



Delft University of Technology

Elastic wavefield inversion by the alternating update method

Rizzuti, Gabrio

DOI

[10.4233/uuid:9c9cddf0-5bae-4869-854f-fda61439f27d](https://doi.org/10.4233/uuid:9c9cddf0-5bae-4869-854f-fda61439f27d)

Publication date

2017

Document Version

Final published version

Citation (APA)

Rizzuti, G. (2017). *Elastic wavefield inversion by the alternating update method*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:9c9cddf0-5bae-4869-854f-fda61439f27d>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Elastic wavefield inversion by the alternating update method

Elastic wavefield inversion by the alternating update method

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben;
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op donderdag 7 december 2017 om 10:00 uur

door

Gabrio RIZZUTI

Master of Science in Mathematics,
Università degli studi di Milano–Bicocca, Italië
geboren te Milaan, Italië

This dissertation has been approved by the
promotor: prof.dr. W.A. Mulder

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof.dr. W.A. Mulder	Delft University of Technology

Independent members:

Prof.dr.ir. C.W. Oosterlee	Delft University of Technology
Prof.dr.ir. E.C. Slob	Delft University of Technology
Prof.dr. H.P. Urbach	Delft University of Technology
Prof.dr. K.J. Batenburg	Leiden University
Dr. L. Hörchens	Applus+ RTD

Other members:

Prof.dr.ir. A. Gisolf	Delft University of Technology
-----------------------	--------------------------------

Copyright ©2017, by G. Rizzuti, Delft Inversion B.V., Delft, the Netherlands.
All rights reserved. No part of this publication may be reproduced, stored in a
retrieval system or transmitted in any form or by any means, electronic, mechanical,
photocopying, recording or otherwise, without the prior written permission of the
author G. Rizzuti, Delft Inversion B.V., 2611 PA, Delft, the Netherlands.

Contents

1	Introduction	1
1.1	Drawing a picture of what cannot be seen	1
1.2	Seismic imaging	3
1.3	Traditional seismic imaging	3
1.4	Full-waveform seismic inversion	5
1.4.1	Main advantages over traditional imaging techniques . .	7
1.4.2	Challenges	8
1.5	Mitigating non-linearity by search space extension	9
1.5.1	The extended modeling principle	10
1.5.2	Treating properties and wavefields independently	11
1.5.3	Brief review of the literature on extended search space methods	11
1.6	Application of seismic inversion	15
1.6.1	From the surface to the reservoir	15
1.6.2	The role of seismic inversion: present and future	16
1.6.3	Other applications	17
1.7	Scope of this thesis and outline	17
2	An iterative method for 2-D inverse scattering problems	21
2.1	Introduction	22
2.1.1	Full-waveform inversion	23
2.1.2	An alternating inversion method	25
2.1.3	Summary of this paper	26

2.2	The scattering integral equation	27
2.2.1	The elastic differential system	27
2.2.2	The elastic wave equation in the integral form	28
2.2.3	The domain equation	29
2.2.4	The data equation	31
2.3	An alternating update scheme for the inverse scattering problem	32
2.3.1	The alternating inversion algorithm	32
2.4	Numerical implementation of the inversion scheme	36
2.4.1	Discretization of the scattering integral equation	37
2.4.2	Treatment of the data equation	40
2.4.3	Complexity analysis	41
2.4.4	Inversion strategy	43
2.5	Numerical experiments	44
2.5.1	The Delphi temple	45
2.5.2	The curly model	47
2.6	Summary and conclusions	54
3	A multigrid method for the elastic wave equation	61
3.1	Introduction	62
3.2	Problem formulation and numerical discretization	64
3.3	A ‘shifted-Laplacian’ preconditioner for the elastic wave equation	70
3.4	Analysis of multigrid applied to the damped elastic wave equation	72
3.4.1	Multigrid for the acoustic wave equation	73
3.4.2	Multigrid for the elastic wave equation	74
3.5	Numerical examples	88
3.5.1	Additional settings and multigrid implementation	89
3.5.2	The homogenous case	90
3.5.3	A highly heterogeneous problem: the elastic Marmousi model	91
3.6	Conclusions	93
4	Ultrasound imaging in non-destructive testing	97
4.1	Linear vs non-linear imaging in non-destructive testing	98
4.2	Assumptions	99
4.3	Full-waveform inversion	100
4.4	Alternating inversion scheme	102
4.4.1	Scattering integral equation	102
4.4.2	Algorithm	103
4.4.3	The Born approximation	105

4.4.4	Regularization and constraints	105
4.4.5	Numerical implementation and efficiency	107
4.5	Results	108
4.5.1	Synthetic crack experiment	109
4.5.2	Imaging from laboratory measurements	111
4.6	Summary and conclusion	114
5	Conclusions and recommendations	119
5.1	Summary	119
5.2	Future directions	120
A	An optimized staggered-grid finite-difference method	125
A.1	The standard staggered-grid scheme	126
A.2	Optimization of the staggered-grid method	128
A.3	Numerical accuracy	130
	Bibliography	133

Introduction

1.1 Drawing a picture of what cannot be seen

‘La Clairvoyance’, by René Magritte (Figure 1.1a), portrays an austere man in the act of painting: a bird is being drawn on the canvas. It is suggested that the man is using an egg as a model, being able to foresee beyond the current state of the object and display its future potential. What it is interesting, here, is the parallel between the scene represented in ‘La Clairvoyance’ and the subject of this thesis, which goes with the name of *Imaging*.

In Imaging, we make a picture out of some observations. For example, we can display the womb of a pregnant woman from ultrasonic measurements, and yet this is precluded to direct vision. It is possible because of the clear understanding of the physical phenomena that lead this type of measurements. In order to create an image of an object, or the interior thereof, we exploit the physics of *waves*. A notable example is offered by nature itself with the echo-location system of bats, used to navigate and forage in the dark. The sound waves transmitted by the bat get reflected back when they meet an obstacle, giving a clue about the position of the target.

In this thesis we are particularly interested in the application of seismic imaging, even though the principles of the methodology discussed in the next chapters are general and applicable to different settings (e.g., electromagnetics). The resolution and penetration provided by elastic waves is useful for

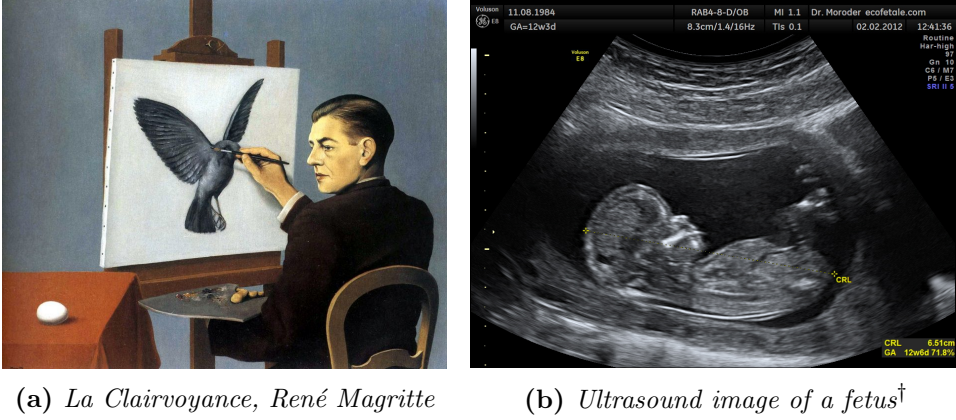


Figure 1.1: *Artistic and scientific rendition of imaging*

detecting details of the object at large range of scales. For acoustical imaging, sundry applications are listed in Table 1.1.

Frequency	Notable applications	Depth (max)
10 MHz	} Material inspection } Welding inspection } Medical diagnostics	0.5 cm
1 MHz		5 cm
100 kHz		20 cm
10 kHz	} Sand exploration, wreck detection } Ground-water, heat, waste } Under-water acoustics (sonar)	20 m
1 kHz		500 m
100 Hz	} Oil and gas exploration	5 km
10 Hz		
	} Tectonics and major faults } Earthquakes, earth crust } Infra-sound, atmosphere	25 km
1 Hz		> 50 km
0.1 Hz		30 km (altitude)

Table 1.1: *A list of traditional acoustical imaging applications, ordered by the frequency range and depth of penetration*

[†]Courtesy of W. Moroder / CC BY-SA 3.0

1.2 Seismic imaging

In this thesis, we will be focusing on the specific settings of *seismic imaging*, of which the goal is to infer a model of the subsurface from the data collected at the surface in a certain set of experiments. In each of these experiments, seismic waves are generated at the surface and travel into the earth where they are scattered, due to the heterogeneities of the earth's interior. These scattering data, recorded at the surface, provide the information to reconstruct the earth's model. Here, in particular, we are interested in the type of experimental setup and data useful for hydrocarbon detection for the oil and gas industry (see Table 1.1).

A canonical seismic acquisition setup is depicted in Figure 1.2, with sources and receivers positioned at the surface. In the marine case, the source is an airgun, while the receivers (hydrophones) are towed by a vessel or lie on the ocean bottom. On land, the waves are excited by a vibrator truck and the receivers are geophones stuck into the ground.

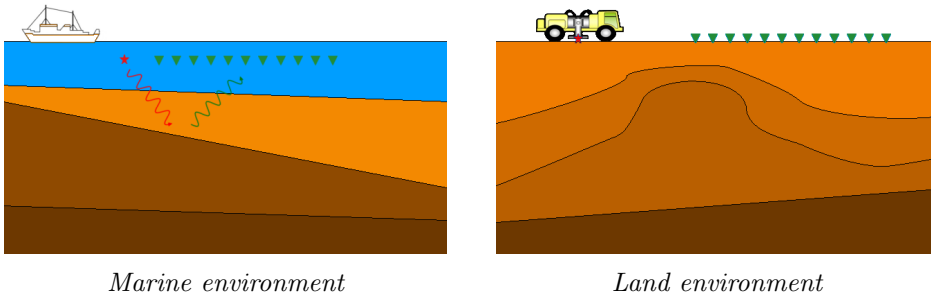


Figure 1.2: Typical acquisition setting for marine and land data. Seismic sources and receivers are denoted by red stars and green triangles, respectively.

In Figure 1.3, we show a typical result of these measurements.

1.3 Traditional seismic imaging

Seismic data (as in Figure 1.3) have been traditionally exploited to obtain a qualitative or structural image of the subsurface. A typical product of *structural imaging* is exemplified by the ‘black and white’ picture in Figure 1.4. Apparently, the only information provided by this process is the discontinuity

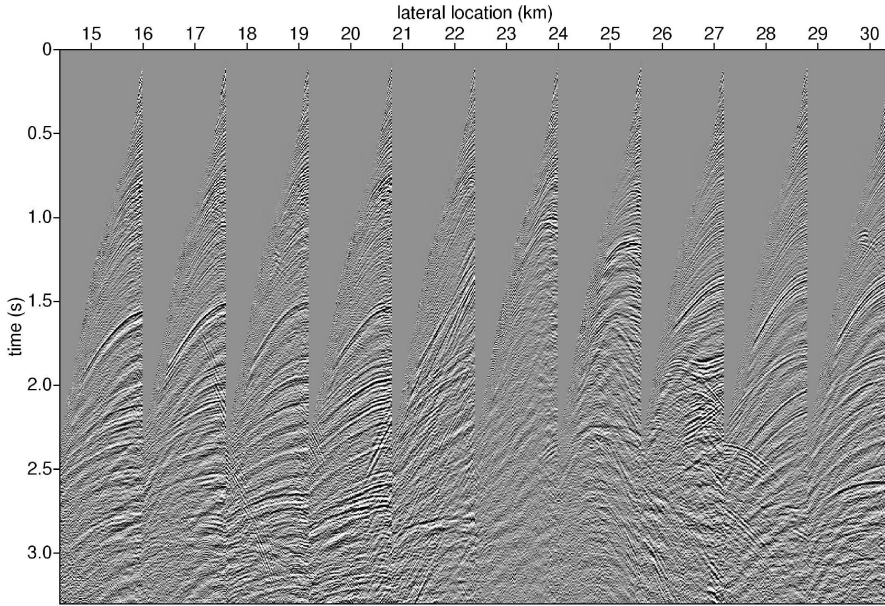


Figure 1.3: *Seismic data organized in common shot records*

distribution of the subsurface, but the relation with physical properties such as P- and S-wave propagation velocities and density is not immediately evident. The knowledge of these properties is of capital importance to determine whether hydrocarbons are present or not.

The kind of output shown in Figure 1.4 is the result of a technique called *migration*, based on the creation of a virtual dataset in each point of the subsurface (through the so-called *back-propagation* of sources and receivers) and the application of an *imaging condition*, which somehow measures whether scattering has occurred or not [Claerbout, 1985].

In geophysics, there is a tradition of dealing with *reflectivity* coefficients, rather than the properties themselves. With reflectivities, we refer to the relative amplitude of the reflected plane waves generated by material discontinuities, for a horizontally homogeneous model. The image amplitudes of Figure 1.4 can be linearly related to reflectivity. Reflectivity, in turn, is linked to the material properties by the so-called Zoeppritz equations, whose linearized version is generally (though unnecessarily) adopted [Aki and Richards, 2002; Gisolf and Verschuur, 2010], so that, in principle, a structural image can be converted to velocities and density.

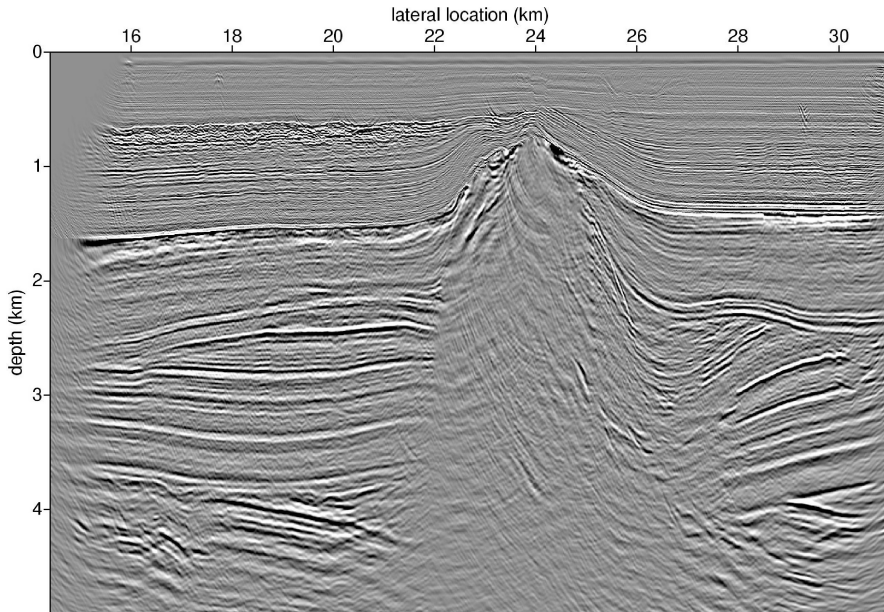


Figure 1.4: *Typical output of structural imaging*

There are multiple issues within this work-flow. To begin with, the previous definition of reflectivity is based on the horizontal stratification of the medium, which is seldom met in practice. This will result, for example, in a troublesome recovery of near vertical reflectors. Note, however, that there are alternative notions of ‘reflectivity’ not vitiated by the horizontal homogeneity assumption (see, for example, Mulder and Plessix [2004b]). Another problem is the underlying assumption concerning the linear relationship of data and properties which affects the back-propagation step and the linearized Zoeppritz relations. This simplification entails the negligence of transmission effects, elastic mode conversion and multiple reflections. The scattering effects in the data due to pronounced non-linearity — and not properly addressed by linear imaging — will likely affect the final result.

1.4 Full-waveform seismic inversion

A seemingly straightforward method of obtaining subsurface properties from seismic data is based on the direct comparison of recorded and simulated waveforms, the latter obtained through the theoretical knowledge of the wave

propagation and an assumed set of subsurface properties. The mismatch can be measured by the least-squares norm and the problem set up in the numerical optimization framework. This route has been popularized in geophysics by the pioneering works of Lailly [1983] and Tarantola [1984].

We denote by \mathbf{d} the measurements taken at the surface and by \mathbf{m} the model parameters, which can be represented by the P- and S-wave velocities and density (for every grid point of the discretized objective volume), in the elastic case. With $\mathbf{u} = \mathbf{u}(\mathbf{m})$, we indicate the solution of the wave equation relative to \mathbf{m} :

$$\mathcal{H}[\mathbf{m}] \mathbf{u} = \mathbf{f}, \quad (1.1)$$

Here, $\mathcal{H}[\mathbf{m}]$ is the linear system associated with the wave equation (in time or frequency domain) and the source term is denoted by \mathbf{f} . Note that we highlighted the dependency of this system with respect to the model parameters \mathbf{m} : $\mathcal{H} = \mathcal{H}[\mathbf{m}]$. Finally, an operator \mathcal{R} is introduced that restricts the values of the wave solution $\mathbf{u}(\mathbf{m})$ to the specified receiver locations \mathbf{x}_r . In short-hand we denote by $\mathcal{F} = \mathcal{F}(\mathbf{m})$ the modeling function defined by $\mathcal{F}(\mathbf{m}) = \mathcal{R}(\mathbf{u}(\mathbf{m}))$. We then seek to minimize the following objective functional:

$$\mathcal{J}(\mathbf{m}) = \|\mathbf{d} - \mathcal{F}(\mathbf{m})\|^2, \quad (1.2)$$

where $\|\cdot\|$ is the least-squares norm (which implicitly includes a summation over receivers, sources and time-steps/frequencies). This approach is widely known in the geophysical community as *full-waveform inversion*. A comprehensive account on this subject can be found in Virieux and Operto [2009].

The terminology ‘full-waveform inversion’ can be explained in the sense that every type of wave ‘mode’ included in the data can be exploited to refine the output model (see also Fichtner [2010] for a brief discussion on this name). Here, we are referring to the common classification of the features typically recognized in a seismic dataset: primary reflections, diving waves, internal multiple scattering, surface waves and so on. We should point out that, although this classification is deeply rooted in the geophysical parlance and culture, its definition is rather ambiguous. The notion of a primary event, for example, as a singly reflected wave, is undermined by the negligence of transmission losses and multiple reflection ‘peg legs’ which will contribute to distort the original source pulse as observed in correctly modeled synthetic seismograms or field data (see O’Doherty and Anstey [1971] for an interesting discussion on this ‘stratigraphic filtering’ effect). Since we resolved to work with the wave equation, which naturally accounts for the — generally inextricable — scattering effects of wave propagation, we do not particularly bother

with such distinctions. Occasionally, we might resort to the above terminology as a mean to distinguish applications where certain type of events are predominant in the data, e.g. surface-wave inversion which refers to those problems where the only apparent phenomenon in the data is the dispersive interference pattern often observed in the presence of a free surface.

The physics of waves can be reasonably approximated by adopting an elastic wave propagation scheme, including anisotropy, eventually.

1.4.1 Main advantages over traditional imaging techniques

An important feature of full-waveform inversion is the higher achievable resolution for the reconstructed model, compared to linear imaging. It can be shown that, e.g. under the Born approximation and typical circumstances, the shortest spatial wavelength for the reconstructed model is $\lambda/2$, where λ is the propagation wavelength [Sirgue and Pratt, 2004].

Note that, in spite of this analysis, super-resolution (achieved when wavelength shorter than $\lambda/2$ are reconstructed), under linear approximation, might be achieved by exploiting the evanescent field, when the object to image is close to the source-receiver apparatus [Simonetti, 2006] (generally not the case in the seismic world), or by analytic continuation of the spectrum of the reconstructed model outside the bandwidth [Bertero and Boccacci, 2003] (although it is an unstable process, and sensitive to noise). In the context of migration, it is clear that beyond-Nyquist wavenumbers can be recovered if, given a certain time sampling rate, the aliased frequency components are not removed prior to digitization of the analog data. Indeed, merging normal moveout corrected traces, instead of stacking, will effectively enhance the sampling rate (as justified in Stark [2013]).

Whatever resolution can be achieved by linear imaging, non-linearity has the potential to improve on it [Simonetti, 2006; Guo et al., 2016].

Another fundamental aspect of this technique, especially regarding the applications we have in mind in this thesis (Section 1.6), is the direct link between elastic properties and seismic recordings, provided by the wave equation. Therefore, the output can be directly processed in order to determine the presence of oil and gas in the subsurface. The relationship between properties and data is not tempered by linear assumption of sort, contrary to traditional imaging.

1.4.2 Challenges: ill-posedness, non-linearity, and computational complexity

A successful implementation of full-waveform inversion is, however, not as straightforward as it would seem from the simple form of equation (1.2).

The first issue is shared with any other imaging technique and is related to the incompleteness of the acquired data in terms of temporal bandwidth and acquisition geometry. Then, there is the issue of inaccuracy, in the form of corruption by measurement errors and noise, in particular coherent noise. Finally, there is the issue of modeling errors, such as an inadequate description of the physics. Incompleteness, in particular, will result in non-uniqueness of the solution to the problem (1.2). The arising non-trivial set of the minimizers of the objective functional should then be dealt with by proper *regularization* techniques to restore well-posedness. This is traditionally achieved by adding to equation (1.2) an extra penalty term which measures the least-squares norm of the model parameters [Tikhonov and Arsenin, 1977] or their derivatives, as in the total variation regularization (see, for example, Esser et al. [2016]), in order to penalize ‘unphysical’ models. These techniques also enforces robustness against data noise.

Another important issue is the non-linearity of the functional (1.2) and, in particular, the existence of local minima which would prevent the algorithm to converge to the global minimum when local search optimization is employed [Mulder and Plessix, 2008]. A simple way to characterize local minima is by looking at ‘loop-skipping’. Seismic data is band-limited, due to the difficulty of recording low and high frequencies, and will consist of ‘oscillatory’ waveforms (as in Figure 1.3). Loop-skipping occurs when the wavefield propagating through a certain medium results in a modeled time trace which is shifted by more than half a period with respect to the given data. In this case, local optimization may stagnate in local minima (Figure 1.5). To mitigate this problem, a sufficiently accurate starting model should be provided and/or a functional less sensitive to local minima should be adopted.

Full-waveform inversion is a computationally-heavy technique, which may have to handle hundreds of millions of variables for large 3D problems. The evaluation of the objective functional (1.2) requires the solution of the wave equation for each source position. Furthermore, when a local search scheme is employed, the calculation of the gradient and the Hessian necessitates many other ‘adjoint’ wave equation solutions. The increase in computing power in

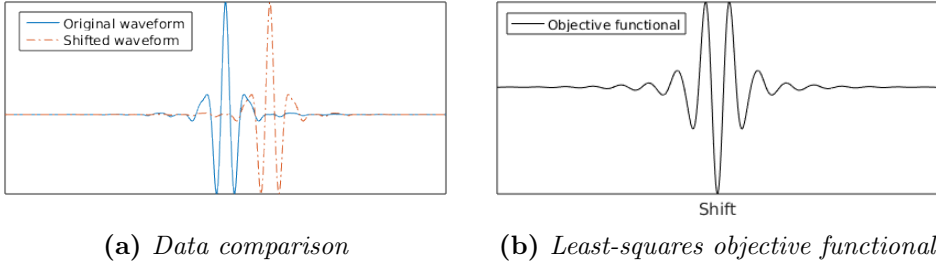


Figure 1.5: *Simplistic depiction of loop-skipping: (a) data (in blue) and example of a shifted waveform (in red), (b) objective functional measuring the least-squares norm of the difference between the two traces, as a function of the shift. The functional has many local minima.*

the last 30 years, however, has made it a feasible technique for single parameter 3D acoustic imaging problems.

The extension of full-waveform inversion to the multi-parameter case (e.g. elastic) still remains challenging due to the larger model space. In the elastic case, wave equation results in a bigger linear system to solve, and slow S-wave regions require a finer discretization. Furthermore, multiple parameters have different imprints on data which can hamper the inversion with crosstalk artifacts. To resolve coupling, a Hessian-based approach seems necessary [Operto et al., 2013].

1.5 Mitigating non-linearity by search space extension

As already pointed out, gradient-based optimization applied to the functional (1.2) suffers from local minima. Consequently, a great deal of research has been devoted to tackle this problem. This includes mitigation of non-linearity by:

- hierarchical strategies such as multiscale methods, where data of increasingly frequency content is being inverted [Bunks et al., 1995], or layer stripping, which consists of imaging shallow-to-deep portions of the sub-surface [Shipp and Singh, 2002];
- choice of proper transform domain for the data, e.g. Laplace(–Fourier) transforms [Shin and Cha, 2008, 2009] or Hilbert transforms [Bozdag et al., 2011];

- alternative objective functionals, opposed to the conventional least-squares formulation [Shin and Min, 2006; Warner and Guasch, 2014; Bharadwaj et al., 2016; Métivier et al., 2016, and many others...].

Obviously, these ideas often overlap. In this section, however, we will focus on a alternative route which will be the basis for the inversion method proposed in Chapter 2.

1.5.1 The extended modeling principle

Many recent works share a viewpoint which can be, perhaps, better understood from the abstract mathematical framework of *extended modeling*, introduced by Symes [2008]. The basic idea consists of an extension of the original model space to include *unphysical* models in the optimization process. The clear intent of this procedure is to relax, somehow, the non-linearity of the original objective functional. Therefore, we include a further item to the previous list: search space extension.

Formally, if we denote the model space by $\mathfrak{M} \ni \mathbf{m}$ and the data space as $\mathfrak{D} \ni \mathbf{d}$, a model extension $\tilde{\mathfrak{M}}$ is introduced via a map $i : \mathfrak{M} \hookrightarrow \tilde{\mathfrak{M}}$ such that the following diagram is ‘commutative’:

$$\begin{array}{ccc}
 \mathfrak{M} & \xrightarrow{\mathcal{F}} & \mathfrak{D} \\
 i \downarrow & \nearrow \tilde{\mathcal{F}} & \\
 \tilde{\mathfrak{M}} & &
 \end{array} \tag{1.3}$$

Commutative means, in this context, that the following property holds: $\mathcal{F} = \tilde{\mathcal{F}} \circ i$, where \circ denotes composition of maps. The injective inclusion map i is possibly non-linear, and allows the identification of \mathfrak{M} with its image — the set of *physical* models. The forward map \mathcal{F} is then extended to $\tilde{\mathcal{F}}$, with domain $\tilde{\mathfrak{M}}$, in a consistent way. This consistency is translated, in mathematical terms, by the commutative property. Now, a new objective on the space $\tilde{\mathfrak{M}}$ can be defined as:

$$\tilde{\mathcal{J}}(\tilde{\mathbf{m}}) = \|\mathbf{d} - \tilde{\mathcal{F}}(\tilde{\mathbf{m}})\|^2 + \mathcal{A}(\tilde{\mathbf{m}}). \tag{1.4}$$

The additional term \mathcal{A} is designed to deal with the large solution space of the first addend of $\tilde{\mathcal{J}}$, arising from the extension (the so-called ‘annihilator’ in Symes [2008]). The obvious goal of this apparatus is to obtain an objective functional $\tilde{\mathcal{J}}$ more amenable to local-search optimization.

1.5.2 Treating properties and wavefields independently

In this thesis, we will focus on a particular model extension allowing for the simultaneous solution of properties \mathbf{m} and wavefields \mathbf{u} as independent variables. This is the route followed in the contrast-source inversion method of van den Berg and Kleinman [1997] and the wavefield reconstruction method of van Leeuwen and Herrmann [2013]. Note that, in the classical least-squares formulation of full-waveform inversion, these two types of variable are linked by the wave equation (1.1). If we denote the model space by \mathfrak{M} and the wavefield space by \mathfrak{U} , we then are looking for a generalized solution (\mathbf{m}, \mathbf{u}) in the product space $\tilde{\mathfrak{M}} = \mathfrak{M} \times \mathfrak{U}$.

A qualitative argument to understand why such an extension can be beneficial in mitigating non-linearity is given as follows. The objective functional \mathcal{J} in equation (1.2) can be rewritten as

$$\mathcal{J}(\mathbf{m}) = \bar{\mathcal{J}}(\mathbf{u}(\mathbf{m})), \quad \bar{\mathcal{J}}(\mathbf{u}) = \|\mathbf{d} - \mathcal{R}\mathbf{u}\|^2$$

(\mathcal{R} is the restriction operator to the receiver positions). Note that $\bar{\mathcal{J}}$ is a quadratic functional of \mathbf{u} and does not have isolated local minima (although it does suffer from a large zero set). The problematic source of non-linearity for \mathcal{J} was introduced by the wave equation, which implicitly defines a subset in the space $\tilde{\mathfrak{M}}$ of admissible states. This constraint can be thought as a parameterized manifold $\mathbf{m} \mapsto (\mathbf{m}, \mathbf{u}(\mathbf{m}))$, which is actually the inclusion map $i: \mathfrak{M} \hookrightarrow \tilde{\mathfrak{M}}$ described in diagram (1.3). It is the restriction of $\bar{\mathcal{J}}$ to this manifold that will produce the local minimum issue. A pictorial description of this argument (suggested by van Leeuwen, oral presentation) is given in Figure 1.6.

1.5.3 Brief review of the literature on extended search space methods

In the following, we briefly review some methods based on the extended modeling principle. We give a more technical treatise for the contrast-source and wavefield reconstruction methods, since they are more similar to the proposed inversion algorithm in Chapter 2.

The contrast-source inversion method

The main inspiration for the inversion algorithm discussed in this thesis (Chapter 2) comes from the *contrast-source inversion* method of van den Berg and

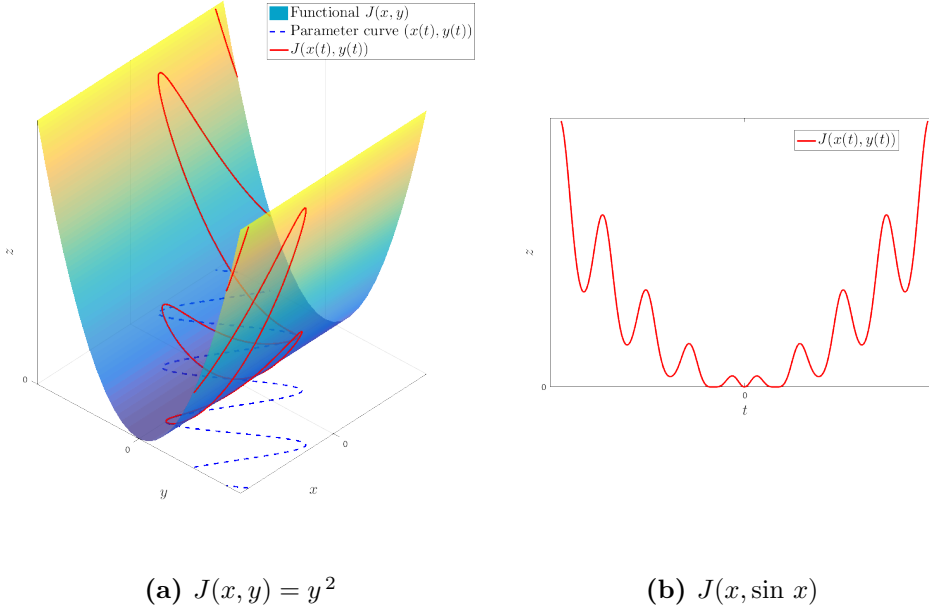


Figure 1.6: Qualitative depiction of non-linearity for classical full-waveform inversion. The wave equation (1.1) imposes a constraint on the space of models \times wavefields, here abstractly represented by the (x, y) plane in (a). The constraint can be visualized as a parameterized curve $t \mapsto (x(t), y(t))$ (the dashed blue curve in (a)). Note that the functional $J(x, y) = y^2$ is quadratic and does not present isolated singular points (although it has a large solution space corresponding to $y = 0$). However, the restriction of J on the parameterized curve in (b) does show many local minima, corresponding to the points whose curve tangent is orthogonal to the gradient of J .

Kleinman [1997]. Here, instead of the differential wave equation (1.1), an integral reformulation is considered. For simplicity, we will restrict ourselves to the acoustic constant-density case. In this setting, we need to consider the Helmholtz equation:

$$-\frac{\omega^2}{c^2} P - \Delta P = f, \quad (1.5)$$

where P is the time-harmonic pressure wavefield related to the angular frequency ω . The term f represents a source. The parameter c represents the spatially varying propagation velocity. An integral representation of equation (1.5) is given by:

$$(I - \mathcal{G}_\Omega \mathcal{W}) P = P^{\text{inc}}, \quad (1.6)$$

where the operators \mathcal{G}_Ω and \mathcal{W} are defined, for any function w and p , by:

$$(\mathcal{G}_\Omega w)(\mathbf{x}) = \int_\Omega G(\mathbf{x}, \mathbf{x}') w(\mathbf{x}') d\mathbf{x}', \quad (\mathcal{W}p)(\mathbf{x}) = \chi(\mathbf{x}) p(\mathbf{x}),$$

The Green function G is relative to a known background medium c_b and P^{inc} is the incident wavefield propagating through the known background. The contrast χ is defined by $\chi = c_b^2/c^2 - 1$ and is assumed to vanish outside the domain of interest Ω . This integral equation states that the total wavefield P can be decomposed as the sum of the incident field and a field which propagates through the background medium and is generated by ‘contrast sources’ $W = \chi P$.

The contrast-source inversion is characterized by the inclusion of both properties χ and wavefields W as independent optimization variables, not forcibly constrained by the wave equation. An objective functional is then set up as follows:

$$\tilde{\mathcal{J}}(\chi, W) = \|\Delta \mathbf{d} - \mathcal{G}_R W\|^2 + \eta \|\chi P^{\text{inc}} - W + \chi \mathcal{G}_\Omega W\|^2. \quad (1.7)$$

Here,

$$(\mathcal{G}_R W)(\mathbf{x}_r) = \int_\Omega G(\mathbf{x}_r, \mathbf{x}') W(\mathbf{x}') d\mathbf{x}'$$

is the domain-to-receiver Green function, where \mathbf{x}_r indicates the receiver position, and η is a trade-off parameter. The known term $\Delta \mathbf{d} = \mathbf{d} - \mathbf{d}^{\text{inc}}$ is the difference between the given data \mathbf{d} and \mathbf{d}^{inc} , which is the restriction of the incident field to the receiver positions.

The inversion algorithm for the contrast-source method consists in a *alternating* update of contrasts and wavefields. Starting from a given estimate χ_n and W_n , an update for W is given by partial local optimization of (1.7) (using for instance the steepest-descent direction or a conjugate gradient scheme), while χ_n is kept fixed. If we denote this update by W_{n+1} , the next step consists in the estimation of χ via global optimization of (1.7) with respect to χ : $\chi_{n+1} = \text{Re} \{ \bar{P}_{n+1} W_{n+1} \} / |P_{n+1}|^2$ (‘Re’ denoting the real part of a complex number), where $P_{n+1} = P^{\text{inc}} + \mathcal{G}_\Omega W_{n+1}$. Note that, in van den Berg and Kleinman [1997], the weight η is adapted at every iteration.

The interesting aspect of the contrast-source inversion is the fact that, unlike equation (1.2), the degree of non-linearity of the functional (1.7) is only polynomial of degree 4, as one might infer from a simple analysis of

$P(\alpha, \beta) = \mathcal{J}(\chi_0 + \alpha \Delta\chi, W_0 + \beta \Delta W)$. Also, the method is computationally advantageous over classical full-waveform inversion, because it does not require the explicit solution of the wave equation for the current model estimate, but only the evaluation of the Green function pertaining to a fixed background (which can be sped up considerably as shown in Abubakar et al. [2009]).

The wavefield reconstruction method

In van Leeuwen and Herrmann [2013], the problem is set up similarly to van den Berg and Kleinman [1997]. The following objective is considered:

$$\tilde{\mathcal{J}}(\mathbf{m}, \mathbf{u}) = \|\mathbf{d} - \mathcal{R}\mathbf{u}\|^2 + \eta \|\mathbf{f} - \mathcal{H}[\mathbf{m}]\mathbf{u}\|^2, \quad (1.8)$$

where \mathcal{R} is the restriction operator to the receiver positions. The inversion algorithm proceeds by projecting out the wavefield variable \mathbf{u} by computing the minimizer of $\tilde{\mathcal{J}}(\mathbf{m}, \cdot)$ for a fixed \mathbf{m} . This corresponds to the solution of an ‘augmented’ wave equation, represented by the rectangular linear system:

$$\begin{pmatrix} \mathcal{R} \\ \eta^{1/2} \mathcal{H}[\mathbf{m}] \end{pmatrix} \mathbf{u} = \begin{pmatrix} \mathbf{d} \\ \eta^{1/2} \mathbf{f} \end{pmatrix}$$

This system, to be solved in a least-squares sense, has a well-defined solution. The property model \mathbf{m} is then updated by a gradient-based optimization of (1.8), while keeping the variable \mathbf{u} fixed. Note how the contrast-source method is based on a projection of the model variable \mathbf{m} , contrary to the wavefield reconstruction scheme where \mathbf{u} is the projected variable.

A distinctive advantage of this method over the contrast-source inversion is the fact that the wavefields do not need to be kept in memory all at the same time. The gradient computation of $\tilde{\mathcal{J}}$ with respect to \mathbf{m} results in a summation of different terms indexed by sources and frequencies. Each of these term only involves the previously computed wavefield specified by the source and frequency index. Therefore, for each term, the wavefields can be calculated independently and discarded immediately. The drawback of this method lies in the need for solutions of the (augmented) wave equation at every iteration which will result in a higher time complexity when compared to the contrast-source method.

1.6 Application of seismic inversion

In this section, we outline the main current applications of full-waveform inversion in the oil and gas industry and we present some ideas which, in our opinion, exploit its true potential. It is interesting to start with highlighting some of the procedural steps taken in a particular segment of the work flow adopted by the oil and gas industry: the exploration phase for detecting potential hydrocarbon reservoirs.

1.6.1 From the surface to the reservoir

We assume that an interesting lead has been identified and that seismic data has been acquired. The first step, after the data acquisition process, is a preliminary assessment of the geological scenario of this particular portion of the subsurface. This is achieved with the help of structural imaging of which the output is a characteristic picture like Figure 1.4, as previously discussed. The imaging method routinely applied to this purpose is migration, and is based on the simulation of virtual sources and receivers placed deep in the subsurface. Then, an imaging condition is applied to the simulated measurements, which detects whether scattering has occurred in that specific location [Gisolf and Verschuur, 2010].

From the result of migration, a potential reservoir can be identified, and now it is time to obtain more specific information regarding this subset of the subsurface. At this stage, a picture like Figure 1.4 is only useful to determine the structure of the reservoir, but it cannot provide pore pressure and rock properties, which are sensitive indicators for the presence of oil or gas. In order to obtain these properties we want to apply more specialized methods. Some preliminary processing can be applied to create ‘localized’ data right on top of the target zone, to relieve the data from unnecessary information pertaining to locations far from the target. This can be achieved, for example, by redatuming, with the back-propagation of sources and receivers (physically located at the surface) to the area of interest (virtually located near the target) [Berryhill, 1979; Gisolf et al., 2015], with a process similar to migration.

At this point, we want to estimate the properties of the reservoir from the localized data. An industry-standard solution to this question is based on the linearization of the relation between data and properties, and is generally termed amplitude vs slowness/offset/angle analysis (see the previous

Section 1.3). Full-waveform inversion is another, better, option, because it does not rely on linearization of any sort. As previously noted, the wave equation links physical properties to data and the inversion output directly consists of the estimated properties. In the elastic case, these properties are the bulk modulus, shear modulus and the density, which can be related more naturally to rock and pore properties than the image amplitudes obtained from migration.

1.6.2 The role of seismic inversion: present and future

Despite the potential of seismic inversion to deliver high resolution physical properties, its main industrial application, to date, seems confined to obtain a global model for the subsurface consisting of a slowly varying function, compared to the actual property variability. This result can be exploited by structural imaging, because the method relies on the accuracy of the background model to focus the data. The resolution of the output model obtained by full-waveform inversion is, in principle, superior to current traveltimes-based tomography methods [Nolet, 1987]. However, if too much detail is imposed in the background model, it could violate the linearity requirement of the migration process and the model has to be smoothed before it can be used by migration.

Since its main purpose is to provide a velocity model for migration, long-offset data (and, in particular, the ones capturing diving waves) are mainly used. This seems to mitigate the loop-skipping limitation described in Section 1.4, as the analysis of Sirgue and Pratt [2004] on the relation between model wavenumber coverage and acquisition setting suggests. The same analysis also clarifies why reflection data pose a challenge for full-waveform inversion. Full-waveform inversion could greatly benefit from low-frequency data, but this kind of information is not available in conventional recordings yet.

The full-waveform inversion described above does not seem to be fit for the quantitative analysis needed for reservoir characterization, where actual physical properties are sought but only reflection data are available. However, if the non-linearity of the problem is correctly handled, the additional illumination produced by multiple scattering can help in resolving the low wavenumbers of the model. Non-linearity is then seen as an opportunity rather than a nuisance. The ‘local 1.5-D’ version of the inversion algorithm presented in this thesis, based on the assumption of locally stratified media, is being routinely

applied for reservoir characterization [Gisolf et al., 2014; Beller et al., 2015].

1.6.3 Other applications

Besides reservoir characterization, there are numerous possible applications of seismic inversion. For example, in the presence of a complex *near-surface*, or near ocean-bottom, the data from a deeper target zone (e.g. the reservoir) is likely to be affected by the multiple scattering occurring in the near-surface. This phenomenon might affect the back-propagation step needed for data localization at the reservoir level. This imprint is a well-known issue in the land environment, due to the unconsolidated nature of the shallow geology, but it can also be problematic for the marine case due, for example, to the presence of gas pockets in the near-surface. To be able to remove these effects from the deeper data, the near-surface must first be imaged to a certain degree. Imaging of the near ocean-bottom might also be important to avoid drilling hazards.

1.7 Scope of this thesis and outline

The objective of this thesis is to discuss an algorithm for elastic full-waveform inversion. As anticipated, the method is based on the same research flavor of the extended modeling principle outlined in Section 1.5. In particular, it employs the decoupling of property and wavefield variables. Its essential mechanism hinges on the *alternating* update of these two types of unknowns by using distinct objective functionals: (i) a measure of the adherence to the recorded data for the update of properties, and (ii) an error function to steer the wavefield towards the physical solution of the wave equation, without imposing it as a hard constraint at each iteration. The *alternating inversion method* is thoroughly analyzed in Chapter 2.

A great deal of the research involved in the design of the inversion algorithm has been about the numerical discretization of the wave equation. The need for flexibility of a-priori property distributions (‘background’ models) and the geometry of the computational domain (mostly dictated by near-surface applications) has driven our investigations towards finite-difference and finite-element methods. An example of a finite-difference scheme is presented in Appendix A. Originally, this work started out with *semi-analytical* discretizations, which are characterized by integral representations of the wave equation, as in (1.6), and

the analytical knowledge of the related Green's function. This change of perspective (which is, in turn, the mainstream choice for traditional full-waveform inversion) raised the important question on how to solve the linear system arising from such a discretization of the wave equation. For 2-D, classical direct methods are readily available [George and Liu, 1981]. For 3-D, however, this issue is highly non-trivial. This stimulated efforts towards iterative methods for the elastic wave equation. Here, we employ the successful idea of the *shifted-Laplacian preconditioning* by multigrid [Erlangga et al., 2006], developed for the Helmholtz equation, and adapt it to the elastic case. Details can be found in Chapter 3.

Our research represents the natural evolution of the local 1.5-D elastic waveform inversion proposed in Gisolf et al. [2012] and the 2-D acoustic version described in Haffinger et al. [2013]. In this thesis, we will discuss the 2-D elastic version. The real-world applications for a 2-D inversion are limited compared to 3-D, or even the local 1.5-D case. However, ultrasound imaging for material inspection (e.g. crack detection in steel pipes) seem to be reasonably fit for the 2-D assumption. Obviously, it also represents a stepping stone for the 3-D case as a 'proof of concept'.

This thesis is organized as follows.

Chapter 2 In this chapter, we discuss the alternating algorithm for elastic waveform inversion in full details. Here, we develop its theoretical ground with emphasis on the point of view of the extended search space method. A complexity analysis is also carried out with a direct comparison with the contrast-source inversion method of van den Berg and Kleinman [1997]. Different numerical examples show that this method is rather successful for synthetic examples. This chapter is adapted after the publication referenced by Rizzuti and Gisolf [2017].

Chapter 3 The numerical machinery involved in the algorithm presented in Chapter 2 greatly benefits from an efficient solution of the wave equation linear system, when methods like finite-differences or finite-elements are employed. Here, we analyze an iterative method to solve the elastic wave equation which is based on the preconditioning technique by Laplacian-shifting [Erlangga et al., 2006]. This chapter reproduces the paper of Rizzuti and Mulder [2016].

Chapter 4 This chapter is dedicated to a particular real-world application of elastic imaging, which slightly departs from the main theme of this thesis — seismic imaging — but is identical in its underlying physical principles: ultrasound imaging. We discuss, in particular, the non-destructive inspection of stress-corrosion cracks in steel pipes, which is a valuable tool to ensure the integrity of oil and gas pipe lines.

Chapter 5 In the final chapter, we summarize the main findings of our research and lay out some ideas for future work.

An iterative method for 2-D inverse scattering problems by alternating reconstruction of medium properties and wavefields: theory and application to the inversion of elastic waveforms[†]

Abstract

We study a reconstruction algorithm for the general inverse scattering problem based on the estimate of not only medium properties, as in more conventional approaches, but also wavefields propagating inside the computational domain. This extended set of unknowns is justified as a way to prevent local minimum stagnation, which is a common issue for standard methods. At each iteration of the algorithm, (i) the model parameters are obtained by solution of a convex problem, formulated from a special bilinear relationship of the data with respect to properties and wavefields (where the wavefield is kept fixed), and (ii) a better estimate of the wavefield is calculated, based on the previously recon-

[†]This chapter is the reproduction of the paper Rizzuti and Gisolf [2017], published in Inverse Problems, 2017

structured properties. The resulting scheme is computationally convenient since step (i) can greatly benefit from parallelization and the wavefield update (ii) requires modeling only in the known background model, which can be sped up considerably by factorization-based direct methods. The inversion method is successfully tested on synthetic elastic datasets.

2.1 Introduction

The problem of reconstructing the medium parameters from the measurement of electromagnetic, acoustic or elastic waves is a central subject of many fields of research including medical imaging, non-destructive testing and seismic exploration (for geological studies or oil and gas prospecting). The principal motivation for the present work comes from seismic imaging, and in the following discussion we will frequently refer to the specific settings of this problem. However, the principles here described are applicable to more general imaging scenarios.

A controlled seismic experiment is generally set up as follows: the Earth is artificially excited near or at the surface by seismic sources (symbolized by red stars in Fig. 2.1) — typically an airgun for marine environment or vibrator for land — and the waves are recorded by detectors located at the surface (green triangles in Fig. 2.1) or the ocean-bottom. The recorders measure

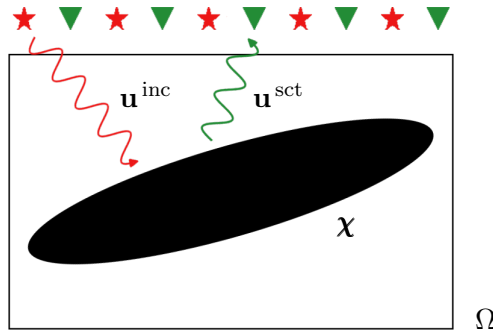


Figure 2.1: *Scattering object inside a domain Ω and typical acquisition setting (sources and receivers respectively denoted by red stars and green triangles). Outside the domain, the contrasts vanish.*

the displacement of the ground caused by a seismic wave. The solution of the inverse scattering problem is then represented by the physical parameters governing the wave propagation that lead to the measured data. These data consist of the waves scattered by the heterogeneity of the subsurface.

Throughout this paper, we will assume the prior knowledge of a background model. This model is often obtained directly from well-logs or, indirectly, by traveltime-based methods. We should remark, however, that under many circumstances the application of these techniques is not so straightforward, and obtaining a background model is, in general, a difficult matter. The background typically comprises the ‘non-reflective’ component of the true model and is a slowly modulated function, relatively to the spatial wavelength of propagation. The original problem is then reduced to the retrieval of the scattering perturbation of the model with respect to the a-priori background.

2.1.1 Full-waveform inversion

In geophysics, a popular approach to the inverse scattering problem is based on least-squares optimization, originally proposed in Tarantola and Valette [1982a], that generally requires gradient-based methods due to the considerable amount of unknowns (hundreds of millions of variables, for 3-D problems). Full-waveform inversion, as this technique is widely known in the geophysical community, has been given significant attention due to the high resolution of the output model (theoretically even higher than $\lambda/2$, where λ is the propagation wavelength). A general account on this subject can be found in, e.g., Virieux and Operto [2009] or Fichtner [2010].

Full-waveform inversion is set up as follows: if we denote the recorded data by \mathbf{d} , we look for a model \mathbf{m} , representing the elastic parameters of the medium, that minimizes the misfit functional:

$$\mathcal{J}(\mathbf{m}) = \|\mathbf{d} - \mathcal{F}(\mathbf{m})\|^2. \quad (2.1)$$

Here, \mathcal{F} is the modeling operator, which typically selects the values of the wavefield $\mathbf{u} = \mathbf{u}(\mathbf{m})$ at the receiver locations. The relationship between the wavefield \mathbf{u} and the model \mathbf{m} is given by the wave equation:

$$\mathcal{H}[\mathbf{m}] \mathbf{u} = \mathbf{f}, \quad (2.2)$$

where \mathbf{f} is a known source term. For convenience, we assume that a specific choice of model parameters has been made such that the wave operator $\mathcal{H}[\mathbf{m}]$

depends linearly on \mathbf{m} , or, in other words, that $\mathcal{H}[\mathbf{m}_1 + \mathbf{m}_2] = \mathcal{H}[\mathbf{m}_1] + \mathcal{H}[\mathbf{m}_2]$ for any \mathbf{m}_1 and \mathbf{m}_2 . We note that the functional (2.1) is non-linear due to the matrix inversion of the wave equation: $\mathbf{u}(\mathbf{m}) = (\mathcal{H}[\mathbf{m}])^{-1} \mathbf{f}$.

The problem (2.1) is notoriously ill-posed and non-uniqueness, in particular, is a fundamental issue [Tarantola and Valette, 1982b]. In this paper, however, we are exclusively interested in the problem of finding a model which optimizes (2.1), making use of regularization techniques to restore well-posedness.

Computational aspects

When a local-search scheme is employed, the gradient of \mathcal{J} may be computed by the adjoint-state method (see Plessix [2006], for example), stemming from optimal control theory, which requires the numerical simulation of a ‘forward’ and ‘backward’ wave equation. With ‘forward’, we mean solving the causal wave equation where the source term is prescribed by the specifics of the seismic experiment, while with ‘backward’ we refer to the wave equation solved back in time where the source is given by the residual defined in the cost functional (2.1). The gradient of \mathcal{J} is then obtained from the forward and backward solution and can be utilized by any gradient-based iterative minimization scheme (as the non-linear conjugate gradient method).

Information about the Hessian of \mathcal{J} might also be included in the optimization, leading to Newton-like methods. Among this class of algorithms, it is worth mentioning the Gauss–Newton (using only the diagonal of the Hessian), quasi-Newton (based on a iterative approximation of the inverse of the Hessian, as in the Broyden–Fletcher–Goldfarb–Shanno algorithm or low-storage version thereof [Nocedal and Wright, 2006]) and full-Newton methods (which utilizes the exact Hessian, as in Haber et al. [2000]). For a general overview, see Nocedal and Wright [2006].

All these optimization techniques, which must be combined with line-search algorithms, require many solutions of the wave equation. In summary, full-waveform inversion is a computationally intensive method whose performance heavily depends on the efficiency of the wave equation solver.

Non-linearity

Besides the computational issues, local optimization methods applied to a non-linear functional like (2.1) are likely hindered by local minima. This is a notorious problem and considerable efforts have been devoted to the mitigation of this problem. Interestingly, many recent developments share a common trait that can be discussed from the abstract point of view of *extended modeling*, a principle of which the main idea has been formalized in Symes [2008]. This will also serve as the starting point of the reconstruction algorithm proposed in this work. It basically consists of a mathematical procedure which involves the inclusion of *unphysical* models in the optimization process, with the goal to relax of the non-linearity issues of the original objective functional. Formally, this means that the search space $\mathfrak{M} \ni \mathbf{m}$ is replaced by an extended model set $\tilde{\mathfrak{M}} \supseteq \mathfrak{M}$. Typically, the objective functional (2.1) is adjusted to $\tilde{\mathfrak{M}}$ in order to be more amenable to local-search optimization.

Without going into the details, we deemed the following works significant, with respect to the extended modeling principle: in van den Berg and Kleinman [1997], the extended search space is represented by properties \mathbf{m} and wavefields \mathbf{u} , which are not strictly constrained by the wave equation (2.2); in Biondi and Almomin [2014], the property models are extended along an extra ‘time’ dimension and a modified wave equation is introduced, in order to facilitate the fitting of large time shifts between acquired data and modeled data generated by inaccurate starting models (which is generally recognized as the main culprit for the failure of full-waveform inversion); in van Leeuwen and Herrmann [2013], the problem is set up similarly to van den Berg and Kleinman [1997], but the wavefield variables are eliminated from the optimization by solving a ‘data-weighted’ wave equation (achieved by a variable projection scheme [Golub and Pereyra, 2003]).

2.1.2 An alternating inversion method

In this paper, we will adopt the view proposed by the contrast-source inversion method of van den Berg and Kleinman [1997], for which the search space $\tilde{\mathfrak{M}}$ consists of property models \mathbf{m} and wavefields \mathbf{u} , treated as independent variables. The method here proposed is, however, based on distinct updating schemes for these two variable sets. In very general terms, the medium update is governed by a data-misfit optimization, similarly to (2.1), where a linear version of the modeling operator is used instead, and the wavefield is up-

dated using the wave equation. The inversion, therefore, produces a sequence $(\mathbf{m}_n, \mathbf{u}_n)$ arising from the alternation of these two processes.

From the computational point of view, the two phases of the alternating inversion are designed to benefit from parallelization and efficient numerical routines. The model update step (i) is based on the solution of a linear system by least-squares optimization, and its evaluation is particularly suited for parallelization. Furthermore, the wave modeling step (ii) will only need the solution of the wave equation relative to the known background model \mathbf{m}_b . Since \mathbf{m}_b is ideally kept fixed throughout the inversion, the LU decomposition of a given discretization of $\mathcal{H}[\mathbf{m}_b]$ can be effectively employed to speed up this step. The same idea was exploited in Abubakar et al. [2009]. From the analysis of synthetic and field data, we experienced that only a limited number of wavefield updates are actually necessary for a satisfactory result and there might be benefits from concentrating the computational work on the model update corresponding to step (i) . For this reason, we believe the method can be competitive with the contrast-source inversion Abubakar et al. [2009], which requires the solution of the wave equation at each iteration. It should be noted that the memory requirements suffer from the need for the factorization of the wave operator to be stored (as it was the case for Abubakar et al. [2009]). Nonetheless, this approach might be even suited for 3-D problems, as argued in Abubakar et al. [2011], when a single-frequency inversion strategy is employed.

The effectiveness of this scheme will be demonstrated with many synthetic experiments, focusing on the inversion of elastic wave data. The adoption of a (isotropically) elastic modeling scheme is, by itself, another interesting aspect of our approach, whereas an acoustic model of the Earth is usually assumed in seismic imaging.

2.1.3 Summary of this paper

This work is organized as follows.

Section 2.2 We start by presenting a classical integral-based reformulation of the wave equation. This alternative point of view provides a useful hint on how to treat model parameters and wavefield variables as independent quantities for the inversion.

Section 2.3 Based on the insights developed in the previous section, we propose an alternating inversion routine. This is achieved by the introduction of a data-misfit functional and an equation-error functional, used to update model parameters and the wavefield, independently.

Section 2.4 In this section, we discuss the actual numerical treatment of the alternating inversion algorithm. The underlying finite-difference scheme is detailed in A. We will discuss the computational advantage of the scheme given by parallelization of the data-misfit optimization and wave modeling by standard factorization methods.

Section 2.5 A variety of numerical examples is presented to display the features of the alternating inversion scheme. We focus on 2-D synthetic elastic datasets.

2.2 The scattering integral equation

In this section, we briefly discuss and present the notation of the classical scattering integral reformulation of the wave equation. The scattering integral equation provides the basis of our inversion method and allows us to explore its main features, as it will be seen in Section 2.3. The treatise contained in this section and the following will be on a very abstract level. The exposition will be formal, in that we implicitly assume that the relevant variables belong to appropriate Hilbert spaces (or subsets) without rigorous specification.

2.2.1 The elastic differential system

We start from the second-order differential equation where the wavefield quantities are represented by the particle displacement $\mathbf{u} = (u^x, u^z)$. For later convenience, we assume that $\mathbf{u} = \mathbf{u}(\mathbf{x})$ is a function of the unbounded domain

$\mathbb{R}^2 \ni \mathbf{x}$. The differential system reads:

$$\begin{cases} -\rho\omega^2 u^x - \partial_x[(K + \frac{4}{3}\mu)\partial_x u^x + (K - \frac{2}{3}\mu)\partial_z u^z] \\ \qquad \qquad \qquad -\partial_z(\mu\partial_z u^x + \mu\partial_x u^z) = f^x \\ -\rho\omega^2 u^z - \partial_x(\mu\partial_z u^x + \mu\partial_x u^z) \\ \qquad \qquad \qquad -\partial_z[(K - \frac{2}{3}\mu)\partial_x u^x + (K + \frac{4}{3}\mu)\partial_z u^z] = f^z, \end{cases} \quad (2.3)$$

where ω is the angular frequency and $\mathbf{f} = (f^x, f^z)$ the source term — typically a function with compact support (as a delta function or derivatives). We adopt a frequency domain approach due to computational efficiency (as it will be discussed later on). The system (2.3) is parameterized by the following elastic properties:

$$\text{bulk modulus } K, \quad \text{shear modulus } \mu, \quad \text{density } \rho.$$

The elastic models will be summarized by the variable $\mathbf{m} = (K, \mu, \rho)$. As anticipated, we assume that a reference medium $\mathbf{m}_b = (K_b, \mu_b, \rho_b)$ is known. Typically, but not necessarily, it represents a smooth background, relatively to the propagation wavelength and is, therefore, ‘non-scattering’. The medium perturbation can then be formulated in terms of contrasts:

$$K = K_b(1 + \chi_K), \quad \mu = \mu_b(1 + \chi_\mu), \quad \rho = \rho_b(1 + \chi_\rho). \quad (2.4)$$

Similarly to \mathbf{m} , we will use the short-hand $\boldsymbol{\chi} = (\chi_K, \chi_\mu, \chi_\rho)$ to denote the contrast model. As a general requirement, the contrast functions vanish outside a bounded domain of interest $\Omega \subseteq \mathbb{R}^2$. This geometric setting is sketched in Fig. 2.1.

2.2.2 The elastic wave equation in the integral form

In order to present the integral equation we introduce the wave operator underlying equation (2.3):

$$\begin{aligned} \mathcal{H} &= -\rho\omega^2 \mathcal{I} - \mathcal{D}, \\ \mathcal{D} &= \begin{pmatrix} \partial_x(K + \frac{4}{3}\mu)\partial_x + \partial_z\mu\partial_z & \partial_x(K - \frac{2}{3}\mu)\partial_z + \partial_z\mu\partial_x \\ \partial_z(K - \frac{2}{3}\mu)\partial_x + \partial_x\mu\partial_z & \partial_x\mu\partial_x + \partial_z(K + \frac{4}{3}\mu)\partial_z \end{pmatrix}, \end{aligned} \quad (2.5)$$

being \mathcal{I} the identity. We will explicitly highlight the linear dependency of the wave operator on the model parameter \mathbf{m} by $\mathcal{H} = \mathcal{H}[\mathbf{m}]$. Indeed, it can be easily seen that $\mathcal{H}[\mathbf{m}_1 + \mathbf{m}_2] = \mathcal{H}[\mathbf{m}_1] + \mathcal{H}[\mathbf{m}_2]$, for any $\mathbf{m}_1, \mathbf{m}_2$. The total and incident wavefield are defined to solve the wave equations in the respective medium:

$$\mathcal{H}[\mathbf{m}]\mathbf{u} = \mathbf{f}, \quad \mathcal{H}[\mathbf{m}_b]\mathbf{u}^{\text{inc}} = \mathbf{f}. \quad (2.6)$$

The incident field is a known quantity when the source term \mathbf{f} is given. We will always assume that this is the case but, in practice, it has to be determined beforehand or treated as an additional variable of the problem. We can now consider the following identity, arising from an elementary algebraic manipulation:

$$\mathcal{H}[\mathbf{m}_b]\Delta\mathbf{u} = -\mathcal{H}[\Delta\mathbf{m}]\mathbf{u}, \quad \Delta\mathbf{u} = \mathbf{u} - \mathbf{u}^{\text{inc}}, \quad \Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_b. \quad (2.7)$$

This identity has been obtained from the discrete differentiation of (2.6) with respect to \mathbf{m} . From a physical point of view, this equation states that the scattered wavefield $\Delta\mathbf{u}$ is generated by a ‘distributed’ source which depends on the material contrast. Indeed, the right-hand side of equation (2.7) is sometime called ‘contrast source’ (as in the contrast-source inversion method of van den Berg and Kleinman [1997]). Once we introduce the Green’s operator:

$$\mathcal{G} = (\mathcal{H}[\mathbf{m}_b])^{-1}, \quad (2.8)$$

we might rewrite (2.7) as:

$$\mathbf{u} = \mathbf{u}^{\text{inc}} + \mathcal{G}\mathcal{W}[\boldsymbol{\chi}]\mathbf{u}, \quad (2.9)$$

where

$$\mathcal{W}[\boldsymbol{\chi}] = -\mathcal{H}[\mathbf{m}_b \cdot \boldsymbol{\chi}] \quad (2.10)$$

is the contrast-source operator. The operation $\mathbf{m}_b \cdot \boldsymbol{\chi}$ is intended as a component-wise multiplication. $\mathcal{W}[\boldsymbol{\chi}]$ acts linearly on the wavefield \mathbf{u} and depends on the contrast properties $\boldsymbol{\chi}$. We stress the fact that this dependency is also linear. This will be exploited by the inversion procedure as discussed in the following sections.

2.2.3 The domain equation

Equation (2.9) is the so-called *scattering integral equation* (also known as Lippman–Schwinger equation in quantum mechanics), and is an equivalent

reformulation of the differential equation (2.3). Note that (2.9) is based on a second-order formulation of the wave equation, but a formally identical treatment is possible for the first-order equation (as in de Hoop [1995]). This equation can be further refined by noticing that, since the contrast properties vanish outside the scattering domain Ω , the wavefield \mathbf{u} is entirely determined by its values on Ω (see (2.9)). We then introduce the restriction operator \mathcal{I}_Ω , defined by

$$\mathcal{I}_\Omega \mathbf{v} : \Omega \rightarrow \mathbb{C}^2, \quad (\mathcal{I}_\Omega \mathbf{v})(\mathbf{x}) = \mathbf{v}(\mathbf{x}),$$

for any complex-valued function \mathbf{v} defined on \mathbb{R}^2 . We will use Ω as a subscript to indicate transformed fields and operators through \mathcal{I}_Ω , e.g. $\mathbf{v}_\Omega = \mathcal{I}_\Omega \mathbf{v}$ and $\mathcal{T}_\Omega = \mathcal{I}_\Omega \mathcal{T} \mathcal{I}_\Omega^*$ (where \mathcal{T} is a generic linear operator). Here, the adjoint operation (denoted by an asterisk) is relative to the conventional least-squares scalar products on \mathbb{R}^2 and Ω :

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_i \int_{\mathbb{R}^2} u_i(\mathbf{x}) \overline{v_i(\mathbf{x})} d\mathbf{x}, \quad \langle \mathbf{u}, \mathbf{v} \rangle_\Omega = \sum_i \int_\Omega u_i(\mathbf{x}) \overline{v_i(\mathbf{x})} d\mathbf{x}, \quad (2.11)$$

where complex conjugation is denoted by a bar. The adjoint operator \mathcal{I}_Ω^* extends a function defined on Ω by setting its value to 0 outside Ω . The restriction of (2.9) to Ω produces the domain equation:

$$\mathbf{u}_\Omega = \mathbf{u}_\Omega^{\text{inc}} + \mathcal{G}_\Omega \mathcal{W}_\Omega[\boldsymbol{\chi}] \mathbf{u}_\Omega \quad (\text{domain equation}). \quad (2.12)$$

Equation (2.12) expresses \mathbf{u}_Ω as the sum of the incident field and the interaction of the total wavefield itself with the heterogeneity of the medium, governed by the scattering operator $\mathcal{K}_\Omega[\boldsymbol{\chi}] = \mathcal{G}_\Omega \mathcal{W}_\Omega[\boldsymbol{\chi}]$. The equation (2.12) can be finally restated in terms of a linear system as:

$$\mathcal{L}_\Omega[\boldsymbol{\chi}] \mathbf{u}_\Omega = \mathbf{u}_\Omega^{\text{inc}}, \quad \mathcal{L}_\Omega[\boldsymbol{\chi}] = \mathcal{I} - \mathcal{K}_\Omega[\boldsymbol{\chi}], \quad \mathcal{K}_\Omega[\boldsymbol{\chi}] = \mathcal{G}_\Omega \mathcal{W}_\Omega[\boldsymbol{\chi}]. \quad (2.13)$$

When the background is simple enough (e.g., homogeneous or horizontally invariant) the structure of the Green's operator (2.8) is advantageous for numerical computation and its evaluation can be carried out efficiently, making (2.12) a viable tool for both modeling (see, for example, Kleinman et al. [1990b,a]; Kleinman and van den Berg [1991], Yang et al. [2008], Abubakar and Habashy [2013]) and some specific applications of inverse scattering [van den Berg and Kleinman, 1997]. When \mathbf{m}_b does not display any particular symmetry, however, we are forced to revert to the implicit definition given by (2.8) and solve the associated linear system (arising from a given discretization),

when the evaluation of the operators involved in (2.12) is needed. Since in many cases, the discretization of the differential equation leads to a sparse system, the action of the Green's operator can be sped up by employing factorization algorithms (as exploited in Abubakar et al. [2009]). In Section 2.4 we will discuss different options for the numerical discretization of (2.12).

2.2.4 The data equation

An important relationship between recorded data and medium properties can be derived when equation (2.9) is specified at the receiver locations, in a similar way to the restriction process that led to the domain equation (2.12). We indicate the set of receiver locations by R , and the restricted Green's operator by:

$$\mathcal{G}_R = \mathcal{I}_R \mathcal{G} \mathcal{I}_\Omega^*. \quad (2.14)$$

At the receiver locations, the recorded data \mathbf{d} consists of the wavefield values assumed on R , that is $\mathbf{d} = \mathcal{I}_R \mathbf{u}$. Likewise, the restriction of the incident field on R is indicated by \mathbf{d}^{inc} . The scattering equation (2.9) then reads:

$$\Delta \mathbf{d} = \mathcal{G}_R \mathcal{W}_\Omega [\boldsymbol{\chi}] \mathbf{u}_\Omega, \quad \Delta \mathbf{d} = \mathbf{d} - \mathbf{d}^{\text{inc}}. \quad (2.15)$$

The backbone of the inversion method that will be presented in detail in the next section relies on a generalization of the relation (2.15). If the solution wavefield \mathbf{u}_Ω is a known quantity, then an estimate of $\boldsymbol{\chi}$ can be obtained by inverting (2.15): this would be tantamount to the inversion of a rectangular linear system (when no other non-linear effects are introduced, as it is the case for certain types of regularization). More generally, we might replace \mathbf{u}_Ω with only an estimate of the solution and proceed likewise to obtain an approximation of $\boldsymbol{\chi}$. By doing so, we are treating the wavefield independently from $\boldsymbol{\chi}$. Evidently, the biggest advantage of this generalization is the fact that the inversion for medium properties boils down to a more tractable linear problem than (2.15), where the wave equation enforces the constraint $\mathbf{u}_\Omega = \mathbf{u}_\Omega(\boldsymbol{\chi})$.

The generalized identity (2.15) expresses a linear relationship between contrasts and data. In order to make this more apparent, we introduce the following auxiliary operator, by a commutative procedure:

$$\mathcal{W}_\Omega [\mathbf{u}_\Omega] \boldsymbol{\chi} := \mathcal{W}_\Omega [\boldsymbol{\chi}] \mathbf{u}_\Omega \quad (2.16)$$

(with a slight abuse of notation). We remind that the contrast-source operator $\mathcal{W}_\Omega[\boldsymbol{\chi}]$ depends linearly on $\boldsymbol{\chi}$. Equation (2.15) can then be conveniently rewritten as:

$$\Delta \mathbf{d} = \mathcal{K}_R[\mathbf{u}_\Omega] \boldsymbol{\chi}, \quad \mathcal{K}_R[\mathbf{u}_\Omega] = \mathcal{G}_R \mathcal{W}_\Omega[\mathbf{u}_\Omega] \quad (\text{data equation}), \quad (2.17)$$

where $\mathcal{K}_R[\mathbf{u}_\Omega]$ is the data kernel.

2.3 An alternating update scheme for the inverse scattering problem

Now that we introduced the domain equation (2.12) and data equation (2.17), we are ready to outline a simple alternating inversion scheme for the inverse scattering problem.

Using the notation of Section 2.1, we are looking for a contrast model $\boldsymbol{\chi}$ that minimizes the objective functional:

$$\mathcal{J}(\boldsymbol{\chi}) = \|\Delta \mathbf{d} - \mathcal{F}(\boldsymbol{\chi})\|^2, \quad \mathcal{F}(\boldsymbol{\chi}) = \mathcal{K}_R[\mathbf{u}_\Omega(\boldsymbol{\chi})] \boldsymbol{\chi}, \quad (2.18)$$

where $\mathbf{u}_\Omega(\boldsymbol{\chi})$ is the solution of the wave equation (2.13). The inversion algorithm is based on the linear relationship (2.17) between data and contrasts when the wavefield is considered as an independent variable from the contrasts. This fact will be fully exploited by designing an inversion algorithm which aims for reconstruction of not only medium properties but also wavefields.

2.3.1 The alternating inversion algorithm

The inversion algorithm will perform a leap-frog reconstruction of contrast properties $\boldsymbol{\chi}$ and wavefields \mathbf{u}_Ω where we alternate the

- (i) update of the contrast model $\boldsymbol{\chi}$ by using the data equation (2.17), and
- (ii) update of the wavefield \mathbf{u}_Ω using the domain equation (2.12),

as schematically shown in Fig. 2.2. Steps (i) and (ii) are described more in detail in the following Sections 2.3.1 and 2.3.1.

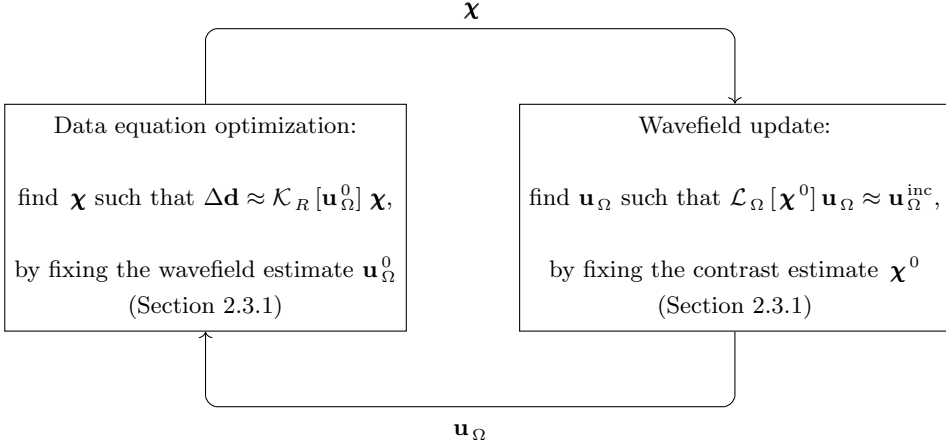


Figure 2.2: Simple inversion algorithm based on the alternating update of contrast properties χ and wavefields \mathbf{u} (see Section 2.3).

Update of contrasts by means of data equation optimization

To obtain an updating scheme for the contrasts we can reformulate the relation given by the data equation (2.17) as a least-squares optimization problem. For this, we will assume the knowledge of a suitable estimate of the wavefield \mathbf{u}_Ω^0 . Therefore, we set the objective functional:

$$\mathcal{J}_{\text{dat}}(\chi) = \|\Delta \mathbf{d} - \mathcal{K}_R[\mathbf{u}_\Omega^0] \chi\|^2, \quad (2.19)$$

employing the data kernel \mathcal{K}_R defined in (2.17).

It is well established that the problem of finding a minimum of (2.19) is an ill-posed problem (see, for example, Devaney [1978] and Devaney and Sherman [1982]). Due to this and the presence of noise in the recorded data $\Delta \mathbf{d}$, regularization techniques must be used. A typical adjustment is obtained by adding a regularization term \mathcal{N} to the objective (2.19):

$$\mathcal{J}_{\text{dat}}^{\text{reg}}(\chi) = \mathcal{J}_{\text{dat}}(\chi) + \alpha_{\text{reg}} \mathcal{N}(\chi).$$

Often, $\mathcal{N}(\chi)$ measures the least-squares norm of χ or derivatives. The determination of the weighting parameter α_{reg} is a non-trivial issue and it is often found through extensive numerical experiments. In this paper, we will adopt the multiplicative regularization technique developed in van den Berg et al. [2003]. Its main feature consists in the regularization of the functional (2.19)

by multiplication:

$$\mathcal{J}_{\text{dat}}^{\text{reg}}(\boldsymbol{\chi}; \boldsymbol{\chi}_{n-1}) = \mathcal{J}_{\text{dat}}(\boldsymbol{\chi}) \mathcal{N}(\boldsymbol{\chi}; \boldsymbol{\chi}_{n-1}). \quad (2.20)$$

Here, the regularization factor $\mathcal{N}(\cdot; \boldsymbol{\chi}_{n-1})$ depends on a given contrast model $\boldsymbol{\chi}_{n-1}$. The minimization strategy of (2.20) explored in van den Berg et al. [2003] employs an iterative scheme for which the functional to minimize $\mathcal{J}_{\text{data}}^{\text{reg}}(\cdot; \boldsymbol{\chi}_{n-1})$ changes at every iteration based on a previously computed model $\boldsymbol{\chi}_{n-1}$. In van den Berg et al. [2003], a non-linear conjugate-gradient method is considered, and verified that it is capable of taking into account this variability. This setup aims to enforce $\mathcal{N}(\boldsymbol{\chi}_n; \boldsymbol{\chi}_{n-1}) \rightarrow 1$ as the method converges, that is $\boldsymbol{\chi}_n - \boldsymbol{\chi}_{n-1} \rightarrow 0$. The regularization (2.20) only depends on tuning parameters that do not require lengthy experimentation, as opposed to additive regularizations. It is important to note that, when some weak assumptions on these parameters are fulfilled, the regularized functional $\mathcal{J}_{\text{dat}}^{\text{reg}}(\cdot; \boldsymbol{\chi}_{n-1})$ remains convex. For a thorough explanation of the method we refer to van den Berg et al. [2003].

Update of the wavefield by adding a order of scattering

As in the previous section, we discuss an updating scheme for the wavefield by fixing a known contrast model $\boldsymbol{\chi}^0$. Analogously to (2.19), we aim to minimize the wave equation error functional:

$$\mathcal{J}_{\text{wav}}(\mathbf{u}_{\Omega}) = \|\mathbf{u}_{\Omega}^{\text{inc}} - \mathcal{L}_{\Omega}[\boldsymbol{\chi}^0] \mathbf{u}_{\Omega}\|^2, \quad (2.21)$$

defined by the norm of the residual of the scattering integral operator (2.13). In the following, we will discuss a specific class of Krylov subspace methods for (2.21) (see Saad [2003] for a general introduction).

We start with the classical forward scattering series:

$$\mathbf{u}_{\Omega} = \mathbf{u}_{\Omega}^{\text{inc}} + \mathcal{K}_{\Omega}[\boldsymbol{\chi}^0] \mathbf{u}_{\Omega}^{\text{inc}} + \mathcal{K}_{\Omega}[\boldsymbol{\chi}^0]^2 \mathbf{u}_{\Omega}^{\text{inc}} + \mathcal{K}_{\Omega}[\boldsymbol{\chi}^0]^3 \mathbf{u}_{\Omega}^{\text{inc}} + \dots \quad (2.22)$$

This expansion can be restated in the following terms: if we denote the estimates of the solution by $\mathbf{u}_{\Omega}^{N-1}$, $\mathbf{u}_{\Omega}^{N-2}$, $\mathbf{u}_{\Omega}^{N-3}$, ..., $\mathbf{u}_{\Omega}^{N-n}$ computed in the previous n iterations, the addition of a term in (2.22) corresponds to the update scheme:

$$\mathbf{u}_{\Omega}^N = \mathbf{u}_{\Omega}^{N-1} + \mathbf{r}_{\Omega}^{N-1}, \quad \mathbf{r}_{\Omega}^{N-1} = \mathbf{u}_{\Omega}^{\text{inc}} - \mathcal{L}_{\Omega}[\boldsymbol{\chi}^0] \mathbf{u}_{\Omega}^{N-1}, \quad (2.23a)$$

where \mathbf{r}_Ω^{N-1} is the residual of the corresponding iteration. Note that $\mathbf{u}_\Omega^{N-2}, \dots, \mathbf{u}_\Omega^{N-n}$ are not utilized. The convergence of this sequence, however, is guaranteed only if the spectral radius of the operator $\mathcal{K}_\Omega[\chi^0]$ is smaller than 1. Typically, for large contrasts, this condition does not hold. Convergence can be achieved for a practical range of contrast size by employing a more general Krylov subspace scheme

$$\mathbf{u}_\Omega^N = \mathbf{u}_\Omega^{N-1} + \sum_{i=1}^n \alpha_i^N \mathcal{P} \mathbf{r}_\Omega^{N-i}, \quad (2.23b)$$

determined by how the coefficients $\alpha_1^N, \dots, \alpha_n^N$ are calculated, by a choice for the preconditioner \mathcal{P} (typically the identity operator or the adjoint of $\mathcal{L}_\Omega[\chi^0]$) and the number n of stored iterations. In the numerical experiments of Section 2.5 we will make use of the formula

$$(\alpha_1^N, \dots, \alpha_n^N) = \arg \min_{\alpha_i^N} \|\mathbf{r}_\Omega^{N-1} - \sum_{i=1}^n \alpha_i^N \mathcal{L}_\Omega[\chi^0] \mathbf{r}_\Omega^{N-i}\|^2.$$

The scheme is convergent whenever a certain $\tilde{\alpha}$ exists for which the spectral radius of $\mathcal{I} - \tilde{\alpha} \mathcal{L}_\Omega[\chi^0]$ is smaller than 1 [Kleinman and van den Berg, 1991] (see also Pratapa et al. [2016] for a comparison of this method with other Krylov subspace schemes). This condition is less stringent than the requirement for the convergence of the forward scattering series. For computational reasons, we will consider the case $n = 1$. Obviously, any other choice for the Krylov method could be employed to minimize (2.21) (e.g., conjugate gradient).

In general, the wavefield update can be obtained in many different ways. In the extreme case, one could fully solve the wave system (2.13) for the current contrast estimate. However, when the starting model is not close to the correct model, committing to such an erroneous model can be detrimental for the convergence of the inversion. This behavior has been observed experimentally. Notably, in van den Berg and Kleinman [1997] and van Leeuwen and Herrmann [2013], the wavefield update is designed to optimize a ‘data-weighted’ wave equation, where the data misfit plays the role of a regularization term and automatically prevents a close solution of the wave equation.

Starting point: Born approximation

The starting point of the algorithm relies on the so-called Born approximation. It essentially consists of a linearization of the inverse scattering problem (2.18)

with respect to the known background model. In other words, we consider the linearized relationship:

$$\mathbf{u}_\Omega(\mathbf{m}) \approx \mathbf{u}_\Omega^{\text{inc}} + d_{\mathbf{m}_b} \mathbf{u}_\Omega[\Delta\mathbf{m}],$$

where the linear operator $d_{\mathbf{m}_b} \mathbf{u}_\Omega$ is the differential of the map $\mathbf{u}_\Omega = \mathbf{u}_\Omega(\mathbf{m})$. A simple analysis of the differential yields the approximation

$$\Delta\mathbf{d} \approx \mathcal{K}_R[\mathbf{u}_\Omega^{\text{inc}}] \chi. \quad (2.24)$$

The corresponding estimate of the contrast is obtained following the instructions contained in Section 2.3.1.

2.4 Numerical implementation of the inversion scheme

In the previous section, we developed an alternating inversion algorithm for inverse scattering problems. So far, however, we avoided an explicit numerical treatise of wavefield and model unknowns and subsequent discretization of the linear operators involved in the data and domain equations.

In this section, we begin by investigating different alternatives for the discretization of the differential wave equation (2.3), and its solution by numerical methods. It is clear that a specific choice for the discretization also determines the corresponding integral equation (2.9) and the operators involved in the domain equation (2.12) and data equation (2.17) (to be discussed, respectively, in Sections 2.4.1 and 2.4.2). Therefore, the numerical scheme underlying the wave equation prescribes every other detail needed for the implementation of the alternating inversion of Section 2.3.

This section is integrated with a brief review of the complexity of the algorithm (Section 2.4.3). For the optimization of the data equation, we outline some general strategies that take advantage from parallelization. For the solution of the wave equation, we must distinguish the 2-D and 3-D case. For 2-D, factorization-based direct methods, such as LU [George and Liu, 1981], are readily available. For large 3-D problems, however, the storage of the LU decomposition is highly impractical when each of the frequencies contained in the data is needed simultaneously. In Section 2.4.4, therefore, we present a simple strategy to overcome this issue, by selecting a sequence of limited frequency sets and applying a sequential inversion which also involves the update of the background model. In this case, when only few frequencies are needed

at the same time, we might still resort to direct methods [Abubakar et al., 2011].

2.4.1 Discretization of the scattering integral equation

The scattering integral domain equation (2.12) has been used as a basis for the wavefield update phase of the alternating inversion algorithm. Clearly, once this equation is specified by a discretization method, the Krylov subspace update described in Section 2.3.1 can be applied in a straightforward manner.

As anticipated, equation (2.12) is determined by the choice for the discretization of the differential system $\mathcal{H}_\Omega[\mathbf{m}]$, for a generic model \mathbf{m} . This, in turn, defines the contrast-source operator $\mathcal{W}_\Omega[\boldsymbol{\chi}]$ and the Green's operator \mathcal{G}_Ω , following the definitions (2.10) and (2.8). We assume that the computational domain Ω is a subset of the 2-D orthogonal grid $G = \{(ih, jh) : i, j \in \mathbb{Z}\}$ (\mathbb{Z} being the set of integers), where h is the grid spacing. Wavefield and medium unknowns will be represented by grid functions $\mathbf{u}_{i,j}$, $\mathbf{m}_{i,j}$ approximating the continuous quantities on the grid points, e.g., $\mathbf{u}_{i,j} \approx \mathbf{u}(ih, jh)$, $\mathbf{m}_{i,j} \approx \mathbf{m}(ih, jh)$. For ease of exposition, we will indicate discretized and continuous quantities with the same notation used in the earlier sections.

A number of different options can be considered for the discretization of $\mathcal{H}_\Omega[\mathbf{m}]$, among which we will primarily focus on two broad classes: *analytical* and *stencil-based* methods. The first class refers to those methods which make use, in some form or another, of the (semi-)analytical expression of the Green's functions for the background model \mathbf{m}_b . The latter class comprises those schemes for which the evaluation of the discretized operator $\mathcal{H}_\Omega[\mathbf{m}]$ can be represented by compact stencils (cf. A). Notable examples are finite-differences and finite-elements. Relative merits of one class of discretization over another is determined by the regularity of \mathbf{m}_b .

Here, we give a general overview of analytical and stencil-based methods. The method of choice for this paper will be a finite-difference scheme based on the numerical dispersion optimization of the well-known staggered-grid method [Virieux, 1986], and is described in great detail in A.

Analytical methods

When the background model \mathbf{m}_b displays some form of symmetry, the related Green's functions might admit a convenient analytical representation. In the extreme case of an homogeneous model, for example, the Green's operator possesses a convolutional structure that can be exploited via fast Fourier transform [Yang et al., 2008]. Another interesting situation is when the medium is 1-D — that is, constant along the horizontal direction x . In this case, the Green's operator retains a 1-D convolutional structure. When we also assume that the 1-D background is smoothly varying with respect to the propagation wavelength, a WKB approximation [Griffiths, 2004] can be used to obtain a semi-analytical estimate of the Green's functions. The 1-D case is particularly useful because, in some cases, the true medium \mathbf{m} is well approximated by a locally-varying 1-D model. The inverse problem can then be parallelized over horizontal locations, after a proper selection and rearrangement of the data. Furthermore, these subproblems are more easily analyzed in the horizontal wavenumber-frequency domain, since the wave equation can be solved independently for any given horizontal wavenumber k_x and frequency ω .

Under the above circumstances, the linear operator underlying the domain equation (2.12) is evaluated efficiently, and (2.12) can be solved iteratively by Krylov subspace methods. Notably, when the model \mathbf{m}_b is close to \mathbf{m} , the integral reformulation (2.12) of the wave equation is more suited for iterative methods than a direct discretization of the differential system $\mathcal{H}[\mathbf{m}]$. This is due to the implicit preconditioning highlighted by the identity $\mathcal{L}_\Omega[\boldsymbol{\chi}] = (\mathcal{H}_\Omega[\mathbf{m}_b])^{-1} \mathcal{H}_\Omega[\mathbf{m}]$. When the contrast is high, however, the system $\mathcal{L}_\Omega[\boldsymbol{\chi}]$ should be preconditioned (see, for example, the renormalization technique in Abubakar and Habashy [2013]).

The analytical implementation of the scattering integral equation has been the basis for the contrast-source inversion method [van den Berg and Kleinman, 1997]. A version of the alternating inversion scheme of Section 2.3 based on analytical methods has already been worked out under the 2-D acoustic assumption [Haffinger et al., 2013] and for the locally 1-D elastic case (Gisolf et al. [2012], e.g.).

Stencil-based methods

If \mathbf{m}_b does not exhibit any particular symmetry, analytical methods are not available. The evaluation of the scattering integral operator (2.12) can still be computed indirectly, by solving the differential system:

$$\mathcal{G}_\Omega \mathbf{w} = \mathbf{v} \quad \Longleftrightarrow \quad \mathcal{H}_\Omega [\mathbf{m}_b] \mathbf{v} = \mathbf{w}.$$

Since the storage of the full matrix $(\mathcal{H}_\Omega [\mathbf{m}_b])^{-1}$ is impractical, one needs to solve the linear system every time the evaluation of the Green's operator is needed, for which an efficient discretization and solver of $\mathcal{H}_\Omega [\mathbf{m}_b]$ is essential. Under these conditions, stencil-based methods are an interesting choice.

Many finite-difference or finite-element schemes have been developed for the time-harmonic elastic wave equation. Classical 9-point finite-difference schemes are by Kelly et al. [1976] and the staggered-grid scheme by Virieux [1986]. These methods often provided the basis for more accurate compact schemes, based on the minimization of the numerical dispersion (e.g., the 9-point scheme of Štekl and Pratt [1998] and the 25-point finite-difference and finite-element schemes of Min et al. [2000, 2003, 2004]). For this paper, we developed a 9-point finite-difference method based on the optimization of the staggered-grid scheme. The group velocity relative error is essentially independent of the P-to-S velocity ratio and is smaller than 1 % for a discretization corresponding to more than 5 samples per wavelength. The method is discussed in A.

Any discretization of the wave equation by stencil-based schemes leads to a large, but sparse, linear system $\mathcal{H}_\Omega [\mathbf{m}_b]$. For sizable problems, direct solution methods as the classical LU factorization can be computationally intensive. This is the reason why time-domain and explicit time-marching solvers are generally preferred. On the other hand, the frequency domain offers many advantages: only a limited number of frequencies — far fewer than what is prescribed by the Nyquist criterion — are actually required to obtain a satisfactory imaging result [Sirgue and Pratt, 2004; Mulder and Plessix, 2004a]. Also, a further level of parallelization is available (over sources and frequencies) with respect to time-domain imaging. An obvious alternative to the direct solution of a linear system is offered by iterative methods: recent developments make this a competitive route [Knibbe et al., 2014; Rizzuti and Mulder, 2016].

2.4.2 Treatment of the data equation

The different discretization schemes described in the previous section also defines the data kernel operator (2.17), $\mathcal{K}_R[\mathbf{u}_\Omega]$. We base the data optimization phase (detailed in Section 2.3.1) on the non-linear conjugate-gradient scheme. This requires the evaluation of the data kernel and its adjoint to compute the search direction. Due to computational reasons, we wish to exploit the decomposition $\mathcal{K}_R[\mathbf{u}_\Omega] = \mathcal{G}_R \mathcal{W}_\Omega[\mathbf{u}_\Omega]$, using the definitions (2.14) for the receiver-to-domain Green's operator \mathcal{G}_R and (2.10), (2.16) for the contrast-source operator $\mathcal{W}_\Omega[\mathbf{u}_\Omega]$. In the following, we describe how to compute these two operators.

Contrast-source operator

By definition (2.16), $\mathcal{W}_\Omega[\mathbf{u}_\Omega]$ is an operator acting on contrast variables $\boldsymbol{\chi}$. This notation explicitly highlights the linear dependency with respect to the wavefield variable \mathbf{u}_Ω . In other words, the operator $\mathcal{W}_\Omega(\boldsymbol{\chi}, \mathbf{u}_\Omega) := \mathcal{W}_\Omega[\mathbf{u}_\Omega] \boldsymbol{\chi}$, acting on both $\boldsymbol{\chi}$ and \mathbf{u}_Ω , is bilinear. When a stencil-based discretization is set, it is clear, from (2.10), that the operator $\mathbf{u}_\Omega \mapsto \mathcal{W}_\Omega(\boldsymbol{\chi}, \mathbf{u}_\Omega)$ is represented by a sparse linear system (showing the same sparsity pattern as $\mathcal{H}_\Omega[\mathbf{m}]$, for a generic model \mathbf{m}). Likewise, the operator $\boldsymbol{\chi} \mapsto \mathcal{W}_\Omega(\boldsymbol{\chi}, \mathbf{u}_\Omega) = \mathcal{W}_\Omega[\mathbf{u}_\Omega] \boldsymbol{\chi}$, acting now on $\boldsymbol{\chi}$, can be conveniently represented by a sparse linear system (and consequently, also its adjoint). Eventually, the action of the contrast-source operator can be implemented in a matrix-free fashion.

Receiver-to-domain Green's operator

The operator \mathcal{G}_R , defined by (2.14), can be readily evaluated by storing the Green's functions generated by source pulses located at each point of the domain Ω and recorded in R . Given a grid point $\mathbf{x} \in G$, we consider a discretized version of the delta pulse function defined by:

$$\langle \delta_{\mathbf{x}}, u \rangle = u(\mathbf{x}),$$

for any scalar grid function u . In 2-D, the type of source appearing in the wave equation (2.3) will be denoted by the vectors $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$. Now, we introduce the Green's functions:

$$g_{k,l}(\mathbf{x}, \mathbf{x}') = \langle \mathcal{G} \delta_{\mathbf{x}'} \mathbf{e}_l, \delta_{\mathbf{x}} \mathbf{e}_k \rangle, \quad (2.25)$$

for generic grid points \mathbf{x}, \mathbf{x}' . Clearly:

$$\begin{aligned} (\mathcal{G}_R \mathbf{w})_k(\mathbf{x}_r) &= \sum_{\mathbf{x} \in \Omega} \sum_l g_{k,l}(\mathbf{x}_r, \mathbf{x}) w_l(\mathbf{x}), \quad \forall \mathbf{x}_r \in R, \\ (\mathcal{G}_R^* \mathbf{d})_k(\mathbf{x}) &= \sum_{\mathbf{x}_r \in R} \sum_l \overline{g_{l,k}(\mathbf{x}_r, \mathbf{x})} d_l(\mathbf{x}_r), \quad \forall \mathbf{x} \in \Omega, \end{aligned} \quad (2.26)$$

for any contrast-source function \mathbf{w} defined on Ω and data vector \mathbf{d} .

Equation (2.26) provides a practical way to compute the action of \mathcal{G}_R and \mathcal{G}_R^* when the functions $g_{k,l}(\mathbf{x}_r, \cdot)$ are pre-computed and stored, for each of the receiver position \mathbf{x}_r . It is tempting to compute $g_{k,l}(\mathbf{x}_r, \cdot)$ by applying the classical reciprocity relation $g_{k,l}(\mathbf{x}, \mathbf{x}_r) = g_{l,k}(\mathbf{x}_r, \mathbf{x})$. This, indeed, would just require the solution of the wave equation where the source term is a delta pulse located at the receiver position. This equivalence holds true in the continuous case because the differential system $\mathcal{H}[\mathbf{m}]$ is real symmetric. However, the discretization of the wave equation does not necessarily lead to a symmetric system. In this case, a pseudo-reciprocity principle can still be exploited by considering the ‘transposed’ Green’s functions $g_{k,l}^{\text{tr}}(\mathbf{x}, \mathbf{x}') = \langle \mathcal{G}^t \delta_{\mathbf{x}'} \mathbf{e}_l, \delta_{\mathbf{x}} \mathbf{e}_k \rangle$ (where real transposition is indicated by the superscript t). The following identity holds: $g_{k,l}(\mathbf{x}', \mathbf{x}) = g_{l,k}^{\text{tr}}(\mathbf{x}, \mathbf{x}')$. Therefore, in general, one should compute the Green’s functions $g_{k,l}(\mathbf{x}_r, \cdot)$ by solving the transposed of the wave linear system. In any case, the numerical scheme of A used in this paper leads to a real-symmetric system for which $\mathcal{G}^t = \mathcal{G}$, hence $g_{k,l}(\mathbf{x}, \mathbf{x}_r) = g_{k,l}^{\text{tr}}(\mathbf{x}, \mathbf{x}_r) = g_{l,k}(\mathbf{x}_r, \mathbf{x})$.

2.4.3 Complexity analysis

The goal of this section is to analyze the complexity of the alternating inversion algorithm for each of its components: data equation optimization and wavefield update. We assume that the discretization method of choice is a stencil-based scheme. For simplicity, the analysis is carried out for a single frequency and a comparison is made with the contrast-source inversion method [Abubakar et al., 2009], which shares many similarities with the current algorithm. In the following, N will indicate the number of grid points of Ω . N_r and N_s are, respectively, the number of receiver and source locations.

For the data equation optimization (Section 2.3.1), the complexity is determined by the calculation of the data kernel $\mathcal{K}_R[\mathbf{u}_\Omega]$ and the total number $N_{\text{dat}}^{\text{iter}}$ of iterations required by the conjugate-gradient scheme. Following

the strategy contained in Section 2.4.2, the evaluation of the operator requires $O(N)$ operations, for each receiver and source. This leads to an overall $O(c_{N_r, N_s} N)$ complexity per iteration, where the factor c_{N_r, N_s} depends on N_r and N_s . With a serial implementation of the matrix-vector product for the data kernel, this factor is equal to $c_{N_r, N_s} = N_r N_s$. Clearly, in an abundance of compute cores, this step can be well parallelized leading to a small c_{N_r, N_s} (ideally, $c_{N_r, N_s} = 1$). A further improvement can be obtained, in principle, with stochastic gradient descent techniques, for which only a limited number of randomly chosen sources and receivers are needed at each iteration to compute the gradient for the local search optimization. Another approach has been followed by [Krebs et al., 2009], where the optimization is driven by data generated with a linear combination of all the source terms, whose coefficients change randomly at each iteration. The computational gain would amount to $c_{N_r, N_s} = N_r + N_s$ per iteration, with a direct serial implementation. In the numerical experiments of this paper, however, we will not resort to these strategies.

For the wavefield update (Section 2.3.1), the application of the integral operator \mathcal{L}_Ω in (2.13) requires the solution of the wave equation by the LU solver. In 2-D, the LU decomposition with nested dissection reordering [George and Liu, 1981] is computed in $O(N^{3/2})$ operations and the LU-based solver requires $O(N \log N)$ operations per source. In 3-D, the factorization takes $O(N^2)$ operations while the solution has a cost of $O(N^{4/3})$.

With a proper parallelization, the data kernel evaluation is a relatively cheap computation with respect to the wavefield update. Hence, it might be beneficial to focus the computational effort on this particular stage of the alternating scheme. Indeed, as it will be shown in the numerical experiments of Section 2.5, the number of wavefield update iterations $N_{\text{wav}}^{\text{iter}}$ needed for a satisfactory result is relatively small.

In terms of memory complexity, the requirements are determined by the storage of the LU factorization and the Green's functions (2.26) and wavefields. Thus, the memory amounts to $O(N \log N + (N_r + N_s) N)$ in 2-D, and $O(N^{4/3} \log N + (N_r + N_s) N)$ in 3-D.

In Table 2.1 we summarized the analysis carried out in this section. A comparison is drawn with the contrast-source inversion method, which shows that the alternating update scheme can be competitive when adequately parallelized, at the cost of slightly higher memory demands.

Factorization		Alternating update	
		Data optimization	Wavefield update
2-D	$N^{3/2}$	$c_{N_r, N_s} N_{\text{dat}}^{\text{iter}} N$	$+ N_{\text{wav}}^{\text{iter}} N_s N \log N$
3-D	N^2	$c_{N_r, N_s} N_{\text{dat}}^{\text{iter}} N$	$+ N_{\text{wav}}^{\text{iter}} N_s N^{4/3} \log N$

	Factorization	Contrast-source inversion
2-D	$N^{3/2}$	$N_{\text{csi}}^{\text{iter}} N_s N \log N$
3-D	N^2	$N_{\text{csi}}^{\text{iter}} N_s N^{4/3} \log N$

(a) Time complexity for a single frequency (big O notation)

	Factorization	Alternating update	Contrast-source inversion
2-D	$N \log N$	$(N_r + N_s) N$	$N_s N$
3-D	$N^{4/3} \log N$	$(N_r + N_s) N$	$N_s N$

(b) Memory complexity for a single frequency (big O notation)

Table 2.1: Complexity analysis of the alternating update scheme and comparison with the contrast-source inversion method: (a) time complexity, (b) memory complexity. N is the size of the problem, N_r and N_s are the number of receivers and sources. For this analysis, we considered a single frequency. For the alternating scheme, $N_{\text{dat}}^{\text{iter}}$ is the number of iterations required for the data equation optimization and $N_{\text{wav}}^{\text{iter}}$ is the number of wavefield updates, while $N_{\text{csi}}^{\text{iter}}$ is the number of iterations employed by the contrast-source inversion. The data optimization stage can greatly benefit from parallelization for which the factor c_{N_r, N_s} becomes small. The alternating update scheme can be quite advantageous since, as it is experimentally observed, $N_{\text{wav}}^{\text{iter}} \ll N_{\text{csi}}^{\text{iter}}$. Note that the time complexity for the factorization in (a) refers to the number of operations used to compute such decomposition, while the memory complexity is the storage needed to keep it in memory. The complexity of the factorization is shown only once since it is the same for both the alternating update and contrast-source inversion methods.

2.4.4 Inversion strategy: all-frequencies-at-once vs frequency-by-frequency

From the analysis of the previous section, it is clear that the LU decomposition plays an important role in the memory complexity of the method. When

N_f frequencies are considered simultaneously, we are forced to store N_f LU factorizations. This makes an all-at-once inversion approach impractical for large problems or when many frequencies are needed.

Alternatively, we can consider the *frequency-by-frequency* strategy proposed for the contrast-source inversion method in Abubakar et al. [2009] (see also He et al. [2016]) which requires the storage of LU factorizations only for a limited number of frequencies at the same time (down to a single frequency). Note that, in what follows, we will always choose an equally sampled frequency set. However, one might consider the more convenient approach described by [Sirgue and Pratt, 2004] (or Mulder and Plessix [2004a]). The strategy adopted in [Abubakar et al., 2009] can be summarized as follows. We start by picking a small subset of the lowest frequencies contained in the data and we compute the corresponding LU factorizations of the wave operator relative to the current background model. We then apply the alternating inversion scheme as described in Section 2.3. The next phase consists in the update of the background model, based on the previous result, a choice of a new higher frequency set and the corresponding LU computation. This will result in another inversion run. The same logic is carried on, until the full spectrum of the data is covered. A simple sketch of the inversion algorithm endowed with a frequency-by-frequency approach is depicted in Figure 2.3.

The frequency-by-frequency inversion strategy might be suited even for 3-D, as argued in Abubakar et al. [2011]. This idea has been explored for the contrast-source inversion method in Abubakar et al. [2011]. In the following section, we will include a comparison between the all-at-once and frequency-by-frequency approaches for 2-D synthetic problems.

2.5 Numerical experiments

We present some 2-D synthetic numerical experiments to demonstrate the capabilities of the alternating inversion algorithm. The first example aims at giving a practical understanding of how the inversion result improves by using increasingly accurate estimates of the wavefield. The model considered in this attempt is the so-called Delphi temple, designed after the logo of the DELPHI consortium[†]. The second example depicts a more geologically plausible scenario and is structurally similar to the synthetic model studied in Gray and

[†]www.delphi-consortium.com

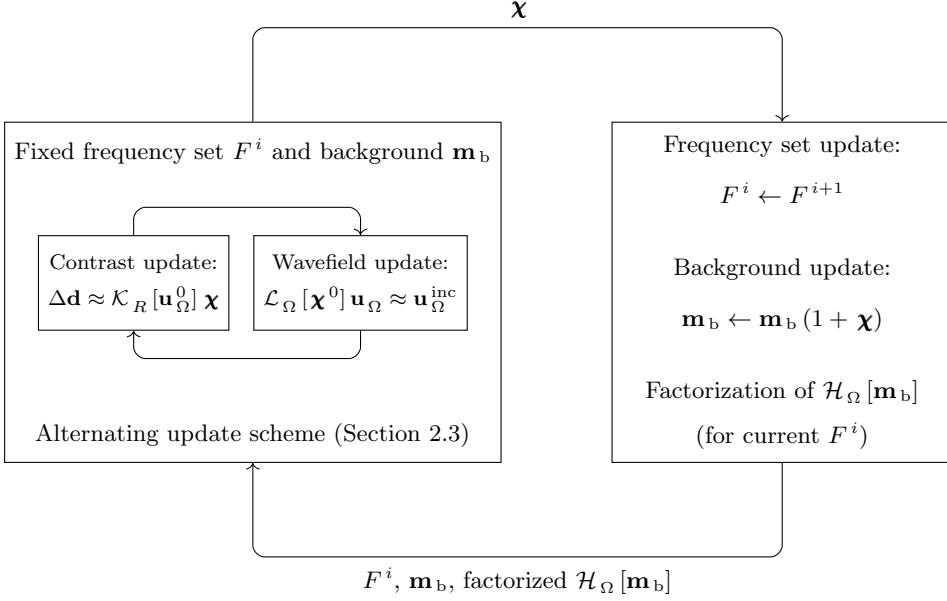


Figure 2.3: *Frequency-by-frequency inversion strategy (to be compared with the all-at-once approach in Figure 2.2). First, a frequency strategy is chosen: $F^1 \rightarrow F^2 \rightarrow F^3 \rightarrow \dots$, where $F^i = \{f_1^i, f_2^i, \dots\}$ are sets of frequencies, possibly consisting of a single element. This choice is made in such a way that F^i contains increasingly higher frequencies. Starting from F^1 and the given background model, the alternating update inversion described in Section 2.3 produces a new contrast model. This model is used now as a replacement for the background and the next frequency set F^2 is considered for another inversion run, after the LU factorization of $\mathcal{H}_\Omega[\mathbf{m}_b]$ is computed.*

Marfurt [1995]. In this case we will test the frequency-by-frequency inversion strategy against the all-at-once approach (Section 2.4.4). For each of the experiments, the synthetic data are generated using the same forward modeling scheme used in the inversion and 5% Gaussian white noise is added to ease the ‘inverse crime’ [Wirgin, 2008].

2.5.1 The Delphi temple

The Delphi model and the source-receiver configuration of the experiment are depicted in Figure 2.4. The model size is 300 m by 600 m. As for physical properties, the homogeneous background corresponds to $K_b = 5.33 \times 10^9$ Pa,

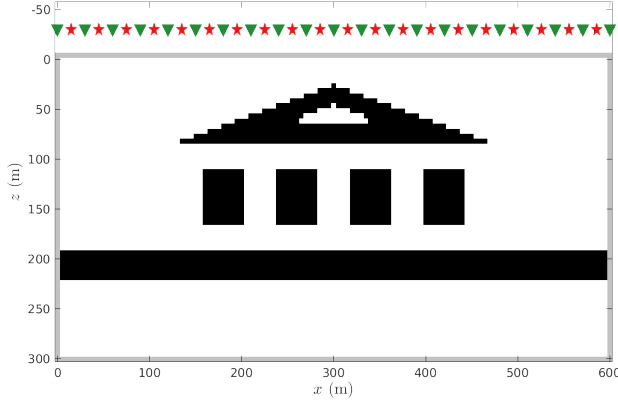


Figure 2.4: *Delphi temple model and source-receiver configuration. The computational domain Ω is delimited by black borders. The non-vanishing contrast of the model (with respect to the homogeneous background) is highlighted in black.*

$\mu_b = 2 \times 10^9$ Pa and $\rho_b = 2000 \text{ kg m}^{-3}$ (in other words, the P-wave velocity is $c_P = 2000 \text{ m s}^{-1}$ and the S-wave velocity is $c_S = 1000 \text{ m s}^{-1}$). For the heterogeneity, we have $K = 6.93 \times 10^9$ Pa, $\mu = 2.6 \times 10^9$ Pa and $\rho = 2200 \text{ kg m}^{-3}$ ($c_P = 2174 \text{ m s}^{-1}$, $c_S = 1087 \text{ m s}^{-1}$). This corresponds to the contrast values $\chi_K = 0.3$, $\chi_\mu = 0.3$ and $\chi_\rho = 0.1$.

We employ 21 sources and 21 receivers evenly distributed along the top of the model. The source type considered is that of a vertical force delta pulse (see eq. (2.3)). The inversion is carried out by considering all the frequencies at once, and inverting both the components of the particle displacement: u^x , u^z . We invert for the frequency spectrum 10–40 Hz, which corresponds, for the current model discretization, to a sampling of 5 points per minimum S-wavelength. At the lowest frequency 10 Hz, the model dimension is $3 \lambda_S$ by $6 \lambda_S$, where λ_S is the corresponding S-wavelength.

Results The output models are depicted in Figures 2.5–2.8. We show the result after every data equation optimization step, starting from the Born approximation. Note that each data optimization step consists of an inner loop where the maximum number of iterations is set to 300. The final result in Figure 2.8 is obtained after 10 wavefield updates. While for the very first

iteration in Figure 2.5 the pillars of the model are not recovered yet, due to a lack of illumination, the estimation of a more accurate wavefield provides the necessary information to gradually reconstruct this part of the model, as it can be seen in Figures 2.6–2.8.

It is interesting to remark that the recovered μ displays a better spatial resolution than K , as it can be qualitatively assessed from Figure 2.8. This can be observed, even more clearly, in the experiments of the next section 2.5.2 (see Figures 2.14–2.15). This is evident, given the fact that the term of the data due to the propagation of S-waves shows much lower sensitivity to K than μ . For a fixed temporal frequency, the S-wave propagation wavelength is shorter than the P-wave and this will translate into better resolution for μ when such data is actually available. Furthermore, we observe that the inverted ρ in Figure 2.8 is affected by long wavelength artifacts, a well-known physical limitation of seismic inversion.

2.5.2 The curly model

For a comparison of the different inversion strategies in Section 2.4.4, we consider the model depicted in Figure 2.9. The size of the computational domain is 310 m by 520 m. The P-wave velocity ranges from 1800 m s^{-1} (top of the model) to 3000 m s^{-1} (bottom), while the S-wave velocity from 1200 m s^{-1} to 2100 m s^{-1} .

We place 27 sources and 26 receivers above the model, similarly to the previous example, at the relative depth of $z = -25 \text{ m}$ (cf. Figure 2.9). As in the Delphi temple, we invert multi-component data generated by known vertical force sources. The frequency spectrum considered is 10–50 Hz, which corresponds to a minimal discretization rate of 5 points per S-wavelength, for the chosen grid spacing.

This experiment aims to compare the result after the two inversion strategies discussed in Section 2.4.4: all-at-once inversion vs frequency-by-frequency inversion. The starting point for both strategies is the simple layered model in Figure 2.10, which will be used as an (initial) background for the inversion. We remark that this model is relatively far from the truth not only in terms of short wavelengths, but also for long components (compare, for example, the vertical sections in Figures 2.14 and 2.15). An example of synthetic shot gather in the frequency-receiver domain can be seen in Figure 2.13a.

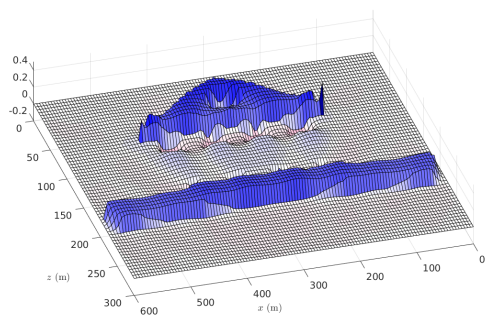
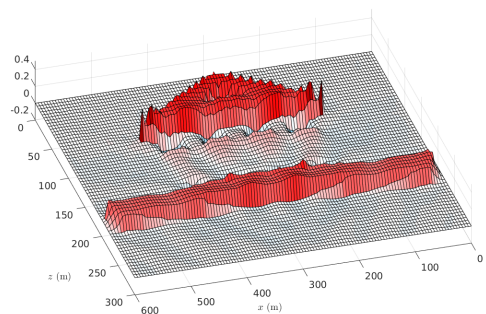
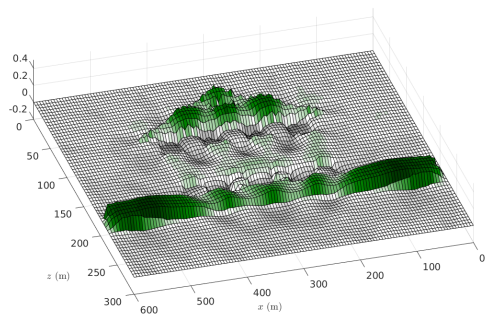

 χ_K

 χ_μ

 χ_ρ

Figure 2.5: Inversion results for the Delphi temple: 1st iteration (Born inversion).

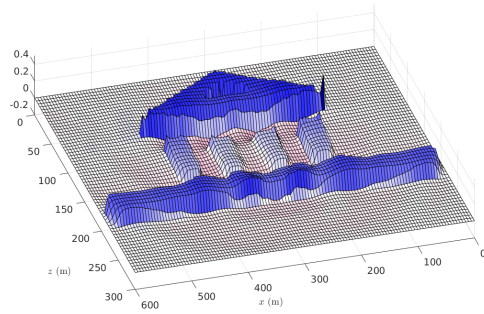
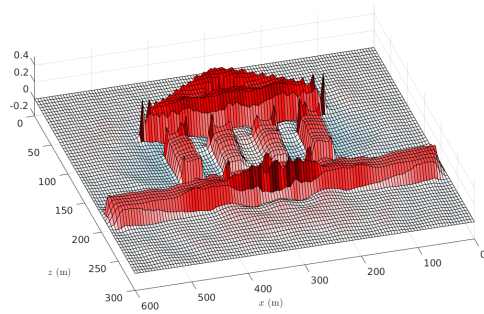
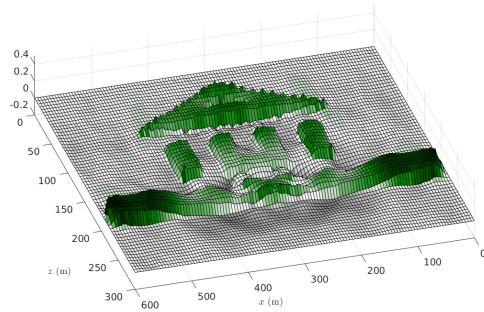

 χ_K

 χ_μ

 χ_ρ

Figure 2.6: *Inversion results for the Delphi temple: 2nd iteration.*

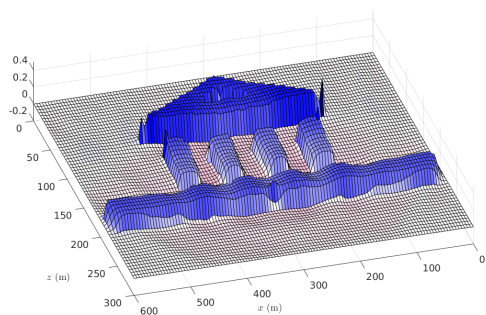
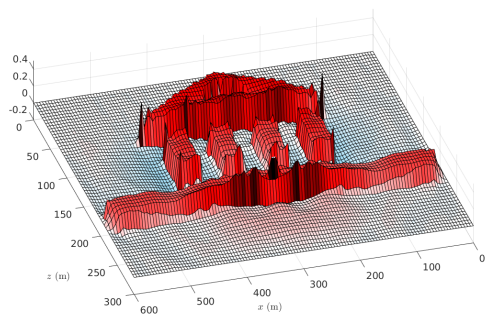
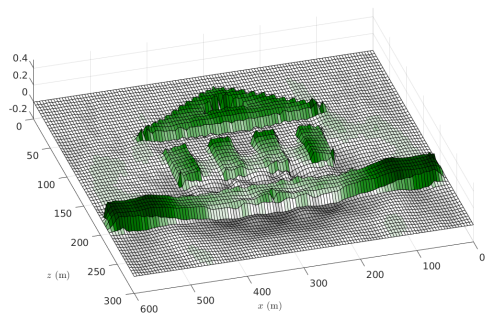

 χ_K

 χ_μ

 χ_ρ

Figure 2.7: *Inversion results for the Delphi temple: 3rd iteration.*

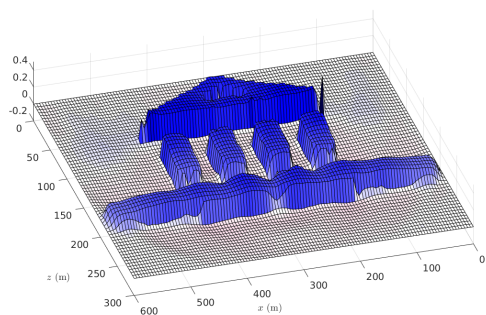
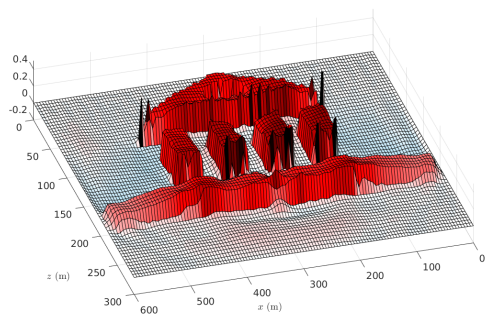
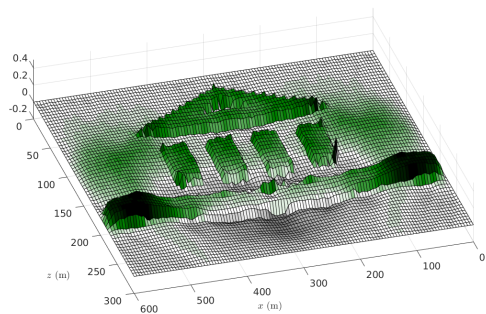

 χ_K

 χ_μ

 χ_ρ

Figure 2.8: *Inversion results for the Delphi temple: 10th iteration.*

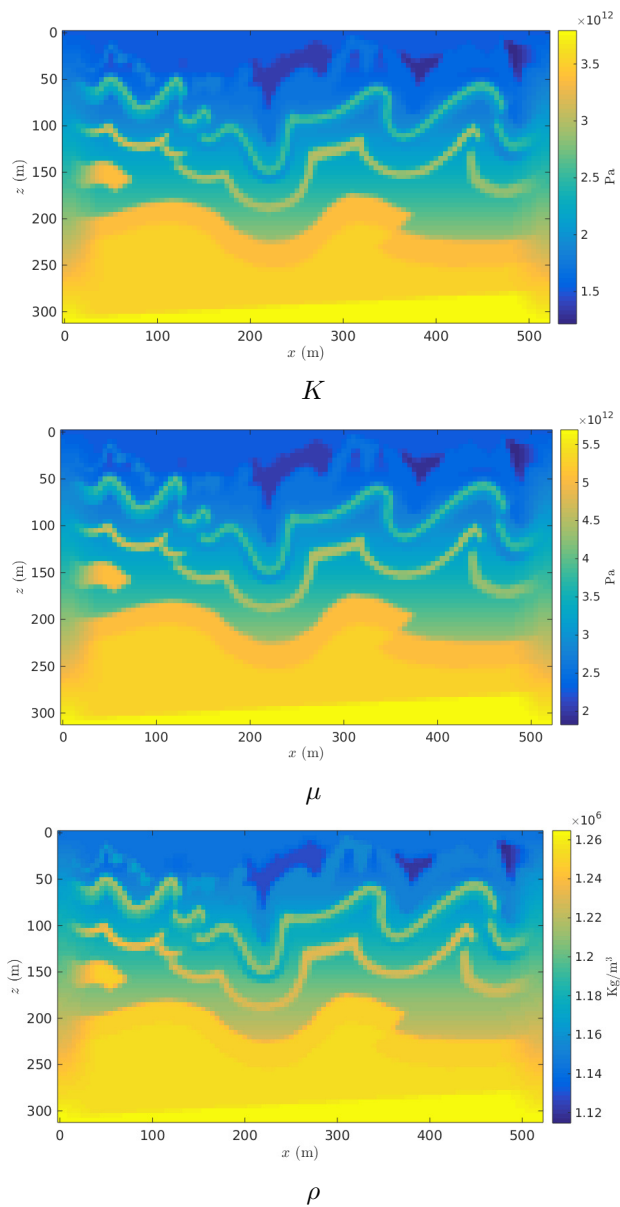


Figure 2.9: Curly model for the elastic parameters: bulk modulus K , shear modulus μ and density ρ (absolute values).

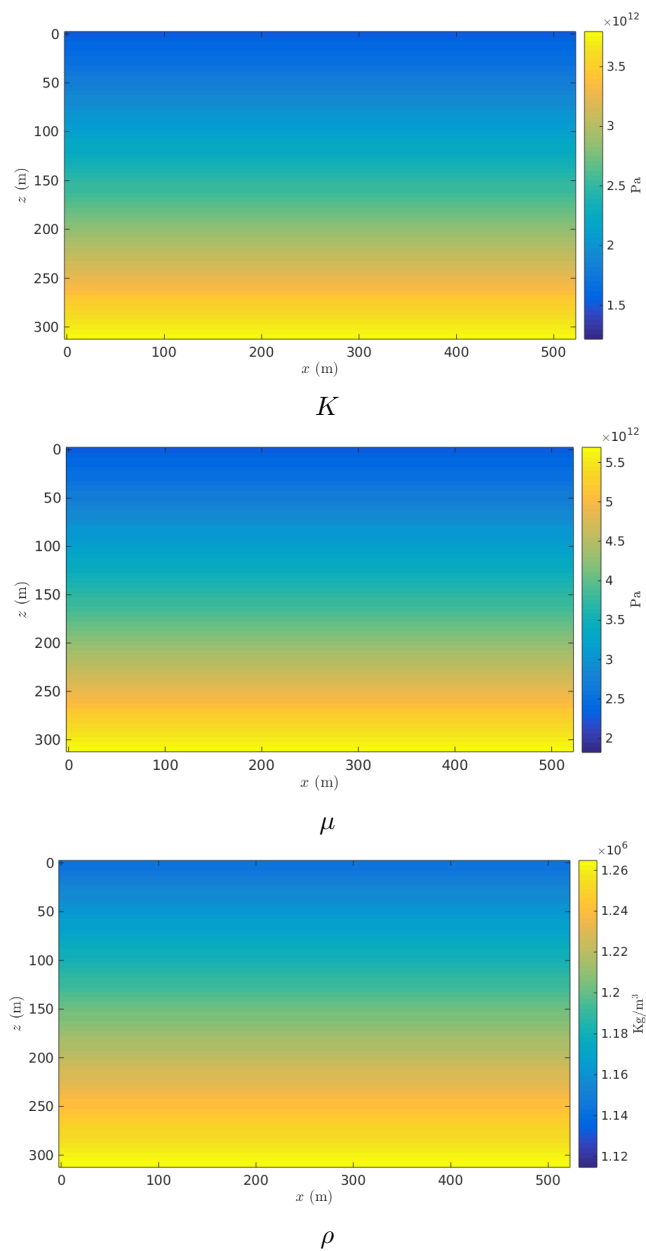


Figure 2.10: Curly background model for the elastic parameters: bulk modulus K , shear modulus μ and density ρ (absolute values).

Results The reconstructed models are collected in Figure 2.11 and Figure 2.12. In Figure 2.11 we show the results obtained from the simultaneous frequency inversion after 10 wavefield updates. We set a maximum number of 300 iterations for the data optimization loop. The results for the frequency-by-frequency approach are depicted in Figure 2.12. For this particular experiment we considered frequency sets of only 3 elements, starting from the lowest available frequencies. For each of these sets, we run the inversion with 10 wavefield update iterations and a maximum number of 300 iterations for the data optimization loop.

In general, the observations made for the Delphi temple also applies in this case: the recovered shear modulus is relatively sharper than the bulk modulus, and the density reconstruction suffers from low frequency instability. From a qualitative point of view, the results of the two inversion strategies compare well with the true model and with respect to each other. In Figure 2.13, the estimated data for both strategies show a good match with the synthetics. In Figures 2.14 and 2.15 we made a more quantitative comparison by considering some selected vertical sections of Figures 2.11 and 2.12, located at, respectively, $x = 210$ m and $x = 300$ m. The section $x = 210$ m shows a very good agreement between the two strategies, especially for the parameters K and μ . However, the all-at-once approach for the section $x = 300$ m delivers a superior inversion. This is especially true for the low wavenumbers of the bulk modulus. As expected, the simultaneous inversion of all the frequencies leads to a more stable approach. This is evident for the density result and it should be considered the true difference between the two results. Indeed, the comparison of P- and S-wave velocities, for both the sections $x = 210$ m and $x = 300$ m, is more favorable than K and μ , and indicates that the two results describe the same kinematic model.

2.6 Summary and conclusions

In this paper, we introduced an inversion algorithm for inverse scattering problems. The main characteristic of this scheme is that both properties and wavefield are estimated inside a target object. This is achieved by an alternating procedure for which: (i) the properties are obtained from the optimization of data mismatch and (ii) the wavefields are updated by minimizing the wave equation residual. Each of these steps corresponds to a convex optimization problem.

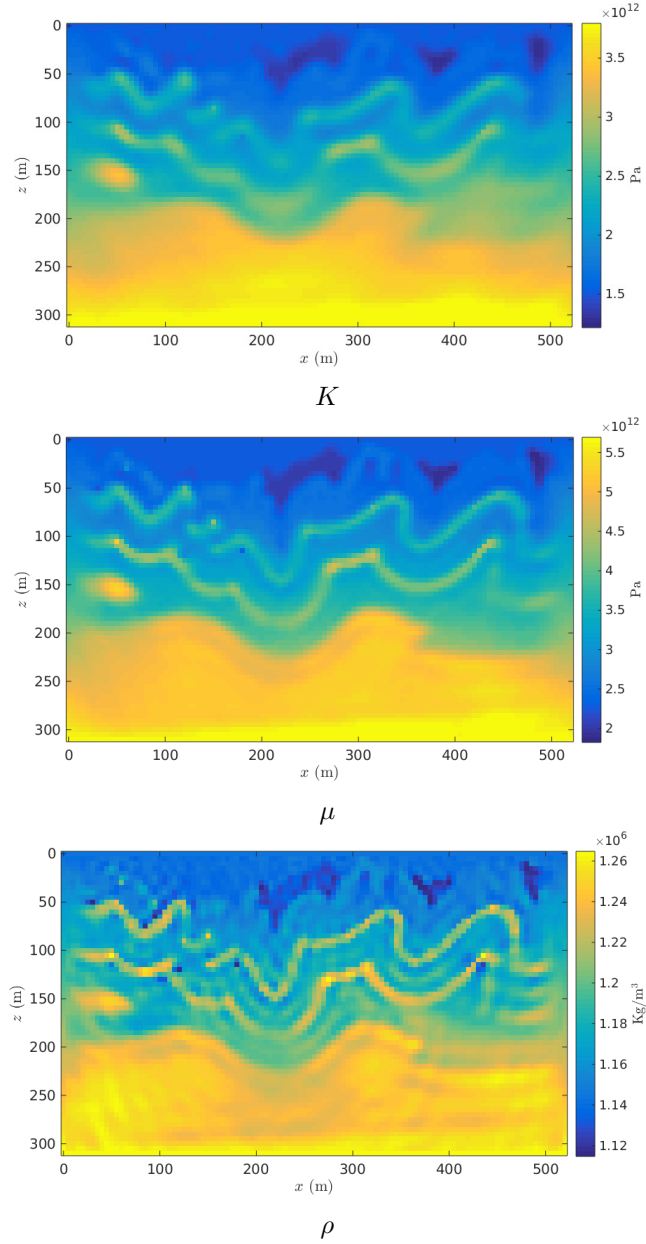


Figure 2.11: Inversion result for the strategy described in Section 2.4.4: all the data frequencies are inverted simultaneously.

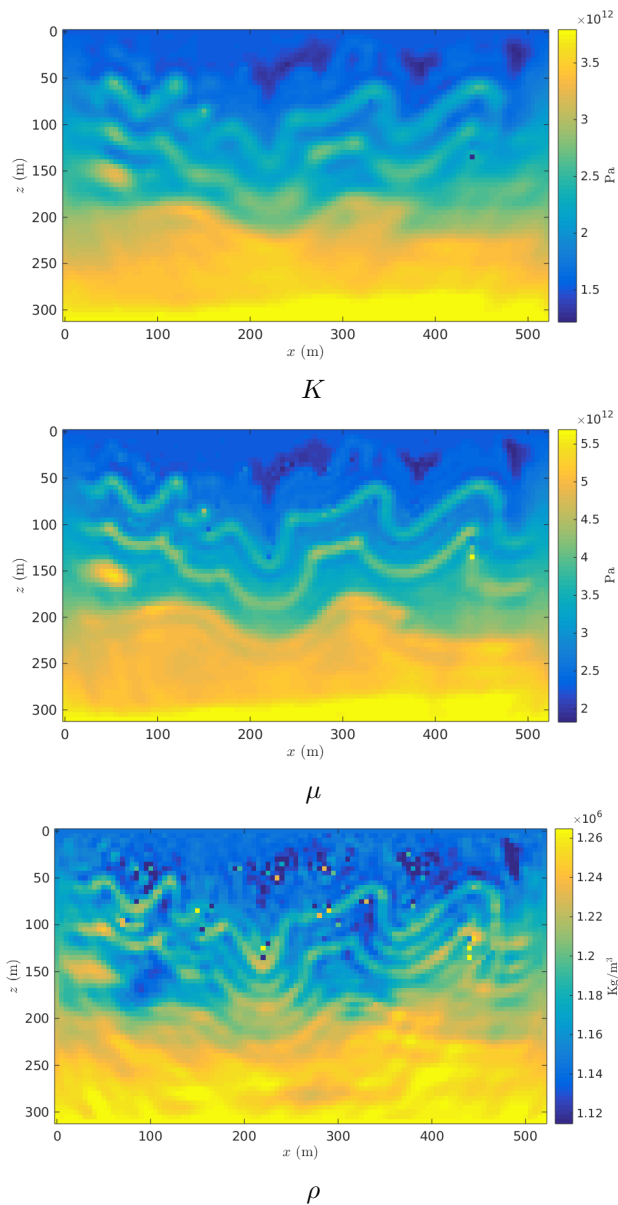


Figure 2.12: Inversion result for the strategy described in Section 2.4.4: few frequencies inverted at the time followed by background update (maximum number of simultaneously inverted frequencies: 3).

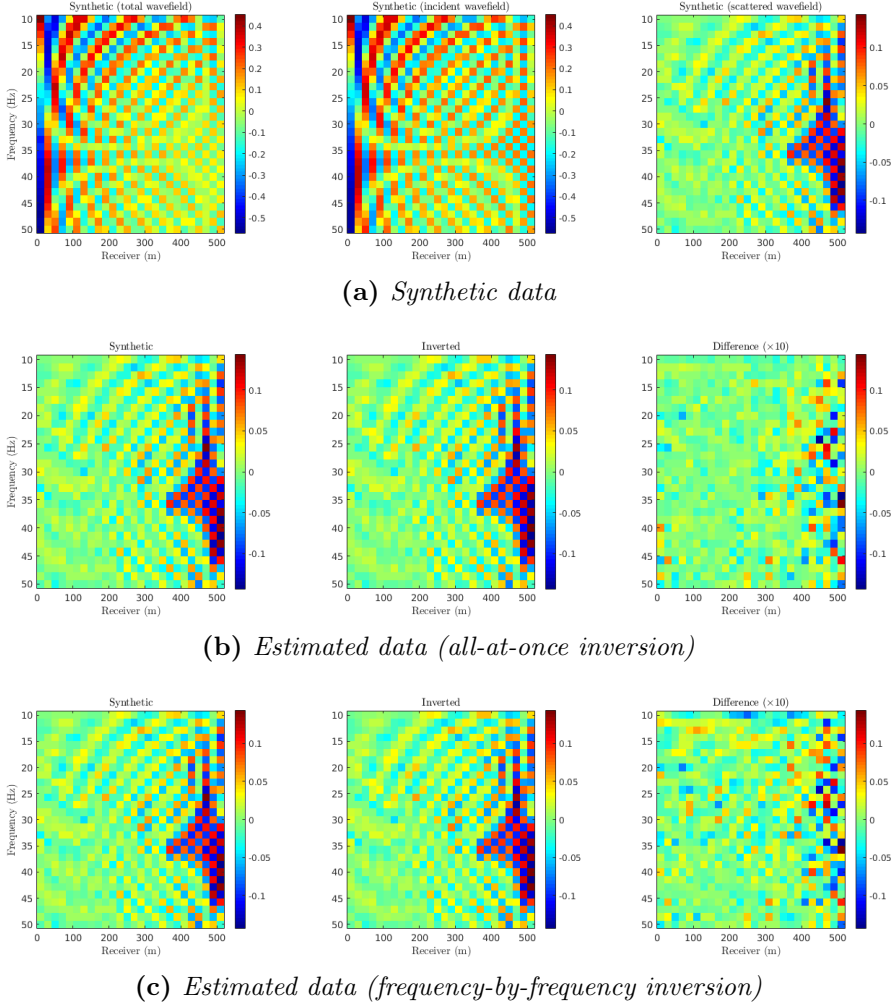


Figure 2.13: Comparison of synthetic and estimated data by the inversion: (a) real part of u^z generated by the shot located at $x = 0$ m, $z = -25$ m (the data values shown here are normalized), (b) comparison of synthetic data and inverted data for the all-at-once approach, (c) comparison of synthetic data and inverted data for the frequency-by-frequency approach.

Vertical section: $x = 210$ m

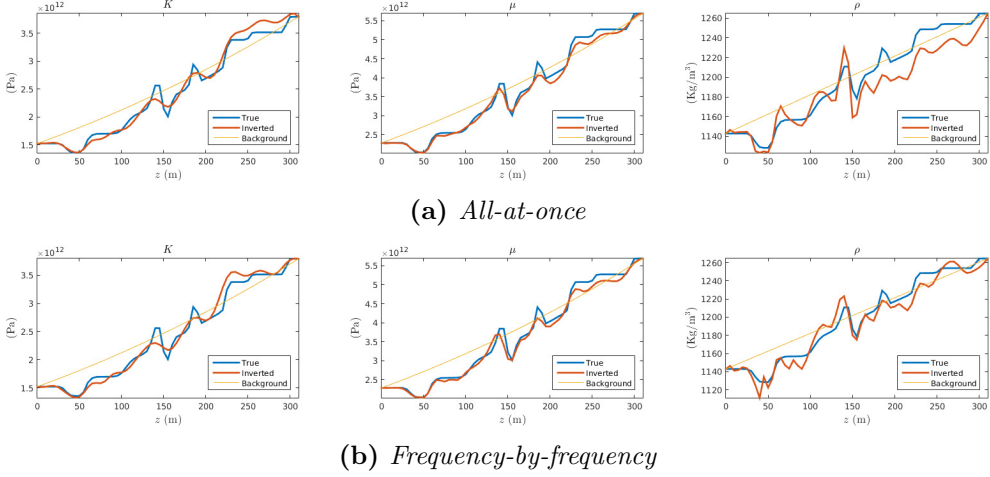


Figure 2.14: Comparison of the results using the: (a) all-at-once and (b) frequency-by-frequency approaches for the vertical section of the curly model located at $x = 210$ m.

Vertical section: $x = 300$ m

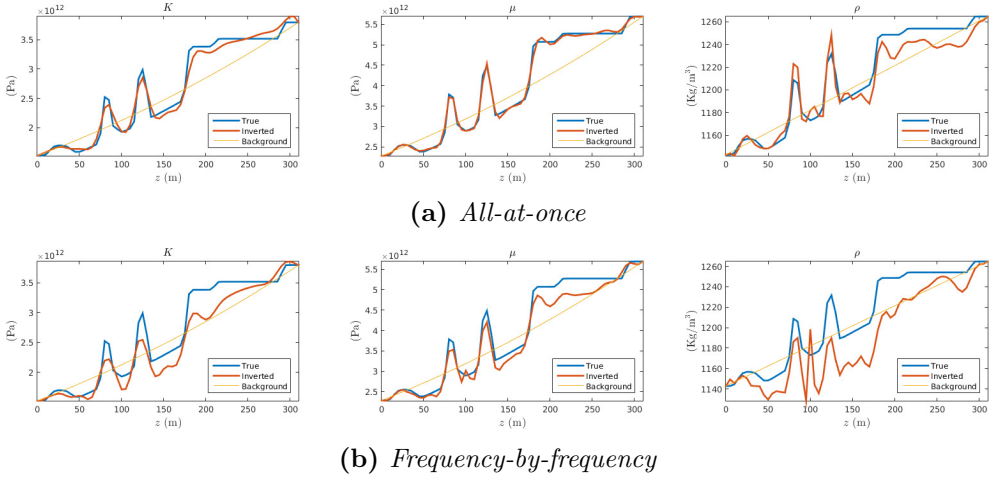


Figure 2.15: Comparison of the results using the: (a) all-at-once and (b) frequency-by-frequency approaches for the vertical section of the curly model located at $x = 300$ m.

From a computational point of view, the scheme can be quite advantageous. Indeed the data optimization step can be efficiently parallelized and the field update exploits direct methods as LU. The complexity analysis highlights a favorable comparison with the similar contrast-source method [Abubakar et al., 2009], which is also based on the alternating update of properties and wave-fields and benefits from the LU factorization to boost the field update step. Contrary to the contrast-source inversion, however, the alternating scheme focus the computational work on the property optimization step and the resulting scheme requires few field update iterations.

We tested the alternating inversion for the 2-D elastic inverse scattering problem, with successful results. We compare two inversion strategies based on an all-frequencies-at-once and frequency-by-frequency approaches, with excellent agreement between the two results. Motivated by the study of van den Berg and Abubakar [2001], we argue that the algorithm might be feasible for 3-D problems, when a frequency-by-frequency strategy is also incorporated. This motivates a future extension of the method to 3-D.

Multigrid-based 'shifted-Laplacian' preconditioning for the time-harmonic elastic wave equation[†]

Abstract

We investigate the numerical performance of an iterative solver for a frequency-domain finite-difference discretization of the isotropic elastic wave equation. The solver is based on the 'shifted-Laplacian' preconditioner, originally designed for the acoustic wave equation. This preconditioner represents a discretization of a heavily damped wave equation and can be efficiently inverted by a multigrid iteration. However, the application of multigrid to the elastic case is not straightforward because standard methods, such as point-Jacobi, fail to smooth the S-wave wavenumber components of the error when high P-to-S velocity ratios are present. We consider line smoothers as an alternative and apply local-mode analysis to evaluate the performance of the various components of the multigrid preconditioner. Numerical examples in 2-D demonstrate the efficacy of our method.

[†]This chapter is the reproduction of the paper Rizzuti and Mulder [2016], published in Computational Physics, 2016

3.1 Introduction

The numerical solution of the wave equation is an important problem in several disciplines. The main motivation for the present work is given by geophysical applications concerning the imaging of the Earth's interior. Seismic waves, excited artificially at the surface or naturally in the subsurface, are recorded to infer its properties. Since the Earth is an elastic object, the elastic wave equation is of particular interest.

Imaging typically requires many wave simulations: for 3-D problems, where one could possibly deal with hundreds of millions of variables, efficient solvers are essential. Time- or frequency-domain methods can be considered. An attractive feature of the frequency domain is the fact that only a limited number of frequencies — well below what is prescribed by the Nyquist criterion — are actually required to obtain a satisfactory imaging result [Pratt, 1990; Mulder and Plessix, 2004a]. Moreover, a further level of parallelization is available since the computation of different sources and frequencies can be carried out independently. Due to the size of the problem, however, factorization methods such as LU decomposition with nested-dissection reordering [George and Liu, 1981], which are quite efficient in 2-D, become prohibitive in terms of required storage. For this reason, explicit time-marching methods are generally preferred, since it is not necessary to solve a large linear problem. For the frequency domain, iterative methods are an obvious alternative but are seriously hindered by the severe indefiniteness of the wave equation, especially for high frequencies. Therefore, a good preconditioner must be provided in order to obtain a reasonable convergence behaviour. In that case, iterative frequency-domain solvers can actually be competitive or even outperform the time-domain counterparts when a large number of compute cores is available, as advocated in recent work [Knibbe et al., 2014].

Several alternatives to iterative methods have been investigated. For example, approximate direct methods can be an efficient option, as shown by parallel multifrontal solver developments [Xia et al., 2009; Wang et al., 2011, 2012] for both the acoustic and elastic equation. One could also consider domain-decomposition methods as worked out for the acoustic case in Stolk [2013] or Zepeda-Núñez and Demanet [2016]. An assessment of the relative merits of these methods is, however, beyond the scope of this paper.

Because of the difficulties mentioned above, the study of iterative solvers for the wave equation in the frequency domain is an active field of research.

The present paper is based on the work of Erlangga et al. [2006, 2004], who proposed preconditioning by a shifted Laplacian, solved by a multigrid iteration. This generalizes the work of Bayliss et al. [1983] and Laird and Giles [2002]. The Helmholtz equation

$$\mathbf{H} p = f, \quad \mathbf{H} = -k^2 - \Delta,$$

where k is the wavenumber and p the pressure wavefield, is replaced by a preconditioned system based on the complex shift of the Laplacian:

$$\mathbf{H} \mathbf{H}_\beta^{-1} \tilde{p} = f, \quad \mathbf{H}_\beta = -k^2 (1 - \beta \iota) - \Delta, \quad (3.1)$$

for some choice of the damping parameter $\beta > 0$. The key aspect of this preconditioning procedure relies on the inversion of the damped wave operator \mathbf{H}_β by multigrid, whose attractive feature is its linear complexity. A standard multigrid method will generally fail when applied to the undamped case because of the large near-null space of the Helmholtz operator, which cannot be approximated very well on the coarser levels [Elman et al., 2001; Tsuji and Tuminaro, 2015]. When strong damping is introduced, however, the computation of $\mathbf{H}_\beta^{-1} \tilde{p}$ can be accomplished by the multigrid method. Standard Krylov subspace methods such as GMRES [Saad, 2003], Bi-CGSTAB [van der Vorst, 1992] or IDR(s) [Sonneveld and van Gijzen, 2008] can then be successfully applied to (3.1).

In this paper, we generalize the techniques presented by Erlangga et al. [2004, 2006] to the elastic wave equation. Multigrid inversion of the damped elastic operator is, as one might expect, a harder problem than Helmholtz because of the different P- and S-wave modes. As P- and S-waves propagate with different velocities and, therefore, have different wavenumbers at a given frequency, smoothing should be tuned to perform well for both. On top of this, the near-null kernel comprises P- and S-waves and is more problematic than in the acoustic case. The value of β should be adjusted accordingly.

We organized this paper as follows: in Section 3.2 we discuss the numerical finite-difference discretization of the elastic wave equation. We will adopt two classical schemes, by Kelly et al. [1976] and by Virieux [1986]. It will be shown that these schemes, besides having different numerical dispersion quality, exhibit distinct multigrid performance. Motivated by the success of the shifted-Laplacian preconditioning for the Helmholtz equation, we introduce an elastic version in Section 3.3 and study how it affects the spectral properties of the elastic wave equation. Similarly to the acoustic case, the preconditioned

P- and S-wave eigenvalues are positioned along a circle in the positive real part of the complex plane, leading to a favourable setting for iterative methods. In Section 3.4, we will discuss the application of the multigrid technique to approximate the evaluation of this preconditioner. The performance of the various multigrid components will be analysed by local-mode analysis, briefly outlined for systems of equations. As it turns out, point-Jacobi smoother does not adequately smooth both the P- and S-wave modes of the error. We therefore introduce a suitable line smoother. This choice will be validated in combination with a multi-level multigrid analysis. Finally, in Section 3.5, we present numerical results for the homogeneous case as well as an elastic version of a highly heterogeneous example, the Marmousi model, showing good convergence properties.

3.2 Problem formulation and numerical discretization

We start this section with a brief discussion on the numerical discretization of the second-order formulation of the 2-D isotropic elastic wave equation. The unknowns will be represented by the particle displacement $\mathbf{u} = (u^x, u^z)$. In the frequency domain, the wave equation reads:

$$\begin{cases} -\rho\omega^2 u^x - \partial_x [(\lambda + 2\mu)\partial_x u^x + \lambda\partial_z u^z] - \partial_z (\mu\partial_z u^x + \mu\partial_x u^z) = f^x, \\ -\rho\omega^2 u^z - \partial_x (\mu\partial_z u^x + \mu\partial_x u^z) - \partial_z [\lambda\partial_x u^x + (\lambda + 2\mu)\partial_z u^z] = f^z, \end{cases} \quad (3.2)$$

where ω is the angular frequency, λ and μ are the Lamé parameters, ρ is the density and $\mathbf{f} = (f^x, f^z)$ a source term. This form will be used for the numerical experiments in Section 3.5. For the smoothing and multi-level multigrid analysis, we consider the homogeneous case. Then, the system (3.2) simplifies to:

$$\begin{cases} -k_S^2 u^x - r^2 \partial_{xx} u^x - \partial_{zz} u^x - (r^2 - 1) \partial_{xz} u^z = g^x, \\ -k_S^2 u^z - \partial_{xx} u^z - r^2 \partial_{zz} u^z - (r^2 - 1) \partial_{xz} u^x = g^z, \end{cases} \quad (3.3)$$

where $k_S = \omega/c_S$ is the wavenumber for the S-waves and $r = c_P/c_S > 1$ the P-to-S velocity ratio. The system (3.3) can be restated in matrix form as

$$\mathbf{H} \mathbf{u} = \mathbf{g}, \quad \mathbf{H} = -k_S^2 \mathbf{I} - \mathcal{D}, \quad \mathcal{D} = \begin{pmatrix} r^2 \partial_{xx} + \partial_{zz} & (r^2 - 1) \partial_{xz} \\ (r^2 - 1) \partial_{xz} & \partial_{xx} + r^2 \partial_{zz} \end{pmatrix}, \quad (3.4)$$

to represent the elastic Helmholtz operator. Since in the 2-D elastic case we deal with a system of equations, \mathbf{H} is a 2×2 -block operator and we specify its scalar blocks by

$$\mathbf{H} = \begin{pmatrix} H_{xx} & H_{xz} \\ H_{zx} & H_{zz} \end{pmatrix}.$$

For the smoothing and multi-level multigrid analysis treated in the following sections, it is convenient to consider an explicit formulation, analogous to (3.4), for the acoustic case with $\mu = 0$. Equation (3.4) should be adjusted by introducing the wavenumber for the P-waves $k_S = r k_P$, where $k_P = \omega/c_P$, and taking the limit $r \rightarrow +\infty$ in (3.4). The resulting elastic operator reads:

$$\mathbf{H} = -k_P^2 \mathbf{I} - \mathcal{D}, \quad \mathcal{D} = \begin{pmatrix} \partial_{xx} & \partial_{xz} \\ \partial_{xz} & \partial_{zz} \end{pmatrix}. \quad (3.5)$$

Both the forms (3.4) and (3.5) have to be considered when dealing with elastic media where acoustic regions with $\mu = 0$ are present.

Several finite-difference schemes are available to discretize the system (3.2). We will represent the unknown functions by the values assumed on the collocation grid $G^h = h\mathbb{Z}^2$ (\mathbb{Z} being the set of integers), where h is the discretization step-length or spacing, and denote a discretized quantity by $u_{k,l} \approx u(kh, lh)$. For simplicity, we will only consider isotropic grids with equal horizontal and vertical spacings. For the treatment of these numerical methods, the stencil notation is particularly convenient. For example, we will indicate a 9-point compact scheme by

$$s \triangleq \begin{bmatrix} s(-1, -1) & s(0, -1) & s(1, -1) \\ s(-1, 0) & s(0, 0) & s(1, 0) \\ s(-1, 1) & s(0, 1) & s(1, 1) \end{bmatrix},$$

meaning that the action of the operator s on a grid function is defined by the discrete cross-correlation:

$$(s \star u)_{i,j} = \sum_{m,n} s(m,n) u_{i+m,j+n}. \quad (3.6)$$

Note that the stencil is expressed with z downwards, following the geophysical rather than the numerical convention. The discretization \mathbf{H}^h of the operator \mathbf{H} , in the homogeneous case, may be represented by the four stencils:

$$\begin{aligned} \mathbf{H}_{xx}^h &\triangleq -k_S^2 - r^2 \partial_{xx}^h - \partial_{zz}^h, & \mathbf{H}_{xz}^h &\triangleq -(r^2 - 1) \partial_{xz;z}^h, \\ \mathbf{H}_{zx}^h &\triangleq -(r^2 - 1) \partial_{xz;x}^h, & \mathbf{H}_{zz}^h &\triangleq -k_S^2 - \partial_{xx}^h - r^2 \partial_{zz}^h, \end{aligned} \quad (3.7)$$

and is determined by the particular choices for the second-order differential operators ∂_{xx}^h , $\partial_{xz;x}^h$, $\partial_{xz;z}^h$, ∂_{zz}^h . A classical 9-point finite-difference scheme by Kelly et al. [1976] is summarized by

$$\begin{aligned} \partial_{xx}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}, & \partial_{xz;x}^h &\triangleq \frac{1}{4h^2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \\ \partial_{xz;z}^h &\triangleq \frac{1}{4h^2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, & \partial_{zz}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \end{aligned} \quad (3.8)$$

The discretized system reads

$$\begin{cases} -k_S^2 h^2 u_{m,n}^x - r^2 (u_{m+1,n}^x - 2u_{m,n}^x + u_{m-1,n}^x) - (u_{m,n+1}^x - 2u_{m,n}^x + u_{m,n-1}^x) \\ \quad - \frac{1}{4}(r^2 - 1)(u_{m+1,n+1}^z - u_{m+1,n-1}^z - u_{m-1,n+1}^z + u_{m-1,n-1}^z) = h^2 g_{m,n}^x \\ -k_S^2 h^2 u_{m,n}^z - (u_{m+1,n}^z - 2u_{m,n}^z + u_{m-1,n}^z) - r^2 (u_{m,n+1}^z - 2u_{m,n}^z + u_{m,n-1}^z) \\ \quad - \frac{1}{4}(r^2 - 1)(u_{m+1,n+1}^x - u_{m+1,n-1}^x - u_{m-1,n+1}^x + u_{m-1,n-1}^x) = h^2 g_{m,n}^z. \end{cases} \quad (3.9)$$

Another popular discretization scheme is based on a staggered-grid formulation, originally developed for the first-order wave equation by Virieux [1986].

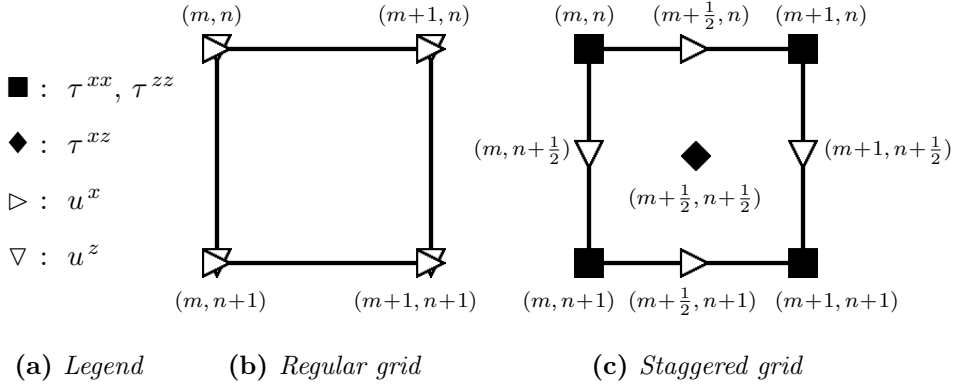


Figure 3.1: Discretization grid according to (b) Kelly's and (c) Virieux's formulation. For the staggered-grid scheme, the elastic wavefields are defined on different grids. The grid points where these wavefield are defined, are denoted by different symbols according to (a). Note that z points downwards.

According to this scheme, the particle displacement components and the stress fields

$$\tau^{ab} = \sum_{cd} C^{ab,cd} \partial_c u^d,$$

$C^{ab,cd}$ being the stiffness tensor, are discretized on different grids that are shifted with respect to G^h . More specifically, the vertical component u^z is defined on the grid $G^h + (0, h/2)$, whereas the horizontal component u^x is positioned on $G^h + (h/2, 0)$. The stress components τ^{xx} and τ^{zz} are defined on the regular grid G^h , while τ^{xz} is on $G^h + (h/2, h/2)$. This configuration, pictured in Fig. 3.1, leads to the finite-difference scheme:

$$\left\{ \begin{array}{l} -k_S^2 h^2 u_{m+\frac{1}{2},n}^x - r^2 (u_{m+\frac{3}{2},n}^x - 2u_{m+\frac{1}{2},n}^x + u_{m-\frac{1}{2},n}^x) \\ \quad - (u_{m+\frac{1}{2},n+1}^x - 2u_{m+\frac{1}{2},n}^x + u_{m+\frac{1}{2},n-1}^x) \\ - (r^2 - 1) (u_{m+1,n+\frac{1}{2}}^z - u_{m+1,n-\frac{1}{2}}^z - u_{m,n+\frac{1}{2}}^z + u_{m,n-\frac{1}{2}}^z) \\ \quad = h^2 g_{m+\frac{1}{2},n}^x \\ \\ -k_S^2 h^2 u_{m,n+\frac{1}{2}}^z - (u_{m+1,n+\frac{1}{2}}^z - 2u_{m,n+\frac{1}{2}}^z + u_{m-1,n+\frac{1}{2}}^z) \\ \quad - r^2 (u_{m,n+\frac{3}{2}}^z - 2u_{m,n+\frac{1}{2}}^z + u_{m,n-\frac{1}{2}}^z) \\ - (r^2 - 1) (u_{m+\frac{1}{2},n+1}^x - u_{m-\frac{1}{2},n+1}^x - u_{m+\frac{1}{2},n}^x + u_{m-\frac{1}{2},n}^x) \\ \quad = h^2 g_{m,n+\frac{1}{2}}^z. \end{array} \right. \quad (3.10)$$

If we define the shifted particle displacement by $\bar{u}_{m,n}^x = u_{m+1/2,n}^x$ and $\bar{u}_{m,n}^z = u_{m,n+1/2}^z$, the stencil representation (3.7) acting on \bar{u}^x and \bar{u}^z can then be summarized by

$$\begin{aligned} \partial_{xx}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}, & \partial_{xz;x}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{bmatrix}, \\ \partial_{xz;z}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, & \partial_{zz}^h &\triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \end{aligned} \quad (3.11)$$

The difference between the mixed-derivative stencils $\partial_{xz;x}^h$ and $\partial_{xz;z}^h$ arises from the definition of the particle components on different grids. In the following sections, we will drop the notation $\bar{\mathbf{u}}$ and use \mathbf{u} instead, to indicate the shifted discretized wavefield when dealing with the staggered-grid scheme.

The numerical dispersion properties of the staggered-grid scheme are significantly better than Kelly's and the S-wave phase and group velocities are independent of the velocity ratio. Indeed, the accuracy of Kelly's scheme gets worse for increasing values of r (Fig. 3.2). These numerical properties are determined by how the scheme realizes, in its discrete version, the decoupling of P- and S-wave modes in a homogeneous medium. The staggered-grid scheme does this 'consistently', meaning that the second-order stencils (3.11) can be derived from forward and backward first-order difference operators, e.g.,

$$\partial_{x+}^h \triangleq \frac{1}{h} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \partial_{x-}^h \triangleq \frac{1}{h} \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (3.12)$$

and

$$\partial_{xx}^h = \partial_{x+}^h * \partial_{x-}^h, \quad \partial_{xz;x}^h = \partial_{x-}^h * \partial_{z+}^h, \quad \partial_{xz;z}^h = \partial_{x+}^h * \partial_{z-}^h, \quad \partial_{zz}^h = \partial_{z+}^h * \partial_{z-}^h, \quad (3.13)$$

where '*' is the convolution operation. As a consequence, the P- and S-wave potentials satisfy a discretized version of the Helmholtz equation, defined by the same stencil in (3.11), and the S-wave phase velocity is therefore independent of r . Kelly's scheme does not have this property.

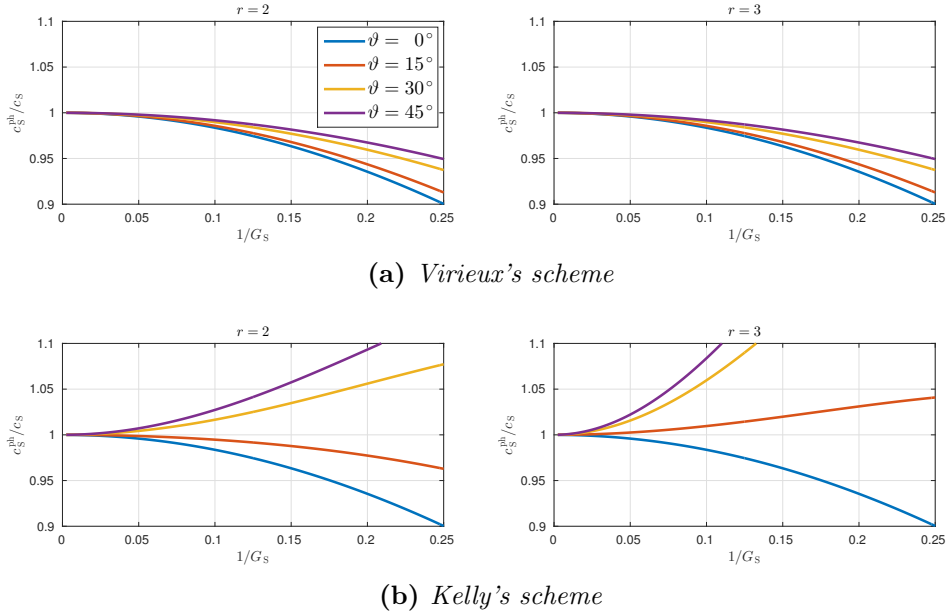


Figure 3.2: Numerical S -wave phase velocity ratios c_S^{ph}/c_S for (a) Virieux's and (b) Kelly's scheme as a function of the inverse of G_S (number of points per S -wave wavelength) and propagation angles ϑ . The numerical dispersion properties of Kelly's scheme become worse for increasing velocity ratios r , while for the staggered-grid scheme, they remain independent of r .

In principle, better options than the one presented here are available in terms of numerical accuracy. In Štekl and Pratt [1998], 9-point compact stencils are specifically designed to minimize numerical dispersion by averaging the differential operator with its rotation, along with lumped mass matrix techniques. A similar rationale has been applied by Min et al. [2000, 2003, 2004] to obtain very accurate 25-point finite-difference and finite-element schemes. The compactness of these stencils is intended to reduce the fill-in of factorization methods such as LU decomposition. It should be noted that this is not necessary when employing iterative methods and one can easily adopt higher-order versions of Kelly's and Virieux's schemes with wider stencils. This will, however, affect the behaviour of the multigrid solver.

In this paper, we will discuss multigrid with the staggered-grid scheme (3.10) in mind. Even though the 'unphysical' coupling of Kelly's scheme is more advantageous for multigrid applications, as we will see later, its use is questionable due to the inaccuracy of the scheme. If there are no acoustic

regions ($\mu = 0$) in the medium, multigrid performs comparably for the two schemes. Generally speaking, the theory and methods developed in the following sections can be straightforwardly applied to any finite-difference or finite-element discretization of the wave equation on structured meshes.

3.3 A ‘shifted-Laplacian’ preconditioner for the elastic wave equation

Given a certain discretization of the wave equation system (3.4),

$$\mathbf{H}^h \mathbf{u}^h = \mathbf{g}^h, \quad \mathbf{H}^h = -k_S^2 \mathbf{I} - \mathcal{D}^h, \quad (3.14)$$

we introduce the following preconditioning

$$\mathbf{H}^h (\mathbf{H}_\beta^h)^{-1} \tilde{\mathbf{u}}^h = \mathbf{g}^h, \quad \mathbf{H}_\beta^h = -k_S^2 (1 - \beta \iota) \mathbf{I} - \mathcal{D}^h, \quad (3.15)$$

based on the very same ideas presented by Erlangga et al. [2006, 2004]. From a physical point of view, the linear system \mathbf{H}_β^h corresponds to a damped version of the elastic wave equation. We begin a preliminary analysis of this preconditioning by testing the spectral properties of the system (3.15), which, ideally, requires the inverse of the damped system. A few iterations of multigrid will achieve only an approximation of this inverse, thus, the following analysis assumes idealized circumstances. We make use of local-mode analysis [Brandt and Livne, 2011; Trottenberg et al., 2001; Briggs et al., 2000], which will be discussed in more detail in Section 3.4.2. Essentially, it consists in a diagonal decomposition of the discrete system (3.14) by the Fourier transform and the canonical P-wave and S-wave decomposition. One can easily find that, for the staggered-grid scheme, the P- and S-wave eigenvalues of the damped system \mathbf{H}_β^h are

$$\begin{aligned} \lambda_{P,\beta}^h &= -k_S^2 (1 - \beta \iota) + \frac{4r^2}{h^2} \left[\sin^2 \left(\frac{\theta_x}{2} \right) + \sin^2 \left(\frac{\theta_z}{2} \right) \right], \\ \lambda_{S,\beta}^h &= -k_S^2 (1 - \beta \iota) + \frac{4}{h^2} \left[\sin^2 \left(\frac{\theta_x}{2} \right) + \sin^2 \left(\frac{\theta_z}{2} \right) \right], \end{aligned} \quad (3.16)$$

and, therefore, the preconditioning transforms the eigenvalues into

$$\frac{\lambda_{P,0}^h}{\lambda_{P,\beta}^h} = \frac{k_S^2 h^2 - 4r^2 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]}{k_S^2 h^2 (1 - \beta \iota) - 4r^2 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]},$$

$$\frac{\lambda_{S,0}^h}{\lambda_{S,\beta}^h} = \frac{k_S^2 h^2 - 4 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]}{k_S^2 h^2 (1 - \beta \iota) - 4 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]}.$$

Note that the second ratio matches the first in the limit for $r \downarrow 1$. These eigenvalue ratios are distributed on the complex plane along a circle with centre $1/2$ and radius $1/2$, exactly as in the acoustic case (Fig. 3.3). The same configuration arises with the Kelly’s scheme.

Similarly to the acoustic case, as β gets smaller, the eigenvalues move away from 0 and cluster around 1, as can be seen in Fig. 3.3. A clustered spectrum is a favourable setting for Krylov subspace methods. However, as it will be shown later, finding the approximate inverse of \mathbf{H}_β^h by multigrid is more efficient when β is large. Thus, the choice of β is a trade-off between the convergence rate of the multigrid solution of the damped system \mathbf{H}_β^h and the desired spectral properties of the preconditioned system (3.15). In the elastic case, the multigrid efficiency will be also be affected by the velocity ratio r . We should point out that ‘complex shifting’ or added damping for the elastic system of equations has a similar effect on the corresponding spectra as with the acoustic equation. This fact provides a case for a successful application of multigrid preconditioning to the elastic wave equation and will be validated experimentally in Section 3.5. There, we will also consider the effect of the inexact multigrid approximation of the preconditioner. Note that our approach is mainly empirical. To the best of our knowledge, a rigorous theory for the convergence of shifted-Laplacian methods is presently lacking.

Recently, some research has been devoted to combining multigrid with deflation-based preconditioning [Nicolaidis, 1987; Frank and Vuik, 2001]. The basic idea is to design a further preconditioner which clusters the eigenvalues in Fig. 3.3 towards 1. We refer to Erlangga and Nabben [2008] and Sheik et

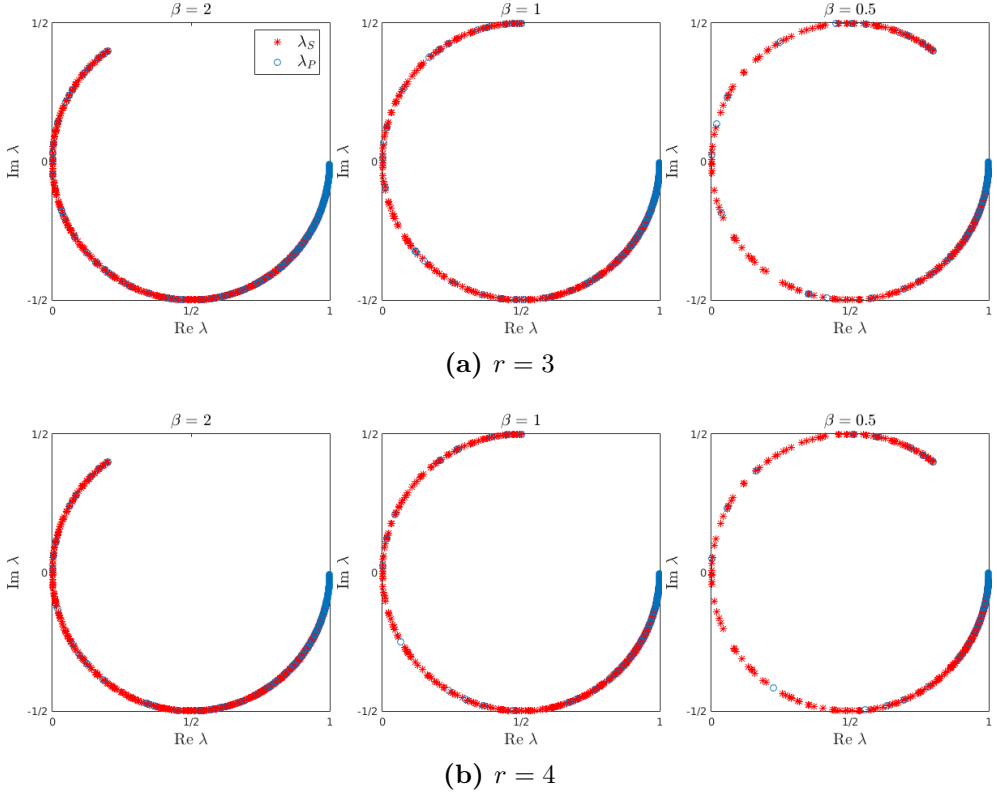


Figure 3.3: Eigenvalue distribution $\lambda_{P,0}^h/\lambda_{P,\beta}^h$, $\lambda_{S,0}^h/\lambda_{S,\beta}^h$ of the preconditioned system (3.15) for different values of wave damping β and velocity ratio: (a) $r = 3$, (b) $r = 4$. The quantity $k_S h$ is fixed to the value $\pi/5$ which corresponds to a sampling of 10 points per wavelength. The plots here displayed correspond to the staggered-grid scheme, but an analogous behaviour can be observed for the Kelly's scheme. Note that for the staggered-grid scheme, the S-wave eigenvalues are independent of r .

al. [2013] for a discussion on this matter. We expect that the same idea can be applied successfully to the elastic case, but in this paper we will focus solely on the multigrid preconditioning (3.15).

3.4 Analysis of multigrid applied to the damped elastic wave equation

In the previous section, we studied preconditioning by the damped wave equation and assumed that the associated linear system could be solved exactly.

Since we intend to obtain only an approximation of its solution by multigrid iterations, the previous analysis should be adapted accordingly. In this section, we will focus on the application of multigrid to the damped wave equation. The starting point is the previous work of Erlangga et al. [2004, 2006], valid for the acoustic case. In order to assess the performance of multigrid, the classical tool of local-mode analysis was considered in Erlangga et al. [2004, 2006]. For the elastic case, we will have to apply this method to a linear system, for which the theory can be found in, e.g., Trottenberg et al. [2001]. To validate the predictions of the following local-mode analysis, we will compare them to numerical experiments.

Multigrid is a well-known numerical technique for solving certain classes of partial differential equations [e.g Briggs et al., 2000; Trottenberg et al., 2001]. It exploits the multiscale behaviour of classical solvers by considering a hierarchy of discretization grids. Multigrid relies on two fundamental ingredients: smoothing and a coarse-grid correction. Given a linear system \mathbf{H}^h , which might arise from the discretization of a partial differential equation on the grid G^h , and starting from an approximation \mathbf{u}_0^h of the solution \mathbf{u}^h , smoothing consists in reducing the high-frequency or short-wavelength components of the error $\mathbf{e}_0^h = \mathbf{u}^h - \mathbf{u}_0^h$. This is a typical trait of many elementary iterative methods, such as Jacobi and Gauss-Seidel, that can easily remove high-frequency solution errors but have difficulty with the low-frequency components. However, once a new approximation $\tilde{\mathbf{u}}_0^h$ is computed for which the error $\tilde{\mathbf{e}}_0^h$ is smooth and satisfies the error equation $\mathbf{H}^h \tilde{\mathbf{e}}_0^h = \tilde{\mathbf{r}}_0^h$ (where $\tilde{\mathbf{r}}_0^h$ is the corresponding residual), the fine-grid problem can be approximated on a coarser grid: $\mathbf{H}^{2h} \tilde{\mathbf{e}}_0^{2h} = \tilde{\mathbf{r}}_0^{2h}$. The solution of the coarse-grid problem is projected or prolonged back to the fine grid to give the so-called coarse-grid correction. Since ‘low frequencies’ on the fine grid become ‘high frequencies’ on the coarse grid after restriction, the same idea can be exploited recursively to solve for the coarse-grid problem.

3.4.1 Multigrid for the acoustic wave equation

Multigrid is typically effective for positive definite linear systems, for example those arising from the discretization of elliptic partial differential equations, and it is known to behave poorly for the undamped wave equation. The reasons are discussed in some detail in Elman et al. [2001] and Erlangga et al. [2006] and the same conclusions hold for the elastic case. Basically, classical smoothers like point-Jacobi will severely amplify the long wavelength com-

ponents of the error on ‘intermediate’ grid levels, effectively hampering the performance of multigrid. Another issue concerns the coarse-grid correction: some eigenvalues corresponding to a certain discretization level h might undergo a change of sign on the coarser grid G^{2h} . It can be shown, as in Elman et al. [2001], that the coarse-grid correction corresponding to this particular component will be updated in the ‘wrong’ direction. Multigrid is thus unable to solve for the space associated with these eigenvalues. When damping with β is introduced, the eigenvalues of \mathbf{H}_β^h are shifted into the complex plane and multigrid can be effectively applied [Erlangga et al., 2006].

3.4.2 Multigrid for the elastic wave equation

In this section, we will show that multigrid can as well be applied to the damped elastic wave equation. We start with a brief introduction on local-mode analysis — the main quantitative tool we will use to assess the numerical behaviour of multigrid. We then proceed by discussing the performance of point- and line-Jacobi smoothers and finally by comparing estimated convergence factors from n -grid analysis to the values obtained by numerical experiments.

Introduction on local-mode analysis for the elastic wave equation

Local-mode analysis is a predictive tool that can be used to effectively measure multigrid performance. In essence, it is the spectral decomposition of the linear operators involved on the space of scalar Fourier grid functions $\varphi(\boldsymbol{\theta})$ defined by

$$\varphi(\boldsymbol{\theta})_{m,n} = \exp[i(\theta_x m + \theta_z n)], \quad \boldsymbol{\theta} = (\theta_x, \theta_z) \in [-\pi, \pi]^2. \quad (3.17)$$

For simplicity, we will only treat the 2-D case. This decomposition is useful in estimating how the various Fourier components of the error of a particular estimate of the solution are reduced after each operation that is part of a multigrid cycle. Since linear operators like \mathbf{H}_β^h are defined by a spatially compact stencil, and its action involves ‘local’ operations, the mode analysis is arguably still valid by setting homogeneous physical parameters in the definition of the discretized linear operators and considering unbounded grids G^h or periodic boundary conditions. The discussion that follows is made under these assumptions. We refer to Trottenberg et al. [2001] for a more extensive introduction on local-mode analysis of linear systems. In this section,

we provide a succinct and formal definition of smoothing and n -grid analysis [Wienands and Oosterlee, 2001] with emphasis on the wave equation and its discretization by the staggered-grid scheme (3.11).

We start with a formal Fourier analysis of the elastic wave equation operator \mathbf{H}_β^h acting on the space of grid function pairs $\mathbb{L}^2(\mathbf{G}^h) \times \mathbb{L}^2(\mathbf{G}^h)$, where $\mathbb{L}^2(\mathbf{G}^h)$ is the set of square-summable functions. Given a vector $\mathbf{u}_0 \in \mathbb{C}^2$, we define the Fourier symbol $\hat{\mathbf{H}}_\beta^h$ as:

$$\mathbf{H}_\beta^h [\varphi(\boldsymbol{\theta}) \mathbf{u}_0] = \hat{\mathbf{H}}_\beta^h(\boldsymbol{\theta}) \mathbf{u}_0, \quad (3.18)$$

which, in other words, is the ‘eigenvalue matrix’ of the Fourier grid function. Since \mathbf{H}_β^h is a 2×2 -block matrix, $\hat{\mathbf{H}}_\beta^h(\boldsymbol{\theta})$ is a 2×2 matrix. The eigenvalues of $\hat{\mathbf{H}}_\beta^h$ can be found by a standard P- and S-wave diagonal decomposition:

$$\hat{\mathbf{H}}_\beta^h(\boldsymbol{\theta}) = \hat{\mathbf{M}}^h(\boldsymbol{\theta}) \hat{\boldsymbol{\Lambda}}_\beta^h(\boldsymbol{\theta}) \hat{\mathbf{M}}^h(\boldsymbol{\theta})^*, \quad (3.19)$$

where the asterisk denotes conjugate transposition. The matrix $\hat{\mathbf{M}}^h(\boldsymbol{\theta}) = [\hat{\mathbf{u}}_P^h(\boldsymbol{\theta}), \hat{\mathbf{u}}_S^h(\boldsymbol{\theta})]$ collects the P- and S-wave eigenvectors, while $\hat{\boldsymbol{\Lambda}}_\beta^h(\boldsymbol{\theta})$ is the eigenvalue diagonal matrix $\text{diag}([\lambda_{P,\beta}^h(\boldsymbol{\theta}), \lambda_{S,\beta}^h(\boldsymbol{\theta})])$. As already noted in Section 3.2, the wave potentials of the staggered-grid scheme are defined by a discretization of the P- and S-wave Helmholtz equation, which is consistent with the stencils (3.11). Their Fourier transforms are

$$\begin{aligned} \hat{\partial}_{xx}^h(\boldsymbol{\theta}) &= -\frac{4}{h^2} \sin\left(\frac{\theta_x}{2}\right)^2, \\ \hat{\partial}_{xz;x}^h(\boldsymbol{\theta}) &= -\frac{4}{h^2} e^{i(-\theta_x + \theta_z)/2} \sin\left(\frac{\theta_x}{2}\right) \sin\left(\frac{\theta_z}{2}\right), \\ \hat{\partial}_{xz;z}^h(\boldsymbol{\theta}) &= -\frac{4}{h^2} e^{i(\theta_x - \theta_z)/2} \sin\left(\frac{\theta_x}{2}\right) \sin\left(\frac{\theta_z}{2}\right), \\ \hat{\partial}_{zz}^h(\boldsymbol{\theta}) &= -\frac{4}{h^2} \sin\left(\frac{\theta_z}{2}\right)^2. \end{aligned}$$

It is easily found that the eigenvectors associated with the eigenvalues $\lambda_{P,\beta}^h$,

$\lambda_{S,\beta}^h$ in (3.16), apart from scaling factors, are

$$\begin{aligned}\hat{\mathbf{u}}_P^h(\boldsymbol{\theta}) &= \frac{2\iota}{h} \left[e^{-\iota\theta_x/2} \sin\left(\frac{\theta_x}{2}\right), e^{-\iota\theta_z/2} \sin\left(\frac{\theta_z}{2}\right) \right], \\ \hat{\mathbf{u}}_S^h(\boldsymbol{\theta}) &= \frac{2\iota}{h} \left[-e^{\iota\theta_z/2} \sin\left(\frac{\theta_z}{2}\right), e^{\iota\theta_x/2} \sin\left(\frac{\theta_x}{2}\right) \right].\end{aligned}\quad (3.20)$$

In the expressions of $\hat{\mathbf{u}}_P^h$ and $\hat{\mathbf{u}}_S^h$, one may recognize the Fourier symbols of the gradient and curl operators, discretized by the centred schemes (3.12) underlying the stencils (3.11):

$$\begin{aligned}\hat{\mathbf{u}}_P^h(\boldsymbol{\theta}) &= \left[\hat{\partial}_{x-}^h(\boldsymbol{\theta}), \hat{\partial}_{z-}^h(\boldsymbol{\theta}) \right] \approx \hat{\nabla}(\boldsymbol{\theta}), \\ \hat{\mathbf{u}}_S^h(\boldsymbol{\theta}) &= \left[-\hat{\partial}_{z+}^h(\boldsymbol{\theta}), \hat{\partial}_{x+}^h(\boldsymbol{\theta}) \right] \approx \hat{\text{curl}}(\boldsymbol{\theta}).\end{aligned}\quad (3.21)$$

The Fourier analysis of the operator \mathbf{H}_β^h allows us to analyse the properties of many smoothing schemes. A simple form of smoothing, which proved to be effective for the damped acoustic wave equation, is given by the classical Jacobi scheme. Given a certain initial guess \mathbf{u}_0^h of the solution of $\mathbf{H}_\beta^h \mathbf{u}^h = \mathbf{f}^h$, the Jacobi iteration consists of the update

$$\mathbf{u}_1^h = \mathbf{u}_0^h + \eta (\mathbf{D}_\beta^h)^{-1} \mathbf{r}_0^h, \quad \mathbf{r}_0^h = \mathbf{f}^h - \mathbf{H}_\beta^h \mathbf{u}_0^h, \quad (3.22)$$

where \mathbf{D}_β^h is the diagonal of \mathbf{H}_β^h and η a relaxation parameter that needs to be adjusted for effective smoothing. To understand how the initial error $\mathbf{e}_0^h = \mathbf{u}_0^h - \mathbf{u}^h$ is reduced, we write

$$\mathbf{e}_1^h = \mathbf{S}^h \mathbf{e}_0^h, \quad \mathbf{S}^h = \mathbf{I}^h - \eta (\mathbf{D}_\beta^h)^{-1} \mathbf{H}_\beta^h, \quad (3.23)$$

where \mathbf{S}^h is the point-Jacobi iteration matrix. The local-mode analysis of \mathbf{S}^h provides a theoretical estimate of how the different wavelength components of the error are damped by the smoothing. Since smoothing is designed to solve for the short-wavelength components, corresponding to

$$|\theta_x| > \frac{\pi}{2} \quad \text{or} \quad |\theta_z| > \frac{\pi}{2},$$

we define the smoothing factor as:

$$\mu_s^h = \max \left\{ |\lambda| : \lambda \text{ eigenvalue of } \hat{\mathbf{S}}^h(\boldsymbol{\theta}), \quad |\theta_x| > \frac{\pi}{2} \quad \text{or} \quad |\theta_z| > \frac{\pi}{2} \right\}. \quad (3.24)$$

Clearly, besides smoothing, a comprehensive analysis of multigrid requires the study of the coarse-grid correction over many grid levels. Any number n of grid levels could be considered, but usually, $n \leq 3$ suffices to analyze multigrid schemes. For the elastic wave equation, we will consider $n = 4$. For ease of exposition, we will summarize the theory of n -grid analysis for the simple two-level case, $n = 2$. We refer to Wienands and Oosterlee [2001] for a more general treatment.

The operations of two-level multigrid are

$$\begin{aligned}
 \mathbf{u}_1^h &= \mathbf{u}_0^h + \eta (\mathbf{D}_\beta^h)^{-1} \mathbf{r}_0^h && \text{(pre-smoothing, e.g., Jacobi),} \\
 \mathbf{r}_1^{2h} &= \mathbf{R}_h^{2h} \mathbf{r}_1^h && \text{(restriction of the residual to the coarse grid),} \\
 \mathbf{H}_\beta^{2h} \mathbf{e}_1^{2h} &= \mathbf{r}_1^{2h} && \text{(solution of the coarse-grid error equation),} \\
 \tilde{\mathbf{e}}_1^h &= \mathbf{P}_{2h}^h \mathbf{e}_1^{2h} && \text{(interpolation of the error onto the fine grid),} \\
 \mathbf{u}_2^h &= \mathbf{u}_1^h + \tilde{\mathbf{e}}_1^h && \text{(coarse-grid correction),} \\
 \mathbf{u}_3^h &= \mathbf{u}_2^h + \eta (\mathbf{D}_\beta^h)^{-1} \mathbf{r}_2^h && \text{(post-smoothing, e.g., Jacobi).}
 \end{aligned}$$

The grid transfer of the functions of $\mathbb{L}^2(\mathbf{G}^h)$ and $\mathbb{L}^2(\mathbf{G}^{2h})$ is managed by the restriction operator \mathbf{R}_h^{2h} and prolongation operator \mathbf{P}_{2h}^h . A natural choice would be

$$\mathbf{R}_h^{2h} = \begin{pmatrix} \mathbf{R}_h^{2h} & 0 \\ 0 & \mathbf{R}_h^{2h} \end{pmatrix}, \quad \mathbf{P}_{2h}^h = \begin{pmatrix} \mathbf{P}_{2h}^h & 0 \\ 0 & \mathbf{P}_{2h}^h \end{pmatrix}, \quad (3.25)$$

where \mathbf{R}_h^{2h} and \mathbf{P}_{2h}^h correspond to scalar restriction and prolongation operators. We assumed standard coarsening. The coarse-grid linear system \mathbf{H}_β^{2h} can be obtained from a direct discretization of the wave equation on the grid \mathbf{G}^{2h} or from the Galerkin condition

$$\mathbf{H}_\beta^{2h} = \mathbf{R}_h^{2h} \mathbf{H}_\beta^h \mathbf{P}_{2h}^h. \quad (3.26)$$

When the restriction and prolongation operators are chosen as in (3.25), it corresponds to a coarse-grid block operator where each block satisfies the scalar Galerkin condition, e.g., $\mathbf{H}_{xx,\beta}^{2h} = \mathbf{R}_h^{2h} \mathbf{H}_{xx,\beta}^h \mathbf{P}_{2h}^h$ for the first block of \mathbf{H}_β^{2h} . Through the two-level multigrid algorithm, the error is transformed

according to

$$\begin{aligned} \mathbf{e}_3^h &= \mathbf{T}^{h,2h} \mathbf{e}_0^h, & \mathbf{T}^{h,2h} &= (\mathbf{S}^h)^{\nu_2} \mathbf{K}^{h,2h} (\mathbf{S}^h)^{\nu_1}, \\ \mathbf{K}^{h,2h} &= \mathbf{I}^h - \mathbf{P}_{2h}^h (\mathbf{H}_\beta^{2h})^{-1} \mathbf{R}_h^{2h} \mathbf{H}^h, \end{aligned}$$

where ν_1 and ν_2 are, respectively, the number of pre and post-smoothing iterations. In the smoothing analysis, we exploit the diagonalization of the operator \mathbf{S}^h over the Fourier grid functions. The complication of the two-level analysis stems from the fact that the coarse-grid operator $\mathbf{K}^{h,2h}$ mixes wavelength components and, therefore, does not diagonalize over the same basis. However, it admits a block diagonal decomposition over the direct sum of the following 8-dimensional subspaces

$$F_2[\boldsymbol{\theta}_0] = \text{span} \left\{ \varphi(\boldsymbol{\theta}) \mathbf{u}_0 : \boldsymbol{\theta} \equiv \boldsymbol{\theta}_0 \pmod{\pi}, \mathbf{u}_0 \in \mathbb{C}^2 \right\}, \quad (3.27)$$

parametrized by the π modulo classes $[\boldsymbol{\theta}_0]$, where $\boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^2$. For a 2-D analysis, these classes consist of four elements:

$$[\boldsymbol{\theta}_0] = \left\{ \boldsymbol{\theta}_0^0 = \boldsymbol{\theta}_0, \boldsymbol{\theta}_0^1, \boldsymbol{\theta}_0^2, \boldsymbol{\theta}_0^3 \right\}.$$

It can be easily proven that $F_2[\boldsymbol{\theta}_0]$ is an invariant subspace for the two-grid operator $\mathbf{T}^{h,2h}$, and the spectral analysis of two-grid multigrid boils down to the matrix representation:

$$\mathbf{T}^{h,2h} \varphi(\boldsymbol{\theta}_0^l) \mathbf{u}_0 = \sum_{k=0}^3 \hat{\mathbf{T}}^{h,2h}(\boldsymbol{\theta}_0^k, \boldsymbol{\theta}_0^l) \varphi(\boldsymbol{\theta}_0^k) \mathbf{u}_0. \quad (3.28)$$

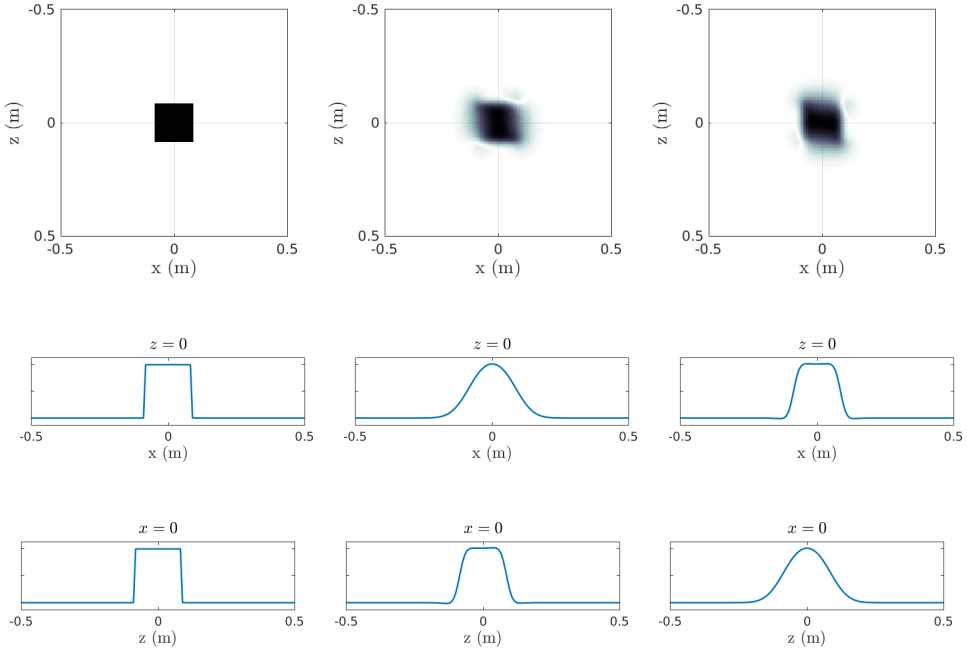
The 4×4 2×2 -block matrix $\hat{\mathbf{T}}^{h,2h}[\boldsymbol{\theta}_0]_{k,l} = \hat{\mathbf{T}}^{h,2h}(\boldsymbol{\theta}_0^k, \boldsymbol{\theta}_0^l)$ is obviously equivalent to an 8×8 matrix. Not only different wavelength components are inter-mixed by the two-grid operator, as in the acoustic (scalar) case, but also P- and S-wave modes. The two-grid convergence factor is then defined by

$$\mu_2^h = \max \left\{ |\lambda| : \lambda \text{ eigenvalue of } \hat{\mathbf{T}}^{h,2h}[\boldsymbol{\theta}_0], \boldsymbol{\theta}_0 \in [-\pi/2, \pi/2]^2 \right\}. \quad (3.29)$$

By induction, the same kind of analysis can be generalized to define n -grid convergence factors μ_n^h [Wienands and Oosterlee, 2001].

Point- and line-relaxation schemes

In the previous section, we discussed smoothing analysis by considering the elementary point-Jacobi smoothing. As already mentioned, this is an effective choice in the context of shifted-Laplacian multigrid for the acoustic case. However, the simple numerical example in Fig. 3.4 shows that point-Jacobi is an anisotropic smoother for the elastic case. This observation can be con-



(a) Initial guess $u^x = u^z$ (b) Smoothing result for u^x (c) Smoothing result for u^z

Figure 3.4: Effect of point-Jacobi smoothing applied to the elastic case for the homogeneous problem $\mathbf{H}_\beta^h \mathbf{u}^h = \mathbf{0}$ for $\beta = 2$, $k_S h = \pi/10$ (corresponding to 20 points per wavelength) and $r = 3$. The damping factor is $\eta = 0.5$. The initial guess \mathbf{u}_0^h , pictured in (a), consists of a unit function whose support is a square set, for both components. The results after 100 iterations show that point-Jacobi is not effectively smoothing in the vertical and horizontal direction, respectively, for the (b) horizontal and (c) vertical particle displacement components.

firmed theoretically by applying the techniques of the previous section. First,

we denote the diagonal operator $\mathbf{D}_\beta^h = \text{diag } \mathbf{H}_\beta^h$, in stencil notation, by

$$\mathbf{D}_{xx,\beta}^h \triangleq \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{H}_{xx,\beta}^h(0,0) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{D}_{zz,\beta}^h \triangleq \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathbf{H}_{zz,\beta}^h(0,0) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (3.30)$$

and $\mathbf{D}_{xz,\beta}^h = \mathbf{D}_{zx,\beta}^h \triangleq \mathbf{0}$, where $\mathbf{H}_{xx,\beta}^h(0,0) = \mathbf{H}_{zz,\beta}^h(0,0) = -k_S^2(1 - \beta\iota) + 2(r^2 + 1)/h^2$. We then apply the P- and S-wave decomposition to the smoothing operator $\hat{\mathbf{S}}^h$ in (3.23). Its eigenvalues are

$$\lambda_{\mathbf{P},\beta}^{s,h}(\boldsymbol{\theta}) = 1 - \eta \frac{k_S^2 h^2 (1 - \beta\iota) - 4r^2 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]}{k_S^2 h^2 (1 - \beta\iota) - 2(r^2 + 1)},$$

$$\lambda_{\mathbf{S},\beta}^{s,h}(\boldsymbol{\theta}) = 1 - \eta \frac{k_S^2 h^2 (1 - \beta\iota) - 4 \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right]}{k_S^2 h^2 (1 - \beta\iota) - 2(r^2 + 1)}. \quad (3.31)$$

In Fig. 3.5, we depict these factors as a function of the Fourier component

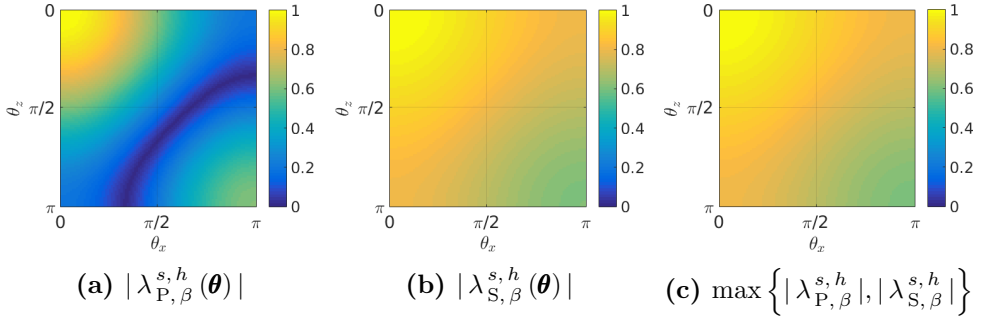


Figure 3.5: Smoothing factors of point-Jacobi for the elastic wave equation as a function of the Fourier component relative to $\boldsymbol{\theta}$ for $\beta = 2$, $k_S h = \pi/10$ (corresponding to 20 points per wavelength), $r = 2$ and $\eta = 0.5$.

relative to $\boldsymbol{\theta}$. From (3.31), it is easily found that

$$\lambda_{\mathbf{P},\beta}^{s,h}(\boldsymbol{\theta}) \approx 1 - 2\eta \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 \right], \quad \lambda_{\mathbf{S},\beta}^{s,h}(\boldsymbol{\theta}) \approx 1, \quad \text{as } r \rightarrow +\infty. \quad (3.32)$$

As evident from (3.32), for large values of the velocity ratio, η can be tuned to only smooth P-wave modes, but the S-wave eigenvalue remains equal to 1 regardless of θ and the relaxation factor η . Recalling the identity (3.21) and considering an initial error decomposed over the P- and S-wave eigenbasis (3.20) as $\hat{\mathbf{e}}_0^h(\theta) = a \hat{\mathbf{u}}_P^h(\theta) + b \hat{\mathbf{u}}_S^h(\theta)$, point-Jacobi transforms the error into

$$\begin{aligned} \hat{\mathbf{e}}_1^h(\theta) &= [\hat{\mathbf{S}}^h(\theta)]^\nu \hat{\mathbf{e}}_0^h(\theta) \\ &= a [\lambda_{P,\beta}^{s,h}(\theta)]^\nu \hat{\mathbf{u}}_P^h(\theta) + b [\lambda_{S,\beta}^{s,h}(\theta)]^\nu \hat{\mathbf{u}}_S^h(\theta) \approx b \hat{\mathbf{u}}_S^h(\theta), \end{aligned}$$

for a sufficiently large number of smoothing steps ν . Because of the inadequate smoothing of the S-wave components for high velocity ratios displayed by equation (3.32), only the S-wave mode component of $\hat{\mathbf{e}}_1^h(\theta)$ survives. From the expression of $\hat{\mathbf{u}}_S^h(\theta)$ in (3.20), we readily observe that the horizontal component of $\hat{\mathbf{e}}_1^h(\theta)$ is effectively damped for $\theta_z \approx 0$, and the vertical component for $\theta_x \approx 0$. Thus, the horizontal and vertical components of $\hat{\mathbf{e}}_1^h(\theta)$ will show a rough profile along, respectively, the vertical and horizontal directions (as in Fig. 3.4).

We showed that point-Jacobi fails to solve for the short-wavelength components of the S-wave mode of the error, and this results in smoothing anisotropy. This is a well-known issue when, for example, dealing with regular grids defined by different horizontal and vertical step-lengths: $G^{h_x, h_z} = h_x \mathbb{Z} \times h_z \mathbb{Z}$. Operators like anisotropic Helmholtz, (e.g., each of the blocks along the diagonal of \mathbf{H}_β in (3.4)) are equivalent to the Helmholtz equation defined on an anisotropic grid, and therefore display the same behaviour. In the literature, the standard solutions for this type of shortcomings are: (i) semi-coarsening, (ii) line smoothing. Semi-coarsening, for example along the horizontal direction x , consists of the discretization hierarchy: $G^{h_x, h_z} \rightarrow G^{2h_x, h_z} \rightarrow G^{4h_x, h_z} \rightarrow \dots$. Semi-coarsening can also be applied simultaneously in multiple directions [Mulder, 1989]. Even though this is, in principle, a valid approach, in this paper we stick to the second option of line smoothing.

A line smoother for the elastic wave equation Line smoothing is based on the idea of treating the unknowns of the problem as a set of ‘block variables’, each block corresponding to the grid points positioned along a row or column of the grid G^h . The operators involved are then reinterpreted as acting on these variable sets as a block, rather than on individual grid points. This point of view concerns the way the inverse of an operator is approximated. For example, when the interaction between variables is completely neglected,

namely when each blocks corresponds to a single grid-point, it leads to the approximation of the inverse of a linear system \mathbf{H}_β^h by its diagonal inverse $(\mathbf{D}_\beta^h)^{-1}$, as in point-Jacobi smoothing (3.23). Based on the observations made about the anisotropy of point-Jacobi, we treat the wavefield variables \mathbf{u} in the following way: the horizontal component u^x is considered as a collection of columns, while the vertical component u^z as a set of rows. The resulting block-diagonal operator of \mathbf{H}_β^h can then be expressed in stencil notation as

$$\begin{aligned} \mathbf{D}_{xx,\beta}^h &\triangleq \begin{bmatrix} 0 & 0 & 0 \\ \mathbf{H}_{xx,\beta}^h(-1, 0) & \mathbf{H}_{xx,\beta}^h(0, 0) & \mathbf{H}_{xx,\beta}^h(1, 0) \\ 0 & 0 & 0 \end{bmatrix}, \\ \mathbf{D}_{zz,\beta}^h &\triangleq \begin{bmatrix} 0 & \mathbf{H}_{zz,\beta}^h(0, -1) & 0 \\ 0 & \mathbf{H}_{zz,\beta}^h(0, 0) & 0 \\ 0 & \mathbf{H}_{zz,\beta}^h(0, 1) & 0 \end{bmatrix}, \end{aligned} \quad (3.33)$$

and $\mathbf{D}_{xz,\beta}^h = \mathbf{D}_{zx,\beta}^h \triangleq \mathbf{0}$. Line-Jacobi smoothing is then defined exactly as in (3.22), by replacing the point-wise diagonal operator with its line-wise counterpart (3.33). For the staggered-grid scheme, its Fourier symbol reads

$$\begin{aligned} \hat{\mathbf{D}}_{xx,\beta}^h(\boldsymbol{\theta}) &= -k_S^2(1 - \beta\iota) + \frac{4r^2}{h^2} \sin\left(\frac{\theta_x}{2}\right)^2 + \frac{2}{h^2}, \\ \hat{\mathbf{D}}_{zz,\beta}^h(\boldsymbol{\theta}) &= -k_S^2(1 - \beta\iota) + \frac{4r^2}{h^2} \sin\left(\frac{\theta_z}{2}\right)^2 + \frac{2}{h^2}. \end{aligned} \quad (3.34)$$

We should observe that the line-smoothing operator does not diagonalize over the same basis of gradient and curl operators (3.20), as in the case of the point smoother, meaning that the P- and S-wave components of the error are mixed. The components relative to this new eigenbasis are damped according to the eigenvalues $\lambda_{1,\beta}$, $\lambda_{2,\beta}$ shown in Fig. 3.6. The expressions for these eigenvalues and eigenvectors are lengthy and therefore omitted. A comparison of the smoothing factors in Fig. 3.5c and Fig. 3.6c demonstrates the effectiveness of this scheme: indeed, the smoothing factors of point-Jacobi take values close to 1 on the short wavelength region $\{|\theta_x| > \pi/2 \text{ or } |\theta_z| > \pi/2\}$ (Fig. 3.5c), while it is reasonably below 1 for line-Jacobi (Fig. 3.6c). For simplicity, we only have analyzed the behaviour of smoothing by direct discretization of the

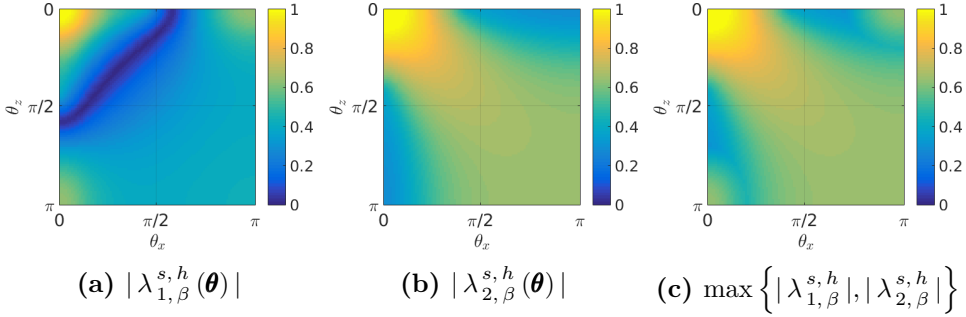


Figure 3.6: Smoothing factors of line-Jacobi for the elastic wave equation as a function of the Fourier component relative to $\boldsymbol{\theta}$ for $\beta = 2$, $k_S h = \pi/10$ (corresponding to 20 points per wavelength), $r = 2$ and $\eta = 0.8$.

elastic wave equation on a certain grid level G^h . In Table 3.1, we report the results of our smoothing analysis for point and line-Jacobi and we observe that line smoothing is also a good choice for the Galerkin coarse-grid discretization (3.26).

From a computational point of view, line-Jacobi requires the solution of the linear system:

$$\mathbf{D}_\beta^h \mathbf{v}^h = \mathbf{f}^h, \quad \mathbf{D}_\beta^h = \begin{pmatrix} \mathbf{D}_{xx,\beta}^h & 0 \\ 0 & \mathbf{D}_{zz,\beta}^h \end{pmatrix},$$

following the definition (3.33). This system can be decomposed in a series of independent tri-diagonal 1-D problems — horizontal-wise for $\mathbf{D}_{xx,\beta}^h$ or vertical-wise for $\mathbf{D}_{zz,\beta}^h$ — that can be solved efficiently with linear complexity (see Thomas algorithm in Press et al. [2007]), resulting in a well-suited scheme for parallel implementation.

Analysis of line-Jacobi for the limit case $r \rightarrow +\infty$ Even though we demonstrated the effectiveness of line smoothing over the point-wise counterpart for the staggered-grid scheme, its smoothing quality deteriorates when $r \rightarrow +\infty$, showing a behaviour similar to equation (3.32). Indeed, it can be proven that

$$\lambda_{1,\beta}^{s,h}(\boldsymbol{\theta}) \approx 1 - 2\eta, \quad \lambda_{2,\beta}^{s,h}(\boldsymbol{\theta}) \approx 1, \quad \text{for } r \rightarrow +\infty. \quad (3.35)$$

The same analysis, however, shows more favourable properties for Kelly' finite-difference scheme (3.9). The eigenvalues related to point-Jacobi, in this case,

Smoothing factors									
Direct					Galerkin				
	$r = 2$	$r = 3$	$r = 4$	$r = 5$		$r = 2$	$r = 3$	$r = 4$	$r = 5$
h	0.90	0.95	0.97	0.98	h	0.90	0.95	0.97	0.98
$2h$	0.91	0.96	0.97	0.98	$2h$	0.82	0.91	0.95	0.97
$4h$	0.93	0.98	0.99	0.99	$4h$	0.87	0.94	0.97	0.98
$8h$	1.00	1.00	1.00	1.00	$8h$	1.00	1.00	1.00	1.00
$16h$	0.71	0.77	0.83	0.84	$16h$	0.93	0.91	0.83	0.94

(a) *Point-Jacobi*

	$r = 2$	$r = 3$	$r = 4$	$r = 5$		$r = 2$	$r = 3$	$r = 4$	$r = 5$
h	0.84	0.84	0.90	0.93	h	0.84	0.84	0.90	0.93
$2h$	0.92	0.92	0.92	0.94	$2h$	0.76	0.76	0.80	0.83
$4h$	0.75	0.87	0.93	0.95	$4h$	0.81	0.81	0.82	0.84
$8h$	0.80	0.94	0.98	1.00	$8h$	0.70	0.71	0.79	0.82
$16h$	0.58	0.66	0.86	0.97	$16h$	0.56	0.56	0.61	0.74

(b) *Line-Jacobi*

Table 3.1: Smoothing factors μ_s^h , as defined in equation (3.24), for: (a) point- and (b) line-Jacobi smoothing with a staggered-grid discretization. The comparison is made for different grid levels, velocity ratios and direct discretization as well as Galerkin coarse-grid operators (defined by full weighting as restriction and bilinear interpolation as prolongation operator). For this numerical comparison, we let the damping factor $\beta = 2$ and S-wave wavenumber $k_S h = \pi/10$ (corresponding to 20 points per wavelength on the finest grid). For each of the four experiments, the relaxation parameters η were tuned to minimize the smoothing factors for the range of velocity ratios $r \in [1, 5]$ on each grid level $h, 2h, 4h, 8h, 16h$.

are

$$\lambda_{P,\beta}^{s,h}(\boldsymbol{\theta}) \approx 1 - 2\eta \left[\sin\left(\frac{\theta_x}{2}\right)^2 + \sin\left(\frac{\theta_z}{2}\right)^2 - \sin\left(\frac{\theta_x}{2}\right)^2 \sin\left(\frac{\theta_z}{2}\right)^2 \right],$$

$$\lambda_{S,\beta}^{s,h}(\boldsymbol{\theta}) \approx 1 - 2\eta \sin\left(\frac{\theta_x}{2}\right)^2 \sin\left(\frac{\theta_z}{2}\right)^2$$

for $r \rightarrow +\infty$, with eigenvectors

$$\begin{aligned}\hat{\mathbf{u}}_P^h(\boldsymbol{\theta}) &\approx \left[\sin\left(\frac{\theta_x}{2}\right) \cos\left(\frac{\theta_z}{2}\right), \cos\left(\frac{\theta_x}{2}\right) \sin\left(\frac{\theta_z}{2}\right) \right], \\ \hat{\mathbf{u}}_S^h(\boldsymbol{\theta}) &\approx \left[-\cos\left(\frac{\theta_x}{2}\right) \sin\left(\frac{\theta_z}{2}\right), \sin\left(\frac{\theta_x}{2}\right) \cos\left(\frac{\theta_z}{2}\right) \right].\end{aligned}$$

So far, we experience the same difficulties seen for point-Jacobi applied to the Virieux's scheme, namely the inefficient smoothing of the S-wave error modes (this is true whenever θ_x or θ_z are close to zero). For the same reasons discussed for the staggered-grid scheme, this produces anisotropic smoothing that can be eased by a line solver. A computation analogous to (3.35) shows that, for the line smoother, we have

$$\lambda_{\cdot,\beta}^{s,h}(\boldsymbol{\theta}) \approx 1 - \eta \left[1 \pm \cos\left(\frac{\theta_x}{2}\right) \cos\left(\frac{\theta_z}{2}\right) \right],$$

meaning that $1 - \eta \leq \lambda_{1,\beta}^{s,h}(\boldsymbol{\theta}) \leq 1 - \eta(1 - \sqrt{2}/2)$ and $1 - \eta(1 + \sqrt{2}/2) \leq \lambda_{2,\beta}^{s,h}(\boldsymbol{\theta}) \leq 1 - \eta$ for short-wavelength components and $\eta > 0$. Therefore, the relaxation parameter η can be chosen such that the smoothing factor is strictly less than 1, contrary to what is the case with the staggered-grid scheme. Now, the smoothing properties of the line relaxation scheme are acceptable, as will appear in the numerical experiments of Section 3.5.

Line relaxation for acoustic regions ($\mu = 0$) As a final remark, we analyse line-Jacobi for the acoustic case $\mu = 0$. Here, we will specifically consider the elastic formulation (3.5), adapted to the acoustic case. We point out that the smoothing schemes just presented for the staggered-grid finite-difference scheme cannot effectively handle acoustic regions inside the medium. Indeed, similarly to (3.35), the eigenvalues of the line-wise iteration operator for the acoustic case are

$$\begin{aligned}\lambda_{\cdot,\beta}^{s,h}(\boldsymbol{\theta}) &= 1 - \eta \pm \\ &\eta \frac{4 \left| \sin\left(\frac{\theta_x}{2}\right) \right| \left| \sin\left(\frac{\theta_z}{2}\right) \right|}{\sqrt{\left[k_P^2 h^2 (1 - \beta \iota) - 4 \sin^2\left(\frac{\theta_x}{2}\right) \right] \left[k_P^2 h^2 (1 - \beta \iota) - 4 \sin^2\left(\frac{\theta_z}{2}\right) \right]}},\end{aligned}$$

which, for small P-wave wavenumbers k_P , reduces to the limit case for S-waves (3.35). The expression for Kelly's scheme is fairly similar, where, in the numerator of the fraction, the arguments $\theta_x/2$, $\theta_z/2$ of the sine function are replaced by θ_x , θ_z . Therefore, the line smoother is able to properly reduce the short-wavelength components of the error. In this paper, we avoid the issue altogether by considering only strictly elastic media with $c_s \gg 0$.

Comparison of n -grid and numerical convergence factors

We conclude the section on elastic multigrid by testing the theoretical findings on the line smoother in combination with a complete multigrid cycle. We refer to Briggs et al. [2000] for an overview on the different types of multigrid scheduling strategies. In our numerical study, we will consider the traditional V- and W-cycles.

As pointed out in Section 3.4.2, it is possible to analytically determine the convergence factors of multigrid under the simplifications discussed earlier. The analysis describes the error reduction by multigrid over a limited number of grid levels on a problem with constant coefficients. On the coarsest grid, the problem is assumed to be solved exactly. In the literature, usually two or three levels are considered. In general, the smaller the number of coarse-grid levels, the more optimistic the estimate. Here, we needed $n = 4$ levels to address the complications of the elastic wave equation.

The comparison is made with numerically computed convergence factors. This requires the spectral radius of

$$\mathbf{I} - \mathbf{M}^h \mathbf{H}_\beta^h, \quad \mathbf{M}^h \approx (\mathbf{H}_\beta^h)^{-1},$$

where \mathbf{M}^h is the approximate inverse of \mathbf{H}_β^h after a multigrid cycle. This computation can be carried out by simply using the power method. In Table 3.2, we list the results of the n -grid analysis for the staggered-grid scheme using a multigrid W-cycle, for different values of the velocity ratio, damping factor and S-wave wavenumber. We observe that we need damping values β that are larger than the one used in Erlangga et al. [2006, 2004] for the acoustic case to obtain a convergent scheme, highlighting a more problematic near-null subspace of modes, caused by the coarse-grid discretization.

As an example, Table 3.2 compares the results of the n -grid analysis to the numerically computed convergence factors using a W-cycle. The numerical

W(1, 1)-cycle convergence factors

$\beta = 1.5$					$\beta = 1.5$				
	$r = 2$	$r = 3$	$r = 4$	$r = 5$		$r = 2$	$r = 3$	$r = 4$	$r = 5$
$\mu_s^{\nu_1+\nu_2}$	0.51	0.72	0.82	0.88	$\mu_s^{\nu_1+\nu_2}$	0.47	0.68	0.80	0.86
μ_2	0.50	0.71	0.82	0.88	μ_2	0.46	0.68	0.80	0.86
μ_3	0.50	0.72	0.82	0.88	μ_3	0.46	0.68	0.80	0.86
μ_4	0.50	0.72	0.91	1.37	μ_4	0.46	0.68	0.80	0.86
μ_{num}	0.50	0.72	0.84	1.26	μ_{num}	0.46	0.68	0.80	0.86

$\beta = 2$					$\beta = 2$				
	$r = 2$	$r = 3$	$r = 4$	$r = 5$		$r = 2$	$r = 3$	$r = 4$	$r = 5$
$\mu_s^{\nu_1+\nu_2}$	0.50	0.71	0.82	0.88	$\mu_s^{\nu_1+\nu_2}$	0.47	0.68	0.80	0.86
μ_2	0.50	0.71	0.82	0.88	μ_2	0.46	0.68	0.80	0.86
μ_3	0.50	0.71	0.82	0.88	μ_3	0.46	0.68	0.80	0.86
μ_4	0.50	0.71	0.82	0.88	μ_4	0.46	0.68	0.80	0.86
μ_{num}	0.49	0.70	0.82	0.88	μ_{num}	0.46	0.67	0.80	0.86

(a) $k_S h = \pi/5$
(b) $k_S h = \pi/10$

Table 3.2: Comparison of the convergence factors, μ_n , computed analytically by n -grid analysis for $n = 2$, $n = 3$, and $n = 4$ as in (3.29) to the numerical convergence factors, μ_{num} , determined by the power method, using the staggered-grid finite-difference scheme. Here we considered a W-cycle with a single pre-smoothing and post-smoothing step. The coarse-grid discretization is obtained by Galerkin coarsening, with full weighting as restriction and bilinear interpolation as prolongation operator. The relaxation parameters are optimized to minimize the smoothing factors. The comparison is made for different values of the velocity ratio r , damping factor β and wavenumber: (a) $k_S h = \pi/5$, (b) $k_S h = \pi/10$.

results are in good agreement with the theoretical estimates, demonstrating that local-mode analysis can be used as a guiding tool to choose and optimize

the different components of multigrid.

Extension of the smoothing scheme to 3-D

The generalization of the proposed line-smoother to 3-D is of paramount importance for seismic imaging applications. As previously noted for the 2-D case, line solvers can benefit from parallelization. The most obvious generalization of line-smoothing to 3-D corresponds to plane-wise smoothing, which amounts to the solution of many 2-D problems. For large problems, however, this might result in a very expensive scheme. To ease the computation in 3-D, we foresee the following alternatives: alternating line-smoothing (x , y and z directions), semi-coarsening, or a combination of both. We might expect better smoothing factors with plane-smoothing than with 3-D alternating line-smoothing. These alternatives can be assessed by local-mode analysis, similar to the 2-D case presented here, and will be the subject of future study.

3.5 Numerical examples

So far, we discussed the performance of different smoothers and their combination with multigrid for the damped elastic wave equation. We validated our theoretical understanding by local-mode analysis and found that these results agree with the actual numerical behavior of the scheme. Now that we have addressed the accuracy and effectiveness of the preconditioner in the system (3.15), estimated by multigrid, we are ready to study the numerical performance of the iterative solver applied to (3.15).

In Section 3.3, we studied the effect of preconditioning with the complex-shifted operator assuming its exact inversion. Clearly, the actual numerical behaviour of a chosen iterative scheme should account for the inexact inversion of the complex-shifted wave operator by multigrid. Consequently, the spectrum configuration as seen in Figure 3.3 and the convergence of the iterative scheme will be affected. The purpose of this section is to perform a series of numerical tests of the preconditioning scheme (3.15) using the actual multigrid implementation described in Section 3.4.

We start with a brief comment on the specific numerical implementation of the various multigrid ingredients introduced in Section 3.4.2 as well as the discrete representation of the problem at hand. We perform two types of

experiments: one to test the numerical scheme for the homogeneous case and the other for an arbitrarily heterogeneous model, based on the well-known benchmark acoustic Marmousi model: the Marmousi2 model [Martin et al., 2006]. Each experiment is accompanied by a discussion of the results, which will show a behaviour — including shortcomings — similar to the acoustic shifted-Laplacian multigrid preconditioning.

3.5.1 Additional settings and multigrid implementation

In the following numerical experiments, we run the tests for the finite-difference schemes of Kelly and Virieux, discussed in Section 3.2. The choice of the discretization of the computational model is made such that the quantity $k_S^{\max} h$, where k_S^{\max} is the maximum wavenumber, is kept constant. This condition is equivalent to a discretization that keeps the number of points per minimum wavelength constant. This means that for a given frequency, the size of the computational grid is adjusted accordingly.

As for boundary conditions, we set up sponge layers to mimic absorption at the edges of the model. Of course, perfectly matched layers can be also considered [Berenger, 1994]. Theoretical and numerical studies on the effect of this type of boundary conditions for acoustic multigrid can be found in Reps and Vanroose [2012] and Pan and Abubakar [2013].

For the multigrid implementation, we stick to traditional choices for the various components. For the coarse-grid operator we can consider: (i) direct discretization of the wave operator on the corresponding grid level, or (ii) Galerkin condition (3.26). Local-mode analysis suggests that (ii) is a more robust approach. For restriction and prolongation operators, we use, respectively, full weighting and bilinear interpolation. In stencil notation,

$$R_h^{2h} \triangleq \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad P_{2h}^h \triangleq \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix},$$

which symbolize the action

$$(R_h^{2h} u^h)_{2i,2j} = (R_h^{2h} \star u^h)_{2i,2j},$$

defined in (3.6), and

$$(P_{2h}^h u^{2h})_{i,j} = (P_{2h}^h \star I_{2h}^h u^{2h})_{i,j},$$

where $I_{2h}^h u^{2h}$ is the injection of u^{2h} on the grid G^h , determined by the values of u^{2h} on $G^{2h} \subseteq G^h$ and vanishing on every other grid point. Operator dependent grid transfer restrictions and prolongations, as studied by de Zeeuw [1990] and extended by Erlangga et al. [2006] to complex matrices, are also an option. As of multigrid grid-level scheduling cycles, we will consider $V(\nu_1, \nu_2)$ -cycles, for computational efficiency. The number of pre and post-smoothing steps will be set to $\nu_1 = \nu_2 = 2$. The relaxation parameters η in (3.22), for the line-Jacobi smoother, are optimized for the range of velocity ratios displayed by the medium, and for all the wavenumbers up to the maximum value of the model. We should point out that the chosen amount of damping, β , is based on numerical investigations rather than rigorous theoretical analysis. As discussed in Section 3.3, large values of β lead to better multigrid convergence of the preconditioner while small values generate a more favourable spectrum for the preconditioned system if that system could be solved exactly.

Lastly, a number of options is available for the outer iterative method. For non-symmetric problems, we could consider any Krylov subspace method such as GMRES, Bi-CGSTAB or IDR(s). In our experience, Bi-CGSTAB performs the best for the problem at hand and will be the choice for the following numerical study.

3.5.2 The homogenous case

We start with testing how the preconditioning scheme of Section 3.3 behaves in the homogeneous case. The computational domain is the 2-D square $[-1/2, 1/2]^2$ and we solve for the wavefield generated by a point source located at the centre $(0, 0)$ of the domain. The type of source considered is a vertical force. We study how the number of iteration increases as a function of the frequency. For each of the experiments, the quantity $k_S h$ will be kept fixed at a value of $\pi/5$ and we vary the wavenumber k_S . Therefore, the grid spacing h is adjusted accordingly, as mentioned before. Also, different velocity ratios r and damping factors β are considered in the analysis.

Results

Table 3.3 lists the results, which highlight a reasonable convergence rate for the considered frequency range. As was the case with the scalar Helmholtz equation, we found experimentally that the required number of iterations in-

creases linearly with the wavenumber k_S . Since the multigrid convergence factors get worse for increasing velocity ratio r , as discussed in Section 3.4, the iterative method will perform worse for high values of r .

3.5.3 A highly heterogeneous problem: the elastic Marmousi model

Now that we have validated the scheme for the homogeneous problem, we test its behaviour for the heterogeneous case. We will adopt an elastic model based on the well-known Marmousi model, originally designed by the Institut Français du Pétrole. This acoustic version has been extended to the elastic case by Martin et al. [2006]. Here, we will only consider the top-central portion of the full model. We discard the water layer to avoid the complications discussed in Section 3.4.2 that will affect the performance of multigrid for the staggered-grid scheme. The corresponding elastic parameters are depicted in Fig. 3.7. For this model, the maximum value of the velocity ratio r is about 5.2.

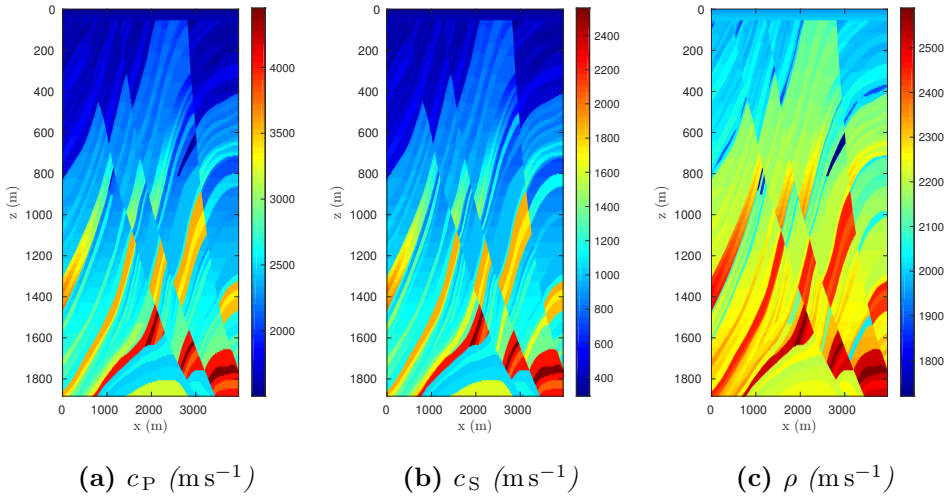


Figure 3.7: *Subset of the elastic Marmousi2 model: (a) P-wave velocity c_P , (b) S-wave velocity c_S and (c) density ρ , used in the numerical experiments in Section 3.5. The water column on the top of the model has been replaced by an elastic layer.*

In the numerical experiments, we are going to model the frequencies 1.45 Hz, 2.9 Hz, 5.8 Hz and 11.6 Hz with different grid spacings. Figure 3.8 shows the real part of the vertical particle displacement generated by a point source.

Results for the homogeneous problem: iteration count

$r = 2$						
	$k_S = 100.5$	$k_S = 181.0$	$k_S = 261.4$	$k_S = 341.8$	$k_S = 422.2$	$k_S = 502.7$
$\beta = 1.5$	106	159	214	268	324	382
$\beta = 2$	139	226	312	402	443	501

$r = 3$						
	$k_S = 100.5$	$k_S = 181.0$	$k_S = 261.4$	$k_S = 341.8$	$k_S = 422.2$	$k_S = 502.7$
$\beta = 1.5$	121	203	286	391	467	562
$\beta = 2$	140	212	285	352	446	522

(a) Virieux's scheme

$r = 2$						
	$k_S = 100.5$	$k_S = 181.0$	$k_S = 261.4$	$k_S = 341.8$	$k_S = 422.2$	$k_S = 502.7$
$\beta = 1.5$	108	150	204	256	327	375
$\beta = 2$	135	208	289	325	430	486

$r = 3$						
	$k_S = 100.5$	$k_S = 181.0$	$k_S = 261.4$	$k_S = 341.8$	$k_S = 422.2$	$k_S = 502.7$
$\beta = 1.5$	108	174	240	299	392	467
$\beta = 2$	124	188	245	320	391	480

(b) Kelly's scheme

Table 3.3: Numerical experiment for a homogeneous model (square computational domain, vertical delta source located at the center). We fixed $k_S h = \pi/5$. We list the iterations required to decrease the residual norm by 6 decades. The selected values of k_S increase linearly and h varies along. We employ a V(2,2) cycle, where each of the components are optimized according to r . The experiments consistently show a linear growth of the iterations with respect to k_S . As the convergence factors decay for increasing r , more iterations are required. A comparison is made between the performance of the (a) Virieux's and (b) Kelly's scheme.

Results

We tested the convergence behaviour of the iterative solver for different frequencies and values of $k_S^{\max} h$. The numerical experiments performed show

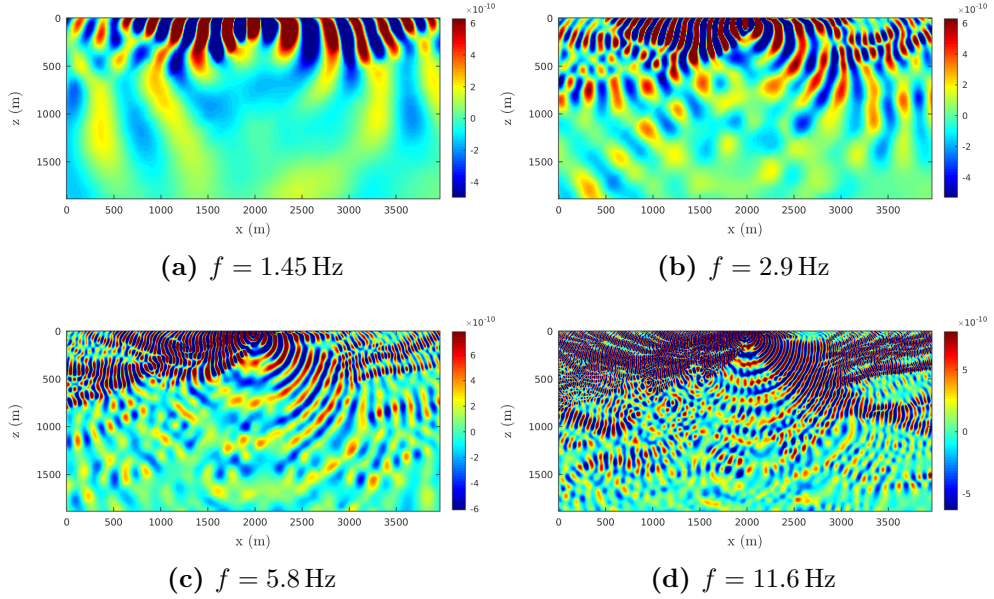


Figure 3.8: *Real part of the vertical component of the wavefield propagating through the Marmousi2 model (Fig. 3.7), for the frequencies (a) 1.45 Hz, (b) 2.9 Hz, (c) 5.8 Hz and (d) 11.6 Hz, generated by a point source.*

good convergence properties. The convergence history of Bi-CGSTAB for the staggered-grid scheme is reported in Fig. 3.9. However, the number of iterations grows linearly with increasing frequency, as was the case for the acoustic wave equation.

Table 3.4 summarizes the numerical experiments. A comparison between the performance of Virieux’s and Kelly’s scheme is included. As highlighted by the local-mode analysis in Section 3.4.2, the multigrid performance for Kelly’s scheme are better than for Virieux’s. However, this comes at the cost of inferior accuracy (Fig. 3.2).

3.6 Conclusions

This paper comprises a first assessment of the applicability of the preconditioning technique by Laplacian shifting, using multigrid, applied to the time-harmonic elastic wave equation. In order to substantiate this analysis, we made use of the classical theoretical tool of local-mode analysis.

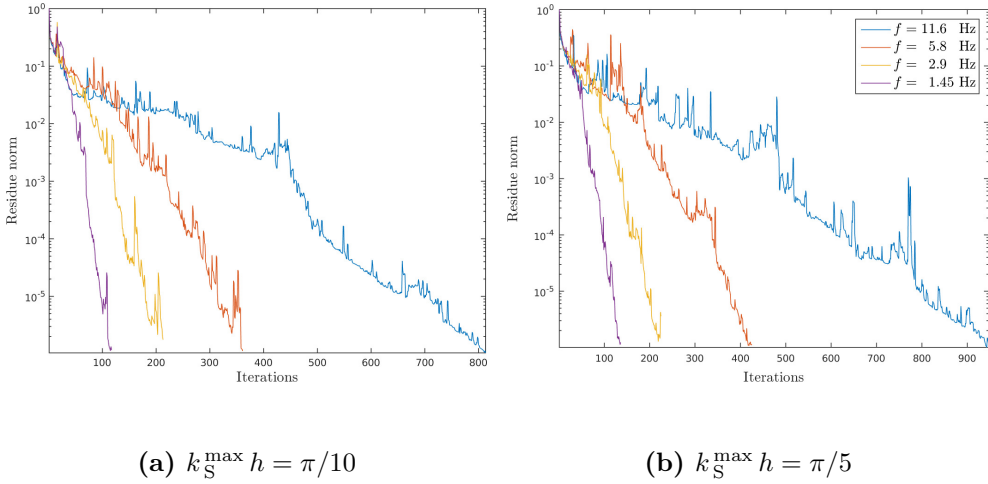


Figure 3.9: Convergence history of the iterative solver applied to the Marmousi2 problem for a number of frequencies and wavenumbers. Here, the wave equation is discretized by the staggered-grid scheme. The discretization for the various frequencies is chosen in such a way that the quantities (a) $k_S^{\max} h = \pi/10$ (equivalent to a discretization of 20 points per minimum wavelength) and (b) $k_S^{\max} h = \pi/5$ (10 points per minimum wavelength) remain constant. In other words, the size of the model is adjusted to the relative frequency as in Table 3.4. The coarse-grid operator is obtained by Galerkin coarsening. We use the damping factor $\beta = 2$ in each experiment. The multigrid cycle is $V(2, 2)$.

By means of local-mode analysis, we were able to point out the shortcomings of the natural choice of point-Jacobi as a smoother, otherwise effective in the acoustic case. This is due to a defective solution of the S-wave modes in the case of large P-to-S velocity ratios. In terms of particle displacement wavefield, this results in smoothing anisotropy. We therefore introduced and validated a suitable line-wise variant of the Jacobi scheme.

In the present experience, different numerical discretization schemes of the wave equation might lead to a different multigrid performance, as it was the case with Kelly's and Virieux's scheme. For Virieux's scheme, in particular, handling acoustic subsets of the model is problematic and should be dealt with in a special way. This leaves the question on how to find the optimal trade-off between accuracy and efficiency.

The results obtained on highly heterogeneous models suggest that the method provides reasonable convergence, in line with what was obtained for the acoustic case. With the elementary multigrid component used in this pa-

Results for the Marmousi2 model: iteration count

f (Hz)	$k_S^{\max} h = \pi/10$			$k_S^{\max} h = \pi/5$		
	Size	Virieux	Kelly	Size	Virieux	Kelly
1.45	385×593	118	102	193×297	137	98
2.9	577×993	214	165	289×497	226	160
5.8	961×1793	362	291	481×897	426	297
11.6	1665×3329	814	694	833×1665	948	647

Table 3.4: Summary of the results of the numerical experiments on the Marmousi2 model. We compare Virieux’s and Kelly’s finite-difference schemes. For the reasons discussed in Section 3.4.2, Kelly’s discretization leads to a slightly better performance, even though the accuracy of the scheme is worse than the staggered-grid’s (see Fig. 3.2). For each of the finite-difference schemes, we choose the Galerkin coarse-grid operator, $\beta = 2$ as damping factor, and a $V(2, 2)$ multigrid cycle.

per, we observed that the damping factors should have higher values than in the acoustic case, highlighting a more problematic near-null kernel. The deflation techniques discussed by Erlangga and Nabben [2008] and Sheik et al. [2013] might be especially helpful in the elastic case.

The elastic case suffers from the same shortcomings as the acoustic shifted-Laplacian multigrid, namely the linear growth of the number of iterations for increasing frequency at fixed number of points per wavelength. As noted in the introduction, this means that frequency-domain imaging methods can only obtain the same computational complexity as their time-domain counterparts, from a theoretical point of view. In practice, the constants involved in the complexity analysis play an important role for the actual performance and the iterative solver here discussed might be a efficient alternative to the time domain, as motivated by Knibbe et al. [2014] for acoustic imaging. While in 2-D a direct LU decomposition with nested dissection will be the preferred choice, if many shots have to be computed in the same model, in 3-D this is no longer the case, and the present study provides guidelines as to how to proceed.

A non-linear full-wavefield inversion method for ultrasound imaging in non-destructive testing[†]

Abstract

Many of the imaging techniques for non-destructive inspection of oil and gas pipe lines are based on a linearized relationship between material properties and the wavefield that propagates inside the medium. This simplification does not take into account multiple scattering nor mode-conversion, and makes the imaging of eventual defects a challenging task. This is especially true for vertically aligned defects, such as stress-corrosion cracks. Here, we treat imaging as a non-linear inverse problem, which honors the correct physics of the problem, by adopting a data-fitting approach. This, in turn, enhances the resolution of the inverted image compared to the linearized approach. Its classic implementation, however, results in a non-linear optimization problem with many local minima that hamper convergence to the correct inversion result. Here, we present an alternating updating scheme where properties and wavefields are both considered independent unknowns of the problem. This enlargement of the model parameter space eases the problem of local minima. We validate the inversion algorithm on synthetic data and show that multiple reflections are

[†]This chapter is the reproduction of a paper submitted for publication (2017)

fundamental for the recovery of vertical cracks. We also present some results on measured data, demonstrating the capability of the method in a realistic setting.

4.1 Linear vs non-linear imaging in non-destructive testing

Ultrasound non-destructive inspection is a well-established procedure to assess the integrity of oil and gas pipe lines. For example, tensile stress as well as the chemical environment may create cracks in steel pipes that ultimately cause mechanical failure. During inspection, a transducer positioned at the external surface of the pipe generates ultrasound elastic waves and their scattered field is measured. The objective is to locate the spatial distribution of material defects from their particular signature recorded in the data.

Current industry-standard imaging methodologies are based on the linearization of the relationship between the excited wavefield and the underlying medium properties. Notable examples are the total focusing method [Felice et al., 2015], the inverse wavefield extrapolation [Pörtzgen et al., 2007, 2008], which employs the same principles as the well-known migration method in seismic imaging, and the classic Born inversion (see, e.g., Ozmen et al. [2015]).

To illustrate the limitations of the linear assumption, we follow the simplified ray-based analysis of wave propagation, which allows a description of the ray-paths in terms of ‘orders of scattering’, each order referring to the number of times the ray has been reflected within the medium. Linearization then entails that only single reflection events in the data can be predicted by this class of methods. Multiple scattering—consisting in the interaction between different scatterers as well as between different parts of the same scatterer—has to be treated as noise. Furthermore, the P- and S-wave mode conversion is not correctly handled. This makes the imaging of vertical cracks, whose effect is mainly visible in data from multiple reflections, a challenging problem (see Fig. 4.1).

The goal of this paper is to present an inversion scheme capable of accounting for all the non-linear scattering effects of wave propagation. For this purpose, we will adopt a method originally presented in Rizzuti and Gisolf [2017] for seismic imaging applications. We will demonstrate the advantages of non-linear imaging on a synthetic model with vertical cracks and confirm its capability on data measured in the laboratory.

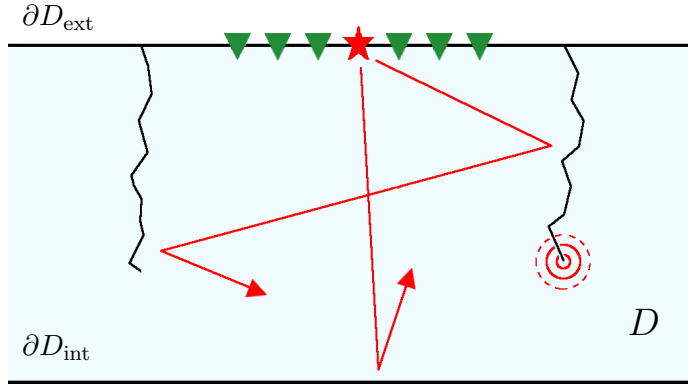


Figure 4.1: Schematic representation of non-destructive testing. We consider the cross-section D of a test sample, delimited by exterior and interior surfaces ∂D_{ext} and ∂D_{int} , with some defects (e.g. cracks). On the external surface ∂D_{ext} , a transducer (denoted by a red star) excites elastic waves. Here, the data is also measured (receivers denoted by green triangles). In red, we highlight different scattering effects, as reflections due to the walls and the multiple bounces between cracks. When the wave-front hits the tips of a crack, we observe diffraction effects (represented by concentric circles).

4.2 Assumptions

Ultrasound imaging for crack detection has several advantages over, for instance, seismic imaging. For one, there are no difficulties in determining a kinematically accurate model for the incident wavefield (background model), as it consists in homogeneous steel. The cracks can be represented as thin void regions inside the computational domain. Because they are usually sparsely distributed, a sparsity-enhancing regularization can help. Furthermore, we can limit the degrees of freedom down to a single-parameter inversion per grid point instead of the three elastic properties: bulk modulus K , shear modulus μ and density ρ . The source wavelet can be estimated with a separate experiment conducted in a section of the pipe devoid of defects. Lastly, the multiple reflections between the pipe walls provide abundant illumination of the scatterers.

Contrary to seismic imaging, we must deal with the additional challenge of simulating two boundaries—to be treated as air-solid or fluid-solid interfaces. Throughout the paper, we will assume the shape of the walls to be known. The problem of determining the geometry of the domain is left for future work. Also, given the fact that cracks behave as void regions inside the medium, the

inversion has to deal with scatterers with extreme contrasts.

4.3 Full-waveform inversion

We have adapted a technique known as ‘full-waveform’ inversion, introduced in the seismic community by Tarantola and Valette [1982a] and Lailly [1983], which attempts to the model parameters by fitting the measured to the modeled data in a least-squares sense. Imaging then becomes an optimization problem on a data-fitting cost functional constrained by the wave equation.

In the following, we will indicate by \mathbf{m} a pre-defined subset of the elastic properties K , μ , and ρ , as spatially-dependent functions defined on a domain D with known geometry. For example, for each of the points corresponding to a crack, we have the values $K = 0$, $\mu = 0$, and $\rho = 0$. We are looking for a model \mathbf{m} that minimizes the least-squares misfit functional

$$\mathcal{J}(\mathbf{m}) = \|\mathbf{d} - \Pi_R \mathbf{u}(\mathbf{m})\|^2, \quad (4.1)$$

where \mathbf{d} is the acquired data and \mathbf{u} is the modeled elastic wavefield, typically described by the particle velocity, as a function of the known and unknown medium parameters (the latter denoted by \mathbf{m}). The operator Π_R samples the values of \mathbf{u} (or a linearly related quantity) at the receiver locations R . The functional (4.1) should be complemented by a regularization procedure to avoid unreasonable solutions, for example by adding to (4.1) a term that measures the least-squares norm of \mathbf{m} or its derivatives.

In equation (4.1), we made explicit the non-linear dependence of the wavefield with respect to the underlying model. This link is provided by the isotropic elastic wave equation,

$$\mathcal{H}[\mathbf{m}] \mathbf{u} = \mathbf{f}, \quad (4.2)$$

equipped with suitable boundary conditions. We denote the source term by \mathbf{f} . The linear system of equations is summarized by $\mathcal{H}[\mathbf{m}]$, highlighting the fact that the operator itself depends on the medium properties \mathbf{m} .

The operator, $\mathcal{H}[\mathbf{m}]$, might represent the wave equation in the time or in the frequency domain. For 2-D inversion work, as considered in this paper, the frequency domain is particularly attractive because of the availability of efficient direct solvers for (4.2) and only a few frequency data points are needed to obtain a satisfactory result [Plessix, 2017].

A review of full-waveform inversion and the role it plays in seismic imaging can be found in Virieux and Operto [2009]. This method has received considerable attention over the last 30–40 years for various reasons. For starters, any feature present in the input data is treated as a genuine effect governed by the wave equation and can contribute to refine the output model. Conversely, when a linearized modeling is adopted, a great deal of effort must be devoted to remove secondary non-linear scattering from the data to avoid false images and artifacts. In addition, the method can handle an arbitrary acquisition configuration of sources and receivers. Another advantage is the high resolution that can in principle be achieved for the final image if local minima can be avoided during the optimization [Simonetti, 2006]. In general, this requires a starting model that is kinematically correct within half a wavelength. Otherwise, manual intervention and more advanced methods are required. For our current problem, this so-called background model is already known.

The minimization of (4.1) is generally handled by local search techniques, due to the large number of unknowns involved in the optimization, especially for 3-D problems. The methods of choice are therefore gradient-based and may be accelerated by the use of the Hessian. The iterations involve the computation of the modeled data and the gradient and require numerous solutions of the wave equation for different parameters, resulting in a computationally heavy scheme. Apart from the computational aspect, the local minima caused by the non-linearity of the functional (4.1) are a serious problem in seismic imaging, where a large part of the data are generated by reflections. The issue with local minima is less severe for transmission data, but even then, the starting model needs to be correct within half a wavelength. In general, these limitations carry over to applications for ultrasound imaging, although there, it may be less difficult to find a good starting model.

Different ideas have been proposed to address the issues resulting from the non-linearity of the functional (4.1). This includes continuation techniques, consisting for example in the inversion of increasing the frequency content data from lower to higher [Bunks et al., 1995] or layer stripping and data windowing by inverting short-to-long offset data [Shipp and Singh, 2002]. Other attempts consider alternative data domains [Shin and Cha, 2008, 2009; Bozdag et al., 2011] and/or alternative objective functionals [Warner and Guasch, 2014; Métivier et al., 2016; Bharadwaj et al., 2016] that are less prone to spurious local minima.

In this paper, we explore yet another route. In Rizzuti and Gisolf [2017],

we describe a 2-D inversion method based on a model space extension. More precisely, the set of unknowns not only consists of the elastic parameters \mathbf{m} but also the wavefields \mathbf{u} . This means that the constraint given by the wave equation (4.2) is not enforced explicitly. The philosophy behind this procedure is to ease the non-linearity by adding extra dimensions to the problem and circumvent any local minima. A similar idea has been followed in van den Berg and Kleinman [1997] and van Leeuwen and Herrmann [2013]. Due to this characteristic, we often refer to our method as full-‘wavefield’ inversion. It will be discussed in depth in the next section.

4.4 Alternating inversion scheme

In this section, we summarize the full-wavefield inversion algorithm recently studied in Rizzuti and Gisolf [2017]. A 1.5-D version of the algorithm, based on the assumption of a locally layered medium, is routinely applied on field data [Gisolf et al., 2014; Beller et al., 2015]. The extension to 2-D has been explored in Haffinger et al. [2013] for the acoustic case. In Rizzuti and Gisolf [2017], we addressed the 2-D elastic case, object of this paper.

Section 4.4.1 starts with the scattering integral equation that represents the foundation of the alternating update inversion strategy. The actual algorithm is described in Section 4.4.2. The first step of the inversion is represented by the Born approximation, the details of which are found in Section 4.4.3. Note that this approximation differs from the traditional ‘single-scattering’ linearization. In Section 4.4.4, we discuss the regularization strategies and constraints. Finally, in Section 4.4.5, we examine aspects of the numerical implementation pertaining to ultrasound imaging.

4.4.1 Scattering integral equation

The method hinges on a Lippman-Schwinger integral reformulation of the wave equation (4.2). We consider a domain $D \subseteq \mathbb{R}^2$ delimited by two air-solid or fluid-solid interfaces ∂D_{ext} , ∂D_{int} , as in Fig. 4.1. We first fix the (homogeneous) background

$$\mathbf{m}_b = (K_b, \mu_b, \rho_b),$$

corresponding to the properties of steel. The wave propagation in the background medium includes the reflections on the interfaces, encoded as boundary

conditions. Indeed, Green's operator

$$\mathcal{G}_b = (\mathcal{H}[\mathbf{m}_b])^{-1}, \quad (4.3)$$

which represents the impulse response of the background model, accounts for the multiple bounces between the external and internal boundaries ∂D_{ext} and ∂D_{int} . Likewise, the incident field \mathbf{u}^{inc} is the wavefield excited by the same source \mathbf{f} as \mathbf{u} , propagating and reflecting through D . These quantities are determined by the a priori knowledge of D and its geometry.

The integral equation reads

$$\mathbf{u} - \mathcal{G}_b \mathcal{W}[\boldsymbol{\chi}] \mathbf{u} = \mathbf{u}^{\text{inc}}, \quad \forall \mathbf{x} \in D. \quad (4.4)$$

Note that, since we favor an operator-based description of this equation, the spatial dependence $\mathbf{u} = \mathbf{u}(\mathbf{x})$ is implicit in (4.4). The contrast operator $\mathcal{W}[\boldsymbol{\chi}]$ is defined by $\mathcal{W}[\boldsymbol{\chi}] = -\mathcal{H}[\mathbf{m}_b \cdot \boldsymbol{\chi}]$, where the variable $\boldsymbol{\chi}$ collects the individual contrasts with respect to the elastic parameters: $\boldsymbol{\chi} = (\chi_K, \chi_\mu, \chi_\rho)$. Its components are defined by

$$\chi_K = \frac{K - K_b}{K_b}, \quad \chi_\mu = \frac{\mu - \mu_b}{\mu_b}, \quad \chi_\rho = \frac{\rho - \rho_b}{\rho_b}. \quad (4.5)$$

In the following, we will assume that $\boldsymbol{\chi}$ vanishes outside a bounded domain $D_{\text{inv}} \subseteq D$. The ultimate goal of the inversion is to determine the contrast functions $\boldsymbol{\chi}$ on D_{inv} .

4.4.2 Algorithm

The scattering integral in equation (4.4) is defined for every point $\mathbf{x} \in D$. We specify two kinds of domains to which this equation can be restricted:

- (i) the data domain $R \subseteq \partial D_{\text{ext}}$, containing the position of the receivers, and
- (ii) the contrast domain D_{inv} , where the inversion unknowns will be estimated.

The restriction or sampling of the wavefield to R enables the comparison with the recorded data and leads to the data equation

$$\Delta \mathbf{d} = \mathcal{G}_R \mathcal{W}[\boldsymbol{\chi}] \mathbf{u}, \quad (4.4i)$$

where

$$\mathcal{G}_R = \Pi_R \mathcal{G}_b \quad (4.6)$$

is the ‘domain-to-receiver’ Green’s operator $\mathcal{G}_R = \mathcal{G}_R(\mathbf{x}_r, \mathbf{x})$ defined for each $\mathbf{x}_r \in R$ and $\mathbf{x} \in D$. The scattering data is given by $\Delta \mathbf{d} = \mathbf{d} - \Pi_R \mathbf{u}^{\text{inc}}$, with the recorded data \mathbf{d} and the incident field $\Pi_R \mathbf{u}^{\text{inc}}$, measured through the data restriction operator Π_R . The restriction to D_{inv} , instead, generates the domain equation

$$\mathbf{u} - \mathcal{G}_b \mathcal{W}[\boldsymbol{\chi}] \mathbf{u} = \mathbf{u}^{\text{inc}}, \quad \forall \mathbf{x} \in D_{\text{inv}}. \quad (4.4ii)$$

The inversion algorithm proposed here consists in a sequence $(\boldsymbol{\chi}_n, \mathbf{u}_n)$ of estimated contrasts and wavefields. At every iteration n , given the previously estimated pairs $(\boldsymbol{\chi}_0, \mathbf{u}_0)$, ..., $(\boldsymbol{\chi}_n, \mathbf{u}_n)$, we produce:

- (i) an update $\boldsymbol{\chi}_{n+1}$, by setting $\mathbf{u} := \mathbf{u}_n$ in (4.4i) and optimizing the fit between the left- and right-hand side of the data equation (4.4i). This is achieved through minimization of the functional

$$\mathcal{J}_R(\boldsymbol{\chi}) = \|\Delta \mathbf{d} - \mathcal{G}_R \mathcal{W}[\boldsymbol{\chi}] \mathbf{u}_n\|^2. \quad (4.1i)$$

To deal with the ill-posedness of this functional and the presence of noise in the data, we modify \mathcal{J}_R by adding or multiplying a regularization term as in (4.1i') or (4.1i'') (see the next Section 4.4.4). Typically, we will employ a non-linear conjugate gradient scheme for the minimization of (4.1i);

- (ii) an update \mathbf{u}_{n+1} , by setting $\boldsymbol{\chi} := \boldsymbol{\chi}_{n+1}$ in (4.4ii) and minimizing

$$\mathcal{J}_D(\mathbf{u}) = \|\mathbf{u}^{\text{inc}} - \mathbf{u} + \mathcal{G}_b \mathcal{W}[\boldsymbol{\chi}_{n+1}] \mathbf{u}\|^2. \quad (4.1ii)$$

We will resort to a Krylov subspace method for this step [Kleinman and van den Berg, 1991].

We remark that each of these two steps corresponds to convex problems, and are amenable to gradient-type minimization.

It is convenient to focus the computational effort on the optimization of the data equation (4.1i), rather than (4.1ii). The reason lies in the need for the evaluation of \mathcal{G}_b when the gradient of the functional (4.1ii) is computed, which is achieved by solving the wave equation in the background medium and requires the relatively costly inversion of the linear system obtained after the discretization of equation (4.3). The optimization of (4.1i), instead, is cheaper and relatively easy to parallelize. We will discuss this further in Section 4.4.5.

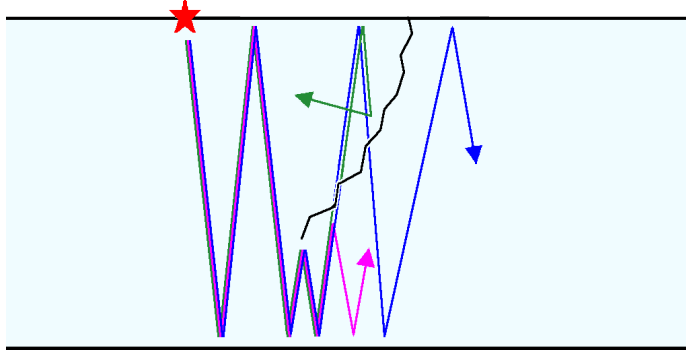


Figure 4.2: A selection of ray-paths undergoing reflections due to the presence of walls and/or an internal scatterer. The linearized modeling associated with Green’s operator (4.3) includes the reflections at the boundaries of the domain but only single scattering due to internal scatterers. In this picture, the blue ray-path is the only admissible in the Born approximation. The violet and green paths interacts at least two times with the scatterer. Note that the blue path is not modeled under the more classical ‘single scattering’ Born approximation.

4.4.3 The Born approximation

When the inversion sequence is started with $\chi_0 \equiv 0$ and $\mathbf{u}_0 = \mathbf{u}^{\text{inc}}$, as will be the case, problem (4.1i) corresponds to a linearized inversion. In a more classic setting, the Born inversion is associated with single-scattering inversion, meaning that a match is sought between the measured data and the synthetics produced by a single reflection against the scatterer. With the term ‘scatterer’, we include the front-wall and the back-wall. In this paper, however, multiple reflections between the walls ∂D_{ext} and ∂D_{int} are automatically encoded, by definition, in Green’s operator (4.3). In terms of ray-paths, the Born approximation will account for any event consisting of: a finite number of wall bounces, a single reflection due to an internal scatterer, and the remaining coda of wall bounces. In Fig. 4.2, we display some of the ray-paths admitted in this specific linear assumption.

4.4.4 Regularization and constraints

As previously stated, we need regularization to resolve the ill-posedness of the inversion problem. Instead of the well-known additive schemes that adds a penalty term to the functional (4.1i), we make use of the multiplicative

regularization originally proposed by van den Berg et al. [2003].

We seek to multiply the data functional (4.1*i*) by the following factor:

$$\mathcal{B}(\chi; \bar{\chi}) = \frac{1}{V_D} \int_D \frac{|\nabla \chi|^2 + \delta^2}{|\nabla \bar{\chi}|^2 + \delta^2}, \quad (4.7)$$

where $\bar{\chi}$ is a given contrast, which ideally represents a certain estimate of the solution contrasts (to be determined throughout the inversion). The operator ∇ provides the gradient of χ , while δ is the so-called steering parameter. V_D is the measure of the set D . The function $(|\nabla \bar{\chi}|^2 + \delta^2)^{-1}$, which appears in (4.7), acts as a regularization weight that promotes the same structure of $\bar{\chi}$ on χ : the cost (4.7) will favor models that are flat in the neighborhood of the points where $\bar{\chi}$ is flat ($\nabla \bar{\chi} \approx 0$), and irregular wherever $\bar{\chi}$ is not flat. This mechanism will enforce ‘blockiness’ whenever $\bar{\chi}$ is a blocky model. This behavior is modulated by the steering parameter: smaller values of δ will enhance the blockiness of the retrieved model.

The optimization problem corresponding to the data functional (4.1*i*) will be now adjusted by minimizing an adaptive functional. This means that, given an estimate χ_n obtained at the previous iterative step n , the update χ_{n+1} will be based on the functional

$$\mathcal{J}_R^{\text{reg}}(\chi; \chi_n) = \mathcal{J}_R(\chi) \cdot \mathcal{B}(\chi; \chi_n). \quad (4.1i')$$

The functional (4.1*i'*) retains convexity for a judicious choice of δ (it should not be too small, see van den Berg et al. [2003]). Note that, for a converging sequence χ_n , $\mathcal{B}(\chi_{n+1}, \chi_n) \rightarrow 1$, so that a decrease of the value of $\mathcal{J}_R^{\text{reg}}$ corresponds to a decrease for the original data functional \mathcal{J}_R .

A multiplicative scheme such as (4.1*i'*) has the advantage over classic methods that it avoids the non-trivial choice of the weight required for additive regularizations. Conversely, the role of the steering parameter δ is only to convey some a priori knowledge about the regularity of the solution.

The same concept expressed by the blocky regularization can be exploited to design a quadratic functional that promotes sparse models. The sparsity-enhancing factor reads:

$$\mathcal{S}(\chi; \bar{\chi}) = \frac{1}{V_D} \int_D \frac{|\chi|^2 + \varepsilon^2}{|\bar{\chi}|^2 + \varepsilon^2}. \quad (4.8)$$

Analogously to (4.7), the weight function $(|\bar{\chi}|^2 + \varepsilon^2)^{-1}$ will promote models χ whose values are close to zero where $\bar{\chi} \approx 0$. If the inversion process leads to a sparse solution, the factor (4.8) will help in enforcing the sparseness of χ .

Finally, both schemes \mathcal{B} and \mathcal{S} may be combined into a functional

$$\mathcal{J}_R^{\text{reg}}(\boldsymbol{\chi}; \boldsymbol{\chi}_n) = \mathcal{J}_R(\boldsymbol{\chi}) \cdot (w_{\mathcal{B}} \mathcal{B}(\boldsymbol{\chi}; \boldsymbol{\chi}_n) + w_{\mathcal{S}} \mathcal{S}(\boldsymbol{\chi}; \boldsymbol{\chi}_n)), \quad (4.1i'')$$

where the weights $w_{\mathcal{B}}$ and $w_{\mathcal{S}}$ satisfy $w_{\mathcal{B}} + w_{\mathcal{S}} = 1$.

Regularization can be thought as a ‘soft’ constraint for the inverse problem, in that it effectively reduces the size of the search space by steering the iterative process towards a smaller set of desirable solutions. At the same time, we might also exploit the a priori assumptions described in Section 4.2 to reduce the number of elastic parameters to be inverted, which also reduces the computational burden of the method. Since we assume that cracks are the only scatterers present inside the medium, the contrast values are either $\chi_{\cdot} = 0$ or $\chi_{\cdot} = -1$ for any elastic parameter (however, we will not force the contrast values to be either 0 or -1). We therefore impose the following equality constraints:

$$\chi_{\mu} = \chi_K, \quad \chi_{\rho} = \chi_K.$$

The inequality $-1 \leq \chi_K \leq 0$ can also complement the set of constraints.

4.4.5 Numerical implementation and efficiency

For the actual implementation of the method, we must first select a suitable discretization scheme of the wave equation (4.2). All the operators defined in (4.1i) and (4.1ii) are subject to this choice. After that, the application of algorithm of Section 4.4.2 is straightforward.

A convenient choice for the discretization of the wave equation is the finite-element method (see, for example, Min et al. [2003]). The elements of the discretized domain are squared cells where the physical properties are set to be constant. The wavefields are represented by \mathbb{Q}_1 polynomials [Quarteroni, 2010] within each square and are uniquely determined by the value assumed on the corners of the element (Fig. 4.3). To improve the accuracy of the method, we replace the consistent mass term of the resulting linear operator with a linear combination of the lump and consistent mass matrices. The linear weights are then chosen to minimize the numerical dispersion of the method.

As argued in Min et al. [2003], free-surface boundary conditions are automatically imposed on the boundary of each region where the elastic parameters K , μ , and ρ are set to zero. This feature is convenient in that not only

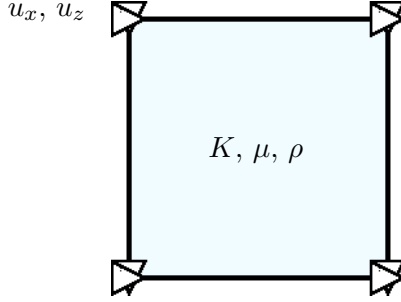


Figure 4.3: Schematic representation of a single finite element of the discretized domain, represented by a squared cell. The elastic properties K, μ, ρ (bulk modulus, shear modulus and density, respectively) are defined to be constant inside the cell. The two components of the particle displacement $\mathbf{u} = (u_x, u_z)$ (denoted by triangle pointing, respectively, right and down) are both defined on the corners of the cell.

front-wall ∂D_{ext} and back-wall ∂D_{int} can be implemented by simply setting proper values for the parameters, but also cracks (whose boundaries are then effectively treated as free surfaces).

Classical direct methods based on factorization, such as the LU decomposition with nested dissection ordering [George and Liu, 1981], are readily available for 2-D problems. A key aspect of our methodology is the fact that the background does not change throughout the inversion and the LU factorization is only computed once. This is the implicit form of Green's operator \mathcal{G}_b . In contrast, traditional full-waveform inversion requires the solution of the wave equation with respect to a changing medium and the factorization must be recomputed each time. The method also requires the pre-computation of the domain-to-receiver Green's operator (4.6) and the incident fields, as well as the storage of the estimated wavefields \mathbf{u} .

4.5 Results

In this section, we demonstrate the capabilities of the alternating inversion algorithm. We set up a synthetic experiment to illustrate the difficulties with the linear imaging of vertical cracks as well as the capability of the non-linear method. We also present an application to measured data and perform imaging of a steel plate with three drill holes. This result represents a preliminary but promising study on the imaging of more complex scatterers (as cracks).

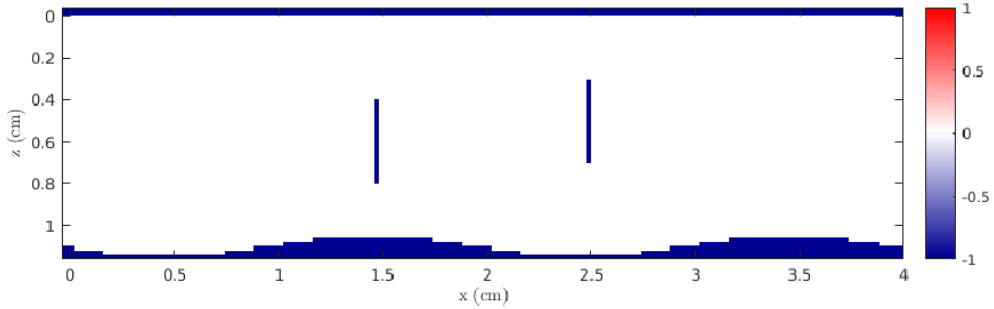


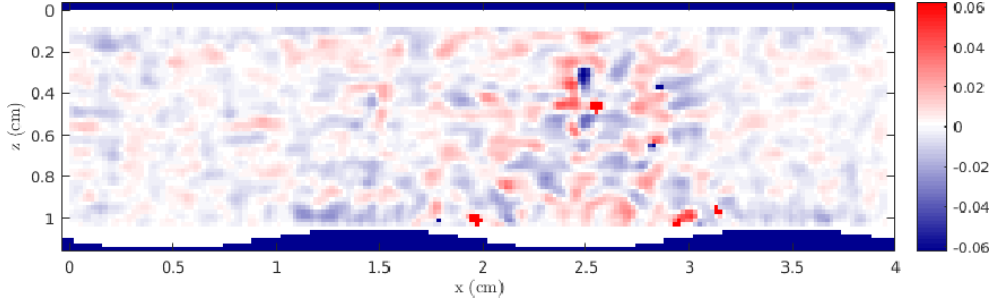
Figure 4.4: *Synthetic model for the experiment described in Section 4.5.1. This picture represents the contrast model for the bulk modulus χ_K . The internal grid points where $\chi_K = 0$ correspond to the steel background, while $\chi_K = -1$ corresponds to the crack position, or a void region.*

4.5.1 Synthetic crack experiment

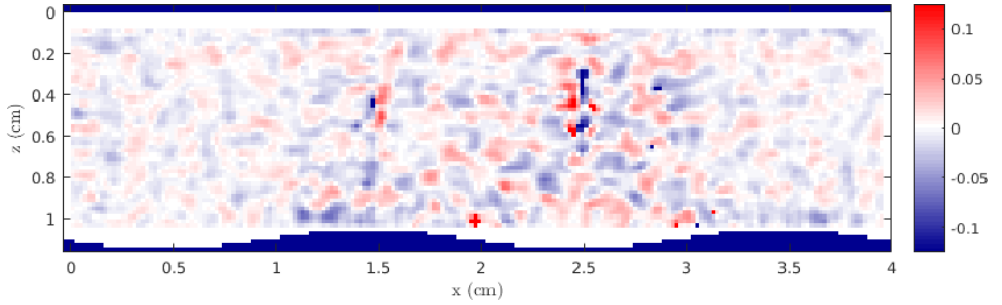
This experiment tests the method on a synthetic crack model, shown in Fig. 4.4. The computational domain measures 1 cm by 4 cm and consists of two vertical cracks embedded in a steel plate with a flat front-wall ∂D_{ext} and a sinusoidal back-wall ∂D_{int} . The shapes of these interfaces are known. The shape of the back-wall is not representative of a realistic pipe, but it intends to illustrate the flexibility of the modeling engine. The background model and its boundaries are given and correspond to homogeneous steel properties: $K_b = 1.66 \times 10^{11}$ Pa, $\mu_b = 8.29 \times 10^{10}$ Pa, $\rho_b = 7850$ kg m $^{-3}$ (equivalently, the P- and S-wave velocities are $c_{P,b} = 5940$ m s $^{-1}$, $c_{S,b} = 3250$ m s $^{-1}$). The cracks are simulated by setting a void region in the cells that correspond to the crack positions ($K = 0$, $\mu = 0$, $\rho = 0$).

We model the synthetics according to 20 line sources evenly distributed along the front-wall. Each source is applied as a normal stress boundary condition exerted on ∂D_{ext} . The receiver configuration consists of 40 stations on ∂D_{ext} that record the vertical particle velocity. We invert a total of 25 frequencies, equally sampled between 1 MHz and 3 MHz. Note that the shortest propagation wavelength roughly equals 1 mm, while the crack width is 0.2 mm. We also add some artificial Gaussian white noise to ease our inverse crime.

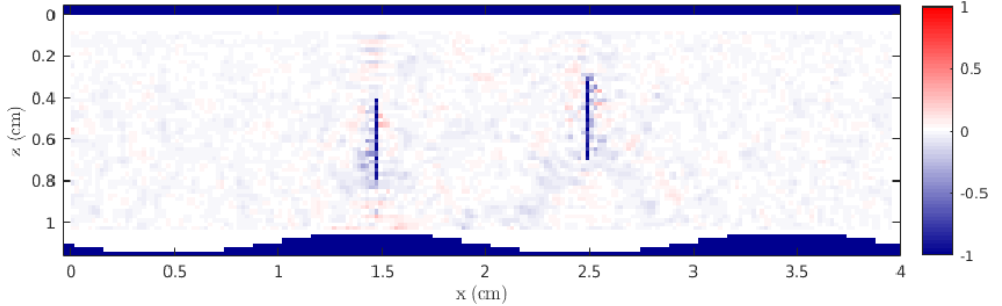
Figure 4.5 depicts the inversion results. We draw an interesting comparison between different stages of the alternating inversion method, corresponding to the Born approximation (Fig. 4.5a), and the results after 1 and 29 wavefield updates (respectively Fig. 4.5b and Fig. 4.5c). The linear assumption underly-



(a) Inverted χ_K from the Born approximation



(b) Inverted χ_K after 1 wavefield update



(c) Inverted χ_K after 29 wavefield updates

Figure 4.5: Results of the experiment presented in Section 4.5.1 at different stages of the alternating inversion algorithm. Fig. 4.5a corresponds to the Born approximation, for which only the tips of the cracks are faintly recovered. After 1 wavefield update, the inverted contrast model of Fig. 4.5a already improves significantly, both in amplitude and shape. The final result in Fig. 4.5c, after 29 wavefield updates, displays a satisfactory inversion of the two cracks.

ing the result of Fig. 4.5a allows only for a faint identification of the tips of the two cracks, since we lack enough illumination for the sides of these scatterers. Once the interplay between the two cracks and the walls are gradually taken into consideration, the inversion starts to build up the correct amplitude and shape of the cracks.

4.5.2 Imaging from laboratory measurements

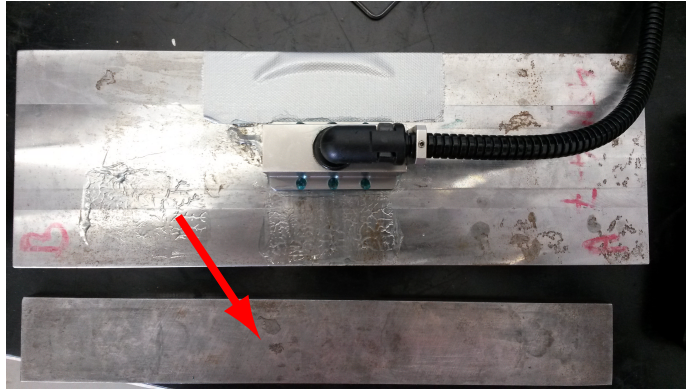
This experiment consists in the inversion of data measured in the laboratory. The equipment and the test sample are shown in Fig. 4.6. The sample is a 10.2 mm thick steel block in which three holes have been drilled side-wise each with a diameter of approximately 0.5 mm. The relative distance between the first and second hole, and the second and the third hole is, respectively, 1 mm and 1.5 mm.

The linear array probe consists of 64 transducer elements. Fig. 4.7 displays a raw shot gather. As shown in the spectral analysis of Fig. 4.8, the effective frequency bandwidth ranges from 1.5 MHz to 3.5 MHz. The nominal peak frequency of the probe is 2 MHz. At the highest frequency, the shortest propagation wavelength is around 1 mm (to be compared with the drill hole size). We assume that the transducers act as normal stress at the front wall and record the vertical particle velocities. For the inversion, we pick a small subset of the data pertaining to just 6 frequencies, equally sampled within the bandwidth 1–3 MHz.

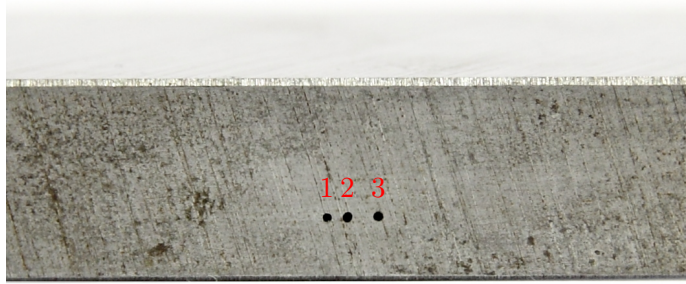
We remark that, in principle, data with higher frequency content are routinely recorded and utilized by many linear imaging algorithms. However, full-waveform inversion is a computer-intensive procedure, and the time and memory complexity increases with frequency (since the discretization of the computational domain is determined by sampling heuristics for the minimum propagating wavelength). This is partially compensated by the higher resolution for the final output. The choice of the highest frequency is therefore limited by the available computational resources.

Some pre-processing steps are necessary to correct for the idealized assumptions behind the acquired data. In doing so, we adopt an approach which is pragmatic in nature, yet effective:

- in our initial comparison with synthetic data, we observed a large mismatch for the S-wave direct arrivals, probably due to the gel coupling



(a) Linear array probe leaning on a steel block (view from the top)



(b) Steel block used for the experiment (view from the side)

Figure 4.6: Laboratory setup for the experiment in Section 4.5.2. In Fig. 4.6a, we show how the measurements are taken with a linear probe array. Note that some coupling gel has been applied on top of the piece. The red arrow indicates the actual test sample in examination. In Fig. 4.6b, the block is viewed from the side, revealing three drill holes of 0.5 mm diameter. The distance between the first and second hole is 1 mm, while the distance between the second and third hole is 1.5 mm.

(a thin film of gel is applied to the block before taking the recordings). We resolve this issue by selectively removing the direct arrivals from the modeled data. This can be easily achieved with a modeling exercise where the back-wall is replaced by an absorbing boundary;

- another complication is given by the 2-D assumption underlying the modeled data, which assumes infinite line sources. In practice, each element of the probe has a finite length (and a finite width!), requiring some form of correction to the data. By modeling finite-length line sources, we observed a transition from cylindrical spreading at early times to geomet-

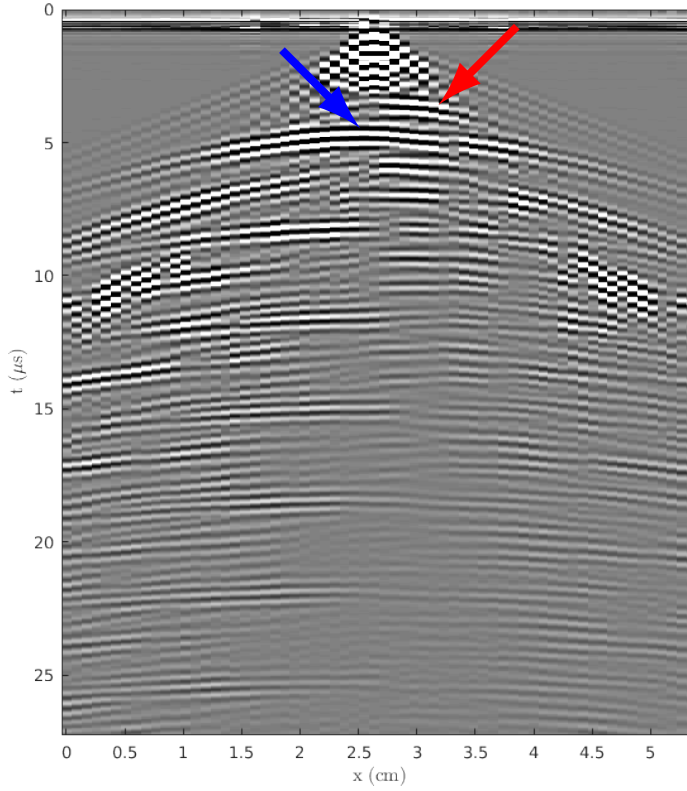


Figure 4.7: *Example of the raw measurements taken for the experiment in Section 4.5.2. Surface waves and the multiple reflections are clearly discernible in the picture. We highlight the reflection event due to the three-hole scatterer with a red arrow, and the first reflection against the back-wall with a blue arrow.*

rical spreading later on. The transition time is determined experimentally, after which a classical cylindrical/geometric time conversion can be applied. Note that this approach is somehow vitiated by an underlying acoustic assumption.

Another pre-processing adjustment is needed to determine the wavelet for the calibration of the synthetics. This can be easily done with a separate experiment where data is acquired over an homogeneous region of the steel block. The source wavelet is then the only unknown, and can be estimated by determining the best fit between measured and modeled data. This is the approach adopted here. Alternatively, the wavelet can be jointly estimated with the material properties, as in Li et al. [2013].

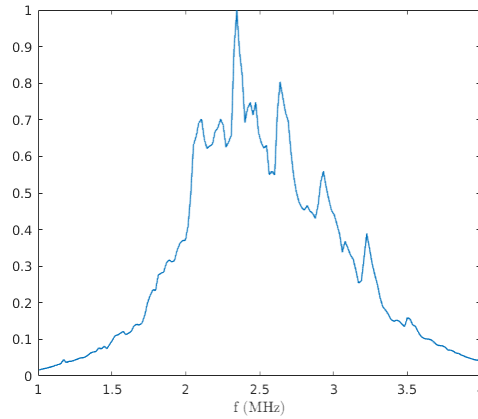


Figure 4.8: *Spectrum of the data shown in Fig. 4.7, averaged over the receiver position*

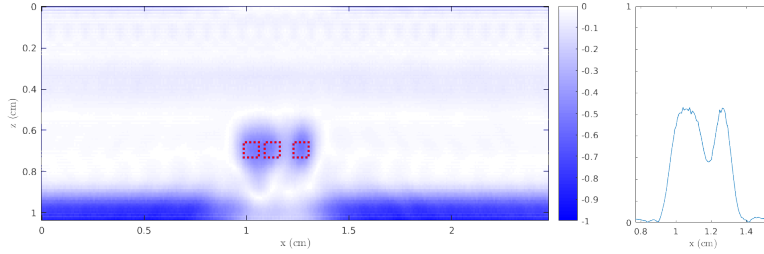
We display the result of the inversion in Fig. 4.9, with a comparison with the inverse wavefield extrapolation method [Pörtzgen et al., 2007], of which the output is the structural image displayed in Fig. 4.9a (properly normalized for a better juxtaposition with the contrast values). In Fig. 4.9b, we show the inverted contrasts after the Born inversion (first stage of the alternating algorithm), and in Fig. 4.9c we show the final result after 4 wavefield updates. In Fig. 4.10, we also provide a comparison between the measured and modeled data, obtained for the final inversion result. We observe that the three drill holes are adequately recovered already under the Born approximation, but the inversion is spoiled by some artifacts. The final result is successful in removing these artifacts and in recovering the scatterers with a much improved amplitude and resolution. Note that, due to the simple nature of the defects, the linear inversion delivers a much better result compared to the synthetic example in Section 4.5.1.

4.6 Summary and conclusion

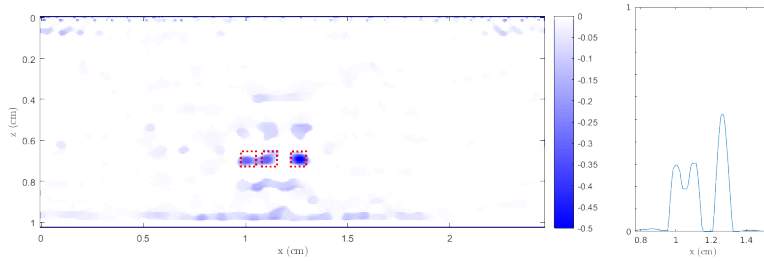
We have advocated imaging methods that exploit the non-linear character of the relationship between ultrasound measurements and material properties. This is especially important for flaws that are difficult to detect by using only single-scattered data. This may be caused by the orientation of the shape/constellation of these flaws in the sample under investigation. In these cases, taking multiple scattering into account is vital for resolving the required

level of detail for accurate characterization and sizing of defects.

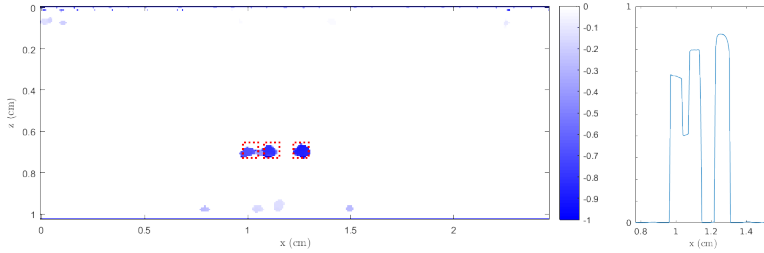
The synthetic experiment presented in this paper confirms that the recovery of vertical defects, while being poor under the linear assumption, can be improved by estimating increasingly higher orders of scattering in the data. Preliminary results on actual measured data suggest that the method should be suited for real-world applications.



(a) Structural image obtained from the inverse wavefield extrapolation method [Pörtzgen et al., 2007] (normalized for comparison), and horizontal section of the absolute value of the amplitudes ($z = 0.7$ cm)



(b) Inverted χ_K from linear inversion, and horizontal section of the absolute value of the amplitudes ($z = 0.7$ cm)



(c) Inverted χ_K from non-linear inversion, and horizontal section of the absolute value of the amplitudes ($z = 0.7$ cm)

Figure 4.9: Results obtained from the laboratory data (Section 4.5.2). The true position and size of the drill holes is indicated by the red squares. In Fig. 4.9a, we show the processed output of the inverse wavefield extrapolation method [Pörtzgen et al., 2007], which represents a structural image of the scatterer. Fig. 4.9b displays the result from the Born inversion, which is the same as the first step of the alternating inversion. Note how the position of the drill holes is essentially recovered, but the image is affected by artifacts and the amplitude $\chi_K = -1$ is not quite correct. Also, compare the scales of the colorbars in Fig. 4.9b and Fig. 4.9c. In Fig. 4.9c we show the final result of the inversion after 4 wavefield updates. Compared to Fig. 4.9b, the resolution of the drill holes is improved and most of the artifacts have been removed.

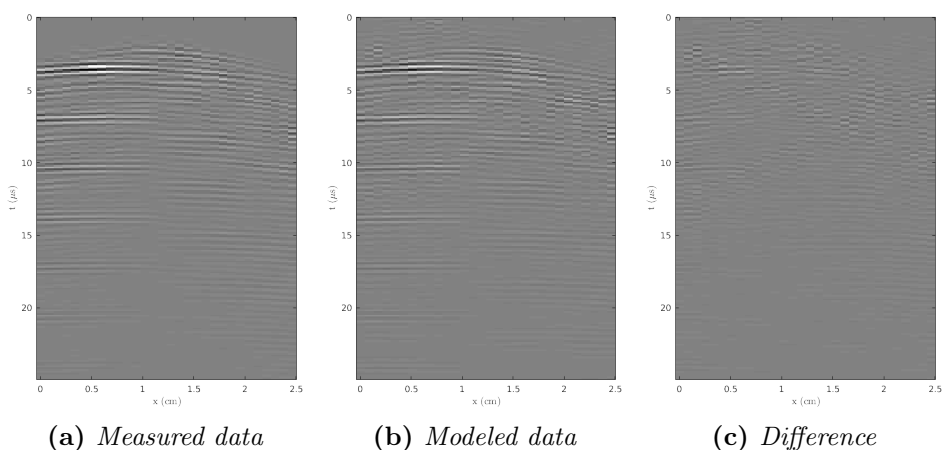


Figure 4.10: Comparison of the measured and modeled data for the final inversion result of the experiment in Section 4.5.2. Note that in Fig. 4.10a, the direct arrivals have been removed and some time damping has been applied. The modeled data in Fig. 4.10b shows a good match with Fig. 4.10a, which is further demonstrated by the difference between the two panels in Fig. 4.10c.

Conclusions and recommendations for further research

5.1 Summary

The main topic of this thesis is the elastic inverse scattering problem. For its solution, we proposed and discussed a novel inversion algorithm. The peculiarity of the method resides in the setup of an optimization process where the search space comprises not only model parameters but also wavefields, whereas traditional full-waveform inversion is based on a cost function depending only on the model's material property variables, and the wave equation is treated as a constraint. Wavefields become, therefore, independent on elastic properties and the wave equation is not directly enforced at each iteration. This point of view leads to an alternating inversion algorithm, which consists of distinct updating procedures for both model properties and wavefield variables.

This method complies with some of the ideas enunciated by the extended modeling principle [Symes, 2008]. Theoretically, this class of methods achieves a relaxation of the non-linearity of the problem by looking for a solution in a extended optimization space. Noteworthy methods akin to this philosophy are the contrast-source inversion of van den Berg and Kleinman [1997] and the wavefield reconstruction method of van Leeuwen and Herrmann [2013]. These works share many similarities with the proposed alternating scheme.

The uncoupling of properties and wavefields also has important implica-

tions from a computational point of view. On the one hand, a linear relationship between properties and data is exploited for the property update step, which greatly benefits from straightforward parallelization. On the other, it has been shown that the algorithm does not require the wave equation to be fully solved for each model estimated during the inversion. This is clearly different for basic full-waveform inversion, whose computational work must leverage on efficient methods to solve the wave equation. The alternating method employs an updating scheme for the wavefield which only involves the solution of the wave equation relative to a known background model. Since the background model does not change (frequently) during the computation, we can employ factorization-based direct methods to speed up the computation. Alternatively, we could employ iterative methods with proper preconditioning (as the multigrid-based method investigated in Chapter 3).

The inversion algorithm has been successfully tested on real-world applications. In particular, we dealt with the realm of ultrasound imaging for the detection of stress-corrosion cracks in oil and gas pipe lines. The method has been able to correctly image size and position of drill holes in a steel block, as shown in Chapter 4, and is a promising tool to predict vertical defects (as stress-corrosion cracks), for which the treatment of non-linear effects of the wave propagation is a fundamental requisite.

5.2 Future directions

In this section, we sketch some of the goals we have in mind to follow up on the research presented in this thesis.

It is clear that 2-D inversion is limited to only a few applications, hence it is crucial to gauge the feasibility of a 3-D extension. As explained in Chapter 2, the main challenge is the need for storage of the matrix factorization and wavefields, for each source and frequency. An alternative to factorization methods is, for example, the iterative solution method presented in Chapter 3 (along with many other attempts, see for instance Gordon and Gordon [2010] or Baumann et al. [2016]), but the storage of the wavefields would remain problematic. A different proposal has been introduced in Abubakar et al. [2009], based on a hierarchical low-to-high single frequency inversion strategy. This in principle requires the storage of the factorized matrix and wavefields only for the current frequency. We preliminarily tested this idea in Section 2.4.4, under the 2-D assumption. Another suggestion is to divide the computational

domain into manageable portions through data localization (e.g., via redatuming over the target subdomain) in order to setup independent local inversion problems [Haffinger et al., 2013; Willemsen et al., 2016].

The alternating inversion scheme has delivered convincing results on synthetic data and the next logical step is the study of field data.

Examples of elastic inversion applications

A first important application is the so-called *near-surface problem*. Due to its unconsolidated nature, the near-surface is a highly heterogeneous portion of the subsurface that has a strong impact on seismic data. Therefore, the removal of these effects is a prerequisite for most of the exploration seismic applications that focus on deep targets. Moreover, it can be a valuable tool to avoid drilling hazards. In order to do so, a detailed knowledge of the near-surface is required. The effect of the surface on wave propagation, which leads to surface waves, must be taken into account. The more apparent difference with the work presented in Chapter 2 is the need to employ proper boundary conditions in the wave equation. Some preliminary synthetic data studies have been carried out. One example is shown in Figures 5.1 and 5.2.

The second application, discussed in Chapter 4, is the non-destructive inspection for stress-corrosion cracking in steel pipes. As previously argued, it is an essential tool to ensure the integrity of pipe lines in the oil and gas industry. This application strays away from the typical seismic scale, but the principles of elastic wave inversion remain exactly the same. We obtained encouraging results for laboratory measurements on a drill hole test sample, and the next step will consist in the imaging of stress-corrosion cracks, for which the account of non-linear effects of the wave propagation will be key.

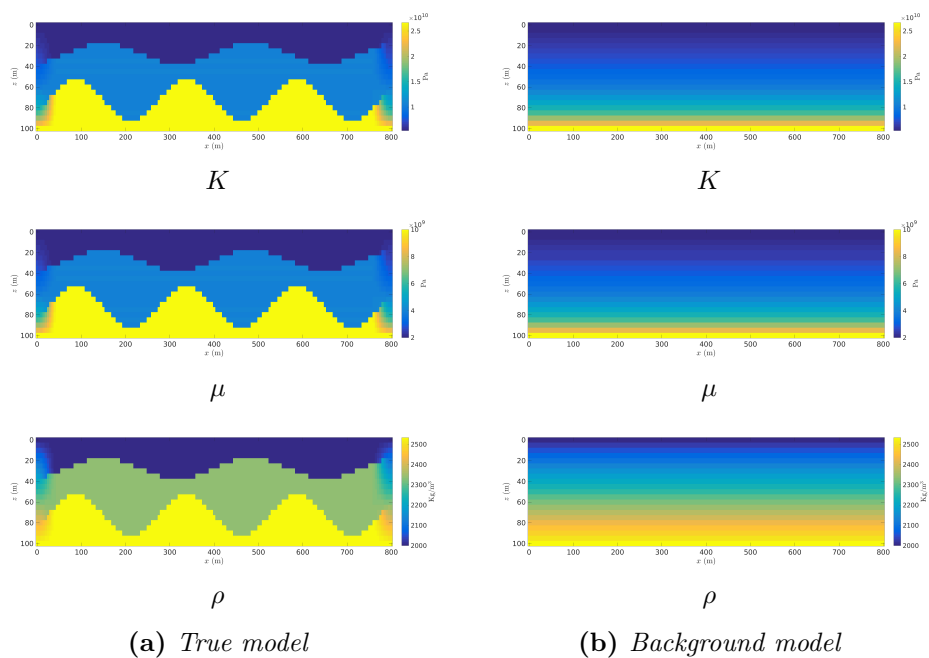


Figure 5.1: Synthetic data inversion exercise for a near-surface model: (a) true model, (b) background model. Sources and receivers are placed along the interface $z = 0$, which corresponds to a free surface.

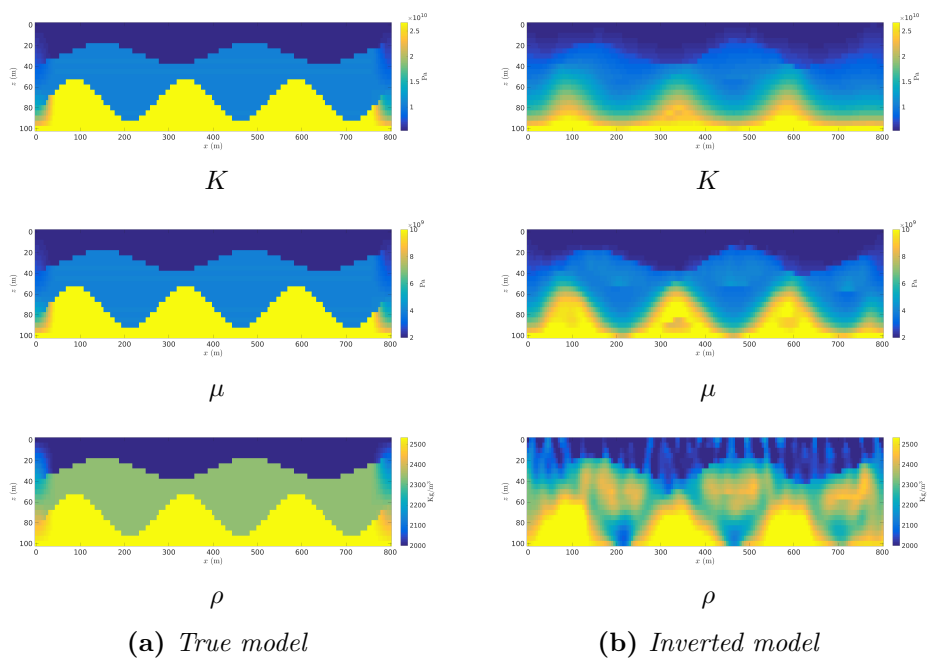


Figure 5.2: Synthetic data inversion exercise for a near-surface model: (a) true model, (b) inverted model. Sources and receivers are placed along the interface $z = 0$, which corresponds to a free surface.

Appendix:

An optimized staggered-grid finite-difference method

In this Appendix, we describe the modeling scheme underlying the inversion method illustrated in the body of the paper. We choose a finite-difference method based on the staggered-grid formulation proposed by Virieux [1986], characterized by the discretization of the wavefield components on different grids.

From a numerical point of view, the most interesting aspect of the staggered-grid scheme is the independence of the numerical dispersion properties with respect to the P-to-S velocity ratio c_P/c_S , contrary to ‘single-grid’ formulations, for which the different components of the wavefield are represented on the same grid. Moreover, its dispersion properties can be sensibly ameliorated if the finite differences representing the partial derivatives in (2.3) are properly optimized. Indeed, we will present a 9-point scheme with a phase and group velocity relative error of less than 1 %, when the propagation wavelength is sampled by a minimum number of 5 points and for a wide range of velocity ratios. This method is quite amenable for standard factorization techniques as the LU decomposition [George and Liu, 1981].

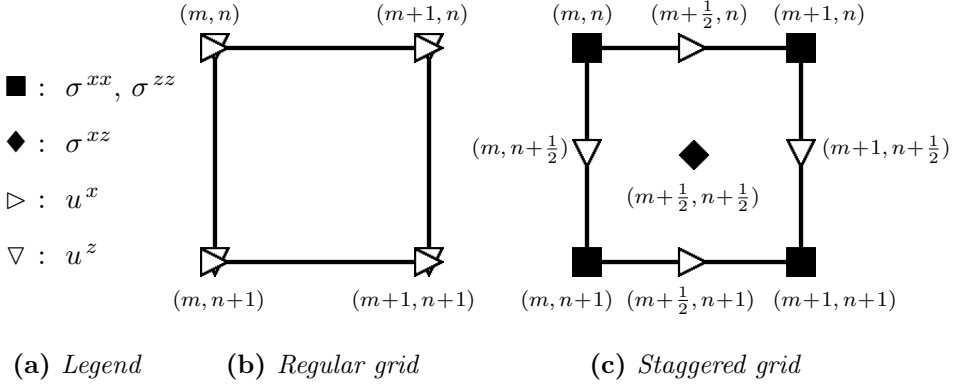


Figure A.1: Discretization for the (b) single-grid and (c) staggered-grid formulations. According to the staggered-grid scheme, the elastic wavefield components (comprising particle displacement (u_x, u_z) and stress field $\sigma^{ab} = \sum_{cd} C^{ab,cd} \partial_c u^d$, where $C^{ab,cd}$ is the stiffness tensor) are defined on different grids. The grid points where each of these wavefield are defined are denoted by different symbols (a). Note that the z -axis points downwards.

A.1 The standard staggered-grid scheme

A key aspect of the staggered-grid scheme is in the approximation of the horizontal and vertical components of the solution of the wave equation (2.5) on *shifted* computational grids (see Fig. 3.1). To better illustrate this idea, we introduce the shift operators:

$$\tau^x u(\mathbf{x}) = u(\mathbf{x} + (1/2, 0)h), \quad \tau^z u(\mathbf{x}) = u(\mathbf{x} + (0, 1/2)h), \quad (\text{A.1})$$

acting on scalar functions u . For vectorial functions $\mathbf{u} = (u^x, u^z)$, we define the shift operator $\mathcal{T} \mathbf{u} = (\tau^x u^x, \tau^z u^z)$. \mathcal{T} entails the following transformation on the wave equation (2.2):

$$\bar{\mathcal{H}} \bar{\mathbf{u}} = \bar{\mathbf{f}}, \quad \bar{\mathcal{H}} = \mathcal{T} \mathcal{H} \mathcal{T}^{-1}, \quad (\text{A.2})$$

where $\bar{\mathbf{u}} = \mathcal{T} \mathbf{u}$ and $\bar{\mathbf{f}} = \mathcal{T} \mathbf{f}$. Note that, throughout this section, we will not denote explicitly the dependency of the linear operator from the model parameters, as it was done, for example, in equation (2.2). We look for an approximate solution $\bar{\mathbf{u}}(\mathbf{x})$ of (A.2) on the set of collocation points $G = \{(i h, j h) : i, j \in \mathbb{Z}\}$, determined by the grid spacing h . In what follows, we will use the bar notation to indicate transformed quantities through \mathcal{T} .

For simplicity, we start with the homogeneous case. The wave equa-

tion (2.3) then simplifies to:

$$\bar{\mathcal{H}} = -\rho \omega^2 \mathcal{I} - \bar{\mathcal{D}},$$

$$\bar{\mathcal{D}} = \begin{pmatrix} (K + \frac{4}{3}\mu) \bar{\partial}_{xx} + \mu \bar{\partial}_{zz} & (K + \frac{1}{3}\mu) \bar{\partial}_{xz} \\ (K + \frac{1}{3}\mu) \bar{\partial}_{xz} & \mu \bar{\partial}_{xx} + (K + \frac{4}{3}\mu) \bar{\partial}_{zz} \end{pmatrix}. \quad (\text{A.3})$$

A finite-difference scheme is determined by a specific choice for the discretization of the second-order derivatives ∂_{xx} , ∂_{xz} and ∂_{zz} . We denote the discretized version of (A.3) by:

$$\bar{\mathbf{H}} = -\rho \omega^2 \mathbf{I} - \bar{\mathbf{D}},$$

$$\bar{\mathbf{D}} = \begin{pmatrix} (K + \frac{4}{3}\mu) \bar{\mathbf{d}}_{xx} + \mu \bar{\mathbf{d}}_{zz} & (K + \frac{1}{3}\mu) \bar{\mathbf{d}}_{xz;x} \\ (K + \frac{1}{3}\mu) \bar{\mathbf{d}}_{xz;z} & \mu \bar{\mathbf{d}}_{xx} + (K + \frac{4}{3}\mu) \bar{\mathbf{d}}_{zz} \end{pmatrix}. \quad (\text{A.4})$$

Note that we explicitly made a distinction between the cross derivatives $\bar{\mathbf{d}}_{xz;x}$ and $\bar{\mathbf{d}}_{xz;z}$. To give a neat expression of $\bar{\mathbf{d}}_{xx}$, $\bar{\mathbf{d}}_{zz}$ and $\bar{\mathbf{d}}_{xz}$, we will employ the stencil notation:

$$S \triangleq \begin{bmatrix} s_{-1,-1} & s_{0,-1} & s_{1,-1} \\ s_{-1,0} & s_{0,0} & s_{1,0} \\ s_{-1,1} & s_{0,1} & s_{1,1} \end{bmatrix}, \quad S u_{\cdot,\cdot} |_{i,j} = \sum_{i',j'} s_{i',j'} u_{i+i',j+j'}. \quad (\text{A.5})$$

which summarizes the action of a linear operator S on a generic grid function $u_{\cdot,\cdot}$. Note that, following the geophysical convention, the z -axis points downwards. With this in mind, the stencils of the staggered-grid scheme are:

$$\mathbf{I} \triangleq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\bar{\mathbf{D}}: \begin{cases} \bar{\mathbf{d}}_{xx} \triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}, & \bar{\mathbf{d}}_{xz;x} \triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{bmatrix}, \\ \bar{\mathbf{d}}_{xz;z} \triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, & \bar{\mathbf{d}}_{zz} \triangleq \frac{1}{h^2} \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \end{cases} \quad (\text{A.6})$$

With a slight abuse of notation, we only indicated one stencil for the 2×2 block identity \mathbf{I} , meaning that the same stencil is used for both the diagonal scalar blocks. The resulting numerical dispersion is depicted in Fig. A.2.

A.2 Optimization of the staggered-grid method

To obtain a more accurate scheme, we use (A.6) as a starting point. We observe that, in the stencils (A.6), many entries correspond to zero. As a general rule of thumb, exploiting all the neighboring grid points of the stencils leads to more accurate schemes. To do so, we borrow some of the ideas presented in Min et al. [2000, 2004], developed for a single-grid formulation:

- treatment of the mass term: a new mass term can be obtained by averaging the neighboring points (i, j) , $(i \pm 1/2, j)$, $(i, j \pm 1/2)$, $(i \pm 1/2, j \pm 1/2)$,

$$\begin{aligned} \bar{\mathbf{I}}^{\text{opt}} \bar{u}_{\cdot,\cdot}^a |_{i,j} = & \\ & \alpha_0 \bar{u}_{i,j}^a + \\ & \alpha_1 (\bar{u}_{i,j-1/2}^a + \bar{u}_{i-1/2,j}^a + \bar{u}_{i+1/2,j}^a + \bar{u}_{i,j+1/2}^a) + \\ & \alpha_2 (\bar{u}_{i-1/2,j-1/2}^a + \bar{u}_{i-1/2,j+1/2}^a + \bar{u}_{i+1/2,j-1/2}^a + \bar{u}_{i+1/2,j+1/2}^a), \end{aligned} \quad (\text{A.7})$$

with weights $\alpha_0, \alpha_1, \alpha_2$. The superscripts denote the different components of the particle displacement: $a = x, a = z$;

- averaging of the double derivatives: the double derivative with respect to x , for instance, might be obtained by averaging over the grid points (i, j) and $(i, j \pm 1/2)$,

$$\bar{d}_{xx}^{\text{opt}} \bar{u}_{\cdot,\cdot}^a |_{i,j} = \beta_1 \bar{d}_{xx} \bar{u}_{\cdot,\cdot}^a |_{i,j-1/2} + \beta_0 \bar{d}_{xx} \bar{u}_{\cdot,\cdot}^a |_{i,j} + \beta_1 \bar{d}_{xx} \bar{u}_{\cdot,\cdot}^a |_{i,j+1/2}. \quad (\text{A.8})$$

Each term of this combination is computed by using the standard stencil \bar{d}_{xx} defined in (A.6) (the treatise of $\bar{d}_{zz}^{\text{opt}}$ follows similarly);

- cross derivatives: these are directly obtained from a special 3-point discretization $\bar{d}_{x\cdot}^{\text{opt}}, \bar{d}_{z\cdot}^{\text{opt}}$ of the first-order derivatives. For example:

$$\begin{aligned} \bar{d}_{xz;x}^{\text{opt}} &= \bar{d}_{x;x}^{\text{opt}} \bar{d}_{z;x}^{\text{opt}}, \\ \bar{d}_{x;x}^{\text{opt}} \bar{u}_{\cdot,\cdot}^x |_{i,j} &= [-\gamma \bar{u}_{i-1,j}^x - (1 - 2\gamma) \bar{u}_{i,j}^x - (\gamma - 1) \bar{u}_{i+1,j}^x] / h, \\ \bar{d}_{z;x}^{\text{opt}} \bar{u}_{\cdot,\cdot}^x |_{i,j} &= [(\gamma - 1) \bar{u}_{i-1,j}^x + (1 - 2\gamma) \bar{u}_{i,j}^x + \gamma \bar{u}_{i+1,j}^x] / h, \end{aligned} \quad (\text{A.9})$$

and analogously for $\bar{d}_{xz;z}^{\text{opt}}$, by antisymmetry. Note that $\gamma = 1$ corresponds to the cross-derivatives in (A.6).

Whenever the value of $\bar{u}_{\cdot,\cdot}^a$ is needed between some grid points where $\bar{u}_{\cdot,\cdot}^a$ is defined, a bilinear interpolation is used. The resulting finite-difference scheme can be readily stated in stencil notation:

$$\bar{\mathbf{I}}^{\text{opt}} \triangleq \begin{bmatrix} \frac{\alpha_2}{4} & \frac{\alpha_1 + \alpha_2}{2} & \frac{\alpha_2}{4} \\ \frac{\alpha_1 + \alpha_2}{2} & \alpha_0 + 2\alpha_1 + \alpha_2 & \frac{\alpha_1 + \alpha_2}{2} \\ \frac{\alpha_2}{4} & \frac{\alpha_1 + \alpha_2}{2} & \frac{\alpha_2}{4} \end{bmatrix},$$

$$\bar{\mathbf{D}}^{\text{opt}}: \begin{cases} \bar{d}_{xx}^{\text{opt}} \triangleq \frac{1}{h^2} \begin{bmatrix} \frac{\beta_1}{2} & -\beta_1 & \frac{\beta_1}{2} \\ \beta_0 + \beta_1 & -2(\beta_0 + \beta_1) & \beta_0 + \beta_1 \\ \frac{\beta_1}{2} & -\beta_1 & \frac{\beta_1}{2} \end{bmatrix}, \\ \bar{d}_{xz;x}^{\text{opt}} \triangleq -\frac{1}{h^2} \begin{bmatrix} \gamma(\gamma - 1) & (1 - 2\gamma)(\gamma - 1) & (\gamma - 1)^2 \\ \gamma(1 - 2\gamma) & (1 - 2\gamma)^2 & (\gamma - 1)(1 - 2\gamma) \\ \gamma^2 & (1 - 2\gamma)\gamma & (\gamma - 1)\gamma \end{bmatrix}, \end{cases} \quad (\text{A.10})$$

and similarly for $\bar{d}_{xz;z}^{\text{opt}}$ and $\bar{d}_{zz}^{\text{opt}}$. We observe that the scheme is consistent for $h \rightarrow 0$ when the following constraints are set:

$$\alpha_2 = (1 - \alpha_0 - 4\alpha_1)/4, \quad \beta_2 = (1 - \beta_0)/2. \quad (\text{A.11})$$

Non-homogeneous case If the elastic properties K , μ and ρ are varying functions, the proposed scheme has to be adapted accordingly. The medium properties will be now discretized by grid functions. In the following, we are adopting the notation $\bar{m}_{\cdot,\cdot}^{abc\dots}$ to represent any of the above gridded properties, shifted correspondingly to the superscripts. For instance, similarly to the shifted particle displacement notation: $\bar{m}_{i,j}^{ab} = \tau^a \tau^b m(ih, jh)$. The scheme is then put in place by the following modifications:

- mass term: the continuous quantity $\tau^a (\rho u^a)$ is discretized by $\bar{\Gamma}^{\text{opt}} (\bar{\rho}_{\cdot,\cdot}^a, \bar{u}_{\cdot,\cdot}^a)$, following the definition (A.7) for $\bar{\Gamma}^{\text{opt}}$;
- double derivatives: to approximate the operator $\partial_x m \partial_x$ we resort to the same idea of averaging employed in (A.8) for the homogeneous case. It suffices, then, to define:

$$\begin{aligned} \tau^a (\partial_x m \partial_x u^a) (i h, j h) &\approx -\bar{m}_{i-1/2,j}^a (-\bar{u}_{i-1,j}^a + \bar{u}_{i,j}^a) + \\ &\quad \bar{m}_{i+1/2,j}^a (-\bar{u}_{i,j}^a + \bar{u}_{i+1,j}^a), \end{aligned}$$

and, in a similar way, $\tau^a (\partial_z m \partial_z u^a)$;

- cross derivatives: the approximation of $\partial_x m \partial_z$ and $\partial_z m \partial_x$ are determined directly by the first-order operators (A.9):

$$\begin{aligned} \tau^z (\partial_x m \partial_z u^x) (i h, j h) &\approx d_{x;x}^{\text{opt}} (\bar{m}_{\cdot,\cdot}^{xz} \cdot d_{z;x}^{\text{opt}} \bar{u}_{\cdot,\cdot}^x) |_{i,j}, \\ \tau^z (\partial_z m \partial_x u^x) (i h, j h) &\approx d_{z;x}^{\text{opt}} (\bar{m}_{\cdot,\cdot} \cdot d_{x;x}^{\text{opt}} \bar{u}_{\cdot,\cdot}^x) |_{i,j}, \end{aligned}$$

where the dot product indicates standard multiplication of grid functions. The quantities $\tau^x (\partial_x m \partial_z u^z)$ and $\tau^x (\partial_z m \partial_x u^z)$ are determined very similarly.

Once again, the value of $\bar{m}_{\cdot,\cdot}^{abc\dots}$ is interpolated between grid points whenever needed.

We remark that, by design, the linear system arising from the discretization is real-symmetric.

A.3 Numerical accuracy

In the finite-difference scheme (A.10), the free parameters $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1$ and γ can be adjusted in order to optimize the dispersion properties. Hence, an objective functional which measures the relative error between analytic and numerical P- and S-wave phase velocities can be set up. It is found that the values

$$\alpha_0 = 0.2597, \quad \alpha_1 = 0.1859, \quad \beta_0 = 0.5789, \quad \gamma = 0.8804,$$

along with (A.11), produce a relative error of phase and group velocities of less than 1 %, for a minimal discretization rate of 5 points per wavelength

and a wide range of velocity ratios r (Fig. A.2). We also tested the scheme with a heterogeneous problem by comparison with the finite-element method of Min et al. [2003]. The numerical experiment shows excellent agreement with accurately computed solutions (Fig. A.3).

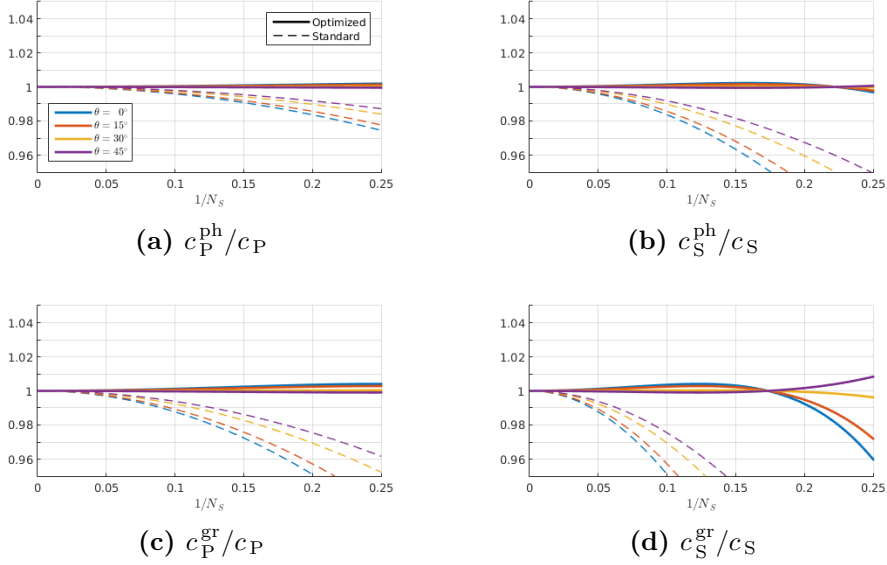
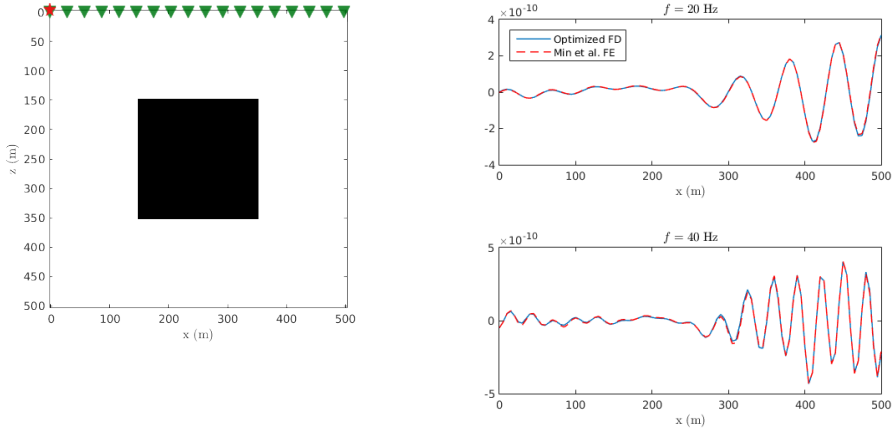


Figure A.2: Numerical dispersion analysis of the optimized staggered-grid finite-difference method and comparison with the standard staggered-grid scheme. We report the P- and S-wave phase and group velocity ratios for different propagation angles θ , and number of samples per S-wavelength N_S . Here, we considered a velocity ratio c_P/c_S of 2, but the analysis is similarly favorable for much higher values (e.g., $r = 100$). The figures display a maximum error of the S-wave group velocity ratio of less than 1% up to 5 points per S-wavelength.



(a) Setting of the numerical experiment (b) Solution comparison (scattered field)

Figure A.3: Comparison of the optimized staggered-grid finite-difference scheme and the finite-element method of Min et al. [2003] for a heterogeneous problem. The computational domain is summarized in Fig. (a). The background model is homogeneous, corresponding to the values $K = 5.33 \times 10^9$ Pa, $\mu = 2 \times 10^9$ Pa and $\rho = 2000 \text{ kg m}^{-3}$. The black region in (a) represents the heterogeneity, defined by $K = 7.27 \times 10^9$ Pa, $\mu = 2.73 \times 10^9$ Pa and $\rho = 2218 \text{ kg m}^{-3}$. The source-receiver setting is also depicted in Fig. (a). For this numerical experiment, we considered a vertical force source. In Fig. (b), a comparison of the real part of u^z (scattered wavefield) is made for the frequencies $f = 20 \text{ Hz}$ and $f = 40 \text{ Hz}$ (which corresponds to a discretization of, respectively, 10 and 5 points per minimum S -wavelength). The solution by finite-elements is obtained for a grid spacing of $h/4$, where h is the step-length corresponding to the grid used for the finite-difference computation.

Bibliography

- Abubakar, A., and Habashy, T. M., 2013, Three-dimensional visco-acoustic modeling using a renormalized integral equation iterative solver: *J. Comput. Phys.*, **249**, 1–12.
- Abubakar, A., Hu, W., Habashy, T. M., and van den Berg, P. M., 2009, Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data: *Geophysics*, **74**, no. 6, WCC47–WCC58.
- Abubakar, A., Pan, G., Li, M., Zhang, L., Habashy, T. M., and van den Berg, P. M., 2011, Three-dimensional seismic full-waveform inversion using the finite-difference contrast source inversion method: *Geophysical Prospecting*, **59**, 874–888.
- Aki, K., and Richards, P. G., 2002, *Quantitative Seismology*: University Science Books, 2 edition.
- Baumann, M., Astudillo, R., Qiu, Y., Ang, E., van Gijzen, M. B., and Plessix, R.-É., An MSSS-Preconditioned Matrix Equation Approach for the Time-Harmonic Elastic Wave Equation at Multiple Frequencies; Technical Report 16-04, DIAM - Delft University of Technology, 2016.
- Bayliss, A., Goldstein, C. I., and Turkel, E., 1983, An iterative method for Helmholtz equation: *J. Comput. Phys.*, **49**, 443–457.
- Beller, M., Doulgeris, P., Gisolf, A., Haffinger, P., Huis in't Veld, R., and Wever, A. K. T., 2015, Resolving carboniferous stacked channel sequences

- with a non-linear AVO technology: 77th EAGE Annual International Conference and Exhibition.
- Berenger, J. P., 1994, A perfectly matched layer for the absorption of electromagnetic waves: *J. Comput. Phys.*, **114**, no. 2, 185–200.
- Berryhill, J. R., 1979, Wave-equation datuming: *Geophysics*, **44**, no. 8, 1329–1344.
- Bertero, M., and Boccacci, P., 2003, Super-resolution in computational imaging: *Micron*, **34**, no. 6, 265–273.
- Bharadwaj, P., Mulder, W., and Drijkoningen, G., 2016, Full waveform inversion with an auxiliary bump functional: *Geophys. J. Int.*, **206**, 1076–1092.
- Biondi, B., and Almomin, A., 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: *Geophysics*, **79**, no. 3, WA129–WA140.
- Bozdag, E., Trampert, J., and Tromp, J., 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: *Geophys. J. Int.*, **185**, no. 2, 845–870.
- Brandt, A., and Livne, O. E., 2011, *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics*: SIAM.
- Briggs, W. L., Henson, V. E., and McCormick, S., 2000, *A multigrid tutorial*: SIAM.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, no. 5, 1457–1473.
- Claerbout, J. F., 1985, *Imaging the earth’s interior*: Blackwell scientific publications Oxford.
- Davydenko, M., and Verschuur, D., 2017, Full-wavefield migration: using surface and internal multiples in imaging: *Geophysical Prospecting*, **65**, no. 1, 7–21.
- de Hoop, A. T., 1995, *The Handbook of Radiation and Scattering of Waves*: Academic Press, Delft.

- de Zeeuw, P. M., 1990, Matrix-dependent prolongations and restrictions in a blackbox multigrid solver: *Journal of Computational and Applied Mathematics*, **33**, no. 1, 1–27.
- Devaney, A., and Sherman, G., 1982, Nonuniqueness in inverse source and scattering problems: *IEEE Transactions on Antennas and Propagation*, **30**, no. 5, 1034–1037.
- Devaney, A. J., 1978, Nonuniqueness in the inverse scattering problem: *J. Math. Phys.*, **19**, no. 7, 1526–1531.
- Elman, H. C., Ernst, G. O., and O’Leary, D. P., 2001, A multigrid method enhanced by Krylov subspace iteration for discrete Helmholtz equations: *SIAM J. Sci. Comput.*, **23**, no. 4, 1291–1315.
- Erlangga, Y. A., and Nabben, R., 2008, Multilevel projection-based nested Krylov iteration for boundary value problems: *J. Sci. Comput.*, **30**, no. 3, 1572–1595.
- Erlangga, Y. A., Vuik, C., and Oosterlee, C. W., 2004, On a class of preconditioners for solving the Helmholtz equation: *Applied Numerical Mathematics*, **50**, 409–425.
- Erlangga, Y. A., Oosterlee, C. W., and Vuik, C., 2006, A novel multigrid based preconditioner for heterogeneous Helmholtz problems: *J. Sci. Comput.*, **27**, no. 4, 1471–1492.
- Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J., 2016, Total-variation regularization strategies in full-waveform inversion: [arXiv:1608.06159](https://arxiv.org/abs/1608.06159).
- Felice, M. V., Velichko, A., Wilcox, P. D., and Lane, C. J. L., 2015, Depth measurement of small surface-breaking cracks using the half-skip total focusing method: *AIP Conference Proceedings*, **1650**, no. 1, 994–1000.
- Feng, R., Luthi, S. M., Gisolf, D., and Sharma, S., 2017, Obtaining a high-resolution geological and petrophysical model from the results of reservoir-orientated elastic wave-equation-based seismic inversion: *Petroleum Geoscience*, **23**, no. 3, 376–385.
- Fichtner, A., 2010, *Full Seismic Waveform Modelling and Inversion*, *Advances in Geophysical and Environmental Mechanics and Mathematics* Springer.

- Frank, J., and Vuik, C., 2001, On the construction of deflation-based preconditioners: *SIAM J. Sci. Comput.*, **23**, no. 2, 442–462.
- George, A., and Liu, J. W. H., 1981, *Computer Solution of Large Sparse Positive Definite Systems*: Prentice Hall.
- Gisolf, A., and Verschuur, E., 2010, *The principles of quantitative acoustical imaging*: EAGE publications.
- Gisolf, A., Tetyukhina, D., and Luthi, S. M., 2012, Full elastic inversion of synthetic seismic data based on an outcrop model (Book Cliffs): *SEG Technical Program Expanded Abstracts 2012*, pages 1–5.
- Gisolf, A., Huis in't Veld, R., Haffinger, P., Hanitzsch, C., Doulgeris, P., and Veeken, P., 2014, Non-linear full wavefield inversion applied to carboniferous reservoirs in the Wingate gas field (SNS, offshore UK): 76th EAGE Annual International Conference and Exhibition.
- Gisolf, A., Haffinger, P. R., and Doulgeris, P., 2015, Redatuming Directly to Angle Gathers, for the Purpose of Target Oriented Quantitative Interpretation: 77th EAGE Conference and Exhibition 2015.
- Gisolf, A., Sharma, S., and Luthi, S. M., 2017, Honouring Prior Geological Information in Wave-equation Based AVO Inversion: Fourth EAGE Exploration Workshop.
- Golub, G., and Pereyra, V., 2003, Separable Nonlinear Least Squares: the Variable Projection Method and its Applications: *Inverse Problems*, **19**, no. 2, R1–R26.
- Gordon, D., and Gordon, R., 2010, CARP-CG: A robust and efficient parallel solver for linear systems, applied to strongly convection dominated PDEs: *Parallel Computing*, **36**, no. 9, 495–515.
- Gray, S. H., and Marfurt, K. J., 1995, Migration from topography: Improving the near-surface image: *Can. J. Expl. Geophys.*, **31**, 18–24.
- Griffiths, D. J., 2004, *Introduction to Quantum Mechanics*: Prentice Hall, 2 edition.
- Guo, B., Huang, Y., Røstad, A., and Schuster, G., 2016, Far-field super-resolution imaging of resonant multiples: *Science Advances*, **2**, no. 5.

- Haber, E., Ascher, U. M., and Oldenburg, D., 2000, On optimization techniques for solving nonlinear inverse problems: *Inverse Problems*, **16**, 1263–1280.
- Haffinger, P., Gisolf, A., and van den Berg, P. M., 2013, Towards high resolution quantitative subsurface models by full waveform inversion: *Geophys. J. Int.*, **193**, no. 2, 788–797.
- He, Q., Chen, Y., Han, B., and Li, Y., 2016, Elastic frequency-domain finite-difference contrast source inversion method: *Inverse Problems*, **32**, no. 3.
- Kelly, K. R., Ward, R. W., Treitel, S., and Alford, R. M., 1976, Synthetic seismograms: a finite-difference approach: *Geophysics*, **41**, no. 1, 2–27.
- Kleinman, R. E., and van den Berg, P. M., 1991, Iterative methods for solving integral equations: *Radio Science*, **26**, no. 1, 175–181.
- Kleinman, R. E., Roach, G. F., Schuetz, L. S., Shirron, J., and van den Berg, P. M., 1990a, An over-relaxation method for the iterative solution of integral equations in scattering problems: *Wave motion*, **12**, no. 2, 161–170.
- 1990b, Convergent Born series for large refractive indices: *J. Opt. Soc. Am.*, **7**, no. 5, 890–897.
- Knibbe, H., Mulder, W. A., Oosterlee, C. W., and Vuik, C., 2014, Closing the performance gap between an iterative frequency-domain solver and an explicit time-domain scheme for 3D migration on parallel architectures: *Geophysics*, **79**, no. 2, 547–561.
- Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M.-D., 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, no. 6, WCC177–WCC188.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: *Conference on Inverse Scattering, Theory and Application*, Society for Industrial and Applied Mathematics, Expanded Abstracts, page 206220.
- Laird, A. L., and Giles, M. B., 2002, Preconditioned Iterative Solution of the 2D Helmholtz Equation: report 02/12, Oxford University Computing Laboratory.

- Li, M., Rickett, J., and Abubakar, A., 2013, Application of the variable projection scheme for frequency-domain full-waveform inversion: *Geophysics*, **78**, no. 6, R249–R257.
- Martin, G. S., Wiley, R., and Marfurt, K. J., 2006, Marmousi2: An elastic upgrade for Marmousi: *The Leading Edge*, **25**, no. 2, 156–166.
- Métivier, L., Brossier, R., Méridot, Q., Oudet, E., and Virieux, J., 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: *Geophys. J. Int.*, **205**, no. 1, 345–377.
- Min, D. J., Shin, C., Kwon, B. D., and Chung, S., 2000, Improved frequency domain elastic wave modeling using weighted averaging difference operators: *Geophysics*, **65**, no. 3, 884–895.
- Min, D.-J., Shin, C., Pratt, R. G., and Yoo, H. S., 2003, Weighted-averaging finite-element method for 2D elastic wave equations in the frequency domain: *Bulletin of the Seismological Society of America*, **93**, no. 2, 904–921.
- Min, D.-J., Shin, C., and Yoo, H. S., 2004, Free-surface boundary condition in finite-difference elastic wave modeling: *Bull. Seism. Soc. Am.*, **94**, no. 1, 237–250.
- Mulder, W. A., and Plessix, R.-E., 2004a, How to choose a subset of frequencies in frequency-domain finite-difference migration: *Geophys. J. Int.*, **158**, 801–812.
- 2004b, A comparison between one-way and two-way wave-equation migration: *Geophysics*, **69**, no. 6, 1491–1504.
- Mulder, W. A., and Plessix, R.-E., 2008, Exploring some issues in acoustic full waveform inversion: *Geophysical Prospecting*, **56**, 827–841.
- Mulder, W. A., 1989, A new multigrid approach to convection problems: *J. Comput. Phys.*, **83**, no. 2, 303–323.
- Nicolaides, R. A., 1987, Deflation of conjugate gradients with applications to boundary value problems: *SIAM J. Numer. Anal.*, **24**, no. 2, 355–365.
- Nocedal, J., and Wright, S. J., 2006, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering Springer, 2 edition.

- Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics: D. Reidel Publ. Co.
- O'Doherty, R. F., and Anstey, N. A., 1971, Reflections on amplitudes: *Geophysical Prospecting*, **19**, no. 3, 430–458.
- Operto, S., Gholami, Y., Prioux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: *The Leading Edge*, **32**, no. 9, 1040–1054.
- Ozmen, N., Dapp, R., Zapf, M., Gemmeke, H., Ruiter, N. V., and van Dongen, K. W. A., 2015, Comparing different ultrasound imaging methods for breast cancer detection: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **62**, no. 4, 637–646.
- Pan, G., and Abubakar, A., 2013, Iterative solution of 3D acoustic wave equation with perfectly matched layer boundary condition and multigrid preconditioner: *Geophysics*, **78**, T133–T140.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophys. J. Int.*, **167**, 495–503.
- Plessix, R.-É., 2017, Some computational aspects of the time and frequency domain formulations of seismic waveform inversion, *in* Lahaye, D., Tang, J., and Vuik, K., Eds., *Modern Solvers for Helmholtz Problems*: Springer International Publishing, 159–187.
- Pörtzgen, N., Gisolf, D., and Blacquiére, G., January 2007, Inverse wave field extrapolation: a different NDI approach to imaging defects: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **54**, no. 1, 118–127.
- Pörtzgen, N., Gisolf, D., and Verschuur, D. J., August 2008, Wave equation-based imaging of mode converted waves in ultrasonic NDI, with suppressed leakage from nonmode converted waves: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **55**, no. 8, 1768–1780.
- Pratapa, P. P., Suryanarayana, P., and Pask, J. E., 2016, Anderson acceleration of the Jacobi iterative method: An efficient alternative to Krylov methods for large, sparse linear systems: *J. Comput. Phys.*, **306**, 43–54.

- Pratt, R. G., 1990, Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging: *Geophysics*, **55**, no. 5, 626–632.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B., 2007, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*: Cambridge University Press.
- Quarteroni, A., 2010, *Numerical models for differential problems*:, volume 2 Springer Science & Business Media.
- Reps, B., and Vanroose, W., 2012, Analyzing the wave number dependency of the convergence rate of a multigrid preconditioned Krylov method for the Helmholtz equation with an absorbing layer: *Numerical Linear Algebra and Applications*, **19**, 232–252.
- Rizzuti, G., and Gisolf, A., 2017, An iterative method for 2d inverse scattering problems by alternating reconstruction of medium properties and wavefields: theory and application to the inversion of elastic waveforms: *Inverse Problems*, **33**, no. 3, 035003.
- Rizzuti, G., and Mulder, W. A., 2016, Multigrid-based ‘shifted-Laplacian’ preconditioning for the time-harmonic elastic wave equation: *J. Comput. Phys.*, **317**, no. 15, 47–65.
- Saad, Y., 2003, *Iterative Methods for Sparse Linear Systems*: SIAM, 2 edition.
- Sheik, A. H., Vuik, C., and Lahaye, D., 2013, On the convergence of shifted Laplace preconditioner combined with multilevel deflation: *Numerical Linear Algebra with Applications*, **20**, 645–662.
- Shin, C., and Cha, Y. H., 2008, Waveform inversion in the Laplace domain: *Geophys. J. Int.*, **173**, no. 3, 922–931.
- Shin, C., and Cha, Y. H., 2009, Waveform inversion in the Laplace–Fourier domain: *Geophys. J. Int.*, **177**, no. 3, 1067–1079.
- Shin, C., and Min, D.-J., 2006, Waveform inversion using a logarithmic wavefield: *Geophysics*, **71**, no. 3, R31–R42.
- Shipp, R. M., and Singh, S. C., 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: *Geophys. J. Int.*, **151**, no. 2, 325–344.

- Simonetti, F., 2006, Multiple scattering: The key to unravel the subwavelength world from the far-field pattern of a scattered wave: *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, **73**, no. 036619.
- Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, no. 1, 231–248.
- Soni, A., and Verschuur, D., 2014, Full-wavefield migration of vertical seismic profiling data: using all multiples to extend the illumination area: *Geophysical Prospecting*, **62**, no. 4, 740–759.
- Sonneveld, P., and van Gijzen, M. B., 2008, IDR(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations: *SIAM J. Sci. Comput.*, **31**, no. 2, 1035–1062.
- Stark, T. J., 2013, Signal recovery beyond conventional Nyquist: The sample rates used for seismic acquisition do not need to limit the maximum recoverable frequencies: *The Leading Edge*, **32**, no. 11, 1334–1339.
- Štekl, I., and Pratt, R. G., 1998, Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators: *Geophysics*, **63**, no. 5, 1779–1794.
- Stolk, C., 2013, A rapidly converging domain decomposition method for the Helmholtz equation: *J. Comput. Phys.*, **241**, 240–252.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790.
- Tarantola, A., and Valette, B., 1982a, Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion: *Rev. Geophys. Space Phys.*, **20**, no. 2, 219–232.
- 1982b, Inverse Problems = Quest for Information: *J. Geophys.*, **50**.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, no. 8, 1259–1266.
- Tikhonov, A. N., and Arsenin, V. Y., 1977, *Solution of Ill-posed Problems*: Winston & Sons.

- Trottenberg, U., Oosterlee, C. W., and Schüller, A., 2001, Multigrid: Academic Press.
- Tsuji, P., and Tuminaro, R., 2015, Augmented AMG-shifted Laplacian preconditioners for indefinite Helmholtz problems: Numer. Linear Algebra Appl.
- van den Berg, P. M., and Abubakar, A., 2001, Contrast source inversion method: State of art: Progr. Electromagn. Res., **34**.
- van den Berg, P. M., and Kleinman, R. E., 1997, A contrast source inversion method: Radio Science, **13**, 1607–1620.
- van den Berg, P. M., Abubakar, A., and Fokkema, J. T., 2003, Multiplicative regularization for contrast profile inversion: Radio Science, **38**, no. 2, 8022.
- van der Vorst, H. A., 1992, Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems: J. Sci. Stat. Comput., **13**, no. 2, 631–644.
- van Leeuwen, T., and Herrmann, F. J., 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophys. J. Int., **195**, 661–667.
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, 1–26.
- Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, no. 4, 889–901.
- Wang, S., de Hoop, M. V., and Xia, J., 2011, On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver: Geophysical Prospecting, **59**, 857–873.
- Wang, S., de Hoop, M. V., Xia, J., and Li, X. S., 2012, Massively parallel structured multifrontal solver for time-harmonic elastic waves in 3-D anisotropic media: Geophys. J. Int., **191**, 346–366.
- Warner, M., and Guasch, L., 2014, Adaptive Waveform Inversion: Theory: Adaptive Waveform Inversion: Theory:, Society of Exploration Geophysicists, SEG Technical Program Expanded Abstracts 2014, 1089–1093.
- Wienands, R., and Oosterlee, C. W., 2001, On three-grid Fourier analysis for multigrid: SIAM J. Sci. Comput., **23**, no. 2, 651–671.

- Willemsen, B., Malcolm, A., and Lewis, W., 2016, A numerically exact local solver applied to salt boundary inversion in seismic full-waveform inversion: *Geophysical Journal International*, **204**, no. 3, 1703–1720.
- Wirgin, A., 2008, The inverse crime: [arXiv:math-ph/0401050](https://arxiv.org/abs/math-ph/0401050).
- Xia, J., Chandrasekaran, S., Gu, M., and Li, X. S., 2009, Superfast multi-frontal method for large structured linear systems of equations: *J. Matrix Anal. Appl.*, **31**, no. 3, 1382–1411.
- Yang, J., Abubakar, A., van den Berg, P. M., Habashy, T. M., and Reitich, F., 2008, A CG-FFT approach to the solution of a stress-velocity formulation of three-dimensional elastic scattering problems: *J. Comput. Phys.*, **227**, 10018–10039.
- Zepeda-Núñez, L., and Demanet, L., 2016, The method of polarized traces for the 2D Helmholtz equation: *Journal of Computational Physics*, **308**, 347–388.

Summary

Full-waveform inversion is a promising tool for a wide range of imaging scenario, in that it has the potential to harness the non-linear relationship between model parameters and data (as opposed to traditional methodologies), in order to produce truly quantitative results. Non-linearity represents an opportunity, in this sense, but it also begets local minimum issues when gradient-based optimization is employed.

In this thesis, we are particularly interested in the quantitative estimation of the elastic parameters of the earth, such as compressibility, shear modulus, and density. If successful, this procedure brings geophysical imaging closer to the ultimate step of seismic exploration: the retrieval of pore pressure and rock properties. This would be of direct use for, say, the oil and gas industry. Other interesting applications are the description of the near surface/near ocean bottom, as a way to reduce drilling hazards, or non-destructive inspection of defects in oil and gas pipes, when ultrasounds are employed.

The scope of the thesis is to devise a reconstruction algorithm which can effectively cope with the non-linearity of the problem, with the aim of providing reliable elastic properties for quantitative studies. We also target a real-world application as a proof-of-concept for the principles enunciated above.

The inversion algorithm is based on the introduction of a slack variable, in addition to the elastic model parameters. This new variable is represented by the elastic wavefields, and is not strictly required to solve the wave equation

relative to the current estimated model. An alternating reconstruction scheme is introduced, where properties and wavefields are updated in a leap-frog manner by optimizing quadratic objective functionals: one related to the data fit, and one related to the wave equation consistency. The success of the scheme, as well as the computational feasibility of a 2-D implementation, is proven with multiple synthetic examples.

A 3-D extension of this work is non-trivial, the most glaring issue being, perhaps, the need for keeping the wavefields in memory for each source and frequency. In this thesis, we investigated yet another aspect of this generalization. In 2-D, given a certain discretization of the wave equation (e.g. by finite-differences or finite-elements), the tool of choice for its solution are factorization-based methods. In 3-D, however, this is not feasible anymore. An alternative, practical solution—although not optimal, computationally-wise—is offered by multigrid-based iterative methods, originally proposed for the Helmholtz equation and here extended to the elastic case.

We conclude this thesis with a practical application to ultrasound imaging of defects in steel pipes. This imaging setting is an excellent benchmark to test the advantages of linear vs non-linear imaging, as it is common to encounter defects, such as stress-corrosion cracks, that are orthogonally aligned with respect to the measurement plane. Non-linear methods, which greatly benefits from multiple illumination, seem to be the natural solution for this scenario. We show that it is possible to image borehole-type defects with great accuracy. This study is the first step towards vertical crack imaging.

Samenvatting

Full-waveform inversie is een veelbelovend hulpmiddel bij het oplossen van een breed scala aan imaging problemen. Om echter kwantitatieve resultaten te produceren, in tegenstelling tot traditionele methodologieën, dient de niet-lineaire relatie tussen modelparameters en gegevens benut te worden. Hoewel de niet-lineariteit ons de kans geeft om kwantitatieve resultaten te verkrijgen, brengt het ook problemen met zich mee wanneer gradient gebaseerde optimalisatie wordt toegepast.

In dit proefschrift zijn we vooral geïnteresseerd in de kwantitatieve schatting van de elastische parameters van de aarde: compressibility, shear modulus en dichtheid. Indien succesvol, dan brengt deze procedure geofysische imaging dichterbij het ultieme doel van seismische exploratie, namelijk, het bepalen van de pore pressure en rotseigenschappen. Deze informatie zou van grote waarde zijn voor de olie- en gasindustrie. Andere interessante toepassing is het karakteriseren van de near surface/near ocean bottom om de risico's te beperken tijdens het boren. Maar de methode kan ook gebruikt worden tijdens niet-destructieve inspectie om defecten op te sporen in olie- en gasleidingen met behulp van ultrasone geluidsgolven.

Het doel van het proefschrift is om een reconstructie-algoritme te ontwikkelen dat de niet-lineariteit van het probleem effectief kan beschrijven, met het doel om betrouwbare elastische eigenschappen te schatten voor kwantitatieve studies. We richten ons ook op een toepassing uit de praktijk als een proof-of-concept voor de hierboven genoemde principes.

Het inversie-algoritme is gebaseerd op een extra slack-variabele, naast de elastische modelparameters. Deze nieuwe variabele wordt beschreven door de elastische golfvelden en is niet strikt vereist om de golfvergelijking ten opzichte van het huidige geschatte model op te lossen. Een alternerend reconstructieschema wordt geïntroduceerd waarbij eigenschappen en golfvelden op een leap-frog manier worden bijgewerkt door het optimaliseren van kwadratische objective functions. De eerste objective function is gerelateerd aan de data fit en de andere is gerelateerd aan de golfvergelijking. Het succes van het voorgestelde schema is bewezen met meerdere 2-D synthetische voorbeelden.

Een 3-D uitbreiding van dit werk is niet-triviaal, het meest in het oog springende probleem is misschien de noodzaak om golfvelden in het geheugen op te slaan voor elke bron en frequentie. In dit proefschrift hebben we nog een ander aspect van deze generalisatie onderzocht. In 2-D, gegeven een zekere discretisatie van de golfvergelijking (bijvoorbeeld door finite-differences or finite-elements), zijn factorisatie gebaseerde methoden de oplossing. In 3-D is dit echter niet meer mogelijk. Een alternatieve praktische oplossing, hoewel rekenkundig niet optimaal, is het gebruik van multigrid gebaseerde iteratieve methoden. Deze methoden waren oorspronkelijk gebruikt om de Helmholtz-vergelijking op te lossen en zijn in dit proefschrift uitgebreid voor het elastische geval.

We sluiten dit proefschrift af met een praktische toepassing in het veld van ultrasone imaging om defecten in stalen buizen te detecteren. Deze toepassing is een uitstekende benchmark om de voordelen van lineaire versus niet-lineaire imaging te testen. De reden hiervoor is omdat het gebruikelijk is om defecten tegen te komen, zoals stress-corrosion cracks, die orthogonaal zijn ten opzichte van het meetvlak. Niet-lineaire methoden, die veel baat hebben bij meervoudige belichting, lijken de natuurlijke oplossing te zijn voor dit scenario. In dit proefschrift laten we zien dat het mogelijk is om borehole-type defecten af te beelden met grote nauwkeurigheid. Deze studie vormt de eerste stap richting de verticale crack-imaging.

Acknowledgements

Before starting my doctorate, the Netherlands were not part of my mental geography. I often use running as a way to connect with a foreign place. So, tracing a running route along the Schie was one of the first thing that I tried, here. I was glad to realize that “it’s not that cold” after all. I would like to start, then, by thanking this country for being my home in the past five years.

I thank Eric for his kind guidance as a supervisor during the first year of my Ph.D.. I am in debt for being so understanding when I announced my sudden departure from the group.

With Dries, my original promotor, I developed a friendship which continued outside the university, especially with the Delft Inversion enterprise. Since the first ‘sign issue’, I believe we develop a proficient dynamics in tackling research problems. I hope that will continue in the near future, and leave you with good memories when I will take on the next adventure.

Alok was the first person of the Delphi group that I met, and will always represent for me a sort of student–mentor figure in the then-unknown field of geophysics. We also share a memorable bike trip in Denver. Biking with a suit? Not the cleverest idea.

I obviously need to mention my fellow student colleagues, with whom I experienced plenty of SEG/EAGEs during the years. In no particular order: Mikhail, Gabriel, Jan-Willem, Xander, Tomohide, Amarjeet, Hannes, Yimin, Apostolos, Runhai, Siddarth, Abdulrahman, Matteo, and Bouchaib. There are

probably many others that I am forgetting: to them, my sincerest apologies.

A very special thanks goes to Prof. Mulder, who lend a hand with one of my research question, and with whom I had the pleasure to work. Wim generously accepted to be the promotor of my thesis when Dries could not do it.

Joining Delft Inversion was easily the most important experience I had during these years. Navigating through rough sea is not comfortable, but it creates a strong bond. There were many important events of life, successes, and failures that I am happy to have lived together: Dries, Panos, Peter, Stefan, Olav, Kostas, Veerle, Anna, Arne, and Mengmeng. Thanks, in particular, to Panos and Peter who created a very peculiar environment with their unique blend of personalities.

Delft Inversion also gave me the opportunity to apply my work with Applus+ RTD: many thanks to Jasper, Lars, Wouter, Wouter, Bernhard, and Daniel for the company in Rotterdam.

E per quanto riguarda la famiglia, madre, padre, fratelli. È difficile esprimere la mia gratitudine in parole, spero che le mie azioni rendano ciò ridondante.

Un ultimo saluto a Cesare Pucci, detto ‘Prašiček’, che ho attirato con un po’ di tonno. Devo ancora capire perché ci visiti solo d’estate.

Ovviamente, la vita mi sarebbe resa miserabile se provassi a non menzionare Sara. Grazie per il sostegno dimostratomi in questi anni. Ne abbiamo passate così tante assieme, che non vedo l’ora di raggiungerci nella prossima avventura.

I would like to dedicate this thesis to anyone, especially students, who found himself in a moment of struggle: many good things in life come from a place of suffering.

Gabrio

Delft, 4/11/2017