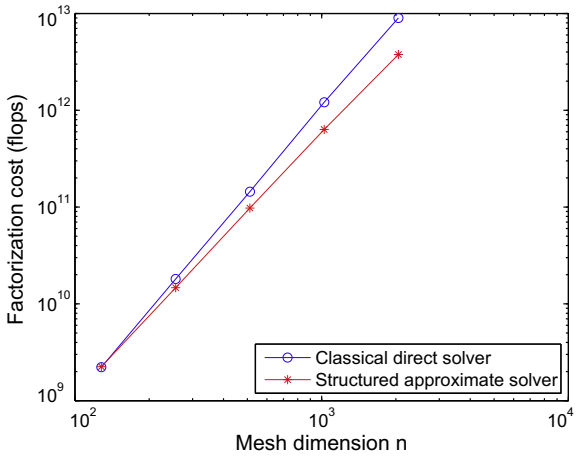

Fig. 8. Factorization cost comparison between the structured factorization and classical direct factorization.

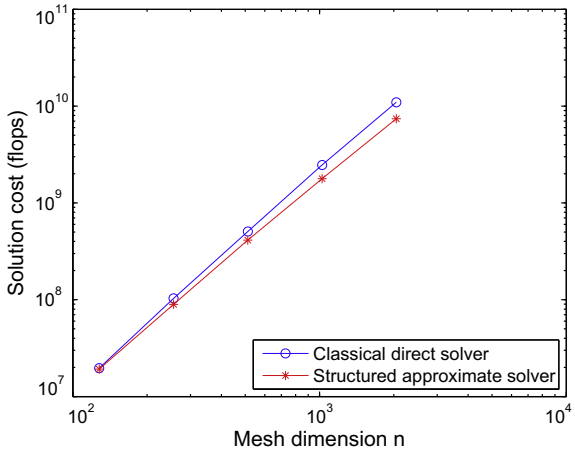

Fig. 9. Solution cost comparison between the structured factorization and classical direct factorization.

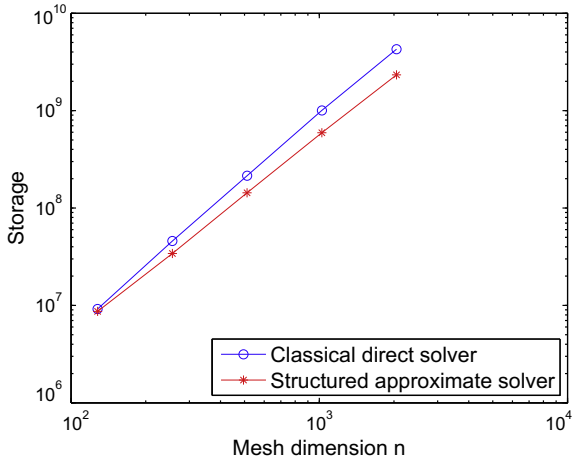

Fig. 10. Storage comparison between the structured factorization and classical direct factorization.

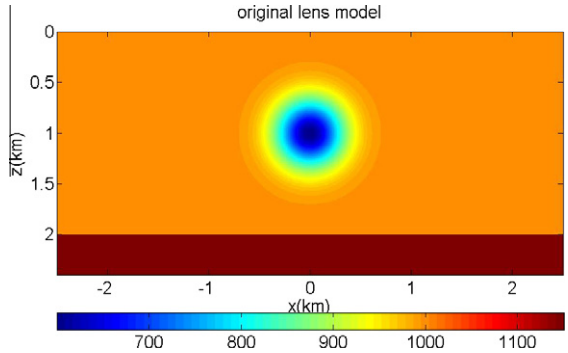

Fig. 11. Original model, containing a low velocity lens (background) and a horizontal reflector (contrast) at a depth of 2 km. The model is sampled on a Cartesian grid with a 5 m stepsize.

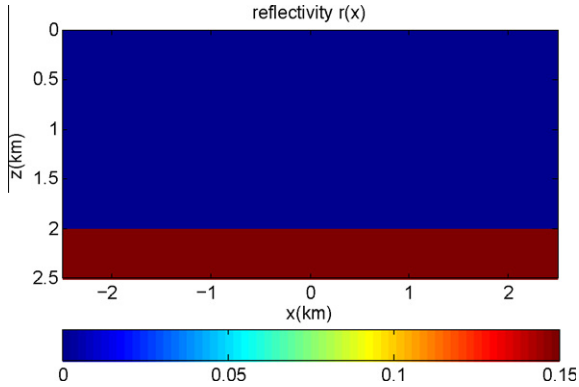

Fig. 12. The reflectivity $r(x) = \delta c(x)/c(x)$ of the lens velocity model displayed in Fig. 11.

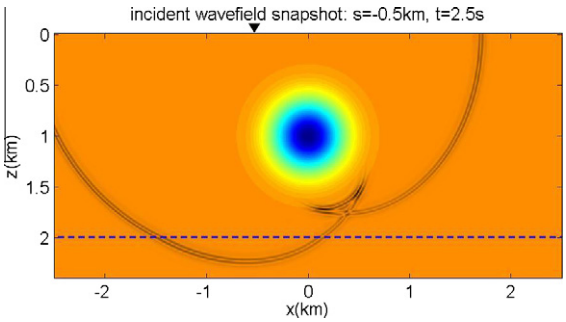

Fig. 13. Snapshot (early time) of an incident wavefield using the model illustrated in Fig. 11; the location of the source is indicated by the solid triangle. The frequency bandwidth is 5–40 Hz.

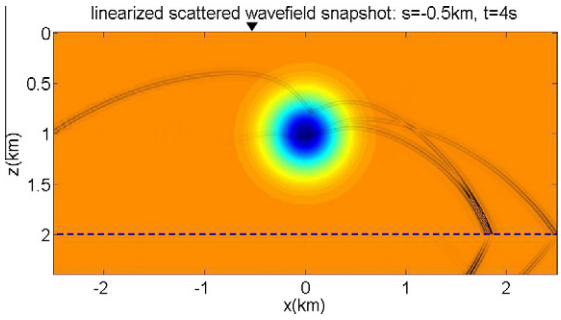

Fig. 14. Snapshot (later time) of the linearized scattered field using the model illustrated in Fig. 11; the location of the source is indicated by the solid triangle. The frequency bandwidth is 5–40 Hz.

5.4. Example: Lens model

Using the structured approximate direct solver for the discrete Helmholtz equation, we illustrate, in dimension $d = 2$, the different components that make up the inverse scattering program presented in Section 2. Throughout, the spatial sampling rate has been set to 10 points per minimum wave length. The original model, shown in Fig. 11, consists of a smooth low

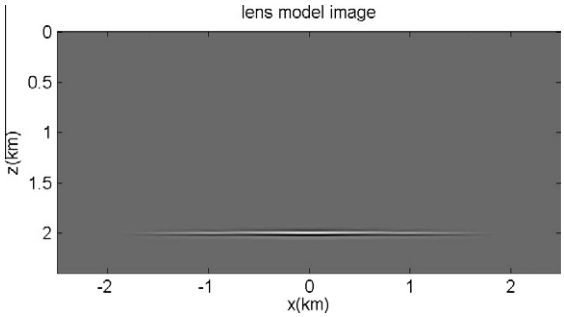

Fig. 15. Image of the reflector in Fig. 11, or inverse scattering gradient, using all the data (multi-source adjoint states computation.)

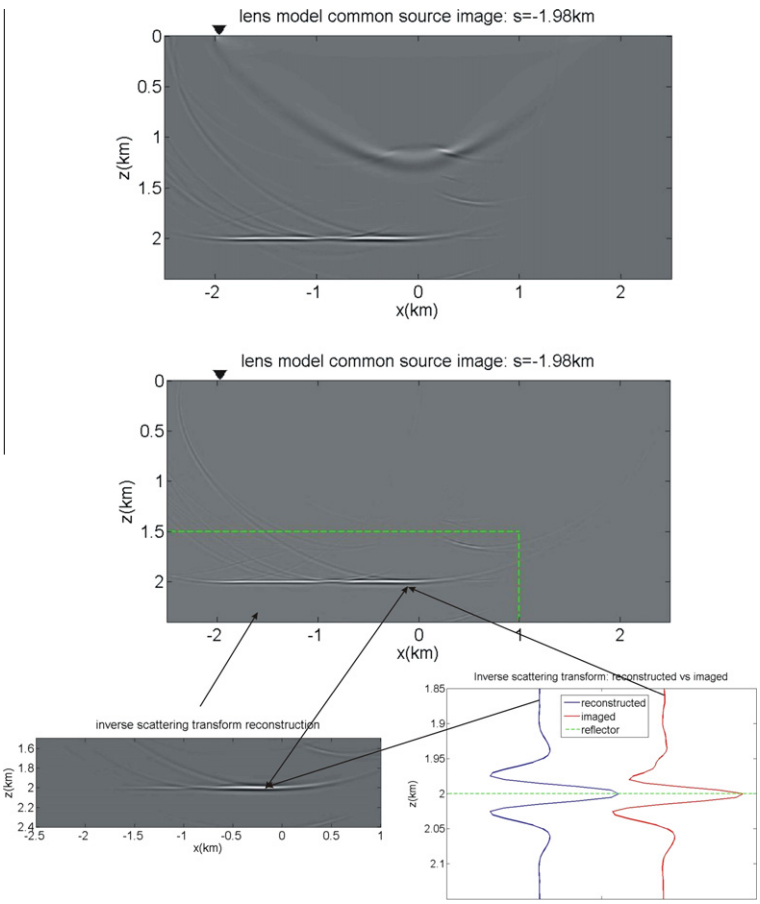

Fig. 16. Top: Single-source image (adjoint states computation using common source data), or inverse scattering gradient, with $x_{s0} = (-1.98, 0)$ km indicated by the solid triangle. Middle: Single-source image obtained after applying a pseudo-differential cutoff removing constituents corresponding with scattering over π; the low-frequency artifact has disappeared. Bottom left: Single-source reconstruction; we note the reduction of limited illumination effects. Bottom right: Reconstructed (left) and imaged (right) regularizations of r below the surface location $x = (-0.1, 0)$ km (with the source at $x_{s0} = (-1.98, 0)$ km); we note that the reconstruction is close zero phase. The incident field does not develop caustics.

wavespeed lens (c) and a flat discontinuity (r), which is displayed in Fig. 12. Using this model, we generated a multi-source synthetic data set.

We begin with illustrating the incident (w_s) and linearized scattered (δG) fields; snapshots are shown in Fig. 13 (incident) and Fig. 14 (linearized scattered). We note the formation of caustics due to the presence of the lens. An image, that is, the inverse scattering gradient (initial iteration, with $r \equiv 0$), using the multi-source synthetic data and the lens (c) as background, is plotted in Fig. 15.

We focus then on images generated from single source data, that is, common-source subsets of the full synthetic data set. If the source position is such that the incident field does not form caustics, the image is artifact-free (Fig. 16 middle); if the incident field does form caustics, the image contains artifacts (Fig. 18) due to violation of the Bolker condition. Also, in Fig. 16 top we show the generation of artifacts from wave constituents corresponding with scattering over π.

The inverse scattering transform yields a direct estimate of r in Fig. 16 bottom, using the single source data used to generate the image of r in Fig. 16 middle. We show the functions representing the reconstructed and imaged regularizations of r below a central surface location.

In Fig. 17, we generate a shot record associated with the source $x_{s0} = (-1.98, 0)$ km to further investigate the artifacts due to the scattering over π.

The anatomy of the inverse scattering gradient is best illustrated by the associated finite-frequency isochrons. We choose two different data points, all asymptotically corresponding with reflections off the flat discontinuity in the original model,

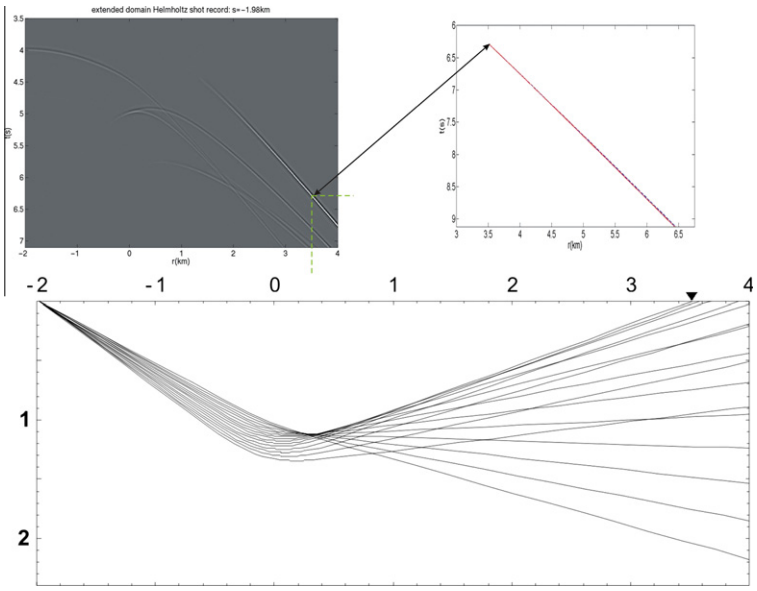

Fig. 17. Top left: shot record (generated by the Helmholtz solver) corresponding with the source $x_{s0} = (-1.98, 0)$ km; note the occurrence of a caustic shadow (indicated by an arrow) associated with turning rays (representing scattering over π). Top right: travel time curve showing the (tip) of the cusped caustic. Bottom: geometrical rays showing the turning rays and the caustic. The triangle indicates the receiver nearest the source where a turning ray reaches the (acquisition) surface.

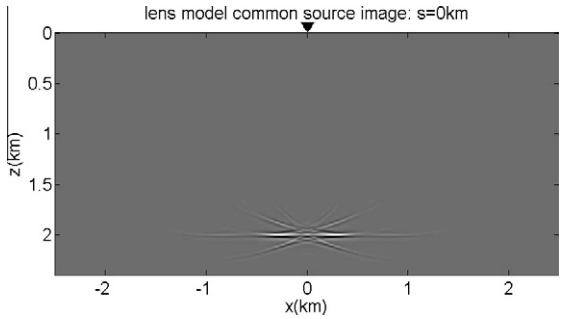

Fig. 18. Single-source image (adjoint states computation using common source data), or inverse scattering gradient, with $x_{s0} = (0, 0)$ km indicated by the solid triangle. The incident field develops caustics, and, hence, the image contains artifacts.

see Figs. 19 and 20; for comparison, we also plot the singular supports derived from (high frequency) asymptotic considerations. The image of the discontinuity follows the envelope of such isochrons.

Finally, we illustrate the gradient for wave-equation reflection tomography (final iteration), using single-source data. The associated kernel, for image point x_0 (and x_{s0}) fixed, is shown in Fig. 21. We choose x_0 to lie on the discontinuity and superimpose an associated finite-frequency isochron. High frequency asymptotic considerations lead to the (broken) ray geometrical counterpart of the finite frequency reflection tomography kernel, which can still be recognized in Fig. 21.

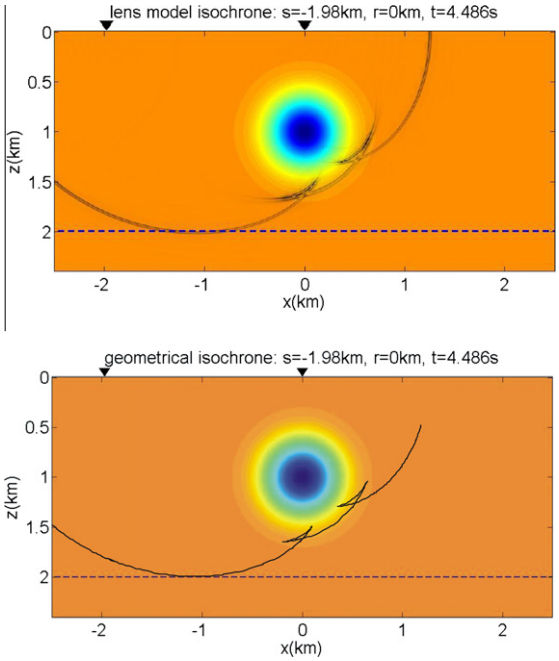

Fig. 19. Top: finite-frequency isochron, or kernel of the inverse scattering gradient, for $x_{r0} = (0,0)$ km, $t_0 = 4.486$ s and $x_{s0} = (-1.98,0)$ km. Bottom: singular support obtained by methods of geometrical acoustics.

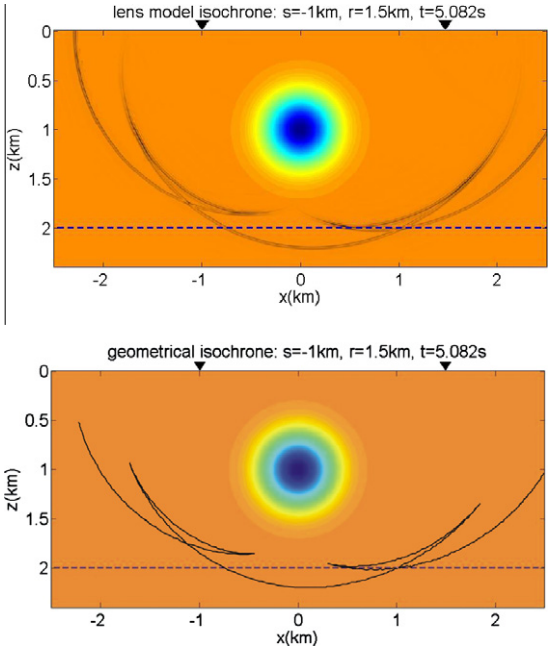

Fig. 20. Top: finite-frequency isochron, or kernel of the inverse scattering gradient, for $x_{r0} = (1.5,0)$ km, $t_0 = 5.082$ s and $x_{s0} = (-1,0)$ km. Bottom: singular support obtained by methods of geometrical acoustics.

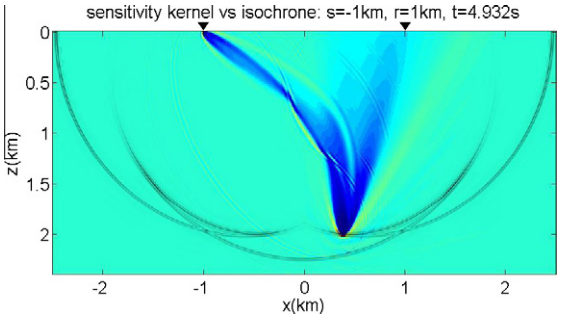

Fig. 21. Kernel of the reflection tomography gradient ($x_{s0} = (-1,0)$ km, $x_0 = (0.38,2)$ km), superimposed on the kernel of an inverse scattering gradient ($x_{r0} = (1,0)$ km and $t_0 = 4.932$ s), illustrating their interplay of sensitivities in the inverse scattering program.

6. Conclusions

We presented a joint acoustic/seismic inverse scattering and finite-frequency (reflection) tomography program, formulated as a coupled set of optimization problems, in terms of inhomogeneous Helmholtz equations. We used a higher (fourth-) order finite difference scheme for these Helmholtz equations to guarantee 3 digits of accuracy at a sampling rate of 10 points per minimum wave length. We applied the second-order absorbing boundary condition, and subjected it to a finite difference approximation of the same order. This yields no complication, since our solution approach does not use the explicit structure of the coefficient matrix defining the resulting system of algebraic equations. The overall complexity for computing the gradients or images in our optimization problems has two parts, viz., the cost for all the matrix factorizations which is roughly $\mathcal{O}(rN)$ (for example $\mathcal{O}(rN \log N)$ when $d = 2$) times the number of frequencies, and the cost for all solutions by substitution which is roughly $\mathcal{O}(N)$ (for example $\mathcal{O}(N \log(r \log N))$ when $d = 2$) times the number of frequencies times the number of sources, where $N = n^d$ if n is the number of grid samples in any direction. With this complexity, the multi-frequency approach to inverse scattering and finite-frequency tomography becomes computationally feasible for dimensions $d = 2$ and 3.

Acknowledgment

The authors thank Anton Duchkov for carrying out the geometrical acoustics computations. The authors are also very grateful to the anonymous referees for their valuable comments.

References

[1] K. Marfurt, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, Geophysics 49 (1984) 533–549.
[2] J. Xia, M. Gu, Robust structured multifrontal factorization and preconditioning for discretized PDEs, preprint, <http://math.purdue.edu/∼xiaj/work/mfhssr.pdf>.
[3] J. Xia, M. Gu, Robust structured factorization and preconditioning for spd matrices, SIAM J. Matrix Anal. Appl., <http://math.purdue.edu/∼xiaj/work/schurmonospd.pdf>, submitted for publication.
[4] H. Elman, O. Ernst, D. O'Leary, A multigrid method enhanced by krylov subspace iteration for discrete helmholtz equations, SIAM J. Scient. Comp. 23 (2001) 1291–1315.
[5] R.-E. Plessix, A helmholtz iterative solver for 3D seismic-imaging problems, Geophysics 72 (2007) SM185–SM194.
[6] Y. Erlangga, C. Vuik, C. Oosterlee, On a class of preconditioners for solving the helmholtz equation, Appl. Numer. Math. 50 (2004) 409–425.
[7] Y. Erlangga, C. Oosterlee, C. Vuik, A novel multigrid based preconditioner for heterogeneous helmholtz problems, SIAM J. Scient. Comput. 27 (2006) 1471–1492.
[8] C. Riyanti, Y. Erlangga, R.-E. Plessix, W. Mulder, C. Vuik, C. Oosterlee, A new iterative solver for the time-harmonic wave equation, Geophysics 71 (2006) E57–E63.
[9] S. Operto, J. Virieux, P. Amestoy, J. L'Excellent, L. Giraud, H. Hadj Ali, A helmholtz iterative solver for 3D seismic-imaging problems, Geophysics 72 (2007) SM185–SM194.
[10] E. Larsson, A domain decomposition method for the helmholtz equation in a multilayer domain, SIAM J. Scient. Comp. 20 (1999) 1713–1731.
[11] L. Sirgue, R. Pratt, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, Geophysics 69 (2004) 231–248.
[12] W. Mulder, R.-E. Plessix, How to choose a subset of frequencies in frequency-domain finite-difference migration, Geophys. J. Int. 158 (2004) 801–812.
[13] C. Shin, W. Ha, A comparison between the behavior of objective functions for waveform inversion in the frequency and laplace domains, Geophysics 73 (2008) VE119–VE133.
[14] J. Virieux, S. Operto, An overview of full-waveform inversion in exploration geophysics, Geophysics 74 (2009) WCC1–WCC26.
[15] P. Lailly, The seismic inverse problem as a sequence of before stack migrations, in: Proceedings of the International Conference on Inverse Scattering, Theory and Applications, SIAM, Tulsa, OK, 1983, pp. 206–220.
[16] A. Tarantola, Inverse Problem Theory, Elsevier, New York, 1987.
[17] H. Kühl, M. Sacchi, Least-squares wave-equation migration for avp/ava inversion, Geophysics 68 (2003) 262–273.
[18] R. Pratt, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model, Geophysics 64 (1999) 888–901.
[19] C. Ravaut, S. Operto, L. Improta, J. Virieux, A. Herrero, 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 (2004) 1032–1056.
[20] C. Stolk, M. De Hoop, Microlocal analysis of seismic inverse scattering in anisotropic elastic media, Commun. Pure Appl. Math. 55 (2002) 261–301.

[21] M. De Hoop, R. Van der Hilst, P. Shen, Wave-equation reflection tomography: annihilators and sensitivity kernels, Geophys. J. Int. 167 (2006) 1332–1352.
[22] X. Xie, H. Yang, The finite-frequency sensitivity kernel for migration residual moveout and its applications in migration velocity analysis, Geophysics 73 (2008) S241–S249.
[23] A. Duchkov, M. De Hoop, A. Sá Barreto, Evolution-equation approach to seismic image, and data, continuation, Wave Motion 45 (2008) 952–969.
[24] S. Wang, M. De Hoop, Illumination analysis of wave-equation imaging with "curvelets", submitted for publication.
[25] V. Brytik, M. De Hoop, M. Salo, Sensitivity analysis of wave equation tomography: a multi-scale approach, Journal of Fourier Analysis and Applications, submitted for publication.
[26] W.W. Symes, Lecture notes: Short Course on Mathematical Foundations of Reflection Seismology, T.R.I.P. consortium.
[27] C. Stolk, M. De Hoop, W. Symes, Kinematics of shot-geophone migration, Geophysics 74 (2009) WCA19–WCA34.
[28] C. Nolan, W. Symes, Global solution of a linearized inverse problem for the wave equation, Commun. Partial Differ. Equat. 22 (1997) 919–952.
[29] B. Levy, C. Esmersoy, Variable background born inversion by wavefield backpropagation, SIAM J. Appl. Math. 48 (1988) 952–972.
[30] C. Stolk, M. De Hoop, T. Op't Root, Inverse scattering of seismic data in the reverse time migration (rtm) approach, <http://www.math.purdue.edu/~mdehoop/10_topics/rtm.pdf>, preprint.
[31] L. Zhao, T. Jordan, C. Chapman, Three-dimensional fréchet differential kernels for seismic delay times, Geophys. J. Int. 141 (2000) 558–576.
[32] M. De Hoop, R. Van der Hilst, On sensitivity kernels for 'wave-equation' transmission tomography, Geophys. J. Int. 160 (2005) 621–633.
[33] Q. Liu, J. Tromp, Finite-frequency kernels based upon adjoint methods, Bull. Seism. Soc. Amer. 96 (2006) 2383–2397.
[34] E. Turkel, A. Yefet, Absorbing pml boundary layers for wave-like equations, Appl. Numer. Math. 27 (1998) 533–557.
[35] I. Singer, E. Turkel, A perfectly matched layer for the Helmholtz equation in a semi-infinite strip, J. Comput. Phys. 201 (2004) 439–465.
[36] A. Bamberger, P. Joly, J. Roberts, Second-order absorbing boundary conditions for the wave equation: a solution for the corner problem, SIAM J. Numer. Anal. 27 (1990) 323–352.
[37] I. Singer, E. Turkel, High-order finite difference methods for the Helmholtz equation, Comput. Methods Appl. Mech. Eng. 163 (1998) 343–358.
[38] S. Lele, Compact finite difference schemes with spectral-like resolution, J. Comput. Phys. 103 (1992) 16–42.
[39] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, H. Van den Vorst, Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide, SIAM, Philadelphia, PA, 2000.
[40] Y. Saad, Iterative methods for sparse linear systems, second ed., SIAM, Philadelpha, PA, 2003.
[41] J. Xia, S. Chandrasekaran, M. Gu, X. Li, Superfast multifrontal method for large structured linear systems of equations, SIAM J. Matrix Anal. Appl. 31 (2009) 1382–1411.
[42] J. George, Nested dissection of a regular finite element mesh, SIAM J. Numer. Anal. 10 (1973) 345–363.
[43] A. Hoffman, M. Martin, D. Rose, Complexity bounds for regular finite difference and finite element grids, SIAM J. Numer. Anal. 10 (1973) 364–369.
[44] I. Duff, J. Reid, The multifrontal solution of indefinite sparse symmetric linear equations, ACM Trans. Math. Softw. 9 (1983) 302–325.
[45] J. Liu, The multifrontal method for sparse matrix solution: theory and practice, SIAM Rev. 34 (1992) 82–109.
[46] S. Eisenstat, J. Liu, The theory of elimination trees for sparse unsymmetric matrices, SIAM J. Matrix Anal. Appl. 26 (2005) 686–705.
[47] E.N.J.R. Gilbert, Predicting Structure in Nonsymmetric Sparse Matrix Factorizations: Graph Theory and Sparse Matrix Computation, Springer-Verlag, 1993.
[48] J. Liu, The role of elimination trees in sparse factorization, SIAM J. Matrix Anal. Appl. 18 (1990) 134–172.
[49] R. Schreiber, A new implementation of sparse gaussian elimination, ACM Trans. Math. Softw. 8 (1982) 256–276.
[50] S.C. Eisenstat, J.W.H. Liu, A tree based dataflow model for the unsymmetric multifrontal method, Electron. Trans. Numer. Anal. 21 (2005) 1–19.
[51] P. Amestoy, C. Puglisi, An unsymmetrized multifrontal lu factorization, SIAM J. Matrix Anal. Appl. 24 (2002) 553–569.
[52] M. Bebendorf, Efficient inversion of galerkin matrices of general second-order elliptic differential operators with nonsmooth coefficients, Math. Comp. 74 (2005) 1179–1199.
[53] M. Bebendorf, W. Hackbusch, Existence of $\mathcal{H}$-matrix approximants to the inverse fe-matrix of elliptic operators with l^∞ coefficients, Numer. Math. 95 (2003) 1–28.
[54] S. Chandrasekaran, P. Dewilde, M. Gu, N. Somasunderam, On the numerical rank of the off-diagonal blocks of schur complements of discretized elliptic PDES, SIAM J. Matrix Anal. Appl. 31 (2010) 2261–2290.
[55] S. Chandrasekaran, P. Dewilde, M. Gu, W. Lyons, T. Pals, A fast solver for HSS representations via sparse matrices, SIAM J. Matrix Anal. Appl. 29 (2006) 67–81.
[56] J. Xia, S. Chandrasekaran, M. Gu, X. Li, Fast algorithms for hierarchically semiseparable matrices, Numer. Linear Algebra Appl. (2010), in press, doi:10.1002/nla.691.
[57] S. Chandrasekaran, M. Gu, T. Pals, A fast *ulv* decomposition solver for hierarchically semiseparable representations, SIAM J. Matrix Anal. Appl. 28 (2006) 603–622.
[58] Y. Erlangga, F. Herrmann, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion, SEG Tech. Prog. Expanded Abstr. 27 (2008) 1956–1960.