PAWAN
BHARADWAJ

SEISMIC WAVEFORM INVERSION

BUMP FUNCTIONAL PARAMETERIZATION ANALYSIS & IMAGING AHEAD OF A TUNNEL-BORING MACHINE

Seismic Waveform Inversion

bump functional, parameterization analysis and imaging ahead of a tunnel-boring machine

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 maandag 22 mei 2017 om 10:00 uur

door

Pawan Bharadwaj Pisupati

Master of Science & Technology (Applied Geophysics),
Indian Institute of Technology (Indian School of Mines), Dhanbad, India,
geboren te Hyderabad, India.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. W. A. Mulder
copromotor: Dr. ir. G. G. Drijkoningen

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. W. A. Mulder	CiTG, Delft University of Technology
Dr. ir. G. G. Drijkoningen	CiTG, Delft University of Technology
Onafhankelijke leden:	
Prof. dr. A. G. Yarovoy	EWI, Delft University of Technology
Dr. ir. W. Broere	EWI, Delft University of Technology
Prof. dr. G. T. Schuster	King Abdullah University of Science and Technology
Prof. dr. ir. C. Vuik	EWI, Delft University of Technology
Prof. dr. ir. C. P. A. Wapenaar	CiTG, Delft University of Technology

Keywords: seismic imaging, full-waveform inversion, bump functional, parameterization, local minima, tunnel-boring machine

Printed by: ProefschriftMaken || Proefschriftmaken.nl

Front & Back: When the seismic waveforms are too complicated to be analyzed during inversion, a simplification of them into envelope-like bumpy waveforms can be useful.

Designed by: Sumedha Vemuri (https://www.behance.net/sumedhavemuri), a Visual and Spatial designer based in Bengaluru, India.

ISBN 978-94-6295-662-9

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

This thesis is dedicated to my parents
for adding high coefficients of curiosity and courage to the equations of my life
and nursing me with their love and affection.

Contents

Summary

During a seismic experiment, mechanical waves are usually generated by various man-made sources. These waves propagate in the subsurface and are recorded at receivers. Modern seismic exploration methods analyze them to infer the mechanical properties of the subsurface; this is commonly referred as quantitative imaging. These properties assist in the determination of the subsurface rock type and structure. Exploration methods are not only useful while looking for the deposits such as crude oil, natural gas and minerals but also for near-surface geotechnical investigation. A motive of this thesis is to adopt these methods to image the subsurface ahead of a tunnel-boring machine for hazard assessment during excavation. Full-waveform inversion (FWI) is a gradient-based optimization problem that is employed in seismic exploration for quantitative imaging of the recorded waves. During FWI, seismic waves are simulated in a computer by using certain physical laws that govern the wave propagation. After inversion, output subsurface properties simulate waves that fit the recorded waves in a least-squares sense. In other words, the gradient-based optimization aims to find the minimum of the least-squares misfit between the simulated and the recorded waves.

Finding such a minimum is not straight forward due to the existence of multiple local minima when using the the least-squares objective. As a result, it might often happen that the optimizer converges to local minima, where the simulated waves only partially explain the recorded waves. The presence of local minima is associated to the strong non-linear dependence of the recorded waves on the subsurface properties. In this thesis, we attempt to overcome this difficulty. We propose a new measure of misfit between the recorded and the simulated waves. This measure compares the waveforms in a simplified form after taking the absolute value and blurring. We show that the new misfit measure suffers less from the local-minima problem. For robust inversion, we use a multi-objective inversion scheme, where the new measure is used as an auxiliary objective to pull the trapped solution out of the least-squares local minimum whenever necessary.

In multi-parameter FWI, more than one kind of subsurface properties are simultaneously estimated. When only the first-order derivatives of the misfit are used during minimization, different choices of subsurface parameterization are not equivalent; they can be interpreted as different preconditioners. Therefore, the choice of parametrization will affect the rate of convergence in multi-parameter FWI and the best choice of parameterization is the one with the highest rate. In this thesis, we also analyse various choices of subsurface parameterization in search of the best one.

It is well known that the local-minima problem in FWI can easily be resolved by reliably generating and recording low-frequency waves in the subsurface. Recently, a seismic source capable of generating such low frequencies is developed based on linear synchronous motors technology. Finally, we demonstrated a shear-wave seismic

ground prediction system using these sources to enable imaging ahead of a tunnel-boring machine (TBM).

Samenvatting

Tijdens een seismisch experiment worden mechanische golven doorgaans opgewekt door middel van diverse kunstmatige bronnen. Deze golven propageren door de ondergrond en worden opgenomen door ontvangers. Moderne seismische onderzoeksmethoden analyseren de opgenomen gegevens om de mechanische eigenschappen van de ondergrond af te leiden; dit wordt meestal aangeduid als kwantitatieve beeldvorming. Deze eigenschappen helpen bij het bepalen van het type ondergrondse gesteente en hun structuur. Opsporingsmethoden zijn niet alleen nuttig bij het zoeken naar voorraden ruwe olie, aardgas en mineralen, maar ook voor geotechnisch onderzoek van het nabije aardoppervlak. Een motief voor dit proefschrift is om deze methoden aan te passen ten behoeve van beeldvorming van de ondergrond net vóór een tunnelboormachine om mogelijke gevaren tijdens het graven vast te stellen. Volledige golfvorminversie is een door de gradiënt gedreven optimalisatieprobleem dat wordt gebruikt bij seismisch onderzoek voor kwantitatieve beeldvorming vanuit het opgenomen golfveld. Tijdens deze inversie worden seismische golven gesimuleerd in een computer, gebruikmakend van bepaalde fysische wetten die de golfvoortplanting beschrijven. Na inversie kunnen de gevonden ondergrondse materiaaleigenschappen de opgenomen golven nabootsen in de zin van de kleinste kwadraten. Met andere woorden, de op de gradiënt gebaseerde optimalisatie beoogt het minimum van de kleinstekwadratenfout tussen de gesimuleerde en de gemeten golven te vinden.

Het vinden van een dergelijke minimum is niet vanzelfsprekend ten gevolge van het bestaan van meerdere lokale minima bij het gebruik van de kleinstekwadratenfout. Als gevolg daarvan kan het vaak gebeuren dat de optimalisatiemethode convergeert naar een lokaal minimum, waarbij de gesimuleerde golven slechts ten dele de opgenomen golven verklaren. De aanwezigheid van lokale minima houdt verband met de sterke niet-lineaire afhankelijkheid tussen de geregistreerde golven en de eigenschappen van de ondergrond. In dit proefschrift proberen we dit probleem aan te pakken. Wij stellen een nieuwe foutenmaat tussen de geregistreerde en de gesimuleerde golven voor. Deze maat vergelijkt de golfvormen in een vereenvoudigde vorm door het nemen van de absolute waarde en het versmeren ervan. We tonen aan dat de nieuwe foutenmaat minder lijdt onder het probleem van de lokale minima. Voor robuuste inversie maken we gebruik van een inversiemethode met meerdere objectieve functies, waar de nieuwe maat wordt gebruikt om de gestrande oplossing uit het lokale minimum van de kleinstekwadratenmethode te halen wanneer dat nodig is.

Bij multiparametergolfvorminversie worden verschillende ondergrondse eigenschappen geschat. Indien alleen de eerste afgeleiden van de foutenmaat worden gebruikt tijdens de minimalisatie, zijn verschillende keuzes van parametrisering niet equivalent; ze kunnen worden opgevat als verschillende preconditioners. Dientengevolge zal de gekozen parametrisering de convergentiesnelheid van multiparametergolfvorminversie

bepalen. De beste keuze van parametrisering is die met de snelste convergentie. In dit proefschrift analyseren we ook verschillende keuzes van parametrisering van de ondergrond, op zoek naar de de beste.

Het is bekend dat het probleem van de lokale minima in golfvorminversie eenvoudig opgelost kan worden door op betrouwbare wijze laag-frequente golven in de ondergrond op te wekken en op te nemen. Onlangs is er een seismische bron ontwikkeld, op basis van technologie voor lineaire synchrone motoren, die in staat is dergelijke lage frequenties op te wekken. Tot slot laten we zien dat met zulke bronnen een systeem voor beeldvorming voorafgaand aan een tunnelboormachine met behulp van schuifgolven mogelijk is.

List of Symbols

In this thesis, different subscripts are used to these symbols whenever necessary. The subscripts are explained in the text.

σ	stress tensor
σ	stress tensor components
u	particle displacement vector field
u	component of displacement field
v	component of velocity field
$\mathfrak{p}$	scalar pressure wavefield in an acoustic medium
κ	external force acting on a medium
κ	external force acting on a medium
x	horizontal coordinate
z	vertical coordinate
x_r	receiver horizontal coordinate
y	out of the plane coordinate
$\mathbf{x}$	sub-surface point, (x, z)
$\mathbf{x}_s$	source position, (x_s, z_s)
$\mathbf{x}_r$	receiver position, (x_r, z_r)
t	time
f	frequency
ω	angular frequency
k_r	wavenumber
S	restriction operator onto the receiver positions
f_d	dominant frequency in observed data
τ_d	dominant period in observed data
τ_{max}	largest period in observed data
λ_d	dominant wavelength at receivers
ν	modelled time-domain wavefield in the medium
p	modelled time-domain wavefield at receivers
χ	adjoint time-domain wavefield in the medium
ψ	adjoint time-domain wavefield in the medium
q	observed time-domain wavefield
P	modelled wavefield in frequency domain
Q	observed wavefield in frequency domain
p_b	modelled time-domain bumpy wavefield
q_b	observed time-domain bumpy wavefield
p_a	modelled time-domain absolute-valued wavefield

$\tilde{q}_b$	observed bumpy wavefield in frequency-wavenumber domain
$\tilde{p}_b$	modelled bumpy wavefield in frequency-wavenumber domain
$\tilde{q}_a$	observed absolute-valued wavefield in frequency-wavenumber domain
$\tilde{p}_a$	modelled absolute-valued wavefield in frequency-wavenumber domain
ϕ_s	sweep signal in time domain
Φ_s	sweep signal in frequency domain
ϕ	source time function
Φ	source function in frequency domain
γ_s	time domain source filter
γ_r	time domain receiver filter
Γ_s	frequency domain source filter
Γ_r	frequency domain receiver filter
J	objective function
J_{ls}	least-squares objective function
J_b	bump objective function
b	blurring function
$\tilde{b}$	blurring function in frequency-wavenumber domain
σ_t	time-blurring parameter
σ_r	receiver-array-blurring parameter
$\mathbb{R}$	real space
$\mathbb{M}$	model space
$\mathbb{D}$	data space
$\mathbb{U}$	space corresponding to time-doimain wave-field in the medium
$\mathbb{D}_b$	bumpy data space
$\mathcal{F}$	forward map to model wavefields
$\mathcal{F}_b$	forward map to create bumpy data
c	wave-speed
c_p	compressional wave-speed
c_s	shear wave-speed
$\mathbf{m}$	vector of medium parameters
c_{po}	reference wave-speed
ρ	mass density
ρ_o	reference mass density
Z_p	P-wave impedance
Z_{po}	reference mass density
K	bulk modulus
K_o	reference bulk modulus
μ	shear modulus
L	wave operator
d	total differential operator
g	constraint
$\mathcal{L}$	Lagrangian for constrained optimization

1

Introduction

Nobody ever figures out what life is all about, and it doesn't matter.
Explore the world.
Nearly everything is really interesting if you go into it deeply enough.

Richard Feynman

1

In seismic exploration, artificially or naturally generated mechanical waves are recorded and processed to infer the structure and mechanical properties of the subsurface. During a seismic experiment, waves generated by various active sources reflect and refract off subsurface formations before travelling to receivers. A motive of this thesis is to develop a seismic system to map the subsurface ahead of a tunnel boring machine for hazard assessment during excavation. Such a system should be able generate the subsurface properties in near real time with very little human interaction.

Conventionally, processing of recorded seismic data is a two-step procedure. First, a background wave velocity of the subsurface is estimated using kinematic information in the data. The background velocity model only contains components that are smoothly varying with location. Secondly, the reflected arrivals in the data are mapped back into depth to produce a *reflectivity* image of the subsurface. This step is called migration. The reflectivity image only contains rough components of the wave velocity and the mass density. It highlights the interfaces between Earth layers and rock formations, each having a different wave impedance, the product of density and velocity.

The kinematic information of seismic data is contained in the traveltimes of both the reflected and refracted arrivals. These traveltimes can be manually picked and then refraction [White, 1989; Zhang and Toksöz, 1998; Schuster and Quintus-Bosz, 1993] or reflection [Luo and Schuster, 1991; Stork, 1992] tomography can be used to produce a background velocity model. With such a model, a suitable migration algorithm [Claerbout, 1971; Stolt, 1978; Berkhout, 1980; Gazdag and Sguazzero, 1984; Bleistein *et al.*, 2000] can generate a reflectivity image of the subsurface.

The migration imaging principle of Claerbout [1971] was recast into a local optimisation problem by Lailly [1983]; Tarantola [1984a]. The latter minimizes the least-squares misfit between the recorded and modelled data that are generated by solving a wave equation, based on simplified laws of physics, assuming a fixed background velocity. As the modelled data depend linearly on the unknown rough components of the wave velocity, this inverse problem is easy to solve. If such an optimisation problem also estimates all the other unknown subsurface properties, including the background velocity, the resulting inversion algorithm is called full waveform inversion (FWI) [Tarantola, 1984b]. Unlike the conventional tomographic or migration methods, FWI uses both the amplitude and the phase of the recorded waveforms to iteratively solve for the elastic properties of the subsurface. The dependence between the modelled data and certain subsurface properties can be highly non-linear. A key assumption, which is often overlooked in practice, of both the linearized and full-waveform inversion is that the simplified wave equation provides a sufficiently accurate description of the true seismic wave propagation. For example, consider an inversion using the two-dimensional isotropic elastic wave equation, where the actual wave propagation has occured in a three-dimensional anisotropic medium.

Over the past 20 years, FWI has gradually replaced the conventional seismic processing and imaging methods, such as traveltime tomography and migration, for the following reasons:

- FWI not only estimates the subsurface wave velocity but also other elastic properties that determine the wave propagation, which leads to multi-parameter quanti-

tative imaging [Djikpéssé and Tarantola, 1999; Plessix *et al.*, 2013; Operto *et al.*, 2013; Prieux *et al.*, 2013].

- FWI has the potential to obtain all wavenumber components of the subsurface models that can be resolved from the information contained in seismic recordings. For example, in the case of the wave-velocity model, both the low- and high-wavenumber components are expected to be reconstructed accurately. The low-wavenumber components mainly affect the transmission of waves and constitute the background velocity, whereas the high wavenumber components add details to the model and are responsible for the reflection of waves. Therefore, full waveform inversion encompasses both migration and tomography [Mora, 1989].
- In principle, data processing with FWI can be done *automatically* with almost negligible human intervention. This is clearly valuable when real-time mapping of the Earth's composition and structure is required. One example is looking ahead of a tunnel boring machine (TBM) for hazard assessment during tunnel excavation, the subject of chapter 5 of this thesis. Another is helping operators to reduce risk and improve efficiency during drilling of wells [Rector III and Marion, 1991; Miranda *et al.*, 1996; Poletto and Miranda, 2004].

Despite the attractive reasons given above, until now the success of the inversion algorithms using only FWI is limited due to the following problems [Santosa and Symes, 1989; Mulder and Plessix, 2008; Symes, 2008; Virieux and Operto, 2009]:

- Convergence to local minima of the least-squares functional. The presence of local minima can be attributed to the well-known non-linear dependence between the modelled data and the unknown background wave velocity. Hence, the local minima make the reconstruction of the low-wavenumber components of the wave velocity difficult. This problem is severe in the absence of reliable low-frequency information in the recorded data. In such situations, FWI is successful only when started from an accurate background velocity model derived from either the conventional tomography or available a priori knowledge.
- The wave equation based on simplified physics is used for modelling; it cannot satisfactorily account for the amplitudes of the recorded arrivals in most of the cases. In addition to this modelling uncertainty, there is always noise in the seismic data due to various external sources in the recording environment.
- During multi-parameter inversion, different parameter types have different sensitivities to the modelled data, making their simultaneous estimation challenging.
- Computational costs associated with the modelling and inversion of 3-D subsurface models are high.

Therefore, successful seismic experiments that aim for automatic quantitative imaging of the subsurface need to

- use automatic procedures that do not rely on manual picking to build low-wavenumber components and/or record reliable low frequencies,

- adopt new minimization criteria to mitigate the sensitivity of inversion to amplitude errors and increase the robustness of FWI when multiple types of parameter are to be estimated, and

- improve computational efficiency, for instance, data-compression techniques can be used when the inversion program is slow due to the lack of memory bandwidth.

The chapters of this thesis can be read independently if the reader already has an understanding of FWI. If not, chapter 2 provides the basics of time-domain FWI, which is prevalent in all the subsequent chapters.

In chapter 3, we focus on the problem of local minima. We first formulate a new data-domain objective function, that we call the bump functional, for inversion. The bump functional can be seen as a generalised envelope-based misfit. We describe various characteristics of the bump functional using simple and illustrative numerical examples. Then, we propose a multi-objective inversion scheme that uses the conventional least-squares functional along with the bump functional. We demonstrate the effectiveness of the proposed multi-objective inversion scheme in reconstructing the low-wavenumber components of the wave-velocity by considering realistic examples.

In chapter 4, we focus on reducing the computational cost of multi-parameter inversion. We analyse various choices of subsurface parameterizations in search of a choice with the fastest convergence for multi-parameter acoustic FWI. We also review two different conventional parameterization analysis methods, i.e., the point-scatterer analysis and diffraction-pattern analysis, which are expected to suggest a parameterization choice with the fastest convergence. Our numerical examples with single-component data do not agree with the suggestions of the conventional analyses and show that we cannot decide on the fastest parameterization choice for multi-parameter acoustic full waveform inversion.

In chapter 5, we undertake the demonstration of FWI's speculative qualification to allow for automatic processing. To enable imaging ahead of a tunnel-boring machine (TBM), we developed a seismic prediction system with a few shear-wave vibrators and horizontal receivers. The horizontal receivers measure the particle velocity, mainly due to horizontally polarised shear (SH) waves. The design of the vibrators is based on linear synchronous motors technology [Noorlandt *et al.*, 2015], which is capable of expanding the source frequency band to the lower frequencies. Hence the local minima problem is less severe, enabling the system to use FWI for processing. Since the acquired data need to be processed in nearly real time with current computing technology, we simplified the SH full waveform inversion problem to 2D. We demonstrated the capabilities of the proposed system by a number of synthetic and field experiments.

Chapter 6 concludes the thesis and provides some recommendations for future research directions.

References

D. White, *Two-dimensional seismic refraction tomography,* Geophysical Journal International **97**, 223 (1989).

J. Zhang and M. N. Toksöz, *Nonlinear refraction traveltime tomography,* Geophysics **63**, 1726 (1998).

G. T. Schuster and A. Quintus-Bosz, *Wavepath eikonal traveltime inversion: Theory,* Geophysics **58**, 1314 (1993).

Y. Luo and G. Schuster, *Wave-equation traveltime inversion,* Geophysics **56**, 645 (1991).

C. Stork, *Reflection tomography in the postmigrated domain,* Geophysics **57**, 680 (1992).

J. F. Claerbout, *Toward a unified theory of reflector mapping,* Geophysics **36**, 467 (1971).

R. Stolt, *Migration by Fourier transform,* Geophysics **43**, 23 (1978).

A. J. Berkhout, *Seismic migration: Imaging of acoustic energy by wave field extrapolation* (Elsevier, 1980).

J. Gazdag and P. Sguazzero, *Migration of seismic data by phase shift plus interpolation,* Geophysics **49**, 124 (1984).

N. Bleistein, J. K. Cohen, and W. John Jr, *Mathematics of multidimensional seismic imaging, migration, and inversion* (Springer Science & Business Media, 2000).

P. Lailly, *The seismic inverse problem as a sequence of before stack migrations,* in *Proc. Conf. on Inverse Scattering, Theory and Applications, Philadelphia* (SIAM, 1983).

A. Tarantola, *Linearized inversion of seismic reflection data,* Geophysical Prospecting **32**, 998 (1984a).

A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation,* Geophysics **49**, 1259 (1984b).

H. A. Djikpéssé and A. Tarantola, *Multiparameter L1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms,* Geophysics **64**, 1023 (1999).

R. Plessix, P. Milcik, H. Rynja, A. Stopin, K. Matson, and S. Abri, *Multiparameter full-waveform inversion: Marine and land examples,* The Leading Edge **32**, 1030 (2013).

S. Operto, Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, *A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice,* The Leading Edge **32**, 1040 (2013).

V. Prieux, R. Brossier, S. Operto, and J. Virieux, *Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: imaging compressional wave speed, density and attenuation,* Geophysical Journal International **194**, 1640 (2013).

P. Mora, *Inversion = migration + tomography,* Geophysics **54**, 1575 (1989).

J. Rector III and B. P. Marion, *The use of drill-bit energy as a downhole seismic source,* Geophysics **56**, 628 (1991).

F. Miranda, L. Aleotti, F. Abramo, F. Poletto, A. Craglietto, S. Persoglia, and F. Rocca, *Impact of the seismic while drilling technique on exploration wells,* First Break **14**, 55 (1996).

F. Poletto and F. Miranda, *Seismic while drilling: Fundamentals of drill-bit seismic for exploration*, Vol. 35 (Elsevier, 2004).

F. Santosa and W. W. Symes, *An analysis of least-squares velocity inversion* (Geophysical Monograph Series, Society of exploration Geophysicists, 1989).

W. A. Mulder and R. E. Plessix, *Exploring some issues in acoustic full waveform inversion,* Geophysical Prospecting **56**, 827 (2008).

W. W. Symes, *Migration velocity analysis and waveform inversion,* Geophysical Prospecting **56**, 765 (2008).

J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics,* Geophysics **74**, WCC1 (2009).

R. Noorlandt, G. Drijkoningen, J. Dams, and R. Jenneskens, *A seismic vertical vibrator driven by linear synchronous motors,* Geophysics **80**, EN57 (2015).

2

Waveform Inversion

I learned very early the difference between knowing the name of something and knowing something.

Richard Feynman

The purpose of this chapter is to briefly review time-domain seismic waveform inversion methods. They are formulated as optimization problems with partial differential equation (PDE) constraints. To aid the understanding of these problems, we also provide a simple two-dimensional constrained optimization example in boxes.

2.1. Notation for Differentiation

We use Euler's notation for differentiation in this chapter. The total differential and the partial differential operators are denoted as d and ∂, respectively. When applied to a function J of two variables x and z,

$$\mathrm{d}_x J = \frac{\mathrm{d}J}{\mathrm{d}x}$$

is the total derivative of J with respect to x and

$$\partial_x J = \frac{\partial J}{\partial x}$$

gives the partial derivative of J with respect to x. The subscript of d or ∂ denotes the derivatives that are being taken. For example, the second partial derivatives of J are $\partial_{xx}J$, $\partial_{xz}J$ and $\partial_{zz}J$. The gradient of J with respect to a vector $\mathbf{x} = (x, z)$ is also a vector, which is symbolically expressed as:

$$\nabla_{\mathbf{x}} J = \left(\frac{\partial J}{\partial x}, \frac{\partial J}{\partial z}\right)^{\mathrm{T}}$$

2.2. PDE-constrained Optimization

In a seismic experiment, we refer to the physical observable recorded at the receivers as *observed or recorded data*. Seismic waveform inversion uses the observed data to estimate the following unknowns:

1. a state variable $v(\mathbf{x}, t)$ at each point $\mathbf{x}$ in the subsurface and evolving as a function of time t;
2. the model parameter vector $\mathbf{m}(\mathbf{x})$ at each point in the subsurface that influences the propagation of seismic waves.

It aims to match the state variable at the receivers to the observed data. The inversion can be formulated as a minimization problem with an objective function

$$J = \sum_s J_s(v, \mathbf{m}), \tag{2.1}$$

where the subscript s denotes a source position. Here, the model parameter vector $\mathbf{m}$ belongs to model space $\mathbb{M}$, which is a set of possible models of the subsurface. For a single-source experiment, the misfit between the state variable $v \in \mathbb{R}$ and the observed data at the receivers is quantified by $J_s : \mathbb{M} \times \mathbb{R} \to \mathbb{R}$. The functional J_s may also contain additional model-dependent regularization terms. We denote time by $t \in [0, t_1]$ and 2-D spatial coordinates with $\mathbf{x} \in \{(x, z) : z \geq z_0\} \subset \mathbb{R}^2$. The horizontal and vertical coordinates are denoted by x and z, respectively. The maximum recording time of the seismic experiment is given by t_1. A horizontal free surface is defined at $z = z_0$ and

a point on the free surface by $\mathbf{x}_0 = (x, z_0)$. The minimization of J_s is subject to the following partial differential equation (PDE) constraints:

$$\begin{cases} g(\mathbf{m}, \nu;\ \mathbf{x}, t) = \mathrm{L}[\mathbf{m}]\nu(\mathbf{x}, t) - \phi(\mathbf{x}, t) = 0; \\ \nu(\mathbf{x}, 0) = 0; \\ \partial_t \nu(\mathbf{x}, 0) = 0; \\ \nu(\mathbf{x}_0, t) = 0. \end{cases} \tag{2.2}$$

These constraints represent the physical laws that the state variable is assumed to obey. We denote the source term by ϕ. The acoustic or shear-horizontal wave operator,

$$\mathrm{L}[\mathbf{m}] = m_a \partial_{tt} - \partial_x m_b \partial_x - \partial_z m_b \partial_z, \tag{2.3}$$

depends on different subsurface medium parameters m_a and m_b, constituting the medium parameter vector $\mathbf{m} = (m_a, m_b)^{\mathrm{T}}$. The partial derivatives with respect to t, x and z are denoted by ∂_t, ∂_x and ∂_z, respectively. The initial boundary conditions imply that the state variable and its time derivative are zero everywhere in the medium at the initial $t = 0$. The Dirichlet spatial boundary condition, $\nu(\mathbf{x}_0, t) = 0$, is necessary for the reflecting free surface at $\mathbf{x} = \mathbf{x}_0$. The theory in the rest of the chapter should be modified accordingly when the Neumann boundary condition, $\partial_z \nu(\mathbf{x}_0, t) = 0$, is imposed at the boundary.

Consider a hypothetical case, where receivers are present everywhere in the subsurface: we can choose the state variable same as the observed data to minimize J_s. However, we can choose to impose additional restrictions on the state variables. We are only interested in those variables that satisfy the partial differential equation (PDE) in 2.2 for subsurface models in $\mathbb{M}$. When $g = 0$, the state variable is a physical realization and it is referred to as *modelled wavefield*. For $\mathbf{m} \in \mathbb{M}$ and $\nu \in \mathbb{R}$, the points in $\mathbb{M} \times \mathbb{R}$ that satisfy the wave-equation constraint belong to what we call the *feasible* region. If the modelled wavefield space is denoted by $\mathbb{U} \subset \mathbb{R}$, then the feasible region is given by $\mathbb{M} \times \mathbb{U}$. A forward simulation $\mathcal{F}$ maps every element of $\mathbb{M}$ to $\mathbb{U}$ using equations 2.2.

Example: 2-D Constrained Optimization

To aid the understanding of PDE-constrained optimization methods discussed in this chapter, we consider a simple two-dimensional constrained optimization example. Let a scalar $\check{m}$ denote a hypothetical subsurface model and a scalar $\check{\nu}$ a hypothetical state variable. In this chapter, the diacritical mark $\check{}$ is added to variables that are related to this simple example. We define an inverse problem that maximizes a functional

$$\check{J}(\check{m}, \check{\nu}) = 2 - \check{m}^2 - 2\check{\nu}^2, \tag{2.4}$$

and subject to a constraint

$$\check{g}(\check{m}, \check{\nu}) = \check{\nu} - 1 + \check{m} = 0. \tag{2.5}$$

Figure 2.1a plots contours or level curves of $\check{J}$. Clearly, the unconstrained maximum is at $(\check{m}, \check{\nu}) = (0, 0)$. Note that this point does not belong to the feasi-

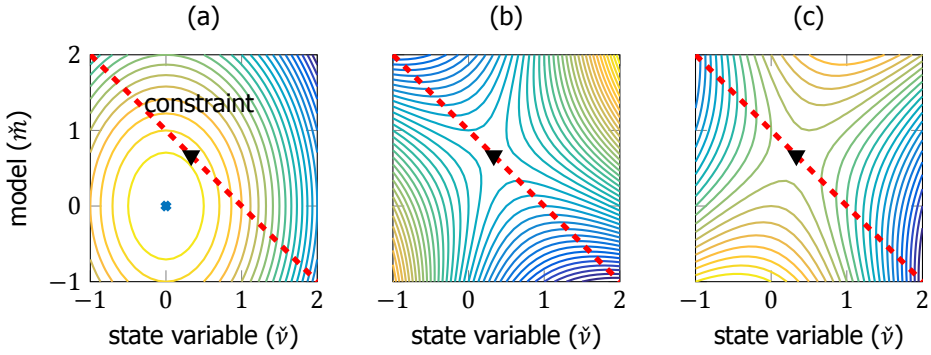


Figure 2.1: A two-dimensional constrained optimization example. (a) Contours or level curves of the objective function are plotted. The feasible region is marked using a dashed red line. (b) Contours of the reduced Lagrangian when the multipliers $\check{\chi}$ are chosen such that $\partial_{\check{v}}\check{\mathcal{L}} = 0$. (c) Contours of the reduced Lagrangian when the multipliers $\check{\chi}$ are chosen such that $\partial_{\check{m}}\check{\mathcal{L}} = 0$.

ble region. The points corresponding to the feasible region are marked by the red-dashed line in Figure 2.1a. To find the constrained maximum, we can first explicitly solve for $\check{v}$ using equation 2.5, yielding

$$\check{v} = 1 - \check{m},$$

and then substitute into $\check{J}$ to get

$$\check{J} = 2 - \check{m}^2 - 2(1 - \check{m})^2.$$

In this way, we arrive at a one-dimensional unconstrained optimization problem. Now, it is easy to find the stationary or saddle point, where the derivative of $\check{J}$ with respect to $\check{m}$ is zero. This point is marked by the triangle in Figure 2.1a. The current substitution technique cannot be used in the case of slightly more complicated constraints such as, for instance, the wave equation.

2.2.1. Method of Lagrange Multipliers

For simplicity, we consider a seismic experiment with one source such that $J = J_s$. We form a function, also called the Lagrangian $\mathcal{L} : \mathbb{M} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, to minimize J_s with the wave-equation constraint in 2.2 as:

$$\begin{aligned}
\mathcal{L}(\mathbf{m}, v, \chi) &= J_s(v, \mathbf{m}) - \langle \chi(\mathbf{x}, t), g(\mathbf{m}, v;\ \mathbf{x}, t)\rangle_{\mathbf{x},t}, \\
&= J_s(v, \mathbf{m}) - \langle \chi(\mathbf{x}, t), \mathrm{L}[\mathbf{m}]v(\mathbf{x}, t) - \phi(\mathbf{x}, t)\rangle_{\mathbf{x},t}, \\
&= J_s(v, \mathbf{m}) - \langle \chi(\mathbf{x}, t), \mathrm{L}[\mathbf{m}]v(\mathbf{x}, t)\rangle_{\mathbf{x},t} + \langle \chi(\mathbf{x}, t), \phi(\mathbf{x}, t)\rangle_{\mathbf{x},t}. \qquad (2.6)
\end{aligned}$$

$\chi(\mathbf{x}, t)$ denotes the Lagrange multipliers. We use $\langle \cdot, \cdot \rangle_{\mathbf{x},t}$ to denote the integration over space and time. For example,

$$\langle \chi(\mathbf{x}, t), \phi(\mathbf{x}, t) \rangle_{\mathbf{x},t} = \int_t \int_{\mathbf{x}} \chi(\mathbf{x}, t) \phi(\mathbf{x}, t) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t.$$

The Lagrangian depends on the state variable v, the medium-parameter vector $\mathbf{m}$ and the Lagrange multipliers χ. Note that the Lagrangian generally contains the initial and boundary-value constraints of equations 2.2. For simplicity, we assume that they have been removed by substitution.

Waveform inversion aims to find saddle points of the Lagrangian given in equation 2.6. At a saddle point the partial derivatives of the Lagrangian with respect to v, χ and $\mathbf{m}$ are zero:

$$\begin{cases} \nabla_{\mathbf{m}} \mathcal{L} = (0, 0)^{\mathrm{T}}; \\ \partial_v \mathcal{L} = 0; \\ \partial_\chi \mathcal{L} = 0. \end{cases} \tag{2.7}$$

Example: Lagrange Multipliers

A geometric way of solving the constrained optimization problem is to a find point where a particular level curve of $\check{J}$ *touches* the line $\check{g} = 0$. At such a point the normal vectors corresponding to the level curve and the constraint line are parallel to each other. This is referred to as the *tangential* condition, which is satisfied at the point marked by a triangle in Figure 2.1a. The method of Lagrange multipliers gives points that satisfy the tangential condition by solving the following equations simultaneously:

$$\begin{cases} \partial_{\check{m}} \check{J} = \check{\chi} \partial_{\check{m}} \check{g}; \\ \partial_{\check{v}} \check{J} = \check{\chi} \partial_{\check{v}} \check{g}; \\ \check{g} = 0. \end{cases} \tag{2.8}$$

Here, $\check{\chi}$ is the Lagrange multiplier. Alternatively, we can compactly represent the above equations at once by writing the Lagrangian,

$$\check{\mathcal{L}}(\check{m}, \check{v}, \check{\chi}) = \check{J} + \check{\chi} \check{g}, \tag{2.9}$$

and finding the saddle points where

$$\begin{cases} \partial_{\check{m}} \check{\mathcal{L}} = 0; \\ \partial_{\check{v}} \check{\mathcal{L}} = 0; \\ \partial_{\check{\chi}} \check{\mathcal{L}} = 0. \end{cases} \tag{2.10}$$

The Lagrangian is seen as a function of independent variables $\check{m}$, $\check{v}$ and $\check{\chi}$. So, we have reformulated the constrained optimization problem in two dimensions into an unconstrained optimization problem in three dimensions.

2.2.2. Full-space Optimization Methods

A *full-space* optimization approach searches the multi-dimensional space spanned by $\mathbf{m}$, v and χ to find the saddle points of the Lagrangian [Haber *et al.*, 2000]. In this approach, the wave-equation constraint is not necessarily satisfied during the optimization. However, the constraint does hold as soon as the optimization has converged to a saddle point. This approach is not feasible for large-scale seismic problems since storing and updating the state variable v and the Lagrange multipliers χ requires a large amount of memory. Contrast-source inversion (CSI) is discussed by van den Berg and Kleinman [1997] and Abubakar *et al.* [2008], where the optimization updates both the state variable and the medium parameters in the subsurface. A full-space optimization method of another kind is considered by van Leeuwen and Herrmann [2013], where an augmented functional is used instead of the Lagrangian.

Example: Full-space Optimization

A *full-space* optimization approach searches the whole space span by $\check{m}$, $\check{v}$ and $\check{\chi}$ for saddle points of the Lagrangian. After using the equations 2.4 and 2.5, the system of equations 2.10 reduces to the following system of linear equations:

$$\begin{cases} -2\check{m} + \check{\chi} = 0; \\ -4\check{v} + \check{\chi} = 0; \\ \check{v} - 1 + \check{m} = 0. \end{cases} \tag{2.11}$$

Example: Reduced-space Optimization

In this case, the search space of the optimization problem is reduced by explicitly solving either one or two equations of 2.10. Now we look at two different illustrations of reduced-space optimization. In the first illustration, the optimizer only searches in a space spanned by $\check{m}$ after the following two steps:

1. choose $\check{\chi}$ such that $\partial_{\check{v}}\check{\mathcal{L}} = 0$
2. and choose $\check{v}$ such that $\partial_{\check{\chi}}\check{\mathcal{L}} = 0$.

The contours of the reduced Lagrangian after the first step are plotted in Figure 2.1b as a function of $\check{v}$ and $\check{m}$. The second step forces the optimizer to only search along the constraint line to find the saddle point. Notice that everywhere along this line, the level curves are horizontal, since the derivative of the reduced Lagrangian with respect to the state variable is set to be zero.
In the second illustration, the optimizer only searches in a space spanned by $\check{v}$ after following the two steps as above except $\partial_{\check{m}}\check{\mathcal{L}} = 0$ during the first step. In this case, the reduced Lagrangian after the first step is plotted in Figure 2.1c. Notice that the contour lines are vertical everywhere on the constraint line, hence the reduced Lagrangian changes only with the state variable $\check{v}$.

2.2.3. Full Waveform Inversion (FWI) as Reduced-space Optimization

In full waveform inversion, instead of solving the equations 2.7 simultaneously, the dimensionality of the full-space Lagrangian optimization is reduced only to the model space by employing the following two steps.

1. We choose the state variable ν such that $\partial_\chi \mathcal{L} = 0$, to obtain

$$\mathrm{L}[\mathbf{m}]\nu(\mathbf{x}, t) - \phi(\mathbf{x}, t) = 0. \tag{2.12}$$

The above equation is known as the *state equation*. This means that, during FWI, we always choose the state variable such that it satisfies the wave-equation constraint. For this reason, the state variable is also referred to as the modelled wavefield. Once equation 2.12 is satisfied, the Lagrangian reduces to

$$\mathcal{L} = J_s. \tag{2.13}$$

2. Next, we differentiate $\mathcal{L}$ with respect to ν. Towards that end, we rewrite the term $\langle \chi, \mathrm{L}[\mathbf{m}]\nu \rangle_{\mathbf{x},t}$ in equation 2.6 by using integration by parts twice:

$$\begin{aligned}
\langle \chi, \mathrm{L}[\mathbf{m}]\nu \rangle_{\mathbf{x},t} &= \langle \chi, m_a \partial_{tt}\nu \rangle_{\mathbf{x},t} - \langle \chi, \partial_x m_b \partial_x \nu \rangle_{\mathbf{x},t} - \langle \chi, \partial_z m_b \partial_z \nu \rangle_{\mathbf{x},t}, && (2.14)\\
&= -\langle \partial_t \chi, m_a \partial_t \nu \rangle_{\mathbf{x},t} + \langle \chi, m_a \partial_t \nu \rangle_{\mathbf{x}} \Big|_{t=0}^{t=t_1} \\
&\quad + \langle \partial_x \chi, m_b \partial_x \nu \rangle_{\mathbf{x},t} - \langle \chi, m_b \partial_x \nu \rangle_t \Big|_{x=-\infty}^{x=\infty} \\
&\quad + \langle \partial_z \chi, m_b \partial_z \nu \rangle_{\mathbf{x},t} - \langle \chi, m_b \partial_z \nu \rangle_t \Big|_{z=z_0}^{z=\infty}, && (2.15)\\
&= \langle \nu, \mathrm{L}[\mathbf{m}]\chi \rangle_{\mathbf{x},t} \\
&\quad + \langle \chi, m_a \partial_t \nu \rangle_{\mathbf{x}} \Big|_{t=0}^{t=t_1} - \langle \partial_t \chi, m_a \nu \rangle_{\mathbf{x}} \Big|_{t=0}^{t=t_1} \\
&\quad + \langle m_b \partial_x \chi, \nu \rangle_t \Big|_{x=-\infty}^{x=\infty} - \langle \chi, m_b \partial_x \nu \rangle_t \Big|_{x=-\infty}^{x=\infty} \\
&\quad + \langle m_b \partial_z \chi, \nu \rangle_t \Big|_{z=z_0}^{z=\infty} - \langle \chi, m_b \partial_z \nu \rangle_t \Big|_{z=z_0}^{z=\infty}, && (2.16)\\
&= \langle \nu, \mathrm{L}[\mathbf{m}]\chi \rangle_{\mathbf{x},t} \\
&\quad + \langle \chi, m_a \partial_t \nu \rangle_{\mathbf{x}} \Big|_{t=t_1} - \langle \partial_t \chi, m_a \nu \rangle_{\mathbf{x}} \Big|_{t=t_1} \\
&\quad - \langle \chi, m_b \partial_z \nu \rangle_t \Big|_{z=z_0}. && (2.17)
\end{aligned}$$

In the equation 2.17, it was possible to have the wave operator L on both the left and right hand sides is the same because it is self-adjoint. We have used the initial and spatial boundary conditions of the equation 2.2 to eliminate the terms containing $\nu(\mathbf{x}, 0)$, $\partial_t \nu(\mathbf{x}, 0)$ and $\nu(\mathbf{x}_0, t)$. The modelled wavefield and its spatial derivatives vanish if $|x|$ or z tend to infinity (Sommerfield radiation conditions, Wapenaar and Haimé [1990]). Again, t_1 corresponds to the maximum modelling time and $z = z_0$ is the free surface. Substituting equation 2.17 into the Lagrangian

of equation 2.6 and then choosing the Lagrange multipliers χ such that $\partial_v \mathcal{L} = 0$ gives:

$$\begin{cases} \mathrm{L}[\mathbf{m}]\chi(\mathbf{x},t) - \partial_v J_s(\mathbf{x},t) = 0, \\ \chi(\mathbf{x},t_1) = 0, \\ \partial_t \chi(\mathbf{x},t_1) = 0, \\ \chi(\mathbf{x}_0,t) = 0. \end{cases} \tag{2.18}$$

The system of equations 2.18 has final conditions at $t = t_1$ and the same spatial boundary conditions as in the system 2.2. It has to be numerically solved in reverse time starting from $t = t_1$. Defining $\chi_1(t)$ as $\chi_1(t) = \chi(t_1 - t)$, and substituting this into equations 2.18 leads to

$$\begin{cases} \mathrm{L}[\mathbf{m}]\chi_1(\mathbf{x},t) - \partial_v J_s(\mathbf{x},t_1 - t) = 0, \\ \chi_1(\mathbf{x},0) = 0, \\ \partial_t \chi_1(\mathbf{x},0) = 0, \\ \chi(\mathbf{x}_0,t) = 0. \end{cases} \tag{2.19}$$

This system can be solved in forward time and it corresponds to the *adjoint state* χ_1, which is also referred to as the adjoint wavefield [Plessix, 2006]. It is modelled by back propagating adjoint source functions, given by $\partial_v J_s(\mathbf{x},t)$, into the medium.

In full waveform inversion, we use an iterative gradient-based optimizer to search in the feasible region for a point where the partial derivative of the Lagrangian with respect to medium parameter vector goes to zero. At every iteration, the gradient-based optimizer first computes a search direction that is used to update $\mathbf{m}$ such that the functional is minimized. In order to determine the search direction, typically the first and sometimes also the second derivatives of the Lagrangian with respect to $\mathbf{m}$ are used. We derive expressions of these derivatives in the following subsections.

First Derivatives of the Lagrangian with Respect to Medium Parameters

Using the relation $\chi_1(t) = \chi(t_1 - t)$ and differentiating the Lagrangian in the equation 2.6 with respect to the model vector $\mathbf{m}$, we obtain

$$\nabla_{\mathbf{m}} \mathcal{L}(\mathbf{x}) = \nabla_{\mathbf{m}} J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x},t_1 - t), \nabla_{\mathbf{m}} \mathrm{L}[\mathbf{m}] v(\mathbf{x},t) \rangle_t. \tag{2.20}$$

If model-dependent regularization terms are absent in J_s, then $\nabla_{\mathbf{m}} J_s = (0,0)$. We can write the derivatives with respect to m_a and m_b individually as

$$\begin{cases} \partial_{m_a} \mathcal{L}(\mathbf{x}) = \partial_{m_a} J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x},t_1 - t), \partial_{tt} v(\mathbf{x},t) \rangle_t; \\ \partial_{m_b} \mathcal{L}(\mathbf{x}) = \partial_{m_b} J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x},t_1 - t), -\partial_{xx} v(\mathbf{x},t) - \partial_{zz} v(\mathbf{x},t) \rangle_t. \end{cases} \tag{2.21}$$

Using equation 2.17, we can alternatively write:

$$\begin{cases} \partial_{m_a} \mathcal{L}(\mathbf{x}) = \partial_{m_a} J_s(\mathbf{x}) - \langle v(\mathbf{x},t), \partial_{tt} \chi_1(\mathbf{x},t_1 - t) \rangle_t; \\ \partial_{m_b} \mathcal{L}(\mathbf{x}) = \partial_{m_b} J_s(\mathbf{x}) - \langle v(\mathbf{x},t), -\partial_{xx} \chi_1(\mathbf{x},t_1 - t) - \partial_{zz} \chi_1(\mathbf{x},t_1 - t) \rangle_t. \end{cases} \tag{2.22}$$

The first derivatives of the Lagrangian with respect to the medium parameters can be computed using the forward modelled wavefield ν and the adjoint wavefield χ_1.

Second Derivatives of the Lagrangian with Respect to Medium Parameters

We refer the reader to Fichtner and Trampert [2011] for a discussion on second derivatives. Consider a point $\mathbf{x}_1$ and its corresponding medium parameter vector $\mathbf{m}_1 = \mathbf{m}(\mathbf{x}_1)$ has parameters $m_{1,a}$ and $m_{1,b}$. Our aim now is to express the derivative of $\partial_{m_{1,a}}\mathcal{L}$ with respect to the medium parameters at all other subsurface points. Towards that end, we use the expression for the first derivatives in equation 2.22. To compute the second derivatives, we introduce a Lagrangian

$$\begin{aligned}\grave{\mathcal{L}}_{1,a}(\nu,\chi,\psi_1,\psi_2,\mathbf{m}) \quad = \quad & \overbrace{\partial_{m_{1,a}}J_s(\mathbf{x}) - \langle\chi_1(\mathbf{x}_1,t_1-t),\partial_{m_{1,a}}\mathrm{L}[\mathbf{m}]\nu(\mathbf{x}_1,t)\rangle_t}^{\partial_{m_{1,a}}\mathcal{L}} \\ & -\langle\psi_1(\mathbf{x},t),\mathrm{L}[\mathbf{m}]\nu(\mathbf{x},t)-\phi(\mathbf{x},t)\rangle_{\mathbf{x},t} \\ & -\langle\psi_2(\mathbf{x},t),\mathrm{L}[\mathbf{m}]\chi_1(\mathbf{x},t)-\partial_\nu J(\mathbf{x},t_1-t)\rangle_{\mathbf{x},t}. \end{aligned} \tag{2.23}$$

The diacritical mark and the subscript in $\grave{\mathcal{L}}_{1,a}$ indicate that it includes the first derivative of the original Lagrangian $\mathcal{L}$ with respect to the medium parameter $m_{1,a}$ at $\mathbf{x}_1$. We aim now to compute the first derivatives of this Lagrangian $\grave{\mathcal{L}}_{1,a}$, which is a function of the state variable ν, the adjoint state variable χ_1, the model vector $\mathbf{m}$ and the Lagrange multipliers ψ_1 and ψ_2. Here, the forward wavefield ν and the adjoint wavefield χ_1 are constrained by equations 2.2 and 2.19, respectively. We follow the way we derived the first derivatives of $\mathcal{L}$ and first choose ν and χ_1 such that $\partial_{\psi_1}\grave{\mathcal{L}}_{1,a}=0$ and $\partial_{\psi_2}\grave{\mathcal{L}}_{1,a}=0$, respectively. In this way, the constraints in equations 2.2 and 2.19 are satisfied. Then, we choose ψ_2 such that $\partial_{\chi_1}\grave{\mathcal{L}}_{1,a}=0$, to obtain

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_2(\mathbf{x},t)+\partial_{m_{1,a}}\mathrm{L}\nu(\mathbf{x}_1,t_1-t)=0;\\ \psi_2(\mathbf{x},t_1)=0;\\ \partial_t\psi_2(\mathbf{x},t_1)=0;\\ \psi_2(\mathbf{x}_0,t)=0.\end{cases} \tag{2.24}$$

We introduce the first secondary adjoint state for the computation of the second derivatives as $\psi_4(\mathbf{x},t)=\psi_2(\mathbf{x},t_1-t)$ and use $\partial_{m_{1,a}}\mathrm{L}=\partial_{tt}$ to rewrite the above system of equations:

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_4(\mathbf{x},t)+\partial_{tt}\nu(\mathbf{x}_1,t)=0;\\ \psi_4(\mathbf{x},0)=0;\\ \partial_t\psi_4(\mathbf{x},0)=0;\\ \psi_2(\mathbf{x}_0,t)=0.\end{cases} \tag{2.25}$$

Similarly, we choose ψ_1 such that $\partial_\nu\grave{\mathcal{L}}_{1,a}(\mathbf{x},t)=0$, yielding

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_1(\mathbf{x},t)+\partial_{m_{1,a}}\mathrm{L}\chi_1(\mathbf{x}_1,t_1-t)-\psi_2(\mathbf{x},t)\partial_{\nu\nu}J_s(\mathbf{x},t)=0;\\ \psi_1(\mathbf{x},t_1)=0;\\ \partial_t\psi_1(\mathbf{x},t_1)=0;\\ \psi_1(\mathbf{x}_0,t)=0.\end{cases} \tag{2.26}$$

We introduce the second secondary adjoint state for the computation second derivatives as $\psi_3(\mathbf{x},t)=\psi_1(\mathbf{x},t_1-t)$ and use $\partial_{m_{1,a}}\mathrm{L}=\partial_{tt}$ to rewrite the above equations:

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_3(\mathbf{x},t)+\partial_{tt}\chi_1(\mathbf{x}_1,t)-\psi_4(\mathbf{x},t)\partial_{vv}J_s(\mathbf{x},t_1-t)=0;\\ \psi_3(\mathbf{x},0)=0;\\ \partial_t\psi_3(\mathbf{x},0)=0;\\ \psi_3(\mathbf{x}_0,t)=0.\end{cases} \tag{2.27}$$

Finally, the expression for the first derivatives of $\grave{\mathcal{L}}_{1,a}$ in terms of the secondary adjoint wavefields, ψ_3 and ψ_4, is given by

$$\begin{aligned}\nabla_{\mathbf{m}}\grave{\mathcal{L}}_{1,a}(\mathbf{x}) &= \nabla_{\mathbf{m}}\partial_{m_{1,a}}J_s(\mathbf{x})\\ &\quad -\langle\chi_1(\mathbf{x}_1,t_1-t),\nabla_{\mathbf{m}}\partial_{m_{1,a}}\mathrm{L}v(\mathbf{x}_1,t)\rangle_t\\ &\quad -\langle\psi_3(\mathbf{x},t_1-t),\nabla_{\mathbf{m}}\mathrm{L}[\mathbf{m}]v(\mathbf{x},t)\rangle_t\\ &\quad -\langle\psi_4(\mathbf{x},t_1-t),\nabla_{\mathbf{m}}\mathrm{L}[\mathbf{m}]\chi_1(\mathbf{x},t)\rangle_t.\end{aligned} \tag{2.28}$$

Again, $\nabla_{\mathbf{m}}\partial_{m_{1,a}}J_s(\mathbf{x})=(0,0)$ if the functional J_s has no model-dependent regularization terms. Using the above expression for the gradient with respect to m_a, we obtain

$$\begin{aligned}\partial_{m_a}\grave{\mathcal{L}}_{1,a}(\mathbf{x}) &= \partial_{m_a}\partial_{m_{1,a}}J_s(\mathbf{x})\\ &\quad -\langle\psi_3(\mathbf{x},t_1-t),\partial_{tt}v(\mathbf{x},t)\rangle_t\\ &\quad -\langle\psi_4(\mathbf{x},t_1-t),\partial_{tt}\chi_1(\mathbf{x},t)\rangle_t.\end{aligned} \tag{2.29}$$

The derivative with respect to m_b is given by

$$\begin{aligned}\partial_{m_b}\grave{\mathcal{L}}_{1,a}(\mathbf{x}) &= \partial_{m_b}\partial_{m_{1,a}}J_s(\mathbf{x})\\ &\quad -\langle\psi_3(\mathbf{x},t_1-t),-\partial_{xx}v(\mathbf{x},t)-\partial_{zz}v(\mathbf{x},t)\rangle_t\\ &\quad -\langle\psi_4(\mathbf{x},t_1-t),-\partial_{xx}\chi_1(\mathbf{x},t)-\partial_{zz}\chi_1(\mathbf{x},t)\rangle_t.\end{aligned} \tag{2.30}$$

In a similar way, we can compute the derivative of $\partial_{m_{1,b}}\mathcal{L}$ with respect to the parameter m_b everywhere in the subsurface by forming another Lagrangian $\grave{\mathcal{L}}_{1,b}$. We use two additional secondary adjoint states, ψ_5 and ψ_6, given by

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_6(\mathbf{x},t)-\partial_{xx}v(\mathbf{x}_1,t)-\partial_{zz}v(\mathbf{x}_1,t)=0;\\ \psi_6(\mathbf{x},0)=0;\\ \partial_t\psi_6(\mathbf{x},0)=0;\\ \psi_6(\mathbf{x}_0,t)=0;\end{cases} \tag{2.31}$$

$$\begin{cases}\mathrm{L}[\mathbf{m}]\psi_5(\mathbf{x},t)-\partial_{xx}\chi_1(\mathbf{x}_1,t)-\partial_{zz}\chi_1(\mathbf{x}_1,t)-\psi_6(\mathbf{x},t)\partial_{vv}J_s(\mathbf{x},t_1-t)=0;\\ \psi_5(\mathbf{x},0)=0;\\ \partial_t\psi_5(\mathbf{x},0)=0;\\ \psi_5(\mathbf{x}_0,t)=0.\end{cases} \tag{2.32}$$

The final expression for the derivative becomes

$$\begin{aligned}\partial_{m_b}\grave{\mathcal{L}}_{1,b}(\mathbf{x}) &= \partial_{m_b}\partial_{m_{1,b}}J_s(\mathbf{x}) \\ &\quad -\langle\psi_5(\mathbf{x},t_1-t),-\partial_{xx}v(\mathbf{x},t)-\partial_{zz}v(\mathbf{x},t)\rangle_t \\ &\quad -\langle\psi_6(\mathbf{x},t_1-t),-\partial_{xx}\chi_1(\mathbf{x},t)-\partial_{zz}\chi_1(\mathbf{x},t)\rangle_t. \end{aligned} \tag{2.33}$$

2.2.4. Interpretation of FWI Without Using the Lagrangian

FWI can also be formulated as a minimization problem without using the Lagrangian. This is due to the fact that the dimensionality of the Lagrangian is finally reduced only to the model space during FWI, as we have already shown. We now minimize a functional $\tilde{J}_s(v(\mathbf{m}),\mathbf{m})$ of only $\mathbf{m}$ such that the modelled wavefield v depends on the model vector $\mathbf{m}$ through equations 2.2. For gradient-based optimization, we need to find the total derivative of $\tilde{J}_s$ with respect to $\mathbf{m}$ given by:

$$\begin{cases}\mathrm{d}_{m_a}\tilde{J}_s = \partial_{m_a}\tilde{J}_s + (\partial_v\tilde{J}_s)(\partial_{m_a}v); \\ \mathrm{d}_{m_b}\tilde{J}_s = \partial_{m_b}\tilde{J}_s + (\partial_v\tilde{J}_s)(\partial_{m_b}v).\end{cases} \tag{2.34}$$

Since $\tilde{J}_s = J_s$, we can replace $\tilde{J}_s$ on the right hand side of the above equation with the reduced Lagrangian, as in equation 2.13, when $\partial_v\mathcal{L}=0$ and $\partial_\chi\mathcal{L}=0$, to obtain

$$\begin{cases}\mathrm{d}_{m_a}\tilde{J}_s = \partial_{m_a}\mathcal{L}; \\ \mathrm{d}_{m_b}\tilde{J}_s = \partial_{m_b}\mathcal{L}.\end{cases} \tag{2.35}$$

Here, $\partial_{m_a}\mathcal{L}$ and $\partial_{m_b}\mathcal{L}$, follow from equations 2.21.

2.3. Specific Wave Operators

Until now, we have considered a more generalized wave operator in equation 2.3 with two different medium parameters, m_a and m_b, such that the propagation of waves is in the x-z plane. In this section, we elaborate on full waveform inversion for two specific wave operators. The first operator corresponds to the propagation of pressure waves in fluid regions of the Earth. It depends on the bulk modulus and mass density of the medium. The second operator corresponds to the propagation of shear horizontal (SH) waves in an elastic medium. It depends on the shear modulus and mass density of the elastic medium. For each of these two operators, we can specialize the expressions 2.21 of the first-order derivatives of the misfit functional with respect to the medium parameters.

2.3.1. Elastic Wave Equation

We now review the equations governing the propagation of elastic waves. At each point in a three-dimensional isotropic elastic medium, the symmetric stress tensor is given by

$$\sigma = \begin{bmatrix}\sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz}\end{bmatrix}. \tag{2.36}$$

Given a normal vector corresponding to an infinitesimal plane, the components of the stress tensor can be used to calculate the force acting on the plane. For example, the force acting on an infinitesimal plane normal to a unit vector along y-axis is given by $\sigma_{yx}\hat{\imath}+\sigma_{yy}\hat{\jmath}+\sigma_{yz}\hat{k}$. Here, $\hat{\imath}$, $\hat{\jmath}$ and $\hat{k}$ are the unit vectors along x, y and z axes, respectively. We denote the vector field of particle displacement using $u = u_x\hat{\imath} + u_y\hat{\jmath} + u_z\hat{k}$. In order to write equations of motion, we now consider an infinitesimal cube in the medium. The external force acting on an infinitesimal cube in the medium is denoted by $\kappa = \kappa_x\hat{\imath} + \kappa_y\hat{\jmath} + \kappa_z\hat{k}$. Using the definition of the stress tensor, the net force acting on the infinitesimal cube in x direction is given by $\partial_x\sigma_{xx} + \partial_x\sigma_{xy} + \partial_x\sigma_{xz} + \kappa_x$. The net force acting on the cube is non-zero not only because of the external force but also because of the presence of gradients in the stress tensor. Newton's second law provides the equation of motion for the continuum, relating the different components of the particle accelerations and the stress tensor:

$$\begin{cases} \rho\partial_{tt}u_x - \partial_x\sigma_{xx} - \partial_x\sigma_{xy} - \partial_x\sigma_{xz} = \kappa_x; \\ \rho\partial_{tt}u_y - \partial_y\sigma_{yx} - \partial_y\sigma_{yy} - \partial_y\sigma_{yz} = \kappa_y; \\ \rho\partial_{tt}u_z - \partial_z\sigma_{zx} - \partial_z\sigma_{zy} - \partial_z\sigma_{zz} = \kappa_z. \end{cases} \tag{2.37}$$

Note that u, κ, σ are functions of position and time. The linear stress-strain relationship for an elastic medium is given by Hooke's law,

$$\sigma_{ij} = \lambda\,\delta_{ij}\,\partial_k u_k + \mu\left(\partial_j u_i + \partial_i u_j\right), \tag{2.38}$$

where λ and μ are the Lamé parameters. Here, the subscripts i, j and k can be either of x, y or z, and δ_{ij} denotes the Kronecker delta function.

2.3.2. 2-D Acoustic Wave Operator

The equations governing the elastic-wave propagation are significantly simplified if we only consider pressure waves by setting $\mu = 0$. The result is the acoustic wave equation that describes sound waves in non-moving fluids but also provides an approximate description of pressure waves in solids. The stress tensor becomes a diagonal matrix, meaning that the force acting on an infinitesimal plane is independent of its orientation. Using the relation $\sigma_{ij} = -\delta_{ij}\mathfrak{p}$ to simplify equations 2.37, we obtain

$$\begin{cases} \rho\partial_{tt}u_x + \partial_x\mathfrak{p} = \kappa_x; \\ \rho\partial_{tt}u_z + \partial_z\mathfrak{p} = \kappa_z, \end{cases} \tag{2.39}$$

where $\mathfrak{p}$ denotes a scalar pressure field in the medium. Hooke's law for fluids states that the pressure $\mathfrak{p}$ at a given point depends on the stiffness or bulk modulus K at that point and the divergence of the particle displacement field, according to

$$\mathfrak{p} = -K\left(\partial_x u_x + \partial_z u_z\right). \tag{2.40}$$

After eliminating the particle displacement from equations 2.39 and 2.40, we obtain a 2-D second-order wave equation for fluids

$$\mathrm{L}_a\mathfrak{p} - \phi_a = K^{-1}\partial_{tt}\mathfrak{p} - \partial_x(\rho^{-1}\partial_x\mathfrak{p}) - \partial_z(\rho^{-1}\partial_z\mathfrak{p}) - \partial_x(\rho^{-1}\kappa_x) - \partial_z(\rho^{-1}\kappa_z) = 0, \tag{2.41}$$

with a wave operator

$$\mathrm{L}_a = K^{-1}\partial_{tt} - \partial_x(\rho^{-1}\partial_x) - \partial_z(\rho^{-1}\partial_z), \tag{2.42}$$

and source term

$$\phi_a = \partial_x(\rho^{-1}\kappa_x) + \partial_z(\rho^{-1}\kappa_z). \tag{2.43}$$

Equation 2.41 is commonly referred to as the 2-D second-order acoustic wave equation and can be used as the constraint for acoustic FWI. Acoustic FWI minimizes the misfit between recorded and modelled pressure wavefield at the receivers. The medium parameter vector that is estimated during inversion consists in the bulk modulus and the mass density. Using $m_a = K^{-1}$ and $m_b = \rho^{-1}$, the equations 2.21 are rewritten as:

$$\begin{cases} \partial_{K^{-1}}\mathcal{L}(\mathbf{x}) = \partial_{K^{-1}}J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x}, t_1 - t), \partial_{tt}v(\mathbf{x}, t)\rangle_t \\ \partial_{\rho^{-1}}\mathcal{L}(\mathbf{x}) = \partial_{\rho^{-1}}J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x}, t_1 - t), -\partial_{xx}v(\mathbf{x}, t) - \partial_{zz}v(\mathbf{x}, t)\rangle_t \end{cases} \tag{2.44}$$

Note that in this case, v is the modelled pressure wavefield. If the subsurface is parameterized by pressure-wave velocity $\bar{c_p} = c_p = \sqrt{K/\rho}$ and density $\bar{\rho} = \rho$ instead of inverse bulk modulus K^{-1} and inverse density ρ^{-1}, the respective derivatives after applying the chain rule become

$$\begin{cases} \partial_{\bar{c_p}}\mathcal{L}(\mathbf{x}) = -2c_p^{-3}\rho^{-1}\partial_{K^{-1}}\mathcal{L}(\mathbf{x}); \\ \partial_{\bar{\rho}}\mathcal{L}(\mathbf{x}) = -\rho^{-2}\partial_{\rho^{-1}}J_s(\mathbf{x}) - c_p^{-2}\rho^{-2}\partial_{K^{-1}}\mathcal{L}(\mathbf{x}). \end{cases} \tag{2.45}$$

Here, the diacritical mark $\bar{}$ denotes the subsurface parameterization.

2.3.3. 2-D SH Wave Operator

In a 2-D SH seismic experiment, horizontally polarized shear waves are used to image the subsurface. Sources and receivers are placed along a transect, say the x-axis. The vibrator sources primarily inject a ground force in the y-direction, perpendicular to the transect. As a result, SH waves propagate in the x, z-plane with particle displacement only in the y-direction. The receivers are assumed to record only the y-component of particle velocity field. For this reason, we substitute $u = u_y\hat{j}$, to write the equation of motion in equations 2.37 as

$$\rho\partial_{tt}u_y - \partial_x\sigma_{yx} - \partial_z\sigma_{yz} = \kappa_y; \tag{2.46}$$

and Hooke's law in equation 2.38 as

$$\begin{cases} \sigma_{yx} = \mu\partial_x u_y; \\ \sigma_{yz} = \mu\partial_z u_y. \end{cases} \tag{2.47}$$

Eliminating the stress components σ_{yx} and σ_{yz} from equations 2.46 and 2.47, we have a second-order equation:

$$\rho\partial_{tt}u_y - \partial_x(\mu\partial_x u_y) - \partial_z(\mu\partial_z u_y) = \kappa_y. \tag{2.48}$$

In terms of the particle velocity in the y-direction v_y, we rewrite the equation above to obtain the 2-D SH wave equation

$$\mathrm{L}_{sh} v_y - \phi_{sh} = \rho \partial_{tt} v_y - \partial_x(\mu \partial_x v_y) - \partial_z(\mu \partial_z v_y) - \partial_t \kappa_y = 0, \tag{2.49}$$

which has a wave operator

$$\mathrm{L}_{sh} = \rho \partial_{tt} - \partial_x(\mu \partial_x) - \partial_z(\mu \partial_z), \tag{2.50}$$

and source term

$$\phi_{sh} = \partial_t \kappa_y. \tag{2.51}$$

We use the wave equation 2.49 as a constraint and choose v_y as the state variable for 2-D SH FWI. In this case, the gradients of the objective function with respect to the shear modulus μ and mass density ρ are given by rewriting equations 2.21 as:

$$\begin{cases} \partial_\rho \mathcal{L}(\mathbf{x}) = \partial_\rho J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x}, t_1 - t), \partial_{tt} v(\mathbf{x}, t) \rangle_t \\ \partial_\mu \mathcal{L}(\mathbf{x}) = \partial_\mu J_s(\mathbf{x}) - \langle \chi_1(\mathbf{x}, t_1 - t), -\partial_{xx} v(\mathbf{x}, t) - \partial_{zz} v(\mathbf{x}, t) \rangle_t \end{cases} \tag{2.52}$$

If the subsurface is parameterized by the shear-wave velocity $\bar{c_s} = c_s = \sqrt{\mu/\rho}$ and density $\bar{\rho} = \rho$ instead of shear modulus μ and density ρ, we can express the derivatives, after applying the chain rule, as

$$\begin{cases} \partial_{\bar{\rho}} \mathcal{L}(\mathbf{x}) = \partial_\rho \mathcal{L}(\mathbf{x}) + c_s^2 \partial_\mu J_s(\mathbf{x}) \\ \partial_{\bar{c_s}} \mathcal{L}(\mathbf{x}) = 2\sqrt{\mu\rho}\, \partial_\mu J_s(\mathbf{x}) \end{cases} \tag{2.53}$$

Again, the diacritical mark $\bar{}$ denotes the subsurface parameterization.

2.4. Data Misfit Functionals

In this section, we discuss various data misfit functionals commonly used in FWI. The modelled wavefield v corresponding to a particular source at $\mathbf{x}_s$ has to be restricted or sampled to the receiver locations prior to comparing it to the observed data. A restriction operator, $S(\mathbf{x}, \mathbf{x}_r) : \mathbb{U} \rightarrow \mathbb{D}$, maps the modelled wavefield into the data space $\mathbb{D}$. In discrete form, S corresponds to an interpolation operator. For every source position $\mathbf{x}_s$, *modelled data* at a particular receiver location $\mathbf{x}_r$ are given by

$$p(\mathbf{x}_r, t;\ \mathbf{x}_s) = S(\mathbf{x}, \mathbf{x}_r) v(\mathbf{x}, t;\ \mathbf{x}_s). \tag{2.54}$$

2.4.1. Least-squares Misfit

The least-squares misfit is given by the sum of squared differences between the observed and modelled data. For an experiment with multiple source positions, it is given by

$$J_{ls} = \tfrac{1}{2} \sum_s J_{ls,s} \tag{2.55}$$

$$= \tfrac{1}{2} \sum_s \sum_r \sum_t \left[p(\mathbf{x}_r, t;\ \mathbf{x}_s) - q(\mathbf{x}_r, t;\ \mathbf{x}_s) \right]^2. \tag{2.56}$$

Here, $q(\mathbf{x}_r, t;\ \mathbf{x}_s) \in \mathbb{D}$ are the pre-processed observed data for a source at $\mathbf{x}_s$ and receiver at $\mathbf{x}_r$ as a function of time t, whereas $p(\mathbf{x}_r, t;\ \mathbf{x}_s) \in \mathbb{D}$ represents the modelled data.

At a particular source position, the adjoint sources necessary for modelling the adjoint wavefield in equation 2.19 are given by $\partial_v J_s(\mathbf{x}, t_1 - t)$. In the case of the least-squares misfit, we use chain rule to obtain

$$\partial_v J_{ls,s}(\mathbf{x}, t) \quad = \quad (\partial_v p)(\partial_p J_{ls,s}). \tag{2.57}$$

Using equation 2.54, we get

$$\partial_v J_{ls,s}(\mathbf{x}, t) \quad = \quad S^{\mathrm{T}}(\mathbf{x}, \mathbf{x}_r)\,[p(\mathbf{x}_r, t; \mathbf{x}_s) - q(\mathbf{x}_r, t)]\,, \tag{2.58}$$

where S^{T} in discrete form is an adjoint interpolation operator that *sprays* the adjoint sources from the receiver locations to the grid points. It is often difficult to achieve the best possible fit between the modelled and the observed data when only the least-squares misfit is used. This is primarily due to cycle-skipping, which is discussed with examples in the following section.

2.4.2. Envelope-based Misfit

The envelope-based misfit measures the difference between the envelopes of the observed and the modelled data [Bozdağ *et al.*, 2011; Wu *et al.*, 2014; Chi *et al.*, 2014; Luo and Wu, 2015]. The envelope of a real-valued signal is found by computing the absolute value of its analytic representation. We denote the envelope operator by E. The envelope of the observed data is given by

$$\mathrm{E}(q) = \sqrt{q^2 + \mathrm{H}^2(q)} \tag{2.59}$$

In the above equation, $\mathrm{H}(q)$ denotes the Hilbert transform of the observed data q. After computing the observed and modelled envelopes, the misfit is given by

$$J_{env} = \sum_s \sum_r \sum_t \left[\mathrm{E}(p(\mathbf{x}_r, t;\ \mathbf{x}_s)) - \mathrm{E}(q(\mathbf{x}_r, t;\ \mathbf{x}_s))\right]^2. \tag{2.60}$$

The pros and cons of using this misfit are given in Chapter 3 of this thesis. The envelope-based misfit suffers less from the cycle-skipping problem than the least-squares objective.

2.4.3. Correlation-based Misfit

In order to compute this misfit, we first need to rescale each receiver record of the data using its root-mean-squared amplitude. For example, the rescaled modelled data are given by

$$p_n(\mathbf{x}_r, t;\ \mathbf{x}_s) = p_n(\mathbf{x}_r, t;\ \mathbf{x}_s) \left[\int_t p^2(\mathbf{x}_r, t;\ \mathbf{x}_s)\mathrm{d}t\right]^{-\frac{1}{2}}. \tag{2.61}$$

Inversion with a cross-correlation based misfit estimates a model such that the rescaled modelled data better correlates with the rescaled observed data at the receivers. Mathematically, it aims to maximize the energy around zero lag of the cross-correlation between the rescaled observed and the rescaled modelled data, denoted by q_n and p_n, respectively. The cross-correlation based inversion minimizes the misfit J_{corr} given by

$$J_{corr} = -\frac{1}{2}\sum_{s}\sum_{r}\sum_{t} w(t)\left[p_n(\mathbf{x}_r, t;\ \mathbf{x}_s) \otimes_t q_n(\mathbf{x}_r, t;\ \mathbf{x}_s)\right]^2 . \tag{2.62}$$

In the above equation, $\otimes_t$ denotes cross-correlation in time and $w(t) > 0$ is a Gaussian window used to quantify energy only close to $t = 0$. This cross-correlation based misfit is useful for kinematically fitting transmitted arrivals. The rescaling prevents its application to complicated datasets with several arrivals.

2.5. Cycle-skipping in Least-squares FWI

Due to the absence of reliable low frequencies in the observed data, the classic least-squares full waveform inversion suffers from cycle-skipping or loop-skipping problem [Mulder and Plessix, 2008]. To illustrate this, we consider two simple synthetic experiments using one source and one receiver. The aim of these experiments is to fit the *observed* data by updating the velocity between the source and receiver.

In the first experiment, a 40-Hz Ricker source wavelet is used to to generate the observed data with only a direct arrival as plotted in Figure 2.2a. The starting velocity model for this experiment is plotted in Figure 2.3a. The modelled direct arrival using the starting model, plotted in Figure 2.2a, has a travel time greater than that of the observed arrival by more than half the dominant period τ_d. At the end of least-squares FWI, the modelled direct arrival does not fit the observed arrival as plotted in Figure 2.2a. The inversion converges to a local minimum due to cycle-skipping. The estimated velocity between the source and the receiver, plotted in Figure 2.3c, is lower than in the starting model.

During the second experiment, we used a 20-Hz Ricker wavelet to generate the observed data. The observed and the initially modelled arrivals are plotted in the Figure 2.2b. The arrival time error in the initially modelled data is less than half the dominant period τ_d. The starting model for inversion, plotted in Figure 2.3b, is same as that of the first experiment. FWI has successfully increased the velocity between the source and receiver as in the final estimated model of Figure 2.3d. This results in a better fit between the modelled data and the observed data as shown in the Figure 2.2b.

2.6. Conclusions

In this chapter, we have outlined the basics of wave-equation constrained optimization. Full waveform inversion can be seen as reduced-space optimization problem, where only the model space is searched to find the minimum. We have given details about 2-D acoustic and SH FWI. Different data misfit functionals that can be minimized during FWI are explained. Finally, we have demonstrated the cycle-skipping problem that limits the applicability of least-squares FWI in the absence of low-frequencies in the data.

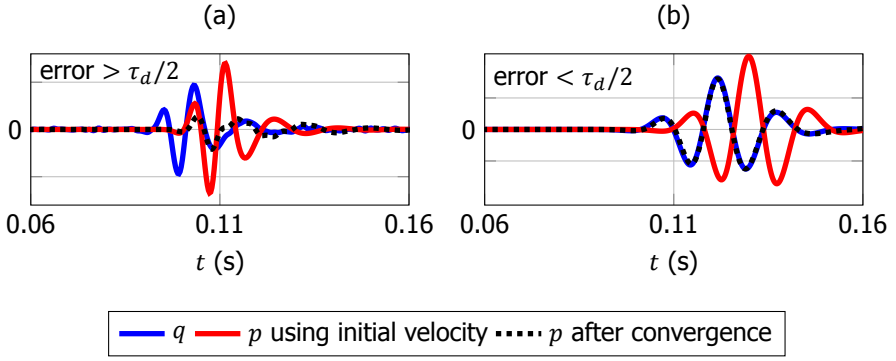


Figure 2.2: (a) The modelled data before and after the least-squares inversion are compared to the observed data in the first experiment with a 40-Hz Ricker source wavelet. The initial traveltime error in this case is greater than $0.5\tau_d$. (b) Same as a) but for the second experiment with a 20-Hz Ricker source wavelet. The initial traveltime error in this case is less than $0.5\tau_d$.

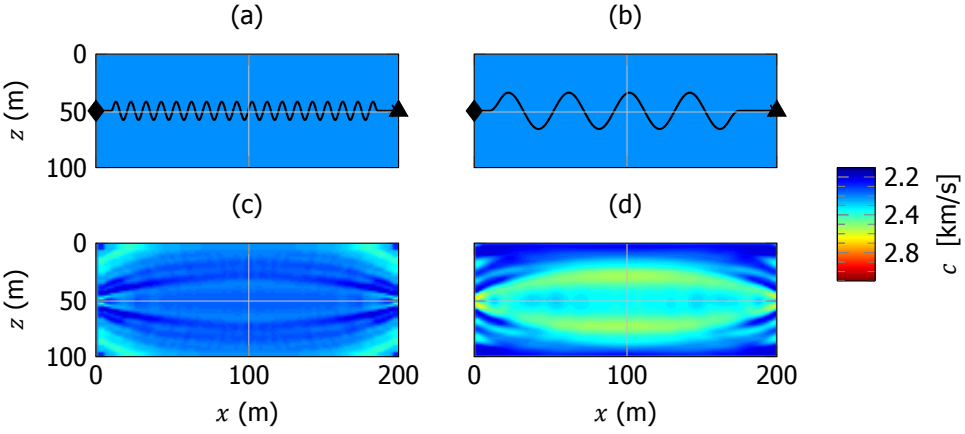


Figure 2.3: A 40-Hz Ricker source wavelet is used to generate the observed data during the first synthetic experiment. In this case, (c) the estimated velocity is lower than (a) the initial velocity between the source (diamond) and receiver (triangle). The second experiment is similar to the first experiment, except for a 20-Hz Ricker source wavelet is used to generate the observed data. Now (d) the estimated velocity is higher than (b) the initial velocity.

References

E. Haber, U. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems,* Inverse Problems **16**, 1263 (2000).

P. M. van den Berg and R. E. Kleinman, *A contrast source inversion method,* Inverse Problems **1607** (1997).

A. Abubakar, W. Hu, P. M. van den Berg, and T. M. Habashy, *A finite-difference contrast source inversion method,* Inverse Problems **24**, 065004 (2008).

T. van Leeuwen and F. J. Herrmann, *Mitigating local minima in full-waveform inversion by expanding the search space,* Geophysical Journal International **195**, 661 (2013).

C. P. A. Wapenaar and G. C. Haimé, *Elastic extrapolation of primary seismic P- and S-waves,* Geophysical Prospecting **38**, 23 (1990).

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

A. Fichtner and J. Trampert, *Hessian kernels of seismic data functionals based upon adjoint techniques,* Geophysical Journal International **185**, 775 (2011).

E. Bozdağ, J. Trampert, and J. Tromp, *Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements,* Geophysical Journal International **185**, 845 (2011).

R.-S. Wu, J. Luo, and B. Wu, *Seismic envelope inversion and modulation signal model,* Geophysics **79**, WA13 (2014).

B. Chi, L. Dong, and Y. Liu, *Full waveform inversion method using envelope objective function without low frequency data,* Journal of Applied Geophysics **109**, 36 (2014).

J. Luo and R.-S. Wu, *Seismic envelope inversion: reduction of local minima and noise resistance,* Geophysical Prospecting **63**, 597 (2015).

W. A. Mulder and R. E. Plessix, *Exploring some issues in acoustic full waveform inversion,* Geophysical Prospecting **56**, 827 (2008).

3

Full Waveform Inversion With An Auxiliary Bump Functional

Take up one idea.
Make that one idea your life — think of it, dream of it, live on that idea.
Let the brain, muscles, nerves, every part of your body, be full of that idea,
and just leave every other idea alone. This is the way to success.

Swami Vivekananda

Least-squares inversion of seismic arrivals can provide remarkably detailed models of the Earth's subsurface. However, cycle skipping associated with these oscillatory arrivals is the main cause for local minima in the least-squares objective function. Therefore, it is often difficult for descent methods to converge to the solution without an accurate initial large-scale velocity estimate. The low frequencies in the arrivals, needed to update the large-scale components in the velocity model, are usually unreliable or absent.

To overcome this difficulty, we propose a multi-objective inversion scheme that uses the conventional least-squares functional along with an auxiliary data-domain objective. As the auxiliary objective effectively replaces the seismic arrivals by bumps, we call it the bump functional. The bump functional minimization can be made far less sensitive to cycle skipping and can deal with multiple arrivals in the data. However, it can only be used as an auxiliary objective since it usually does not provide a unique model after minimization even when the regularized-least-squares functional has a unique global minimum and hence a unique solution. The role of the bump functional during the multi-objective inversion is to guide the optimization towards the global minimum by pulling the trapped solution out

Parts of this chapter have been published in Geophysical Journal International **206 (2)**, 1076 (2016).

of the local minima associated with the least-squares functional whenever necessary. The computational complexity of the bump functional is equivalent to that of the least-squares functional.

In this paper, we describe various characteristics of the bump functional using simple and illustrative numerical examples. We also demonstrate the effectiveness of the proposed multi-objective inversion scheme by considering more realistic examples. These include syntheic and field data from a cross-well experiment, surface-seismic synthetic data with reflections and synthetic data with refracted arrivals at long offsets.

3.1. Introduction

Full waveform inversion (FWI) is a non-linear optimization procedure that estimates the Earth's model parameters by least-squares fitting of the recorded arrivals in the seismic data [Tarantola, 1984; Mora, 1988, 1989; Pratt, 1999; Virieux and Operto, 2009; Fichtner, 2010]. Due to the computational cost of the wave-equation modelling, the optimization is usually performed with gradient-based techniques, although several authors have tried more costly global optimization techniques [Sen and Stoffa, 1991; Stoffa and Sen, 1991; Sambridge and Drijkoningen, 1992; Gao *et al.*, 2014; Datta, 2015]. While fitting the observed and the modelled seismic arrivals with the conventional least-squares objective function, the gradient-based optimization will get trapped in the nearest local minimum when the error in the arrival time exceeds about half a period of the signal [Gauthier *et al.*, 1986; Snieder *et al.*, 1989; Mulder and Plessix, 2008; Symes, 2008]. Here, period is related to the dominant frequency of the data. In other words, the least-squares inversion cannot reconstruct velocity anomalies that cause shifts in the arrival times larger than half a period. Usually, the velocity anomalies accounting for the arrival times have relatively low wave-numbers. Hence, inversion of the low-frequency seismic signals with larger periods is easier and they help in the reconstruction of a kinematically correct velocity model.

3.1.1. Data-domain Objective Functions

Many authors have formulated alternative data-domain functionals to achieve global convergence in the absence of the necessary low frequencies [Luo and Schuster, 1991; Shin and Min, 2006; Shin and Cha, 2008, 2009; Zhang and Wang, 2009; van Leeuwen and Mulder, 2010; van Leeuwen, 2010; Bozdağ *et al.*, 2011; Chauris *et al.*, 2012; Donno *et al.*, 2013; Warner *et al.*, 2014; Engquist and Froese, 2013]. These can be grouped into the following two major classes.

Class 1. Functionals that give more weight to the kinematic than to the amplitude error between the seismic arrivals. They often involve the cross-correlation between the observed and modelled arrivals [Luo and Schuster, 1991; van Leeuwen and Mulder, 2010; van Leeuwen, 2010]. After cross-correlation, the arrival-time error can be picked by hand, which is tedious and to be avoided if possible. Fully automatic methods involve proper normalization of the arrivals prior to cross-correlation in order to ensure that the energy at non-zero time lags quantifies the arrival-time error. These functionals, however, suffer from cross-talk between multiple arrivals. Hence, the data are assumed to have only single arrivals or strong first arrivals. Such functionals are primarily used in cross-well data and tomographic inversion. Luo *et al.* [2011] suggested replacing the cross-correlation with a deconvolution to obtain better convergence.

Class 2. Functionals that aim to fit the data after transforming them into a simpler form. By this transformation, the strong non-linear dependence of the functional with respect to the medium parameters can be avoided [Schuster, 2015]. The functionals using simplified data have a basin of attraction with a larger size than that of the least-squares functional. The simplified data are easier to fit as they have *arti-*

ficial ultra-low frequencies that can circumvent the cycle-skipping problem during the optimization. Creation of these artificial low frequencies by a non-linear operation bears a similarity to other methods such as using deconvolution prior to data fitting [Fei *et al.*, 2012], travel-time picking, and a differential-semblance variant that recovers missing low frequencies [Sun and Symes, 2012]. It should be noted that these simplifying operators act in a non-linear way on the data, unlike the usual scale separation in the multi-scale inversion approach [Bunks *et al.*, 1995], which is linear. This means that they do not solely rely on the low-frequency content of the data. An example is the functional that measures the misfit between the *envelopes* of the observed and the modelled waveforms [Bozdağ *et al.*, 2011; Wu *et al.*, 2014; Chi *et al.*, 2014; Luo and Wu, 2015]. The envelope operator is non-linear and the resulting data have ultra-low frequencies. Another example is the normalized integration method [Chauris *et al.*, 2012; Donno *et al.*, 2013] that uses the normalized time integral of the squared data for least-squares minimization. The normalized integration method, however, might suffer from noise. An advantage of using these functionals, compared to the correlation-type functionals, is that they take multiple arrivals in the data into account.

3.1.2. Problems Using the Envelope-based Misfit

In this paper, we formulate another data-domain objective function that uses the data in a reduced or simplified form. We call it the *bump* functional, after the mathematical definition of a bump function. A bump function or mollifier is a smooth function in the sense of having continuous derivatives of all orders. Intuitively, given a function which is rather irregular or rough, convolution with a mollifier will *mollify* the function, that is, its sharp features are smoothed. It belongs to the second class of functionals mentioned above. The bump functional bears a strong similarity to the envelope-based misfit functional, without suffering from the following issues.

Issue 1. The envelope-based misfit may perform worse than a correlation-based functional, even in the case of a single arrival, because global convergence with a gradient-based method can only be obtained if the modelled and the observed envelopes partially overlap. In other words, this functional cannot reconstruct velocity anomalies that separate the modelled and the observed arrivals in time by roughly more than the dominant period. This is due to the fact that the envelope-based misfit is not sensitive to arrival-time errors that exceed the overall width of the observed and modelled envelopes.

Issue 2. One of the key challenges in waveform inversion is the reconstruction of the smooth background velocity from *reflected* arrivals recorded later in time and at shorter offsets, compared to *transmitted* early arrivals. The envelope-based misfit cannot reconstruct the background velocity of the model using only reflection data. Wu *et al.* [2014] and Luo and Wu [2015] notice that the envelope-based inversion results are much rougher when fitting the reflected arrivals in the data. During inversion, in order to fit only the stronger transmitted arrivals in the records, they use squared envelopes instead of just the envelopes. In this paper, we will analyse the cause of this issue in more detail.

3.1.3. Why the Bump Functional?

The bump functional can be seen as a generalized envelope-based misfit. We show that it not only can be made insensitive to cycle skipping but also has an improved global-convergence robustness compared to the envelope-based misfit. The bump functional is *sensitive* to arrival-time errors in the modelled data that are larger than a period. Furthermore, as the evaluation of the functional does not involve cross-correlation, it is also applicable to data containing multiple arrivals. The price paid is severe non-uniqueness while estimating the solution. In other words, the solution of the bump functional will depend on the initial guess and the chosen optimization method.

3.1.4. Importance of Multi-objective Strategy

In this paper, we discuss different properties of the bump functional. We notice that the bump-functional inversion suffers from the same second issue as the envelope-based misfit. In order to partially overcome this difficulty, we propose a multi-objective inversion strategy using both the bump functional and the least-squares functional. We illustrate the effectiveness of this inversion strategy by a number of examples.

The rest of the paper is organized such that first we describe the conventional least-squares optimization. After that, we formulate the bump functional and demonstrate its characteristics using some numerical experiments. Finally, we describe the proposed multi-objective inversion strategy and show its effectiveness using realistic numerical examples. The last section summarizes the paper.

3.2. Review of the Least-squares Inversion

The classic least-squares inversion aims to find a model in $\mathbb{M}$, a set of possible models of the subsurface, which minimizes the functional [Tarantola, 1984],

$$J_{ls} = \tfrac{1}{2} \sum_s \sum_r \sum_t \left[p(\mathbf{x}_r, \mathbf{x}_s, t) - q(\mathbf{x}_r, \mathbf{x}_s, t) \right]^2 . \tag{3.1}$$

Here, $q(\mathbf{x}_r, \mathbf{x}_s, t)$ is the observed pressure for a source at position $\mathbf{x}_s$ and receiver at $\mathbf{x}_r$ as a function of time t. The modelled data are denoted by $p(\mathbf{x}_r, \mathbf{x}_s, t)$. Both q and p belong to the data space denoted by $\mathbb{D}$. The operator $\mathcal{F} : \mathbb{M} \rightarrow \mathbb{D}$ requires the solution of a wave equation to provide data in the data space $\mathbb{D}$ for a given model in $\mathbb{M}$. For the examples in this paper, we will consider the 2-D acoustic wave equation

$$\frac{1}{\rho c^2} \frac{\partial^2 p}{\partial t} - \nabla \cdot \frac{1}{\rho} \nabla p = \phi(t) \delta(\mathbf{x} - \mathbf{x}_s), \tag{3.2}$$

with pressure $p(\mathbf{x}, \mathbf{x}_s, t)$, sound speed $c(\mathbf{x})$, mass density $\rho(\mathbf{x})$ and source signature $\phi(t)$. In 2D, $\mathbf{x} = (x, z)$. We use a time-domain staggered-grid finite-difference code to model the pressure wavefield required to perform the forward and the adjoint wavefield computations [Tarantola, 1984; Fichtner, 2010]. Absorbing boundary conditions are used on all sides of the computational domain except for a realistic example, where reflections from the boundary at $z = 0$ are included. We assume that the source signature, $\phi(t)$, is known in our examples. In all our examples, we perform full-bandwidth

inversion as opposed to the multi-scale inversion approach [Bunks *et al.*, 1995] that relies on the low-frequency information in the data. The optimizations are performed by the conjugate-gradient method. The gradient at each iteration is preconditioned by the type 1 migration weights of Plessix and Mulder(2004).

3.3. The Bump Functional

3.3.1. Definition

To compute the bump functional, the modelled and the observed data should be mapped into the space $\mathbb{D}_b$ of *bumpy* data. The observed and modelled data in bumpy-data space are denoted by p_b and q_b, respectively. A function, $\mathcal{F}_b : \mathbb{D} \rightarrow \mathbb{D}_b$, maps the modelled data into the bumpy-data space and is given by

$$\begin{aligned} p_b(\mathbf{x}_r, \mathbf{x}_s, t) &= \mathcal{F}_b[\sigma_t, \sigma_r, p, \epsilon](\mathbf{x}_r, \mathbf{x}_s, t) \\ &= b(\mathbf{x}_r, \sigma_r) *_r \left[b(t, \sigma_t) *_t \sqrt{p^2(\mathbf{x}_r, \mathbf{x}_s, t) + \epsilon^2} \right]. \end{aligned} \tag{3.3}$$

The non-zero constant ϵ makes p_b differentiable at $p = 0$. We chose ϵ much smaller than p, such that $\sqrt{p^2(\mathbf{x}_r, \mathbf{x}_s, t) + \epsilon^2} \simeq |p(\mathbf{x}_r, \mathbf{x}_s, t)|$. In the above equation, b is a blurring function for which we choose a Gaussian with standard deviation equal to either σ_t when blurring in time or σ_r when blurring in one of the receiver coordinates. Convolution in time is denoted by $*_t$, and along the receiver position, $\mathbf{x}_r$, by $*_r$.

The non-linear mapping $\mathcal{F}_b$ replaces the arrivals in the data with *bumps* and is non-injective due to the following reasons:

1. $p_b \geq 0$, irrespective of the polarity of arrivals in p,
2. depending on the amount of blurring, p_b is less dependent on the source signature and
3. the blurring operation removes the details from p, depending on the blurring parameters, σ_t and σ_r.

The blurring parameters control the amount of blurring applied to the absolute-valued data and determine the width of the resulting bumps. For example, choosing $\sigma_t = \tau_d$ and $\sigma_r = 0$, where $\tau_d = 1/f_d$ is the period of the data corresponding to the dominant frequency f_d, would blur an impulsive arrival in the absolute-valued data such that it is spread roughly over $2\tau_d$ in time. Similarly, choosing $\sigma_t = 0$ and $\sigma_r = \lambda_d$, where $\lambda_d = c_r/f_d$ is the wavelength corresponding to the velocity, c_r, close to the receivers, would blur an impulsive arrival in the absolute-valued data such that it is spread roughly over receivers within a distance of $2\lambda_d$.

The bump functional, J_b, is the least-squares difference between the observed and the modelled bumpy data:

$$J_b = \tfrac{1}{2} \sum_s \sum_r \sum_t \left[p_b(\mathbf{x}_r, \mathbf{x}_s, t) - q_b(\mathbf{x}_r, \mathbf{x}_s, t) \right]^2. \tag{3.4}$$

Inversion based on J_b produces a model in $\mathbb{M}$ that minimizes the difference between p_b and q_b. Details on the gradient computation of the bump functional with respect to the

medium parameters are given in Appendix A. Note that minimizing their difference does not necessarily reduce the difference between p and q, since the bump functional is insensitive to the polarity of the events in the data. Also note that the bump functional with the parameters $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$ closely resemble the envelope-based misfit [Bozdağ *et al.*, 2011].

3.3.2. Adjoint Source Functions for the Bump Functional

The gradient of the bump functional with respect to the medium parameters, c and ρ, is computed by correlating the forward-propagated source wavefield with the adjoint wavefield at each point in the subsurface. This is similar to the gradient computation for the conventional least-squares inversion, where the adjoint wavefield is generated by injecting the adjoint source functions from the receiver positions [Tarantola, 1984; Plessix, 2006; Fichtner, 2010]. The adjoint source functions for the least-squares functional are the difference between the modelled and the observed data. In order to compute the adjoint source functions for the bump functional, the following steps are performed in order:

1. the difference between the modelled and the observed bumpy data is calculated,
2. time and/or receiver blurring is applied to the difference, based on σ_t and σ_r, and
3. the result is multiplied with the *stabilized* sign of the modelled data in the time domain.

The adjoint source functions for the bump functional are given by $\partial_p J_b$, where ∂_p denotes the derivative with respect to the modelled data p. Using the chain rule, we write:

$$\partial_p J_b \quad = \quad \partial_{p_a} J_b \, \partial_p p_a = \partial_{p_a} J_b \, \underbrace{p \, [p^2 + \epsilon^2]^{-\frac{1}{2}}}_{\mathrm{sgn}(p), \text{ when } \epsilon \text{ is } 0}, \tag{3.5}$$

where ∂_{p_a} denotes the derivative with respect to the absolute-valued-modelled data p_a and sgn is the sign function. In order to derive an expression for $\partial_{p_a} J_b$, we consider the bump functional in frequency-wavenumber domain. In such a domain, time and/or receiver blurring corresponds to a simple multiplication with the blurring function $\tilde{b}$. Without loss of generality, we consider the case of a regular horizontal array of receivers that record waves originating from a given source location. Using Parseval's theorem, we rewrite the expression of the bump functional in Equation 3.4 by Fourier transforming the bumpy data from time t and horizontal receiver coordinate x_r to frequency f and horizontal wavenumber k_r as:

$$\begin{aligned} J_b \quad &= \quad \tfrac{1}{2} \sum_r \sum_t \left[p_b(x_r, t) - q_b(x_r, t) \right]^2 , \\ &= \quad \tfrac{1}{8\pi^2} \sum_r \sum_f |\tilde{p}_b(k_r, f) - \tilde{q}_b(k_r, f)|^2 . \end{aligned} \tag{3.6}$$

Where we used the fact that p_b and q_b are real valued. Here, $\tilde{p}_b$ and $\tilde{q}_b$ denote the modelled and the observed bumpy data in the f-k_r domain, respectively. In the f-k_r

domain, the bumpy data are obtained by multiplying the absolute-valued data, $\tilde{p}_a$ and $\tilde{q}_a$, with the blurring function $\tilde{b}$, so we have:

$$J_b = \frac{1}{8\pi^2}\sum_r\sum_f |\tilde{b}(f,\sigma_t)\tilde{b}(k_r,\sigma_r)[\tilde{p}_a(k_r,f)-\tilde{q}_a(k_r,f)]|^2. \tag{3.7}$$

We rewrite the above equation using the real and imaginary parts of the absolute-valued data as:

$$\begin{aligned} J_b &= \frac{1}{8\pi^2}\sum_r\sum_f \tilde{b}^2(f,\sigma_t)\tilde{b}^2(k_r,\sigma_r) \\ &\quad \{[\Re(\tilde{p}_a)-\Re(\tilde{q}_a)]^2+[\Im(\tilde{p}_a)-\Im(\tilde{q}_a)]^2\}. \end{aligned} \tag{3.8}$$

Where we used the fact that the blurring function $\tilde{b}$ is real valued. We now differentiate the above equation with respect to the real and imaginary parts of the modelled-absolute-valued data $\tilde{p}_a$ to obtain:

$$\begin{aligned} \partial_{\tilde{p}_a} J_b &= \partial_{\Re(\tilde{p}_a)} J_b + \imath\partial_{\Im(\tilde{p}_a)} J_b \\ &= \frac{1}{4\pi^2}\tilde{b}^2(f,\sigma_t)\tilde{b}^2(k_r,\sigma_r)[\tilde{p}_a(k_r,f)-\tilde{q}_a(k_r,f)], \\ &= \frac{1}{4\pi^2}\tilde{b}(f,\sigma_t)\tilde{b}(k_r,\sigma_r)[\tilde{p}_b(k_r,f)-\tilde{q}_b(k_r,f)]. \end{aligned} \tag{3.9}$$

Using the chain rule, we can write the derivative of the bump functional with respect to modelled-absolute-valued data in the time domain as:

$$\begin{aligned} \partial_{p_a} J_b &= \sum_r\sum_f \partial_{\tilde{p}_a} J_b \partial_{p_a}\tilde{p}_a, \\ &= b(x_r,\sigma_r) *_r \{b(t,\sigma_t) *_t [p_b(x_r,t)-q_b(x_r,t)]\}. \end{aligned} \tag{3.10}$$

After substituting $\partial_{p_a} J_b$ in the Equation 3.5, we obtain the final expression to compute the adjoint source functions for the bump functional in the time domain:

$$\partial_p J_b = \{b(x_r,\sigma_r) *_r \{b(t,\sigma_t) *_t [p_b - q_b]\}\}\; p\,[p^2+\epsilon^2]^{-\frac{1}{2}}. \tag{3.11}$$

3.3.3. Characteristics

The main characteristics of inversion with the bump functional are listed next.

Insensitivity to Cycle Skipping

The bump-functional inversion is less sensitive to cycle skipping even in the absence of low frequencies in the data. This is due to the fact that it fits the bumpy data, as in Equation 3.4, which are non-oscillatory if sufficient blurring is applied.

Global-convergence Robustness

While computing the bump functional, the absolute-valued data are blurred in time and/or along a receiver coordinate, as in Equation 3.3. The role of the blurring is to

increase the size of the basin of attraction of the objective function compared to that of the envelope-based misfit. The global-convergence robustness of J_b inversion, when using gradient-based optimization, increases with the amount of blurring applied to the absolute-valued data. Intuitively, after taking the absolute value, a single pulse in the modelled data will still not be able to sense a corresponding pulse in the observed data if they are separated in time. Large amounts of blurring increases the width of the observed and modelled bumps, so that they can sense each other in time during the inversion. In other words, blurring is essential to make the bump functional sensitive to large arrival-time errors. Since the bumpy data become simpler with increased blurring, it is easier to find models that fit them.

Non-uniqueness

The price paid for blurring is a loss of resolution and an increase of non-uniqueness. Since the mapping $\mathcal{F}_b : \mathbb{D} \to \mathbb{D}_b$ is non-injective, the inverse problem of estimating a model that minimizes the bump functional suffers from non-uniqueness, which gets worse with the amount of blurring applied in the Equation 3.3. Note that the least-squares inversion also suffers from non-uniqueness due to band-limitation of the data and lack of illumination, making regularization necessary to obtain an acceptable solution. The bump functional inversion suffers from non-uniqueness even when the same regularization as for the least-squares inversion is used.

Once the bump functional is minimized, the observed and modelled bumpy data match each other. However, this does not imply that the modelled data fit the observed data very well. The modelled arrivals can have the opposite polarity as well as a different wavelet, compared to the observed arrivals. Furthermore, the modelled data can contain ancillary *false* arrivals that are not present in the observed data. This prevents the bump-functional inversion from correctly updating the background velocity while fitting the reflected arrivals. The bump functional inversion might result in a model with many high-wavenumber artefacts that generate the false arrivals. Next, we will discuss some simple examples to illustrate these characteristics.

3.3.4. Illustrative Examples Using the Bump Functional

Two Arrivals

As a very simple example, we assume that both the observed and the modelled data to contain a single record with two arrivals. For the first arrival, $w_1(t)$, we use a Ricker wavelet with a peak frequency of 20 Hz and a band-limitation of 10–15–50–60 Hz. We used the same Ricker wavelet with a band-limitation of 15–17–23–25 Hz for the second arrival, $w_2(t)$. The arrivals do not contain any low-frequency information. The modelled and 'observed' data are then chosen as

$$p(t) = w_1(t + \beta - \beta_0) + 0.6\, w_2(t + \gamma - \gamma_0),$$

$$q(t) = w_1(t - \beta_0) + 0.6\, w_2(t - \gamma_0).$$

To the observed data, we added 20% of white random noise. The model parameters were chosen as $\beta_0 = 2.16$ s and $\gamma_0 = 2.55$ s. Figure 3.1 plots the observed and modelled

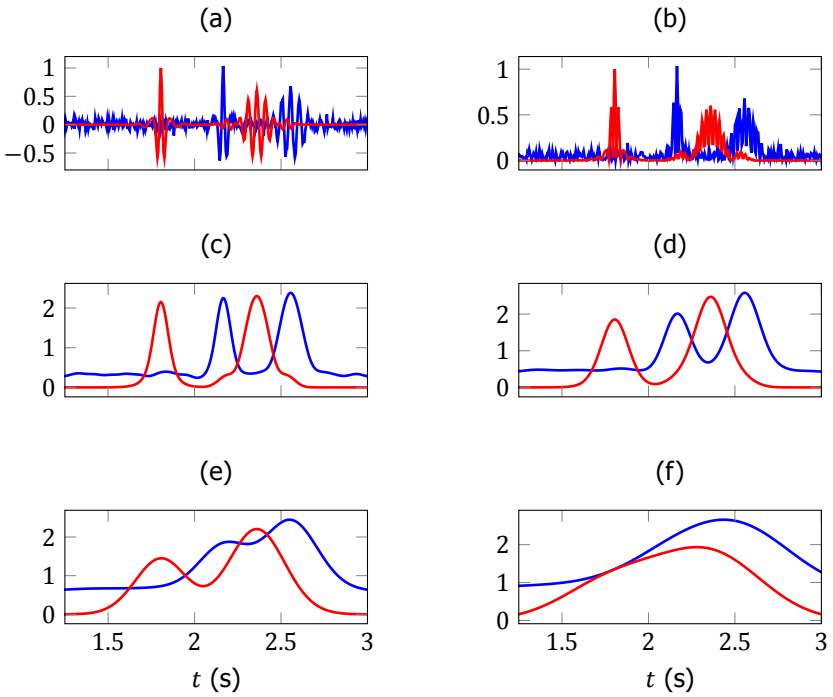


Figure 3.1: The observed and modelled data with two arrivals are plotted in blue and red, respectively. (a) The arrival-time error in the modelled arrivals is more than the dominant period, τ_d. (b) Absolute-valued data. (c) Bumpy data obtained using $\sigma_t/\tau_d = 0.5$, (d) $\sigma_t/\tau_d = 1$, (e) $\sigma_t/\tau_d = 2$ and (f) $\sigma_t/\tau_d = 4$.

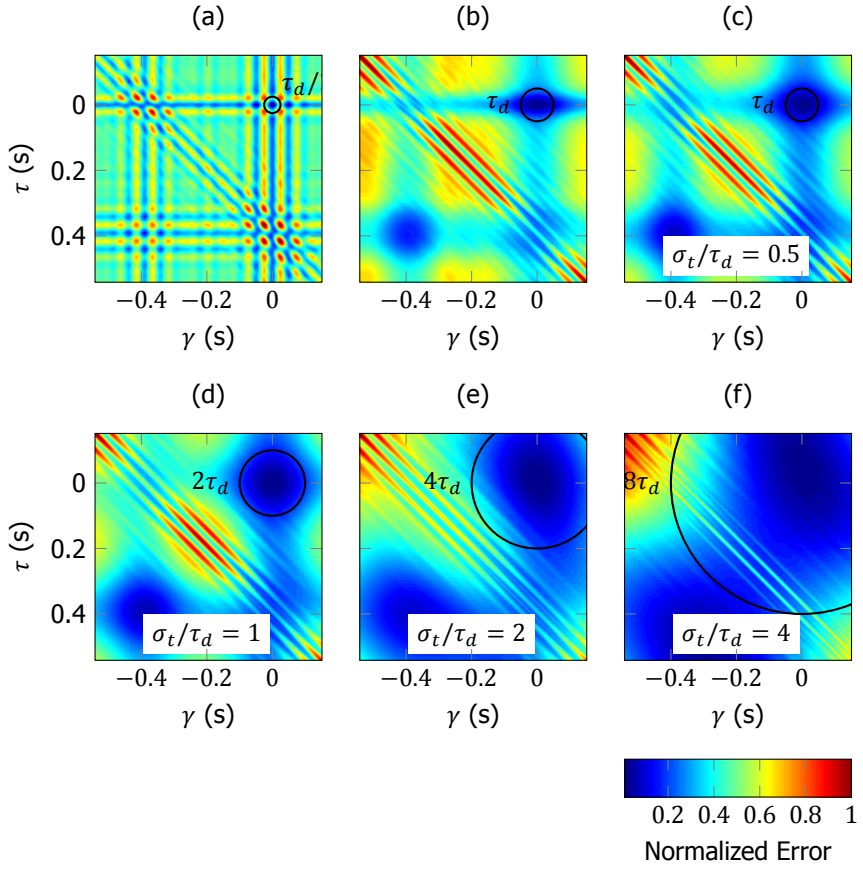


Figure 3.2: Various objective functions are plotted as a function of the arrival-time errors, γ and β, for the observed and modelled data shown in Figure 3.1. The basin of attraction is outlined by a black circle. (a) The least-squares functional, J_{ls}. (b) The envelope-based misfit. (c) The bump functional, J_b, with $\sigma_t/\tau_d = 0.5$, (d) $\sigma_t/\tau_d = 1$, (e) $\sigma_t/\tau_d = 2$, and (f) with $\sigma_t/\tau_d = 4$.

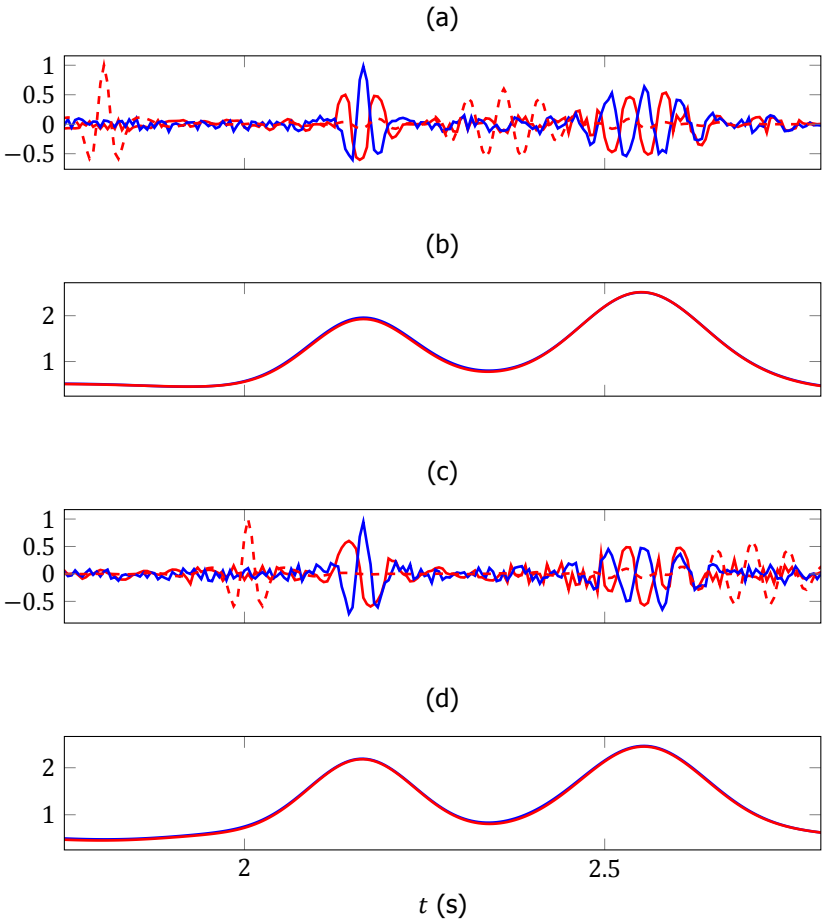


Figure 3.3: Plots corresponding to the two-arrival example as in Figure 3.1. Two inversions (I & II) are performed, starting from different initial values of γ and β using the bump functional with $\sigma_t/\tau_d = 1$. The observed and estimated data are plotted in blue and red respectively. The initially estimated data are drawn with a dashed red line. (a) Data corresponding to inversion I. (b) Bumpy data corresponding to inversion I. (c) Same as (a), but for inversion II. (d) Same as (b), but for inversion II.

data in blue and red, respectively. The inversion parameters are the arrival-time errors, β and γ. Both should be zero at the global minimum. We now consider the inverse problem of estimating the arrival times using different objective functionals.

Least-squares Functional

Figure 3.2a shows the classic least-squares functional, $J_{ls} = \int [p(t) - q(t)]^2 \mathrm{d}t$, as a function of the arrival-time errors. The oscillatory nature of the arrivals causes various local minima to occur around the global minimum, shrinking the radius of the basin of attraction. In this case, the radius of the basin is approximately $\tau_d/2 = 0.025\,\mathrm{s}$, where, again, τ_d denotes the period of the dominant frequency. In the multi-scale inversion approach [Bunks *et al.*, 1995], a low-pass filter is applied to the data to first fit the lowest reliable frequencies. The radius of the basin is then half the period corresponding to the lowest frequency, $\tau_{max} = 0.1\,\mathrm{s}$. This implies that, if the arrival-time error in the initially modelled data exceeds $0.1\,\mathrm{s}$, the least-squares inversion cannot accurately estimate the arrival times.

Envelope-based Misfit

The envelope-based misfit [Bozdağ *et al.*, 2011] computes the least-squares error using the observed and the modelled envelopes. In this example, Figure 3.2b shows that the misfit suffers less from cycle skipping and local minima do not occur close to the global minimum because the envelopes are non-oscillatory. In this case, the radius of attraction roughly equals the dominant period, $\tau_d = 0.05\,\mathrm{s}$. It can be seen that the functional is insensitive to arrival-time shifts larger than τ_d, when there is no overlap between the observed and modelled data envelopes.

Bump Functional

To compute the bump functional, the absolute-valued data are considered, as plotted in Figure 3.1b. The choice of σ_t determines the amount of blurring applied to the absolute-valued data. For $\sigma_t/\tau_d = 0.5$, the observed and modelled bumps, as plotted in Figures 3.1c, are similar to envelopes. In this case, the bump functional has characteristics similar to the envelope-based misfit, as plotted in Figure 3.2c. In general, for any given σ_t, since the observed and the modelled bumps are non-oscillatory, the bump functional suffers less from cycle skipping.

In order to corroborate our claim that the global-convergence robustness of J_b depends on the amount of blurring, we now examine the bump functional for different values of σ_t. The bumpy data for different values of σ_t are plotted in Figures 3.1c–f. As can be seen in Figures 3.2d–f, the radius of the attraction circle of the bump functional increases with σ_t. With increased blurring, the separated bumps of the observed and modelled data can sense each other in time, resulting in an improved global-convergence robustness. Note that the radius of the basin of attraction for J_b, unlike the least-squares functional, is independent of the presence of low frequencies in the data. The radius of basin of attraction in the case of $\sigma_t/\tau_d = 4$ and 8 is larger than period corresponding to the minimum frequency in the data, τ_{max}, which shows that J_b can also be effective in the absence of low frequencies.

Finally, we show that the data minimizing the bump functional cannot be determined uniquely. We perform two inversions (I & II) using the conjugate-gradient method starting from different initial values of β and γ. During each inversion, we estimate the data that minimize the bump functional for $\sigma_t/\tau_d = 1$. We expect that the estimated data output from both the inversions to be different while fitting each other in their bumpy form. This is possible because the non-linear mapping $\mathcal{F}_b$ is non-injective. The observed and initially estimated data for both the inversions are plotted in Figures 3.3a and 3.3c, respectively. The estimated data after the inversions, plotted in Figures 3.3a and 3.3c, are different and they contain false arrivals and have polarity mismatches. For both the inversions, the estimated bumpy data, plotted in Figures 3.3b and 3.3d, at the last iteration do fit the observed bumpy data.

Cross-well Transmission Example

We now consider a cross-well experiment with a Gaussian anomaly. This is a transmission problem for which the background velocity needs to be estimated. We used evenly spaced vertical arrays of sources and receivers at $x = 1$ m and $x = 96$ m, respectively. The source wavelet had a peak frequency of 140 Hz. The assumed Earth model and the corresponding 'observed' data are plotted in Figures 3.4a and 3.5a, respectively. The initial velocity model for inversion is homogeneous with $c = 1800$ m/s. The corresponding modelled data are plotted in Figure 3.5b and have arrival-time errors of more than three times the dominant period, $T = 0.0071$ s. Therefore, larger amounts of blurring are necessary during the J_b inversion to ensure that the bumps in the initially modelled data can sense the bumps in the observed data. We used the J_b functional with $\sigma_t/\tau_d = 0.5$, 1, 2 and 4. Gaussian smoothing is applied to the gradient at each iteration with a standard deviation of 1.5 m in order to avoid high-wavenumber artefacts in the solution. In all the cases, the iterations were stopped when the convergence slowed down too much.

Figures 3.4b–e show the velocity models and Figures 3.5c–f show the modelled arrivals at the last iteration. For $\sigma_t/\tau_d = 0.5$ and 1, the inversion was not able to reconstruct the velocity anomaly, as in Figures 3.4b–c, since the J_b functional has a small basin of attraction. At the last iteration, the observed and the modelled data in bumpy form do not match each other, as mentioned in Table 3.1. Only the error in records with initial arrival-time error less than $2\tau_d$ is minimized (see Figures 3.5c–d). The J_b inversion with $\sigma_t/\tau_d = 0.5$ is similar to the envelope-based inversion, as shown in the Figures 3.1 and 3.2. In the case of $\sigma_t/\tau_d = 2$ and 4, the bump functional inversion has successfully reconstructed the velocity anomaly, as in Figures 3.4d–e, proving that the size of the basin of attraction for J_b increases with the amount of blurring. As shown in the Table 3.1, the inversion is able to fit the observed bumpy data. Also, the modelled data after inversion, shown in Figures 3.5e–f, reasonably fit the observed data in Figure 3.5a. Hence these solutions are acceptable as summarized in the Table 3.1.

To illustrate the non-uniqueness problem, we choose $\sigma_t/\tau_d = 8$ for J_b inversion. At the last iteration, the inversion is able to fit the observed and the modelled data in bumpy form as mentioned in Table 3.1. However, Figure 3.5g shows that the modelled data do not necessarily fit the observed data. There are many models providing data that can fit the observed data in their bumpy form. Since the non-uniqueness becomes

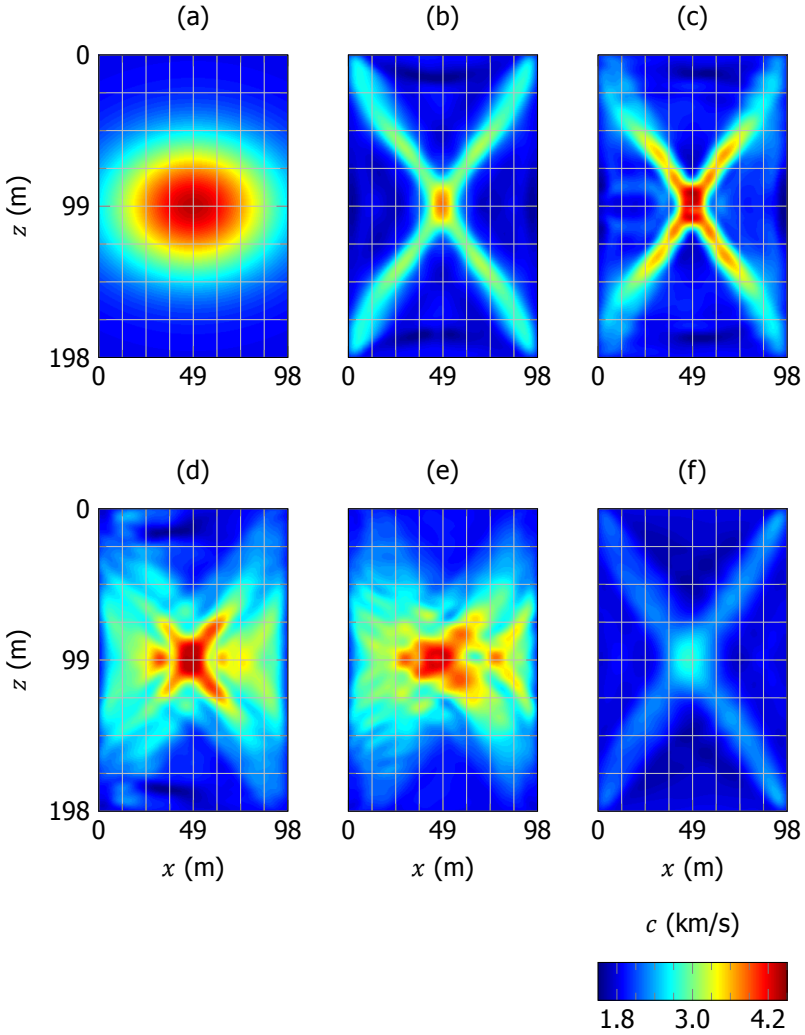


Figure 3.4: For the cross-well experiment with a Gaussian anomaly, J_b inversion is performed with different amounts of time blurring. The initial model before inversion is a homogeneous model with $c = 1800$ m/s. (a) Earth model with a Gaussian anomaly at the center. (b) Result using $\sigma_t/\tau_d = 0.5$, (c) $\sigma_t/\tau_d = 1$, (d) $\sigma_t/\tau_d = 2$, (e) $\sigma_t/\tau_d = 4$ and (f) $\sigma_t/\tau_d = 8$.

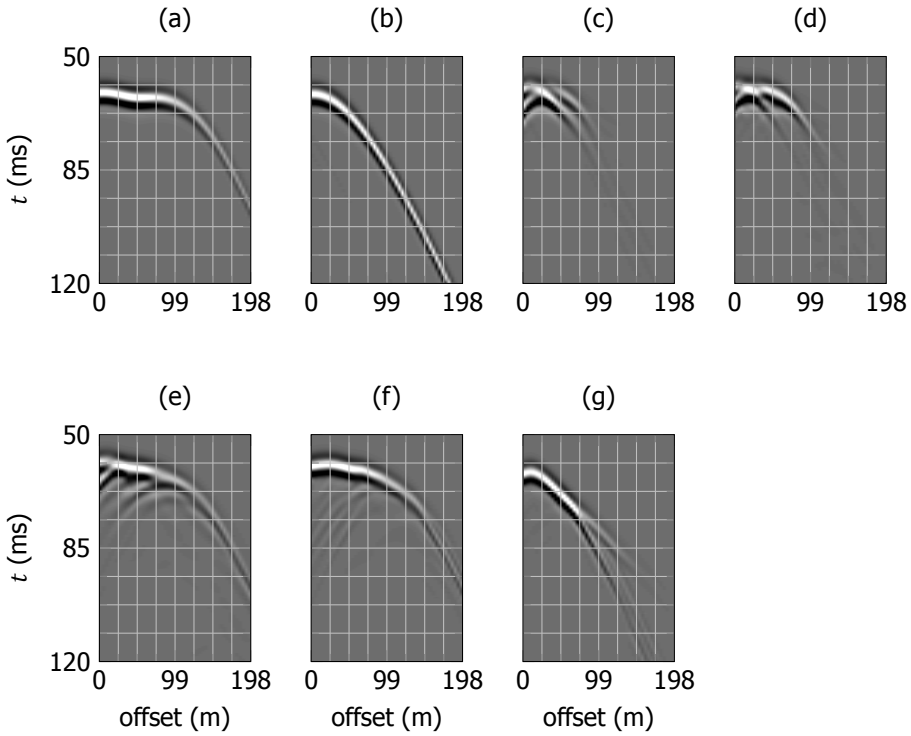


Figure 3.5: Data panels corresponding to the cross-well experiment in Figure 3.4 for a source at $(1, 0)$. J_b inversion is performed using different amounts of time blurring and $\sigma_r = 0$. The shot gathers at the last iteration are plotted with the same gray scale. (a) Observed shot gather. (b) Initially modelled shot gather. (c) Using J_b with $\sigma_t/\tau_d = 0.5$, (d) with $\sigma_t/\tau_d = 1$, (e) with $\sigma_t/\tau_d = 2$, (f) with $\sigma_t/\tau_d = 4$ and (g) with $\sigma_t/\tau_d = 8$.

Table 3.1: Convergence of the bump functional for different amounts of blurring in the case of a cross-well experiment on a Gaussian anomaly.

$\sigma_t\ /\tau_d$	p_b fits q_b	p fits q	comment
0.5	no	no	lack of sensitivity
1	no	no	lack of sensitivity
2	yes	yes	acceptable solution
4	yes	yes	acceptable solution
8	yes	no	non-uniqueness

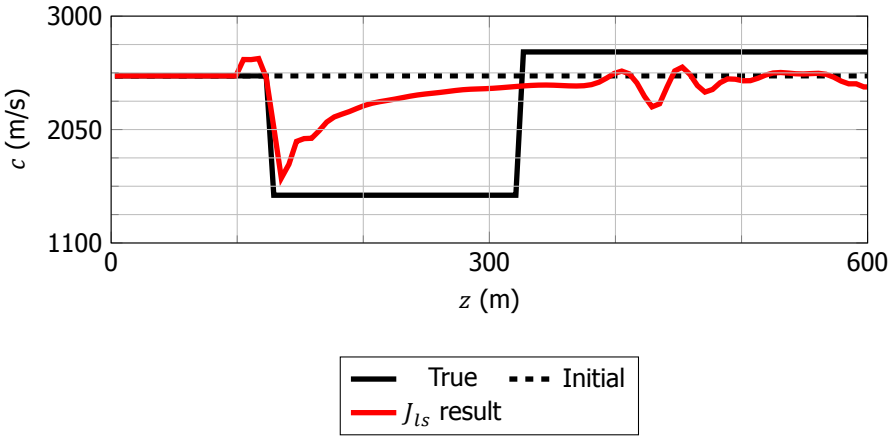


Figure 3.6: Results of single-objective J_{ls} inversion of the three-layer reflection example. The initial model (dashed black) before inversion is also shown.

more severe with larger amounts of blurring, the inversion fails to reach an acceptable solution, as shown in Figure 3.4f. Also, false arrivals appear in the modelled data and they assist in the minimization of the bump functional.

Three-layer Reflection Example

We consider another example in which the 'observed' data contain primarily reflected arrivals. In the case of reflection problems, both the background velocity and the position of the reflectors have to be estimated. The background velocity is determined by the offset-based moveout information in the data. This example illustrates the inability of the bump functional to estimate the background velocity using reflected arrivals. We also show that the velocity model that minimizes the bump functional cannot be uniquely determined.

We consider an assumed Earth model of 1500-m width and 600-m depth with a negative velocity anomaly of approximately 40%. The velocity model, plotted in Figure 3.6, is assumed to only vary with depth and contains two reflectors at $z = 120$ m and $z = 320$ m. We used a horizontal array of 100 evenly spaced receivers at a depth of 30 m and a source at $(0, 20)$ m. We generated the data using a fourth-order minimum-phase Butterworth wavelet of bandwidth 10–30 Hz. Due to the negative velocity anomaly, head waves do not arrive at the receivers before 1.2 s recording time. Figure 3.7a depicts the 'observed' shot gather, where an internal multiple can also be seen. The initial model for inversion is homogeneous with $c = 2500$ m/s. The spatial sampling for the inversion mesh is 6 m.

Figure 3.6 shows the result of the J_{ls} inversion. It fails to reconstruct the background velocity of the model and convergence starts to become very slow after 20 iterations. After updating the reflectivity of the first layer, the J_{ls} inversion updates the reflectivity of the model at depths between 400 and 450 m. This causes the modelled data, plotted

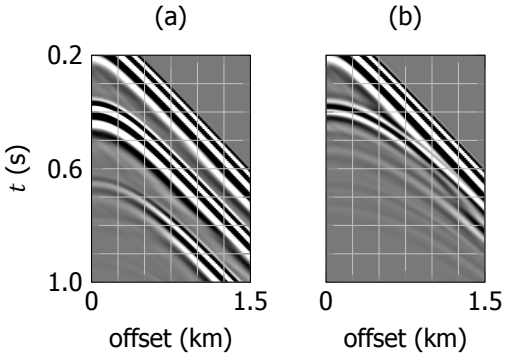


Figure 3.7: The shot gathers corresponding to the three-layer reflection experiment in Figure 3.6. (a) Observed gather (b) Cycle-skipped modelled gather at the last iteration of the J_{ls} inversion.

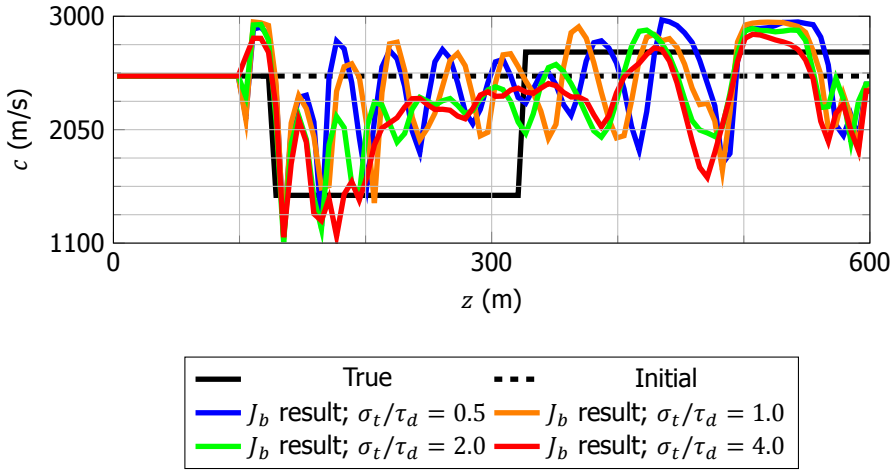


Figure 3.8: Same as Figure 3.6, except that the bump functional is used with different amounts of time blurring and $\sigma_r = 0$.

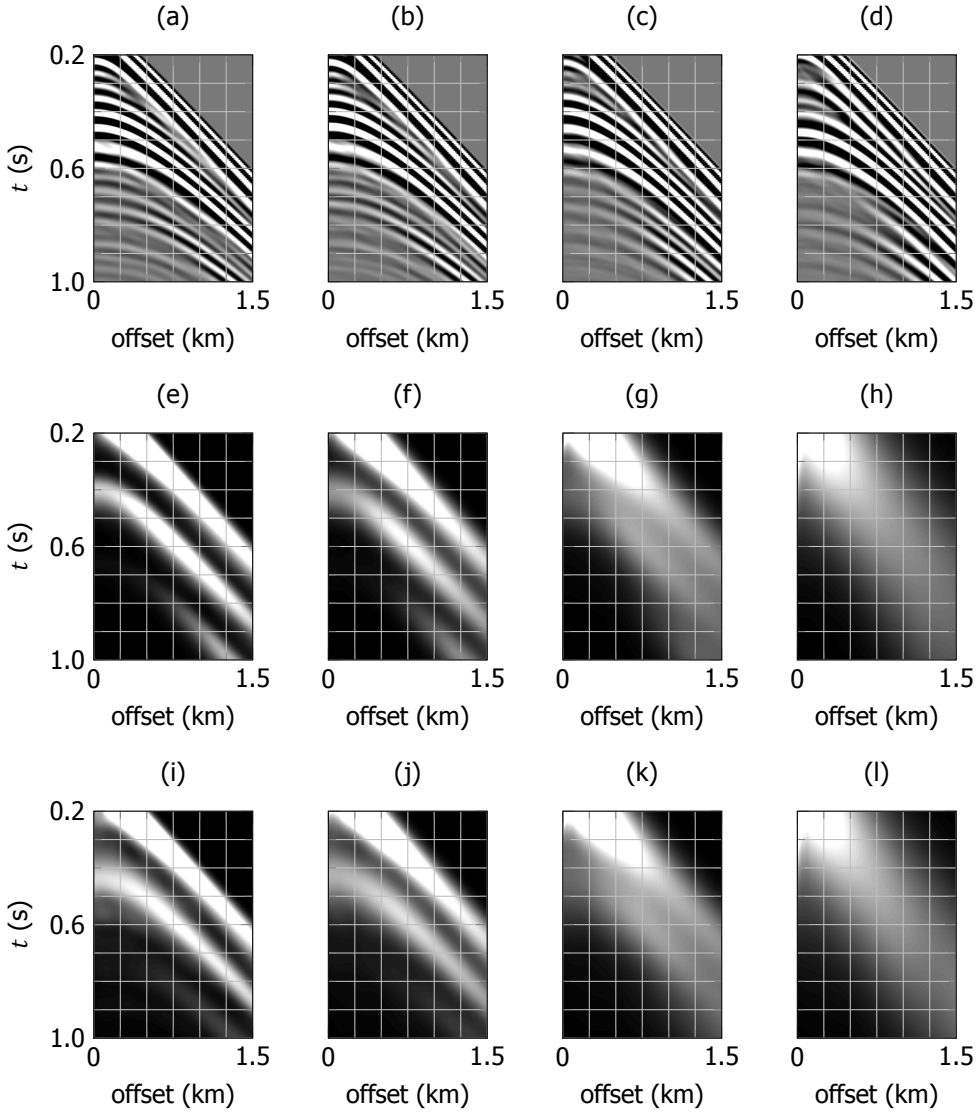


Figure 3.9: Modelled shot gathers after J_b inversion corresponding to the reflection experiment in Figure 3.8. No blurring is applied along the receiver array ($\sigma_r = 0$) in all cases. (a) Data using $\sigma_t/\tau_d = 0.5$, (b) $\sigma_t/\tau_d = 1$, (c) $\sigma_t/\tau_d = 2$ and (d) $\sigma_t/\tau_d = 4$. (e) Observed bumpy data using $\sigma_t/\tau_d = 0.5$, (f) $\sigma_t/\tau_d = 1$, (g) $\sigma_t/\tau_d = 2$ and (h) $\sigma_t/\tau_d = 4$. (i) Modelled bumpy data after J_b inversion using $\sigma_t/\tau_d = 0.5$, (j) $\sigma_t/\tau_d = 1$, (k) $\sigma_t/\tau_d = 2$ and (l) $\sigma_t/\tau_d = 4$.

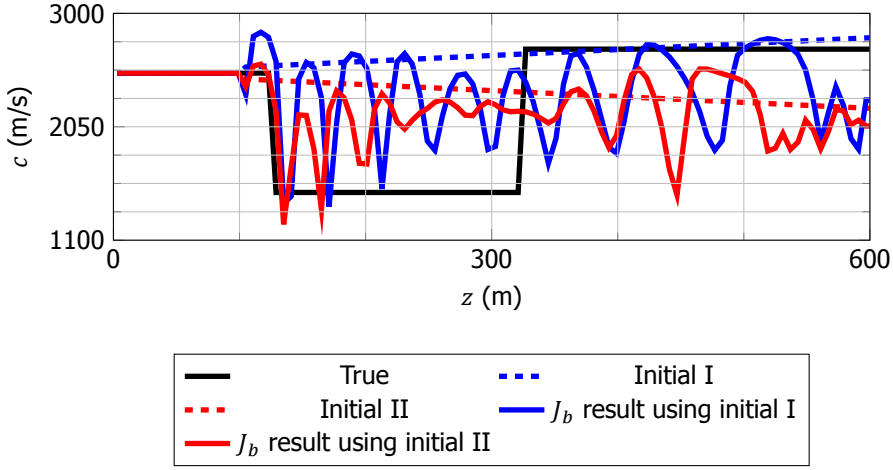


Figure 3.10: Same as Figure 3.6, except that the bump functional is used with $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$. Two different initial models were used for inversion.

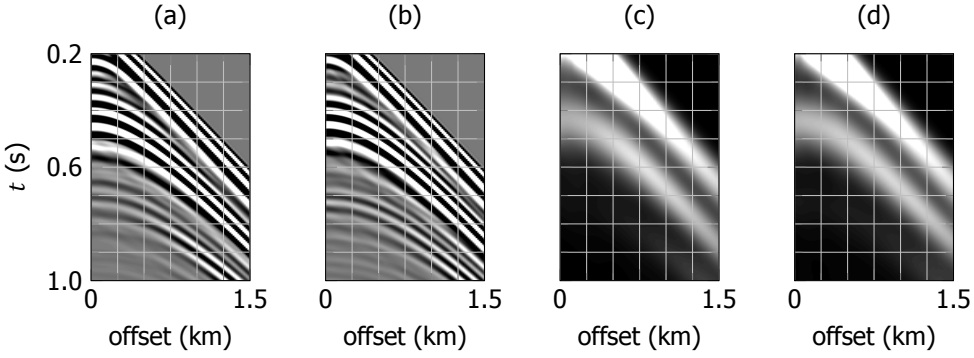


Figure 3.11: Single-objective bump functional inversion using $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$ starting from two different initial models (I & II), as plotted in Figure 3.10. (a) Modelled data at the last iteration using initial model I and (b) initial model II. (c) Modelled bumpy data at last iteration using initial model I and (d) initial model II.

in Figure 3.7b, to partially match the 'observed' data, primarily at the short offsets. We also see that these reflectors, positioned at the wrong depths, generate arrivals in the data that are cycle-skipped at larger offsets. This behaviour of least-squares inversion is well-known. The cycle skipping indicates that the solution is caught in a local minimum. It has to be noted that the results might differ depending on the coarseness of the inversion mesh [Ma *et al.*, 2012]. We did neither apply smoothing to the gradient at each iteration nor use any additional smoothness constraints on the model. In the case of this simple example, we noticed that the least-squares inversion was able to reach the global minimum when the spatial sampling for the inversion mesh was 10 m.

We have seen that the bump functional does not suffer from cycle skipping if enough blurring is applied. In order to see if it can update the background velocity, we performed the bump-functional inversion with $\sigma_t/\tau_d = 0.5,\ 1,\ 2$ and 4. No blurring along the receiver coordinate is applied since this is an illustrative example. The observed bumpy data for these cases are plotted in Figures 3.9e–h. The modelled bumpy data at the final iteration are displayed in Figures 3.9i–l. We notice that the inversion has matched the observed and the modelled data in their bumpy form at the last iteration. However, the resulting velocity models have an incorrect background velocity, as Figure 3.8 shows. The J_b inversion has not updated the background velocity but only boosted the reflectivity of the model at depths between 400 and 550 m. In all cases, the modelled data contain a lot of *false* arrivals and do not fit the observed data, as can be seen in Figures 3.9a–d. These false arrivals help the J_b inversion in matching the bumpy form of the observed and modelled data. The false arrivals are caused by several reflectors at incorrect depths. Even in the presence of moveout information in the observed data, we see that the bump functional cannot update the background velocity of the model.

We now perform the bump functional inversion with $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$ starting from two different initial models (I & II), plotted in Figure 3.10. The output models in both the cases are different in a non-trivial way since the data corresponding to them, plotted in Figures 3.11a and 3.11b, do not match each other. However, the bumpy data corresponding to the output models, plotted in Figures 3.11c and 3.11d, are similar. This shows that the velocity model that explains the observed-bumpy reflections is not unique. Due to this, the bump functional inversion fails to output a model with the correct background velocity, even though it suffers less from cycle-skipping.

3.4. Multi-objective Inversion

As mentioned before, least-squares inversion suffers from cycle skipping and cannot recover velocity errors that cause arrival-time errors larger than $\tau_d/2$. Let $\mathbb{M}_{ls}$ denote a set that includes all the models corresponding to the local minima of the least-squares objective function, excluding the global minimum. We assume that the bump functional does not suffer from cycle skipping, for particular values of σ_t and σ_r, and let $\mathbb{M}_b$ denote a set that contains all the non-unique solutions that minimize it. We state without a proof that the sets $\mathbb{M}_{ls}$ and $\mathbb{M}_b$ are disjoint due to the following intuitive reasons:

- A model that belongs to $\mathbb{M}_{ls}$ generates data that are cycle-skipped compared to

the observed data, but this model does not belong to $\mathbb{M}_b$ since the bump functional does not suffer from cycle skipping for sufficiently large blurring.

- A model that belongs to $\mathbb{M}_b$ generates data that contain false arrivals, but this model does not belong to $\mathbb{M}_{ls}$ because false arrivals increase the least-squares functional.

This motivates the use of both the least-squares and the bump functional in the inversion. In the inversion strategy, the primary objective is to minimize the least-squares difference. It is also used to constrain the model space $\mathbb{M}_b$. Minimization of the bump functional is an auxiliary objective needed to move away from models that belong to $\mathbb{M}_{ls}$.

3.4.1. Strategy

The multi-objective inversion scheme we use throughout this paper is given in Figure 3.12. The inversion consists of several round-trips depending on how far the starting model is from the global minimum. Within each round-trip, we optimize both the J_b and J_{ls} objectives separately. During each individual optimization, we update the subsurface models and the iterations are stopped whenever convergence becomes too slow. The multi-objective inversion stops when the change in the output between consecutive round-trips is negligible. This implies that the inversion has converged to a model that simultaneously minimizes the least-squares and the bump functional.

We start with the bump-functional inversion, where strong blurring is chosen both in time and along the receiver coordinate. Then, we gradually reduce the blurring and, at the same time, perform more iterations. The motivation for starting with strong blurring is that it allows the observed and modelled bumps, if they are far apart, to interact with each other so that the time gap between them can be closed. Once the bumps are close enough, inversion with a smaller amount of blurring will provide a fit with better resolution. Instead of choosing strong initial blurring, the amount of blurring can also be chosen according to the complexity of the problem and hence by considering the initially modelled data. A strong initial blurring, in time and along the receiver coordinate, can be used only if the arrival-time errors are large in the initial comparision between the modelled and observed data. Otherwise, if the arrival-time errors are low, the data are blurred either in time or along the receiver coordinate using a small amount of initial blurring. This will reduce the computational burden of the multi-objective inversion. Note that the purpose of blurring is to just create an overlap between the modelled and the corresponding observed bumpy arrivals. As long as this overlap is created, the exact choice of blurring parameters will not affect the multi-objective inversion. The resulting model that minimizes the bump functional is non-unique and its corresponding data will contain false arrivals. We therefore use the output of the bump-functional optimization as input to the least-squares optimization to complete a round-trip. The least-squares objective removes the false arrivals that are not present in the observed data. However, it could happen that the least-squares optimization now converges to a local minimum. Then we use the output corresponding to the, possibly local, minimum of the least-squares inversion from the first round-trip as input to the bump-functional optimization. The auxiliary bump objective now pulls the solution out of the local minimum, since it

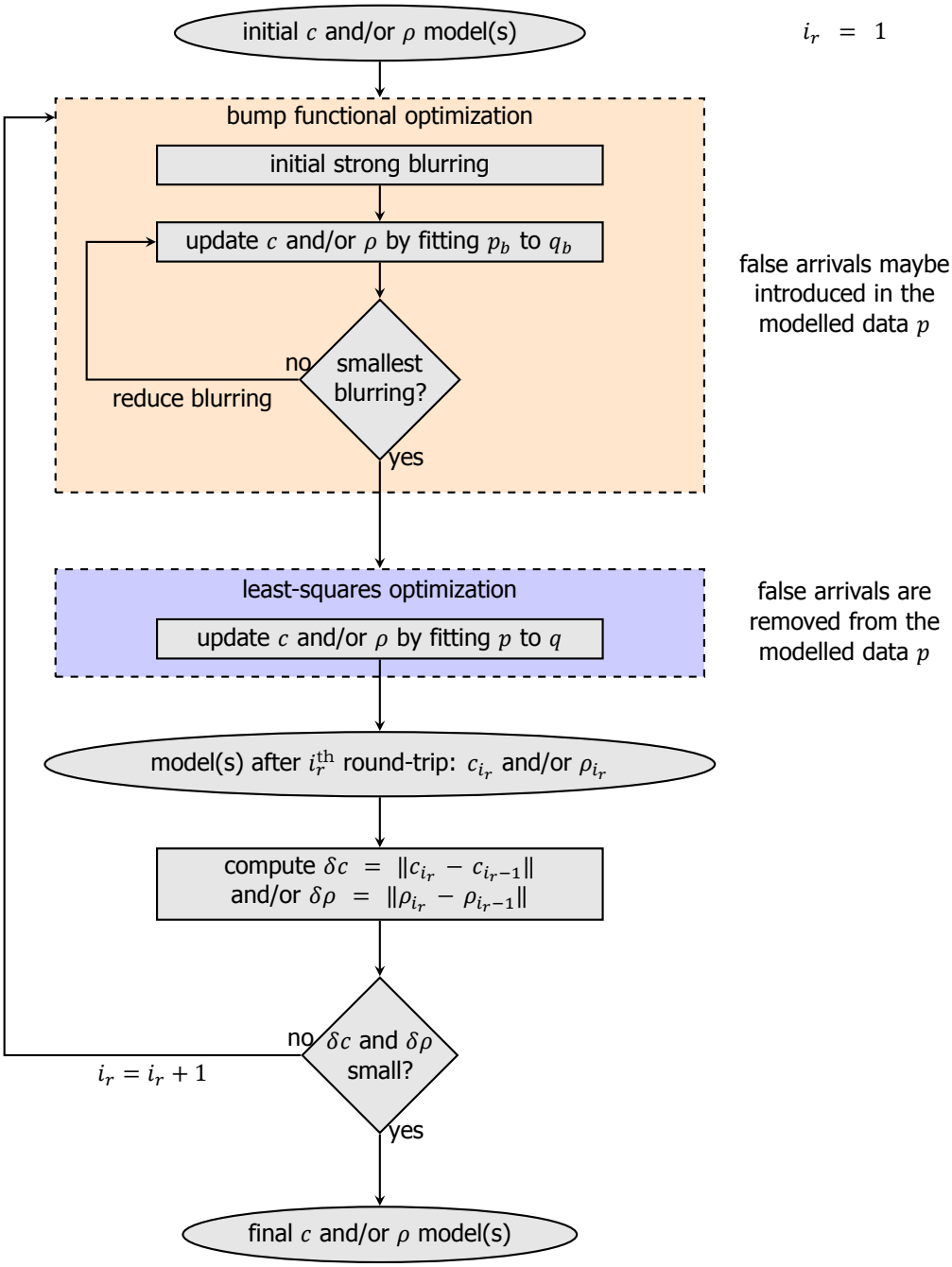


Figure 3.12: Flowchart of multi-objective inversion strategy.

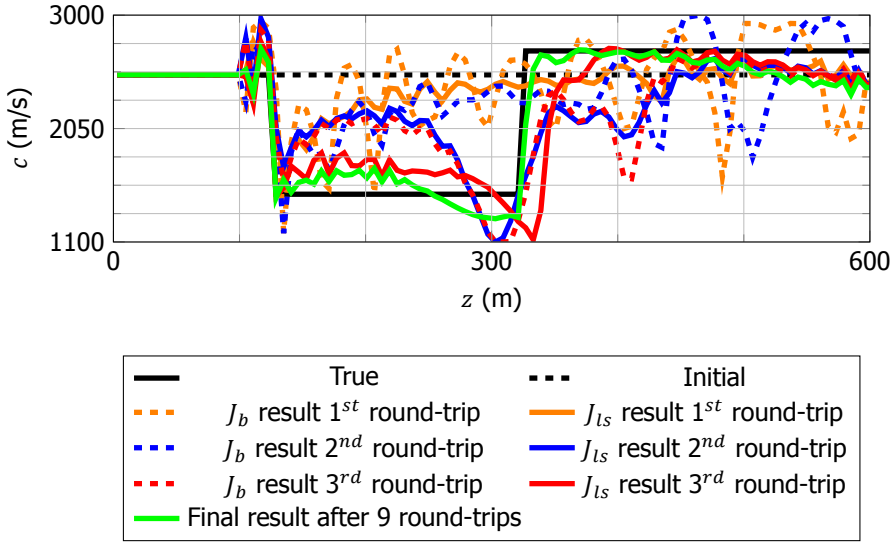


Figure 3.13: Same as Figure 3.6, but now with the proposed multi-objective inversion scheme. The bump functional with $\sigma_t/\tau_d = 1.0$ and no receiver blurring ($\sigma_r = 0$) is used along with the least-squares objective.

does not suffer from cycle skipping. Subsequently, we carry out more round-trips to converge to the global minimum of the least-squares objective.

3.4.2. Three-layer Reflection Example, Again

To demonstrate the effectiveness of our multi-objective inversion scheme, we consider the same three-layer reflection example with a negative velocity anomaly as used before for single-objective inversion with either J_{ls} or J_b. In each round-trip, we first use the bump functional, followed by minimization of the least-squares functional, as shown in the Figure 3.12. The blurring parameters used in this case are given in the Table 3.2. For this simple example, we did not use the bump functional with strong blurring. The final model after 9 round-trips is plotted in green in Figure 3.13. The corresponding modelled shot gather, plotted in Figure 3.14f, is not cycle-skipped when compared to the observed shot gather in the Figure 3.7a.

During the first round-trip, the bump functional inversion outputs a model that belongs to $\mathbb{M}_b$ as shown in Figure 3.13. The modelled data corresponding to this model, in Figure 3.14a, has false arrivals due to strong reflectivity around $z = 450$ m. We notice that the J_{ls} optimization removes those false arrivals that are not present in the observed data, as visible in Figure 3.14b. The J_{ls} inversion converges to a local minimum, where the modelled data are cycle-skipped compared to the observed data. We now use the bump functional to pull the trapped solution out of the local minimum during the second round-trip. The output model and the corresponding modelled data are plotted in Figures 3.13 and 3.14c, respectively. We notice that whenever the least-squares objective removes the false arrivals, it updates the background velocity of the model.

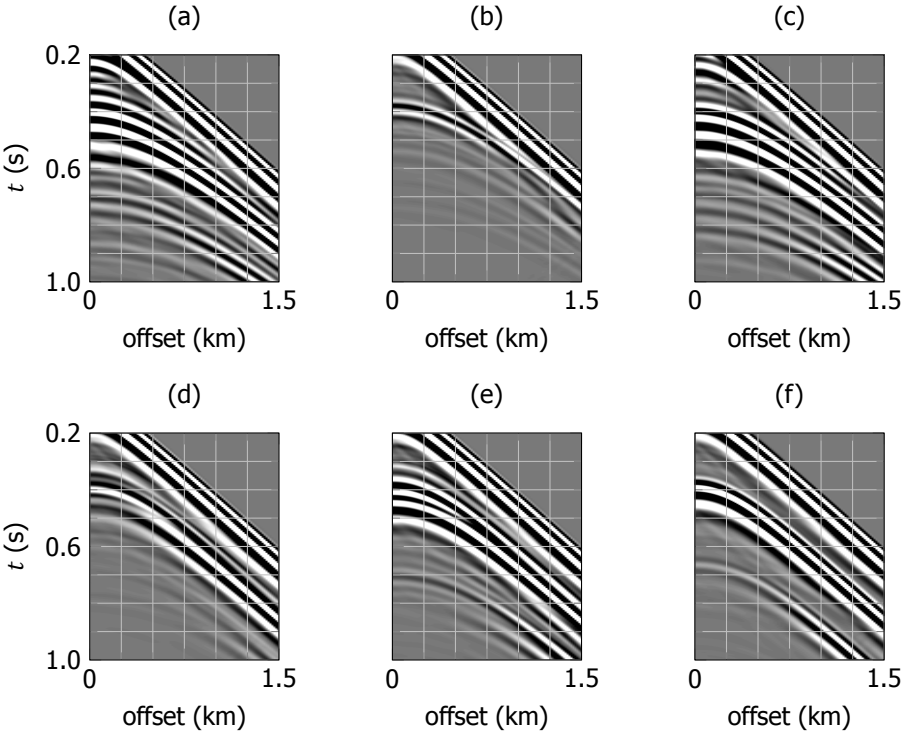


Figure 3.14: Modelled shot gathers at the last iteration corresponding to the multi-objective inversion in Figures 3.13. (a) After the 1^{st} J_b inversion. (b) After the 1^{st} J_{ls} inversion. (c) After the 2^{nd} J_b inversion. (d) After the 2^{nd} J_{ls} inversion. (e) After the 3^{rd} J_b inversion. (f) After the 3^{rd} J_{ls} inversion.

Table 3.2: Blurring parameters used in different examples during the multi-objective inversion. Strong blurring is chosen initially depending on the complexity of the example. The blurring is then reduced to the lowest as shown in Figure 3.12. The multi-objective inversion is not sensitive to the exact blurring parameters chosen.

example	initial strong blurring (σ_t /τ_d , σ_r /λ_d)	reduced lowest blurring (σ_t /τ_d , σ_r /λ_d)
three-layer reflection	skipped	(1, 0)
cross-well with field data	(1.5, 1.0)	(0.5, 0)
five layer: early arrivals	(1.0, 0.5)	(0.5, 0)
five layer: all arrivals	(1.5, 1.0)	(0.5, 0)
complex 2-D model	(1.0, 0.5)	(0.5, 0)

J_{ls} inversion again converges to another local minimum that is closer to the global minimum than after the first round-trip. This behaviour is reminiscent of Mora's (1989) observation that the J_{ls} functional can provide low-wavenumber updates to the velocity model, provided the reflectivity is strong enough to allow for an interaction between the scattered and the direct wavefield. In our case, after each round-trip, the output of the least-squares inversion suffers less and less from cycle skipping and provides convergence to the global minimum.

3.5. Realistic Examples

3.5.1. Cross-well Example with Field Data

We consider a cross-well experiment to demonstrate the applicability of the multi-objective inversion strategy to field data. The field data were also used by van Leeuwen [2010] and van Leeuwen and Mulder [2010] to perform wave-equation based travel-time tomography with a cross-correlation type functional. Two wells are located at $x = 205$ m and $x = 27$ m with 122 sources and 125 receivers, respectively. The source and receiver spacing interval is approximately 3.84 m. The velocity and density Earth models after interpolating the well-log data are plotted in Figures 3.15a and 3.15b. During the inversion, both velocity and density are estimated and a 110-Hz Ricker source wavelet is used. We choose a homogeneous initial velocity model with $c = 2500$ m/s for inversion such that the arrival-time error in the initially modelled data, plotted in Figure 3.17b, exceeds the dominant period by a factor 2 to 3. Due to this reason, J_b inversion with high blurring is used initially in the multi-objective inversion strategy, which is outlined in the Figure 3.12. The blurring parameters used during the multi-objective inversion are given in Table 3.2. It is expected that the single-objective J_{ls} inversion suffers from cycle-skipping.

Synthetic Data Inversion

We first use the interpolated well-log models to generate synthetic pressure data, as plotted in Figure 3.17a. While inverting these data, the output models are only allowed to vary with depth. The bumpy 'observed' data used during the inversion are plotted in

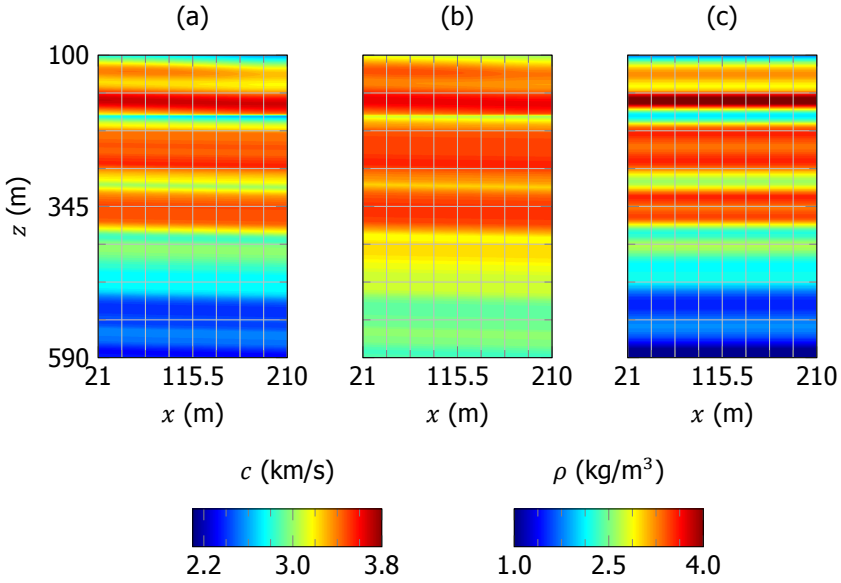


Figure 3.15: A synthetic cross-well example. (a) c and (b) ρ models after interpolating the well-logs used for forward modelling. (c) Multi-objective inversion output c model using synthetic data from the well-log models. The output c and ρ models are only allowed to vary with depth. The ρ model is not shown.

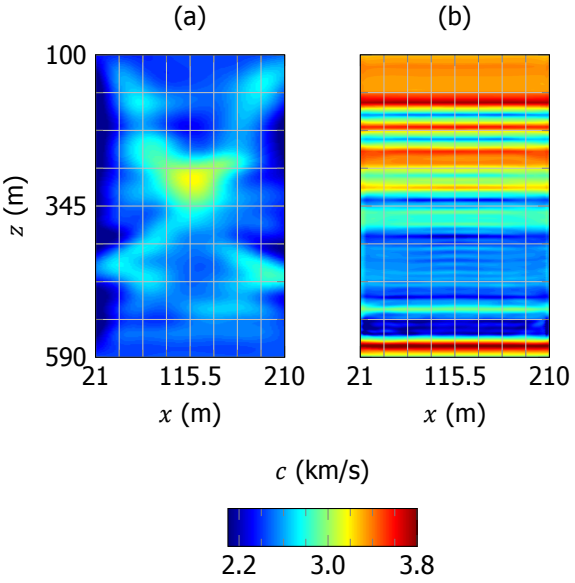


Figure 3.16: A cross-well example with field data and well-logs plotted in Figures 3.15a and 3.15b. The output ρ models are not shown. (a) Reconstructed c model from the field data using the single-objective J_{ls} inversion. (b) Same as (a), but using the multi-objective inversion strategy.

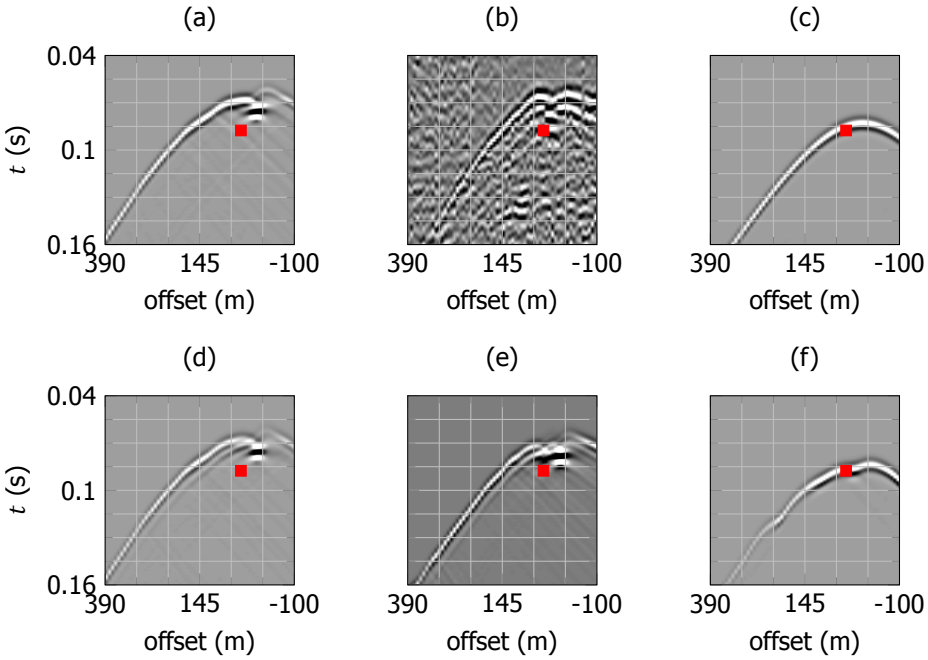


Figure 3.17: Shot gathers corresponding to the cross-well experiment in Figures 3.15 and 3.16 for a source at $(205, 199)$ m. (a) Synthetic data using interpolated well-log models. (b) Observed field data. (c) Initially modelled data for synthetic and field data inversion. (d) Modelled data after synthetic data multi-objective inversion, to be compared to (a). (e) Modelled data after field data inversion, to be compared to (b). (f) Same as (e), but after single-objective J_{ls} inversion.

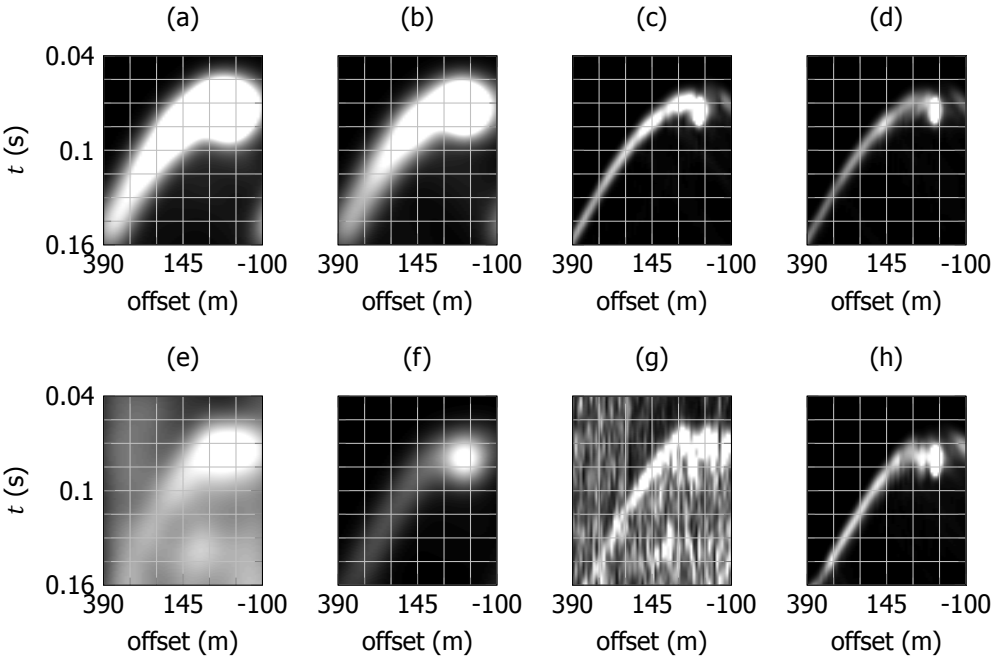


Figure 3.18: Bumpy shot gathers corresponding to the cross-well experiment in Figures 3.15 and 3.16 for a source at $(205, 199)$ m. (a) Observed and (b) modelled gathers after synthetic data inversion using $\sigma_t/\tau_d = 1.5$ and $\sigma_r/\lambda_d = 1.0$. (c) Observed and (d) modelled gathers after synthetic data inversion using $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$. (e) Observed and (f) modelled gathers after field data inversion using $\sigma_t/\tau_d = 1.5$ and $\sigma_r/\lambda_d = 1.0$. (g) Observed and (h) modelled gathers after field data inversion using $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$.

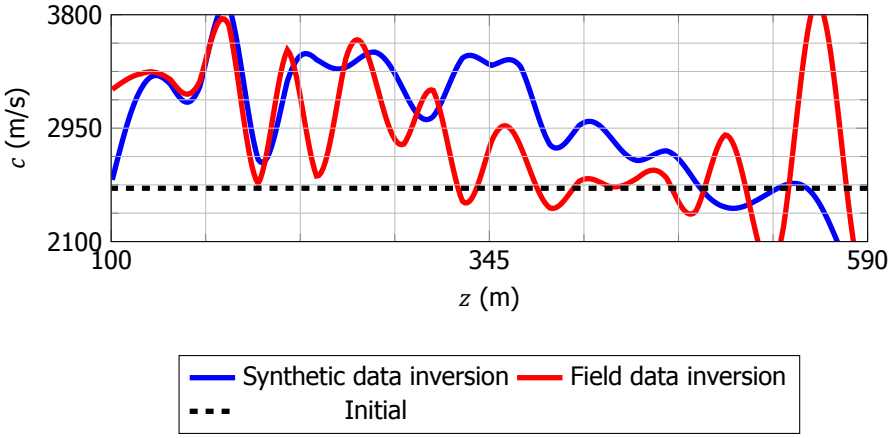


Figure 3.19: Output models for the cross-well experiment. For the synthetic data inversion, the output in Figure 3.15c varies only with depth and is plotted in blue. A cross-section at $x = 110$ m, obtained from the field data and corresponding to the reconstructed model in Figure 3.16b, is plotted in red.

Figures 3.18a and 3.18c. The multi-objective inversion results in the final velocity model plotted in 3.15c. The modelled data after the final iteration, plotted in Figure 3.17d, match the synthetic observed data. And the modelled bumpy data after the inversion, plotted in Figures 3.18b and 3.18d, match the observed bumpy data.

Field Data Inversion

Under the acoustic approximation, we assumed that the recorded x-component data generated by x-component sources are not too different from acoustic pressure data generated by explosive sources. We band-limited the observed data to the range between 80 to 120 Hz, such that the low frequencies are removed. Attenuation is not taken into account, but during pre-processing, an amplitude-versus-offset correction was applied to the observed data using the initially modelled data, following Brenders and Pratt [2007]. The observed shot gather for a source at $(205, 199)$ is shown in Figure 3.17b. Both single- and multi-objective inversions in this case are performed in two stages. During the first stage, the subsurface models are allowed to vary only with depth. Then, the output is used as a starting model to estimate the final results that vary in both the x- and z-coordinates.

Single-objective J_{ls} inversion outputs the velocity model plotted in Figure 3.16a that varies both in the x- and z-coordinates. The corresponding modelled shot gather in Figure 3.17f is cycle skipped compared to the observed shot gather in Figure 3.17b: notice the arrivals around the red square. The multi-objective inversion results in the final velocity model plotted in Figure 3.16b that also varies both in the x- and z-coordinates. The modelled shot gather after inversion, plotted in Figure 3.17e, matches the observed shot gather in Figure 3.17b. The modelled bumpy data, plotted in Figures 3.18f and 3.18h, match the observed bumpy data in Figures 3.18e and 3.18g after the inversion. Finally, Figure 3.19 displays vertical cross-sections of the synthetic- and field-data multi-objective inversion results.

3.5.2. Five-layer Example

We now numerically test the applicability of the multi-objective strategy when:

- the data are acquired at sufficiently long offsets to contain refracted waves and
- the data contain many arrivals including free-surface reflection as well as refraction multiples.

A somewhat more realistic five-layer Earth model of 6000 m width and 2500 m depth is assumed and the velocity varies with depth as depicted in Figure 3.20. We placed an evenly spaced horizontal array of 200 receivers at a depth of 10 m. A source is placed at $(0, 10)$ such that the maximum offset in the data is 6000 m. We generated 'observed' data using a fourth-order minimum-phase Butterworth source wavelet of bandwidth 5–10 Hz. The 'observed' shot gather, plotted in Figure 3.21a, contains refractions due to high-velocity contrasts at the interfaces just below 750 m and 1200 m. We choose a free-surface boundary condition so that free-surface multiples are present in the data. The initial velocity model, plotted in Figure 3.20, has an incorrect background velocity. The modelled data using the initial velocity model are plotted in Figure 3.21b.

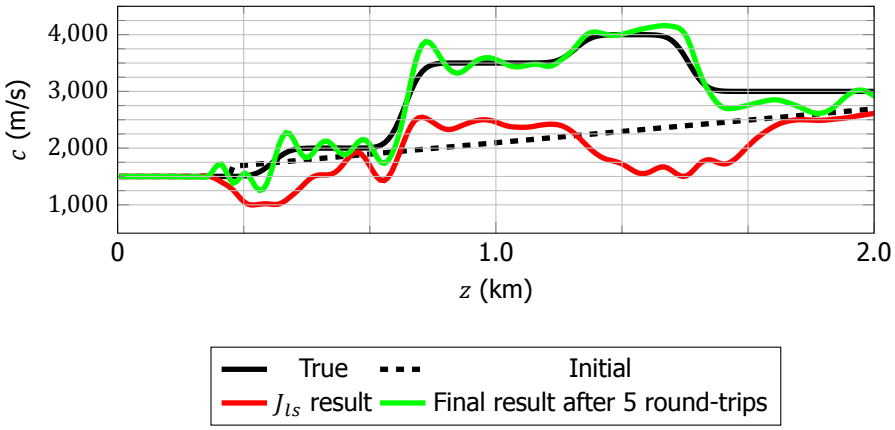


Figure 3.20: Plot corresponding to the five-layer example. The actual (solid black) and initial (dashed black) models are plotted. Single-objective J_{ls} inversion result is plotted in red. The multi-objective inversion scheme, depicted in Figure 3.12, results in a model plotted in green.

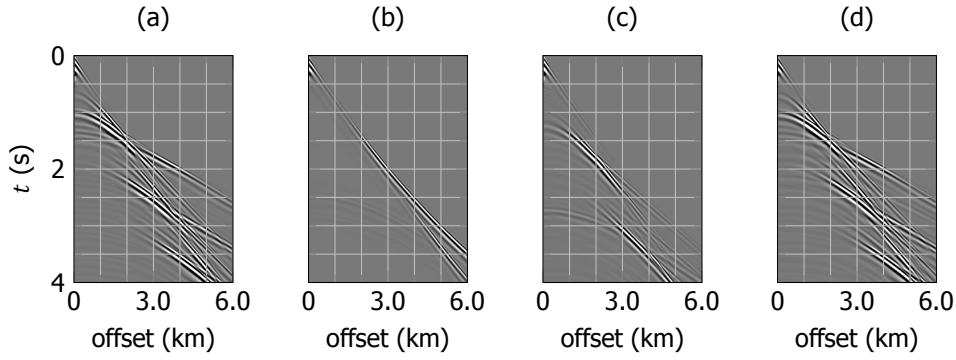


Figure 3.21: The shot gathers corresponding to the results in Figure 3.20. (a) Observed gather. (b) Initially modelled gather. (c) Modelled gather after single-objective J_{ls} inversion. (d) Modelled gather after multi-objective inversion.

We now perform the multi-objective inversion depicted in Figure 3.12. Each round-trip consists of two different optimization stages. During the first stage, we used the blurring parameters as in Table 3.2 with time windowing to fit only the early refracted arrivals at offsets larger than 2000 m. The inversion model grid spacing during these optimizations is 50 m and a Gaussian with standard deviation of 210 m is used for gradient smoothing. In the next stage, the optimizations minimize the error in the entire shot gather without windowing and use the blurring parameters given in Table 3.2. The model spatial sampling is 30 m. The standard deviation of the Gaussian used for gradient smoothing is 60 m. The modelled data after 5 round-trips, plotted in Figure 3.21d, match the synthetic observed data. The corresponding output velocity model is plotted in Figure 3.20.

We now perform the single-objective J_{ls} inversion also in two stages that are described above. Single-objective J_{ls} inversion fails to reconstruct the background velocity of the model, as plotted in Figure 3.20. Also, the modelled data at the last iteration, plotted in Figure 3.21c, do not fit the observed data.

3.5.3. Complex 2-D Model with Reflection Data

Figure 3.22a depicts an assumed Earth model of 4800-m width and 1280-m depth with a high-velocity structure. We placed an evenly spaced horizontal array of 60 sources and 100 receivers at a depth of 20 m. We used a 5–12-Hz fourth-order minimum-phase Butterworth source wavelet. The 'observed' data generated for the actual model do not contain any refracted arrival due to the negative velocity anomaly at shallow depths. The initial velocity model for inversion is homogeneous with $c = 3000$ m/s. The background velocity of the model has to be reconstructed using only the reflected arrivals in the data. In order to reduce high-wavenumber artefacts, we apply Gaussian smoothing to the gradient at each iteration with a standard deviation of 32 m. The single-objective J_{ls} inversion results in a model with an incorrect background velocity, plotted in Figure 3.22b. We started again from the homogeneous model but now with the bump functional, setting $\sigma_t/\tau_d = 1$ and $\sigma_r/\lambda_d = 0.5$. As expected, the single-objective bump functional inversion results in a model that does not have the correct background velocity, as can be seen in Figure 3.22c. Finally, we performed 10 round-trips using the multi-objective inversion strategy, outlined in Figure 3.12. The multi-objective inversion uses the blurring parameters as in Table 3.2. The resulting velocity model is plotted in Figure 3.22d and resembles the actual velocity model fairly well.

3.6. Conclusions

We formulated a data-domain functional that matches the observed and the modelled data in a simplified form. It can be viewed as a generalized envelope-based misfit. The simplification results in bumpy data, obtained by taking the absolute value of the data and subsequent smoothing or blurring with a Gaussian. Using numerical examples involving transmission or reflection data, we illustrated the following characteristics of the bump functional:

1. the functional is less sensitive to cycle skipping and does not rely on the low

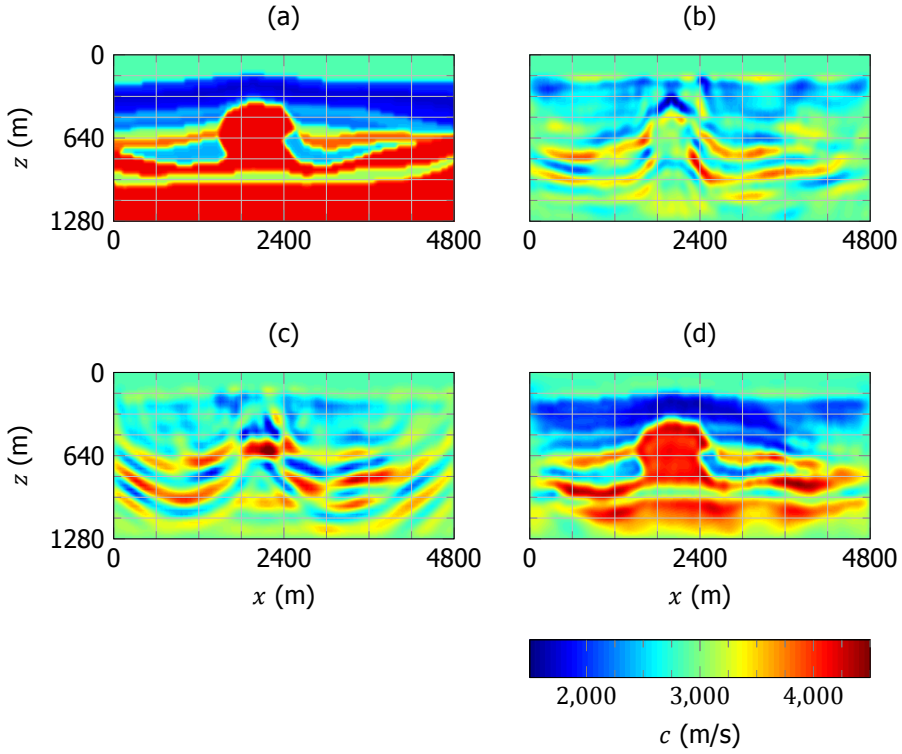


Figure 3.22: A 2-D reflection experiment on a complex model with a high-velocity structure. The initial model for inversion is homogeneous with $c = 3000$ m/s. (a) Actual velocity model. (b) Single-objective J_{ls} inversion result. (c) Single-objective J_b inversion result using $\sigma_t/\tau_d = 1$ and $\sigma_r/\lambda_d = 0.5$. (d) Result from the multi-objective inversion strategy, outlined in Figure 3.12, after 10 inversion round-trips.

frequencies present in the data;

2. blurring increases the size of the basin of attraction that corresponds to the functional and hence its global-convergence robustness;

3. the bump-functional inversion suffers from the fact that the model that matches given bumpy data tends to be highly non-unique.

Single-objective bump-functional inversion can produce acceptable results while fitting transmitted arrivals. While in the case of reflected arrivals, the non-uniqueness prevents the bump functional to update the background velocity of the model. Therefore, in order to reach the global minimum corresponding to the least-squares objective, we proposed a multi-objective inversion scheme that uses the bump functional as an auxiliary functional. We demonstrated the potential of the bump functional to pull the trapped solution out of the least-squares local minimum whenever necessary. Finally, we have tested the applicability of the multi-objective inversion scheme using realistic numerical examples as well as cross-well field data. In all the cases, the scheme found the model that well explains the observed data and hence corresponds to the global minimum of the least-squares functional, even in the absence of low frequencies in the data.

References

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

P. Mora, *Elastic wave-field inversion of reflection and transmission data,* Geophysics **53**, 750 (1988).

P. Mora, *Inversion = migration + tomography,* Geophysics **54**, 1575 (1989).

R. G. Pratt, *Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model,* Geophysics **64**, 888 (1999).

J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics,* Geophysics **74**, WCC1 (2009).

A. Fichtner, *Full seismic waveform modelling and inversion* (Springer, 2010).

M. K. Sen and P. L. Stoffa, *Nonlinear one-dimensional seismic waveform inversion using simulated annealing,* Geophysics **56**, 1624 (1991).

P. L. Stoffa and M. K. Sen, *Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms,* Geophysics **56**, 1794 (1991).

M. Sambridge and G. Drijkoningen, *Genetic algorithms in seismic waveform inversion,* Geophysical Journal International **109**, 323 (1992).

Z. Gao, Z. Pan, J. Gao, X. Zhang, *et al.*, *Building an initial model for full waveform inversion using a global optimization scheme,* in *2014 SEG Annual Meeting* (Society of Exploration Geophysicists, 2014).

D. Datta, *Estimating starting models for full waveform inversion using a global optimization method,* in *77th EAGE Conference and Exhibition 2015* (2015).

O. Gauthier, J. Virieux, and A. Tarantola, *Two-dimensional nonlinear inversion of seismic waveforms: Numerical results,* Geophysics **51**, 1387 (1986).

R. Snieder, M. Xie, A. Pica, and A. Tarantola, *Retrieving both the impedance contrast and background velocity: A global strategy for the seismic reflection problem,* Geophysics **54**, 991 (1989).

W. Mulder and R.-E. Plessix, *Exploring some issues in acoustic full waveform inversion,* Geophysical Prospecting **56**, 827 (2008).

W. Symes, *Migration velocity analysis and waveform inversion,* Geophysical Prospecting **56**, 765 (2008).

Y. Luo and G. Schuster, *Wave-equation traveltime inversion,* Geophysics **56**, 645 (1991).

C. Shin and D.-J. Min, *Waveform inversion using a logarithmic wavefield,* Geophysics **71**, R31 (2006).

C. Shin and Y. H. Cha, *Waveform inversion in the Laplace domain,* Geophysical Journal International **173**, 922 (2008).

C. Shin and Y. H. Cha, *Waveform inversion in the Laplace-Fourier domain,* Geophysical Journal International **177**, 1067 (2009).

Y. Zhang and D. Wang, *Traveltime information-based wave-equation inversion,* Geophysics **74**, WCC27 (2009).

T. van Leeuwen and W. A. Mulder, *A correlation-based misfit criterion for wave-equation traveltime tomography,* Geophysical Journal International **182**, 1383 (2010).

T. van Leeuwen, PhD thesis, Delft University of Technology. (2010).

E. Bozdağ, J. Trampert, and J. Tromp, *Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements,* Geophysical Journal International **185**, 845 (2011).

H. Chauris, D. Donno, and H. Calandra, *Velocity estimation with the normalized integration method,* in *74th EAGE Conference & Exhibition incorporating SPE EUROPEC* (2012).

D. Donno, H. Chauris, and H. Calandra, *Estimating the background velocity model with the normalized integration method,* in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC* (2013).

M. Warner, L. Guasch, *et al.*, *Adaptive waveform inversion: Theory,* in *2014 SEG Annual Meeting* (Society of Exploration Geophysicists, 2014).

B. Engquist and B. D. Froese, *Application of the Wasserstein metric to seismic signals,* arXiv preprint arXiv:1311.4581 (2013).

S. Luo, P. Sava, *et al.*, *A deconvolution-based objective function for wave-equation inversion,* in *2011 SEG Annual Meeting* (Society of Exploration Geophysicists, 2011).

G. T. Schuster, *Seismic Imaging* (SEG Press, 2015).

T. W. Fei, Y. Luo, F. Qin, P. G. Kelamis, *et al.*, *Full waveform inversion without low frequencies: A synthetic study,* in *2012 SEG Annual Meeting* (Society of Exploration Geophysicists, 2012).

D. Sun and W. W. Symes, *Waveform inversion via nonlinear differential semblance optimization,* in *SEG Technical Program Expanded Abstracts* (2012) Chap. 497, pp. 1–7.

C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, *Multiscale seismic waveform inversion,* Geophysics **60**, 1457 (1995).

R.-S. Wu, J. Luo, and B. Wu, *Seismic envelope inversion and modulation signal model,* Geophysics **79**, WA13 (2014).

B. Chi, L. Dong, and Y. Liu, *Full waveform inversion method using envelope objective function without low frequency data,* Journal of Applied Geophysics **109**, 36 (2014).

J. Luo and R.-S. Wu, *Seismic envelope inversion: reduction of local minima and noise resistance,* Geophysical Prospecting **63**, 597 (2015).

R.-É. Plessix and W. A. Mulder, *Frequency-domain finite-difference amplitude-preserving migration,* Geophysical Journal International **157**, 975 (2004).

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

Y. Ma, D. Hale, B. Gong, and Z. Meng, *Image-guided sparse-model full waveform inversion,* Geophysics **77**, R189 (2012).

A. Brenders and R. Pratt, *Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model,* Geophysical Journal International **168**, 133 (2007).

4

A Parameterization Analysis for Acoustic Full Waveform Inversion

We are trying to prove ourselves wrong as quickly as possible,
because only in that way can we find progress.

Richard Feynman

With single-parameter full-waveform inversion (FWI), estimating the inverse of the Hessian matrix will accelerate the convergence but is computationally expensive. Therefore, it is often replaced by the inverse of an approximate Hessian that is easier to compute and serves as a preconditioning matrix. Alternatively, or in combination, a BFGS-type of optimization method that estimates the inverse Hessian from subsequent iterations can be applied. Similarly, in the case of multi-parameter full-waveform inversion, the computation of the Hessian terms that contain derivatives with respect to more than on type of parameter, called cross-parameter Hessian terms, is not usually feasible at each iteration. If a simple gradient-based minimization with, for instance, just scalar weights for each of the parameter types is used, different choices of parametrization can be interpreted as different preconditioners. If the non-linear inverse problem is well-posed, then the inversion should converge to a band-limited version of the true solution irrespective of the parametrization choice, provided we start sufficiently close to the global minimum. However, the choice of parametrization will affect the rate of convergence to the exact solution and the 'best' choice of parametrization is the one with the fastest rate.

Parts of this chapter have been submitted for publication to Geophysical Journal International

In this paper, we are in a search of the best choice for acoustic full-waveform inversion, where 1. anomalies with a size less than a quarter of the dominant wavelength have to be estimated; 2. the scattered wavefield is recorded at all the scattering angles. Towards that end, we review two different conventional analysis methods, i.e., the point-scatterer analysis and diffraction-pattern analysis. To validate these analyses, we consider single-component numerical examples, where the inversion estimates one of the following: 1. only contrast of a point-shaped scatterer; 2. only contrast of a Gaussian-shaped scatterer; 3. both shape and contrast of a Gaussian-shaped scatterer. The numerical examples show that the suggestions of the conventional analyses are only valid while estimating the contrast of point-shaped anomalies at a known location. For extended anomalies, they suggest that the best choice of parameterization depends on the contrast of the subsurface scatterer that the inversion tries to estimate. Therefore, we cannot decide on the best parameterization choice for full-waveform inversion in the general case.

In addition to full-waveform inversion, we consider Born modelling and inversion to learn if the dependence on the contrast of subsurface scatterer is due to the non-linearity in the full-waveform modelling and inversion. We again observe that we cannot decide on the best parameterization choice for Born modelling and inversion in the general case, where both the shape and the size of the anomalies have to be estimated.

4.1. Introduction

Quantitative imaging of various near-surface elastic parameters is essential in many civil engineering applications as well as for hydrocarbon exploration. One approach to this is full-waveform inversion (FWI) of the recorded elastic wavefield [Tarantola, 1986; Virieux and Operto, 2009], which is sensitive to the shear and compressional properties of the subsurface. Multi-parameter FWI is a non-linear procedure that minimises the least-squares misfit between the recorded and the modelled seismic data to estimate various subsurface parameters. Given the size of the seismic problem, i.e., estimating thousands or millions of parameters, it is only feasible in practice to use descent methods for optimisation.

In order to reduce the number of iterations needed to reach an acceptable solution, the gradient at each iteration should be preconditioned using the inverse of the Hessian matrix [Pratt *et al.*, 1998]. The elements of the Hessian matrix are the second-order derivatives of the misfit function with respect to the model parameters. If model parameter of kinds a and b at locations $\mathbf{x}_i$ and $\mathbf{x}_j$ are given by $m_a(\mathbf{x}_i)$ and $m_b(\mathbf{x}_j)$, respectively, the elements of the Hessian matrix $\mathbf{H}$ are

$$\mathbf{H}_{(a,i),(b,j)} = \partial_{m_a(\mathbf{x}_i)} \left(\partial_{m_b(\mathbf{x}_j)} J_{ls} \right), \tag{4.1}$$

where J_{ls} denotes the least-squares misfit functional.

We will refer to the Hessian terms with $a = b$ as *mono-parameter* terms. The terms of the Hessian on its band diagonal with $i \approx j$, $i \neq j$ and $a = b$ will be called *band-diagonal* mono-parameter terms, as sketched in Figure 4.1. During the preconditioning, they deconvolve the gradient such that it is less dependent on source bandwidth and acquisition geometry. The terms of the Hessian where $i = j$ and $a = b$, called *main-diagonal* mono-parameter terms, account for the amplitude effects in wave propagation [Pratt *et al.*, 1998; Virieux and Operto, 2009], for instance, due to geometrical spreading. In the presence of multiple scattering, the change in data due to a model perturbation at $\mathbf{x}_i$ depends on the model perturbation at $\mathbf{x}_j$ even if $i \neq j$ and $i \not\approx j$. This means that these non-band-diagonal terms of the Hessian, $\mathbf{H}_{(a,i),(b,j)}$ with $i \neq j$ and $i \not\approx j$, are non-zero. During the preconditioning, they correct the gradient such that multiple scattering is taken into account.

The non-zero Hessian terms with $a \neq b$ represent cross-talk between the different parameter types. We call them *cross-parameter* terms, which are marked in the Figure 4.1. The cross-parameter terms corresponding to a single subsurface position when $i = j$ are called *main* cross-parameter terms, whereas the cross-parameter terms with $i \approx j$ and $i \neq j$ are called *block* cross-parameter terms. When cross-parameter terms are smallest, a change in the data due to a perturbation of one parameter should be almost independent of the perturbation of another parameter at the same location.

The computation of the Hessian matrix or its inverse is computationally expensive for large-scale acoustic inverse problems. Therefore, in this paper, we restrict ourselves to gradient-based optimisation without the involvement of the Hessian. Since the terms of the Hessian matrix depend on the choice of the subsurface parameterization in the case of multi-parameter full-waveform inversion, the various parameterization choices are not equivalent. If the inverse problem is well-posed, the inversion should converge

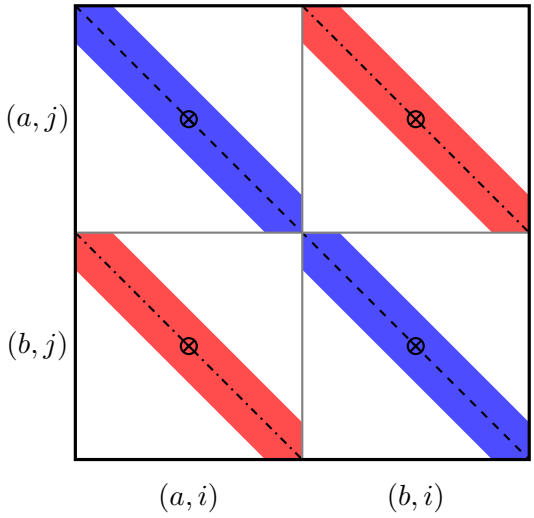


Figure 4.1: This sketch marks different terms of the Hessian matrix in equation 4.1. The main-diagonal mono-parameter terms are marked with a dashed line. The band-diagonal mono-parameter terms are marked in blue and the block cross-parameter terms in red. Dashed-dotted line mark the main cross-parameter terms. The terms of the reduced Hessian, in equation 4.17, are marked with $\otimes$.

to a band-limited version of the true solution irrespective of the parameterization choice, provided we start close to the global minimum. The *best* choice of parameterization is the one that converges the fastest. It should be noted that the rate of convergence of a particular parameterization depends on the background subsurface model, without scatterers, and the acquisition geometry. The background subsurface model serves as the initial or starting model for inversion.

Hence, for a given acquisition geometry and background subsurface model, it is obvious that the fastest convergence will be obtained for a parameterization choice that provides zero cross-parameter terms in the Hessian. We notice, for an acoustic inverse problem, that both the main and block cross-parameter terms are non-zero for any of the parameterization choices involving wave velocity, density, bulk modulus or wave-impedance. Among these options, we, therefore, have to choose a parameterization that provides the smallest condition number of the Hessian matrix. Because the computation of the condition number of the full Hessian matrix for large-scale inverse problems is costly or intractable, we need a simpler method for parameterization analysis. There are two conventional methods to analyse different subsurface parameterization choices with the aim to find the *best* one. The first method, called diffraction-pattern analysis, chooses the best parametrization by examining the radiation patterns of the scattered wavefield generated by different perturbations [Tarantola, 1986; Operto *et al.*, 2013; Prieux *et al.*, 2013]. The second method, called point-scatterer analysis, considers the Hessian of a reduced inverse problem that only estimates the contrast of a point scatterer at a known location. The reduction will result in fewer unknowns compared to

Table 4.1: Various medium parameterizations that are used in this paper for the acoustic inverse problem.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p,\rho}$	$\xi_{c_p} = (c_p - c_{po})/c_{po}$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(B)	$\mathbf{m}_{Z_p,\rho}$	$\xi_{Z_p} = (Z_p - Z_{po})/Z_{po}$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(C)	$\mathbf{m}_{c_p,Z_p}$	$\xi_{c_p} = (c_p - c_{po})/c_{po}$	$\xi_{Z_p} = (Z_p - Z_{po})/Z_{po}$
(D)	$\mathbf{m}_{K,\rho}$	$\xi_K = (K - K_o)/K_o$	$\xi_\rho = (\rho - \rho_o)/\rho_o$
(E)	$\mathbf{m}_{1/K,1/\rho}$	$\xi_{1/K} = (1/K - 1/K_o)/(1/K_o)$	$\xi_{1/\rho} = (1/\rho - 1/\rho_o)/(1/\rho_o)$

the original inverse problem, where both the contrast and shape of the scatterer have to be estimated. The block cross-parameter terms of the original non-reduced Hessian, shown in Figure 4.1, are absent in the reduced Hessian. Therefore, this analysis ignores the block cross-parameter terms of the original non-reduced Hessian matrix and chooses a parameterization which has main cross-parameter terms with the least possible magnitude.

In this paper, we consider an almost well-posed 2-D acoustic inversion example. In the numerical examples, the aim is to reconstruct seven different scatterers surrounded by sources and receivers using the steepest-descent minimization method, which does not involve an expensive computation of the Hessian matrix at each iteration. The steepest-descent method, unlike the conjugate-gradient or L-BFGS quasi-Newton [Byrd *et al.*, 1995] methods, allows us to clearly observe the differences in convergence rates among the various parameterization choices. In this regard, we employ three different modelling and inversion schemes, where the first two schemes use the Born approximation for modelling and inversion, while the third scheme uses a time-domain acoustic staggered-grid finite-difference code. For Born modelling and inversion, we use a homogeneous background model and the analytical expressions of the corresponding 2-D Green functions. Using this example, we show that the findings of the conventional parameterization analysis methods are not valid in general. For elastic waveform inversion, Modrak *et al.* [2016] also numerically demonstrate that the choice of material parameters is more complicated than the current literature suggests.

This paper is organised as follows. We start with the considered parameterization choices. Then, we outline the three modelling and inversion schemes that are employed in the numerical example. Next, we briefly review the two conventional parameterization analysis methods before validating them against the numerical examples. The last section concludes the paper.

4.2. Parameterization Choices

We parameterize the Earth model with two parameters at each point, involving density, ρ, and compressional-wave speed, c_p, or their combinations, for instance, compressional-wave impedance, $Z_p = \rho c_p$, or bulk modulus, $K = \rho c_p^2$. We denote the model parameter

Table 4.2: Conversion formulas used for the non-linear re-parameterization.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p,\rho}$	$\xi_{c_p} = \sqrt{\frac{(\xi_{1/\rho}+1)}{(\xi_{1/\kappa}+1)}} - 1$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$
(B)	$\mathbf{m}_{Z_p,\rho}$	$\xi_{Z_p} = \sqrt{\frac{1}{(\xi_{1/\rho}+1)(\xi_{1/\kappa}+1)}} - 1$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$
(C)	$\mathbf{m}_{c_p,Z_p}$	$\xi_{c_p} = \sqrt{\frac{(\xi_{1/\rho}+1)}{(\xi_{1/\kappa}+1)}} - 1$	$\xi_{Z_p} = \sqrt{\frac{1}{(\xi_{1/\rho}+1)(\xi_{1/\kappa}+1)}} - 1$
(D)	$\mathbf{m}_{K,\rho}$	$\xi_K = -\frac{\xi_{1/\kappa}}{\xi_{1/\kappa}+1}$	$\xi_\rho = -\frac{\xi_{1/\rho}}{\xi_{1/\rho}+1}$

of kind a by m_a and write

$$m_a(\mathbf{x}) = m_{ao}(\mathbf{x})\left[1 + \xi_{m_a}(\mathbf{x})\right], \tag{4.2}$$

where $\xi_{m_a}(\mathbf{x})$ is the contrast function and an additional subscript o is used to denote the parameters corresponding to the reference medium. The subsurface location is denoted by $\mathbf{x}$. When the subsurface is parameterized using ξ_{m_a} and ξ_{m_b}, the model vector at each point is denoted by $\mathbf{m}_{m_a,m_b}$ and the choice of parameterization is indicated by the subscript 'm_a, m_b'. For example, in the case of $\mathbf{m}_{c_p,\rho}$-parameterization, the subsurface is parameterized using the contrast functions of the compressional-wave speed c_p and mass density ρ. The model vector for $\mathbf{m}_{c_p,\rho}$-parameterization is

$$\mathbf{m}_{c,\rho} = \begin{bmatrix} \xi_c \\ \xi_\rho \end{bmatrix}. \tag{4.3}$$

In this paper, we consider the five different parameterization choices given in the Table 4.1. In the case of a $\mathbf{m}_{c_p,Z_p}$-parameterization, perturbations in the first subsurface parameter mainly affect the transmission of waves and perturbations in the second mainly their reflections. We consider the $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization because the acoustic wave-operator is linear in $\xi_{1/\kappa}$ and $\xi_{1/\rho}$, providing zero second derivatives with respect to these medium parameters. The second derivatives, which are non-zero for other parameterization choices, are used during the computation of the full Hessian matrix as shown by Fichtner and Trampert [2011]. Now, we discuss two approaches to re-parameterize the subsurface from one parameterization to another.

4.2.1. Re-parameterization

The non-linear re-parameterization approach involves the exact non-linear transformation. For example, in order to obtain ξ_K from $\xi_{1/\kappa}$, we write

$$\xi_K = \frac{K - K_o}{K_o} = \left(\frac{1}{\xi_{1/\kappa}\,1/\kappa_o + 1/\kappa_o} - K_o\right)\frac{1}{K_o} = \frac{1}{\xi_{1/\kappa}+1} - 1 = -\frac{\xi_{1/\kappa}}{\xi_{1/\kappa}+1}. \tag{4.4}$$

Similarly, Table 4.2 gives non-linear re-parameterization formulas for various parameterization choices.

Table 4.3: Conversion formulas used for the linear re-parameterization.

	Parameterization	First Parameter	Second Parameter
(A)	$\mathbf{m}_{c_p,\rho}$	$\xi_{c_p} = \frac{1}{2}\left(\xi_{1/\rho} - \xi_{1/\kappa}\right)$	$\xi_\rho = -\xi_{1/\rho}$
(B)	$\mathbf{m}_{Z_p,\rho}$	$\xi_{Z_p} = \xi_{1/\rho} - \frac{1}{2}\xi_{1/\kappa}$	$\xi_\rho = -\xi_{1/\rho}$
(C)	$\mathbf{m}_{c_p,Z_p}$	$\xi_{c_p} = \frac{1}{2}\left(\xi_{1/\rho} - \xi_{1/\kappa}\right)$	$\xi_{Z_p} = \xi_{1/\rho} - \frac{1}{2}\xi_{1/\kappa}$
(D)	$\mathbf{m}_{K,\rho}$	$\xi_K = -\xi_{1/\kappa}$	$\xi_\rho = -\xi_{1/\rho}$

The linear re-parameterization approach assumes small contrasts, ignoring higher-order terms in the exact non-linear conversion formula. As an example, we consider a case where the subsurface is initially parameterized by $\mathbf{m}_{K,\rho}$. In order to re-parameterize the subsurface to $\mathbf{m}_{1/\kappa,1/\rho}$, we write the model vector as

$$\begin{aligned}\mathbf{m}_{1/\kappa,1/\rho} = \begin{bmatrix}\xi_{1/\kappa}\\ \xi_{1/\rho}\end{bmatrix} &= \begin{bmatrix}-\frac{K_o}{K} & 0\\ 0 & -\frac{\rho_o}{\rho}\end{bmatrix}\begin{bmatrix}\xi_K\\ \xi_\rho\end{bmatrix}\\ &= \begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\xi_K\\ \xi_\rho\end{bmatrix} + \begin{bmatrix}\frac{\xi_K}{1+\xi_K} & 0\\ 0 & \frac{\xi_\rho}{1+\xi_\rho}\end{bmatrix}\begin{bmatrix}\xi_K\\ \xi_\rho\end{bmatrix}.\end{aligned} \tag{4.5}$$

Now, the second term on the right-hand side of the above equation is ignored since it is of the order ξ_K^2 and ξ_ρ^2 for small ξ_K and ξ_ρ. The linear re-parameterization formula becomes

$$\mathbf{m}_{1/\kappa,1/\rho} \approx \begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}\mathbf{m}_{K,\rho} = -\mathbf{m}_{K,\rho}. \tag{4.6}$$

Similarly, Table 4.3 gives linear re-parameterization formulas for other parameterization choices.

4.3. Modelling and Inversion

We denote the 2-D spatial coordinates by $\mathbf{x} = (x, z)$. The origin is denoted by $\mathbf{x}_0 = (0, 0)$ and the positions of sources and receivers are denoted by $\mathbf{x}_s$ and $\mathbf{x}_r$, respectively. We introduce the reference Green function, $G_o(\mathbf{x}, \omega;\ \mathbf{x}_s)$, satisfying the 2-D acoustic wave equation

$$\mathrm{L}_o G_o = \delta(\mathbf{x} - \mathbf{x}_s). \tag{4.7}$$

Here, ω denotes angular frequency. The wave operator L_o is given by

$$\mathrm{L}_o = \frac{\omega^2}{K_o} + \nabla \cdot \frac{1}{\rho_o}\nabla, \tag{4.8}$$

where K_o denotes the bulk modulus and ρ_o the mass density of the reference medium. The Green function corresponding to the actual inhomogeneous medium is denoted by $G(\mathbf{x}, \omega;\ \mathbf{x}_s)$ and satisfies the wave equation

$$\mathrm{L}\, G = \delta(\mathbf{x} - \mathbf{x}_s), \tag{4.9}$$

with

$$\mathrm{L} = \frac{\omega^2}{K} + \nabla \cdot \frac{1}{\rho}\nabla. \tag{4.10}$$

We write Green's function in the actual inhomogeneous medium as $G = G_o + G_s$, where G_s stands for the scattered component of the total pressure wavefield. The Lippmann-Schwinger equation produces

$$\begin{aligned} G(\mathbf{x}_r, \omega;\ \mathbf{x}_s) \quad = \quad & G_o(\mathbf{x}_r, \omega;\ \mathbf{x}_s) + \\ & \int_{\mathbf{x}} G_o(\mathbf{x}_r, \omega;\ \mathbf{x})\omega^2 \left(\frac{1}{K} - \frac{1}{K_o}\right) G(\mathbf{x}, \omega;\ \mathbf{x}_s)\mathrm{d}\mathbf{x} + \\ & \int_{\mathbf{x}} G_o(\mathbf{x}_r, \omega;\ \mathbf{x}) \left(\nabla \cdot \left(\frac{1}{\rho} - \frac{1}{\rho_o}\right) \nabla G(\mathbf{x}, \omega;\ \mathbf{x}_s)\right) \mathrm{d}\mathbf{x}, \end{aligned} \tag{4.11}$$

where $G_o(\mathbf{x}_r, \omega;\ \mathbf{x})$ denotes the scatterer-to-receiver field in the unperturbed medium and $G(\mathbf{x}, \omega;\ \mathbf{x}_s)$ denotes the source-to-scatterer field in the perturbed medium. We now employ the Born approximation to replace the actual Green function G in the right-hand side of equation 4.11 with G_o. Also, we use integration by parts to obtain the Born approximation, G_b, of Green's function as

$$\begin{aligned} G_b(\mathbf{x}_r, \omega;\ \mathbf{x}_s) \quad = \quad & G_o(\mathbf{x}_r, \omega;\ \mathbf{x}_s) + \\ & \int_{\mathbf{x}} \frac{\omega^2}{K_o} G_o(\mathbf{x}_r, \omega;\ \mathbf{x}) G_o(\mathbf{x}, \omega;\ \mathbf{x}_s) \xi_{1/K} \mathrm{d}\mathbf{x} - \\ & \int_{\mathbf{x}} \frac{1}{\rho_o} \left[\nabla G_o(\mathbf{x}_r, \omega;\ \mathbf{x}) \cdot \nabla G_o(\mathbf{x}, \omega;\ \mathbf{x}_s)\right] \xi_{1/\rho} \mathrm{d}\mathbf{x}, \end{aligned} \tag{4.12}$$

$$\begin{aligned} \quad = \quad & G_o(\mathbf{x}_r, \omega;\ \mathbf{x}_s) + \\ & \int_{\mathbf{x}} \mathbf{W}^{\mathrm{T}}_{1/K, 1/\rho}(\mathbf{x}, \omega;\ \mathbf{x}_r, \mathbf{x}_s)\, \mathbf{m}_{1/K, 1/\rho}(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned} \tag{4.13}$$

where the modelling vector $\mathbf{W}_{1/K, 1/\rho}$ is given by

$$\mathbf{W}_{1/K, 1/\rho}(\mathbf{x}, \omega;\ \mathbf{x}_r, \mathbf{x}_s) = \begin{bmatrix} \frac{\omega^2}{K_o} G_o(\mathbf{x}_r, \mathbf{x}) G_o(\mathbf{x}, \mathbf{x}_s) \\ -\frac{1}{\rho_o} \left(\nabla G_o(\mathbf{x}_r, \mathbf{x})\right) \cdot \left(\nabla G_o(\mathbf{x}, \mathbf{x}_s)\right) \end{bmatrix}. \tag{4.14}$$

We now discuss different modelling and inversion schemes that are used in the numerical experiments in this paper. In each scheme, the least-squares misfit between the artificially-generated *observed* data $Q(\mathbf{x}_r, \omega;\ \mathbf{x}_s)$ and the modelled data $P(\mathbf{x}_r, \omega;\ \mathbf{x}_s)$,

$$J_{ls} = \frac{1}{2} \sum_{\omega \geq 0} \sum_{s,r} \|P(\mathbf{x}_r, \omega;\ \mathbf{x}_s) - Q(\mathbf{x}_r, \omega;\ \mathbf{x}_s)\|^2, \tag{4.15}$$

is minimized. If the subsurface is parameterized by a model vector other than $\mathbf{m}_{1/K, 1/\rho}$, either the linear or the non-linear re-parameterization approach is used. Then, for forward and adjoint modelling, we use either the Born or a full-waveform approach. After the adjoint modelling, the gradient of J_{ls} with respect to $\mathbf{m}_{1/K, 1/\rho}$ is obtained by chain rule. As already mentioned, we use a steepest-descent algorithm to minimize J_{ls}.

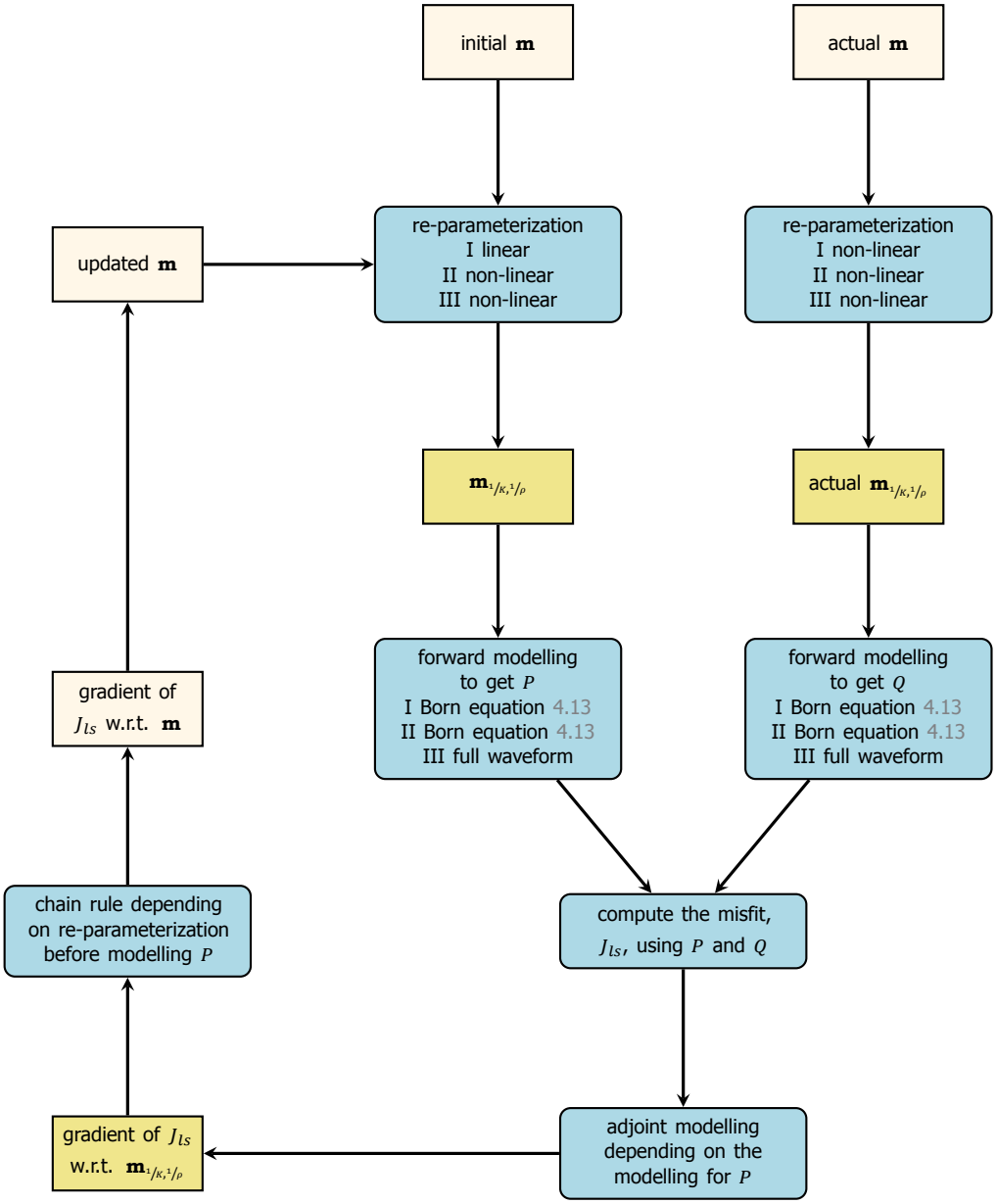


Figure 4.2: A flowchart to illustrate the modelling and inversion schemes I, II and III. Square boxes represent subsurface models and boxes with rounded corners represent an operation. Boxes that depend on the parameterization choice are white. Linear and non-linear re-parameterizations are described by the formulas in Tables 4.3 and 4.2, respectively.

4.3.1. Scheme I

With this scheme, the modelled data are obtained by using the linear re-parameterization, listed in Table 4.3, and the Born-approximated Green functions G_b of equation 4.13:

$$P(\mathbf{x}_r, \omega;\ \mathbf{x}_s) = \Phi(\omega) G_b(\mathbf{x}_r, \omega;\ \mathbf{x}_s). \tag{4.16}$$

Here, $\Phi(\omega)$ denotes the source spectrum, which is assumed to be known. Also, the 'observed' data are generated with G_b, committing an inverse crime, but for those, the non-linear re-parameterization approach of Table 4.2 is used. The flowchart in Figure 4.2 illustrates this scheme.

4.3.2. Scheme II

As shown in Figure 4.2, this scheme is similar to the modelling and inversion scheme I, except that the non-linear re-parameterization, listed in Table 4.2, is adopted while generating both the modelled and the 'observed' data.

4.3.3. Scheme III

The numerical results obtained for this modelling and inversion scheme are decisive because it is identical to conventional acoustic full-waveform modelling and inversion. We used a time-domain acoustic staggered-grid finite-difference code to solve equation 4.9 for the forward as well as the adjoint wavefield computations required for the gradients [Tarantola, 1984, 1986]. Absorbing boundary conditions are used on all sides of the computational domain. The flowchart in the Figure 4.2 also illustrates this scheme.

4.4. Diffraction-pattern Analysis

The most common way to analyse various parameterization choices is with diffraction patterns [Wu and Aki, 1985; Tarantola, 1986; Malinowski *et al.*, 2011; Operto *et al.*, 2013; Gholami *et al.*, 2013; Prieux *et al.*, 2013; He and Plessix, 2016]. In this analysis, for each individual parameterization choice, a perturbation of both the first and the second parameter in the same scatter point at the centre, $\mathbf{x}_0 = (0, 0)$, of the model is considered, leading to an inverse problem with two parameters. When the primary wavefield, incident on these perturbations, is scattered, the contours of the scattered energy produce diffraction patterns. For example, in the case of a $\mathbf{m}_{c_p,\rho}$-parameterization, contours of the scattered energy due to a point scatterer with velocity and density contrasts are plotted in Figures 4.3a and 4.3b, respectively. Here, the incident wavefield is generated by a point source at $(0, 490)$. Note that these plots are insensitive to the sign of the incident wavefield. The diffraction patterns for other parameterization choices are also plotted in Figure 4.3. Alternatively, analytical expressions derived in the framework of asymptotic ray+Born inversion [Forgues and Lambaré, 1997] can be used to obtain the diffraction patterns.

Diffraction pattern analysis chooses subsurface parameterization depending on the recorded arrivals, at a particular scattering angle θ, which are being inverted or fit. A parameterization where the two diffraction patterns at θ differ as much as possible has the fastest convergence because the change in the data due to a perturbation

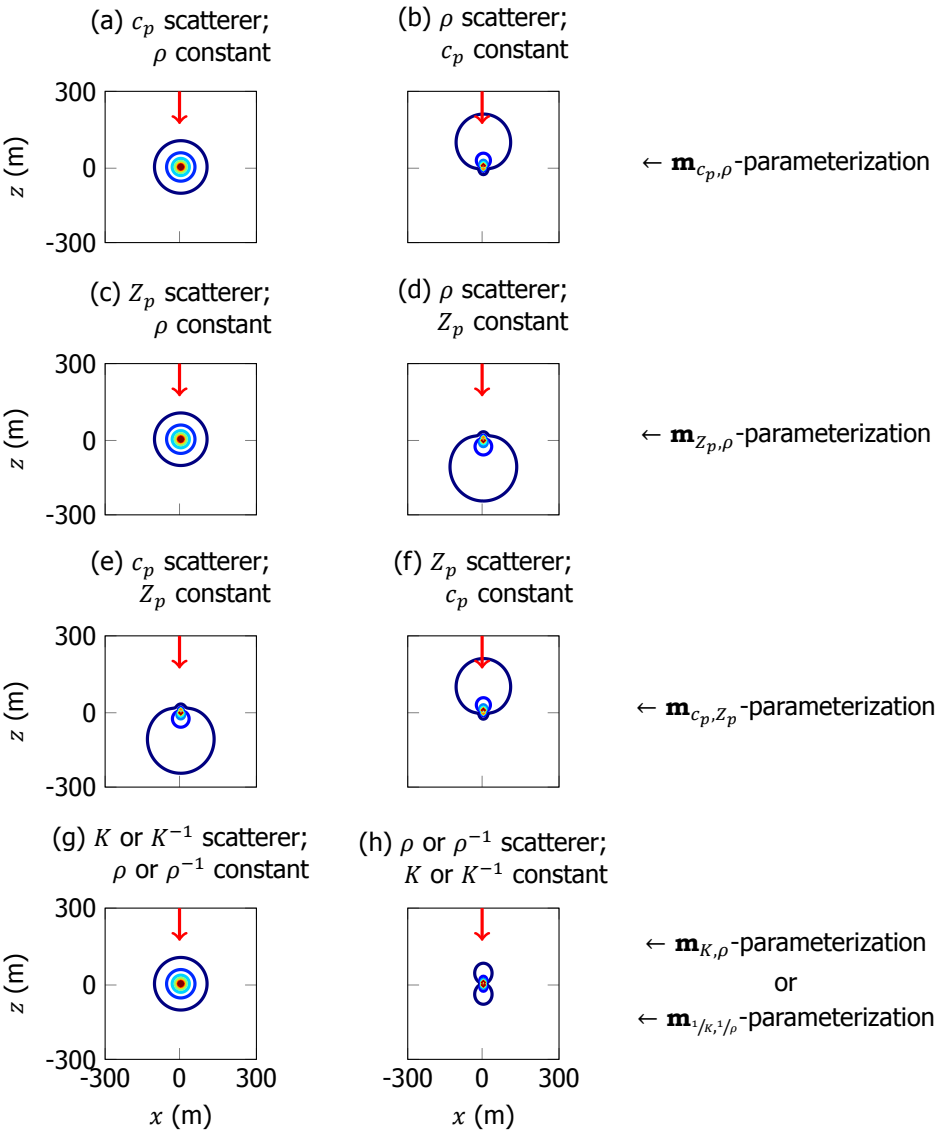


Figure 4.3: Diffraction patterns due to different point scatterers located at the center of the model (0,0). The red arrow indicates the direction of the incident primary wavefield radiated by a source located outside the plotted domain at (0,490). Point scatterers have (a) velocity and (b) density perturbations for a $\mathbf{m}_{c_p,\rho}$-parameterization. (c), (d) For a $\mathbf{m}_{Z_p,\rho}$-parameterization. (e), (f) For a $\mathbf{m}_{c_p,Z_p}$-parameterization. (g), (h) For a $\mathbf{m}_{K,\rho}$-parameterization or $\mathbf{m}_{1/K,1/\rho}$-parameterization. Red-coloured contours are used for higher scattered energy than blue-coloured contours.

in the first parameter is independent of that caused by the second parameter. For example, when mainly inverting arrivals recorded at short-to-intermediate scattering angles, the trade-off between the two parameters is minimum in the case of $\mathbf{m}_{c_p,Z_p}$- or $\mathbf{m}_{Z_p,\rho}$-parameterizations as shown in the Figures 4.3c, 4.3d, 4.3e and 4.3f. This is due to the fact that there is an overlap in the diffraction patterns in the case of other parameterization choices (Figures 4.3a, 4.3b, 4.3g and 4.3h).

In this paper, we are only interested in the case in which the recorded arrivals for all the scattering angles are being inverted simultaneously. For surface-seismic experiments, this corresponds to an acquisition scenario with a wide-angle illumination of the target. In these cases, the contrasts in both the parameters can be uniquely determined as the inverse problem is almost well-posed. This analysis suggests that a $\mathbf{m}_{c_p,Z_p}$-parameterization is the *best* choice for an almost well-posed problem because its two diffraction patterns differ the most. Later in this paper, we want to numerically validate this claim. In the next section, we plot the terms of the Hessian matrix to show that there is a dependence between the two parameters for all the parameterization choices, given that all the scattering angles are considered for inversion.

4.5. Point-scatterer Analysis

During this analysis, the inverse problem is simplified such that only the contrast of a point-shaped scatterer at a known location $\mathbf{x}_0$ has to be estimated. In this case, the terms of the original Hessian matrix in equation 4.1 are reduced to a 2 by 2 symmetric matrix since there are only two unknowns during the inversion:

$$\mathbb{H}_{a,b} = \begin{bmatrix} \mathbf{H}_{(a,0),(a,0)} & \mathbf{H}_{(a,0),(b,0)} \\ \mathbf{H}_{(b,0),(a,0)} & \mathbf{H}_{(b,0),(b,0)} \end{bmatrix}, \tag{4.17}$$

where $\mathbf{H}$ was defined in equation 4.1. We call the matrix $\mathbb{H}$ in equation 4.17 the reduced Hessian. It varies with the parameterization choice and acquisition geometry. As an example, when the modelling and inversion scheme I is employed and the subsurface is parameterized using $\mathbf{m}_{1/\kappa,1/\rho}$, we can write $\mathbb{H}$ in terms of the modelling vector $\mathbf{W}$ as

$$\mathbb{H}_{1/\kappa,1/\rho} = \sum_{\omega} \Phi(\omega)\Phi^*(\omega) \sum_{s,r} \mathbf{W}_{1/\kappa,1/\rho}(\mathbf{x}_0,\omega;\ \mathbf{x}_r,\mathbf{x}_s)\mathbf{W}^{\mathrm{H}}_{1/\kappa,1/\rho}(\mathbf{x}_0,\omega;\ \mathbf{x}_r,\mathbf{x}_s). \tag{4.18}$$

The point-scatterer analysis compares the reduced Hessians $\mathbb{H}$ for different parameterization choices with the objective to find the one with the fastest convergence. As the reduced Hessian also depends on the acquisition geometry, we choose, as an example, a circular acquisition geometry with sources and receivers along a circle with $\mathbf{x}_0$ as the centre. The eigenvalues and eigenvectors of the reduced Hessians [Operto *et al.*, 2013] for the various parameterization choices can be plotted to assess the relative convergence rate for each choice. This analysis was also employed for ray-based inversion by Forgues and Lambaré [1997], where under the high-frequency approximation and in the absence of multiple scattering, the terms of the non-reduced Hessian

matrix, $\mathbf{H}_{(a,i),(b,j)}$, are non-zero only if $i = j$. This analysis suggests that the $\mathbf{m}_{1/\kappa,1/\rho}$ and $\mathbf{m}_{K,\rho}$ parameterization choices are equivalent and have the fastest convergence for the chosen circular acquisition geometry for the following reasons:

Condition Number: Convergence to the exact solution is possible in only one iteration step if all the eigenvalues of the reduced Hessian are equal. Intuitively, this corresponds to the case when the ellipsoidal contours of the objective function become circular. The condition number of $\mathbb{H}_{1/\kappa,1/\rho}$ or $\mathbb{H}_{K,\rho}$, unlike the reduced Hessians for other parameterization choices, is one.

Cross-parameter Terms: The convergence is faster when the coordinate axes, which refer to the parameters, coincide with the eigenvector directions of the reduced Hessian. This corresponds to the case where the cross-parameter terms of the reduced Hessian are zero. This is true only if the subsurface is parameterized using either $\mathbf{m}_{K,\rho}$ or $\mathbf{m}_{1/\kappa,1/\rho}$.

This analysis is not suited for realistic inverse problems, where both the shape and contrast of subwavelength scatterers have to be estimated, unlike the simplified one we considered with the point-shaped scatterer. To show this, we employ the modelling and inversion scheme I and plot some terms of the non-reduced Hessian matrix, $\mathbf{H}_{(a,i),(b,j)}$ with $\mathbf{x}_j = \mathbf{x}_0$, in Figure 4.4. It can be observed that the cross-parameter terms of the non-reduced Hessian matrix are non-zero for any given parameterization demonstrating the dependence between the two parameters. As discussed before, the cross-parameter terms of the reduced Hessian, i.e., the main cross-parameter terms of the non-reduced Hessian matrix, are zero (white colour in Figure 4.4k) for a $\mathbf{m}_{K,\rho}$- or $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization. However, the block cross-parameter terms are non-zero for these parameterization choices, as in Figure 4.4k, resulting in the dependence between an K or $1/\kappa$ contrast at $\mathbf{x}_0$ and a ρ or $1/\rho$ contrast at any of the points neighbouring $\mathbf{x}_0$. Hence, there is a dependency between the parameters for all the parameterization choices.

Anyway, our goal is to validate the suggestions of this analysis with the numerical examples later in this paper.

4.6. Almost Well-posed Example

The acoustic inverse problem is non-unique when the source-receiver aperture is limited and/or the necessary frequencies are lacking from the data. It is known that in the presence of non-uniqueness, different choices of parametrization may lead to different inversion results (see, for example, Bharadwaj *et al.* [2014]), each explaining the data at convergence. In other words, the model obtained after the first few iterations depends on the choice of parameterization. Lack of angle-dependent information in the data prevents convergence to the true solution. Hence, the search for the *best* parametrization in terms of convergence speed might be obfuscated by non-uniqueness problems.

Therefore, we define an almost well-posed numerical example by placing 10 sources and 100 receivers evenly on a circle with a 490-m radius all around $\mathbf{x}_0$ to avoid ill-posedness problems related to incomplete illumination. Figure 4.5 shows the setup. The

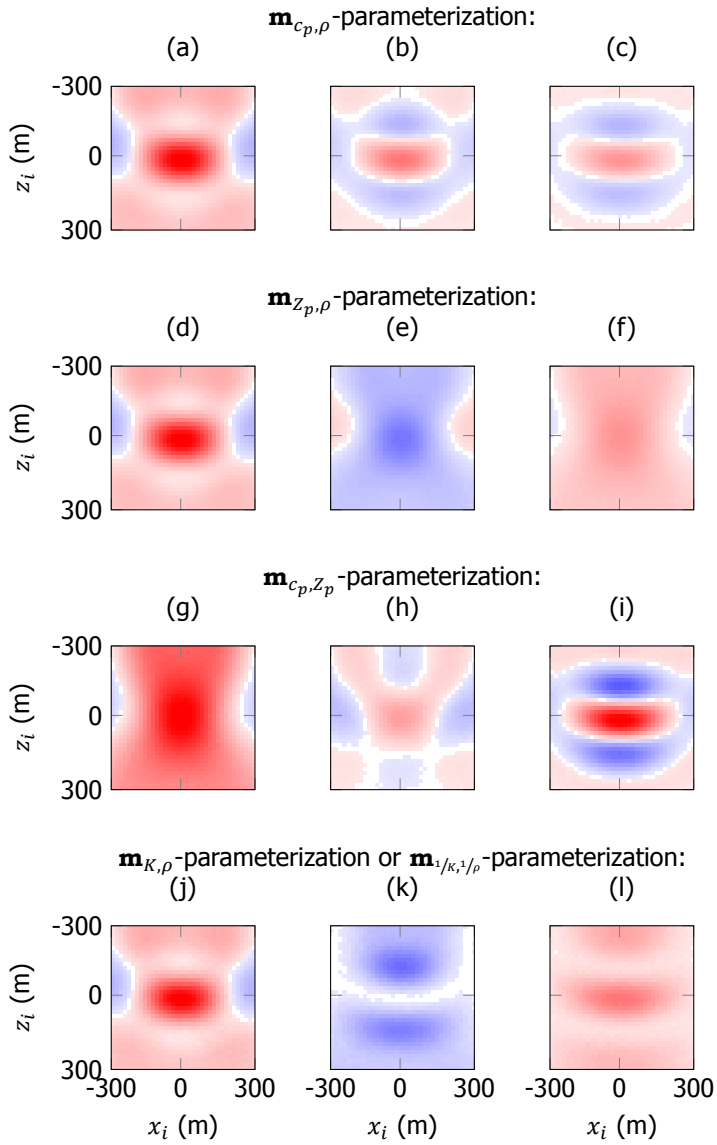


Figure 4.4: The terms of the non-reduced Hessian matrix, $\mathbf{H}_{(a,i),(b,j)}$ such that $\mathbf{x}_j = \mathbf{x}_0$, plotted as a function of $\mathbf{x}_i = (x_i, z_i)$ for the chosen, almost well-posed problem. The modelling and inversion scheme I is adopted. A source is located at (0,−490) and receivers surround $\mathbf{x}_0 = (0, 0)$ along a circle of radius 490 m. In $\mathbf{H}_{(a,i),(b,j)}$, a and b are chosen as: a) $a = c_p$ and $b = c_p$; b) $a = c_p$ and $b = \rho$; c) $a = \rho$ and $b = \rho$; d) $a = Z_p$ and $b = Z_p$; e) $a = Z_p$ and $b = \rho$; f) $a = \rho$ and $b = \rho$; g) $a = c_p$ and $b = c_p$; h) $a = c_p$ and $b = Z_p$; i) $a = Z_p$ and $b = Z_p$; j) $a = K$ and $b = K$ or $a = 1/\kappa$ and $b = 1/\kappa$; k) $a = K$ and $b = \rho$ or $a = 1/\kappa$ and $b = 1/\rho$; l) $a = \rho$ and $b = \rho$ or $a = 1/\rho$ and $b = 1/\rho$. Observe that the block cross-parameter terms of the non-reduced Hessian are non-zero, for any given parameterization choice, showing a dependence between the two parameters. For each parameterization choice, the terms of the reduced Hessian used for the point-scatterer analysis are marked with ⊗. It can be seen that in the case of the $\mathbf{m}_{K,\rho}$- or $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization (j–k), the cross-parameter terms of the reduced Hessian are zero. In this figure, red and blue colours represent positive and negative values, respectively. White color corresponds to zero value.

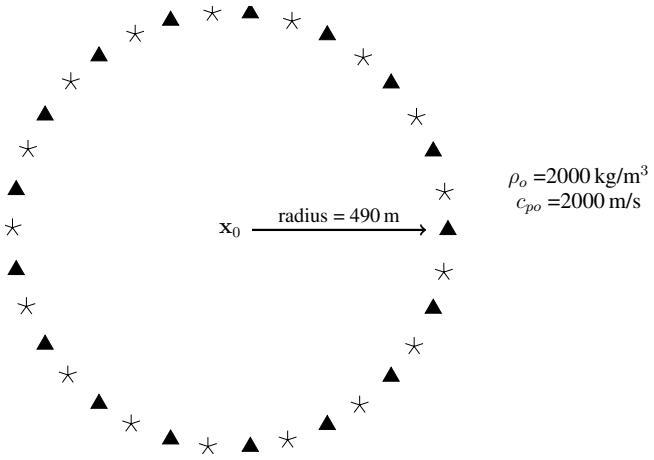


Figure 4.5: A sketch illustrating the almost well-posed example used for parameterization analysis. We place 10 sources, marked by stars, and 100 receivers, marked by triangles, on a circle with a 490-m radius. Different scatterers listed in the Table 4.4 are located at $\mathbf{x}_0$.

starting model for all inversion examples is the same as the background homogeneous model with $c_p = 2000\,\mathrm{m\,s^{-1}}$ and $\rho = 2000\,\mathrm{kg\,m^{-3}}$. The peak frequency of the Ricker source wavelet, which is assumed to be known during inversion, is 5 Hz. The dominant wavelength of the acoustic wavefield used for imaging is 400 m.

If one aims to reconstruct model perturbations of a size comparable to the wavelength that corresponds to the dominant frequency in the data, then the problem may be considered as almost well-posed even with band-limited data. The standard deviation of the extended Gaussian-shaped scatterers is 50 m, corresponding to a size around one-fourth of the dominant wavelength.

For the parameterization analysis, we position one out of the seven different scatterers, listed in Table 4.4, at $\mathbf{x}_0$. We employ the three modelling and inversion schemes to

Table 4.4: Properties of different scatterers located at $\mathbf{x}_0$.

Scatterer Case	Actual Properties (Shape & Contrast)	Shape Known?	Properties to be Estimated
(i)	Point; c_p-only	Yes	Contrast
(ii)	Gaussian-shaped; c_p-only	Yes	Contrast
(iii)	Gaussian-shaped; c_p-only	No	Shape & Contrast
(iv)	Point; ρ-only	Yes	Contrast
(v)	Gaussian-shaped; ρ-only	Yes	Contrast
(vi)	Gaussian-shaped; ρ-only	No	Shape & Contrast
(vii)	Gaussian-shaped; non-K	No	Shape & Contrast

reconstruct the scatterer at $\mathbf{x}_0$ with different subsurface parameterizations (Table 4.1). During inversion, we examine the relative convergence rate of different parameterizations by displaying the least-squares misfit as a function of the iteration count on a log-log scale. We want to observe if

- the suggestions of the point-scatterer analysis and/or radiation pattern analysis can be validated at least by one of the three modelling and inversion schemes used;
- the relative convergence rate of a particular parameterization choice is similar among different inversion schemes;
- the relative convergence rate of a particular parameterization choice depends on the actual contrast of the scatterer — we have chosen scatterers (i)–(iii) with a c_p-only contrast, (iv)–(vi) with a ρ-only contrast and scatterer (vii) such that there is no contrast in K;
- the relative convergence rate of a particular parameterization choice depends on the scatterer properties that are to be estimated — only the contrast of the scatterers (i), (ii), (iv) and (v) has to be estimated, while both the contrast and shape are unknown for other scatterers;
- the relative convergence rate of a particular parameterization choice during reconstructing point-shaped scatterers is different when compared to Gaussian-shaped scatterers — scatterers (i) and (iv) are point-shaped while the others are Gaussian-shaped.

The parameterization choices with the faster and slower convergence are coloured green and orange, respectively, in Table 4.5.

4.6.1. Error-bowl Analysis

Since there are only two unknown variables while reconstructing scatterers (i), (ii), (iv) and (v) with a known shape (see Table 4.4), we could plot two-dimensional logarithmic contours of the least-squares misfit, i.e., contours of $-\log_{10}(J_{ls})$, as a function of the variables. These contours or error bowls can be used to assess the observed relative convergence rate of each parameterization choice. Similar to the analysis using the Hessian matrix, the relative convergence rate for each parameterization can also be assessed based on the (a) ellipticity of the misfit contours; (b) angle between the error vector and one of the principle axis of the contours. We call this an error-bowl analysis.

We are analysing the same simplified inverse problem as that of the point-scatterer analysis, when the error bowls are plotted for the point-shaped scatterers (i) and (iv) after adopting scheme I. Therefore, the error-bowl analysis is the same as the point-scatterer analysis in this case, because the shape and orientation of the error bowls in Figures 4.6 and 4.7 are determined by the reduced Hessians. For example, the error bowls for $\mathbf{m}_{K,\rho}$- or $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization, in Figures 4.6 and 4.7, are circular due to the fact that the condition number of $\mathbb{H}_{1/\kappa,1/\rho}$ or $\mathbb{H}_{K,\rho}$ is one. Furthermore, the error-bowl analysis can be seen as an extension of the point-scatterer analysis, in which

Table 4.5: Reconstruction of seven different scatterers in Table 4.4 when different modelling and inversion schemes are employed. The parameterizations, listed in Table 4.1, with faster and slower convergence relative to each other are in green and orange, respectively.

Scatterer	Scheme I	Scheme II	Scheme III
(i)	(A) (B) (C) (D) (E) Figure 4.6	(A) (B) (C) (D) (E) Figure 4.8	(A) (B) (C) (D) (E) Figure 4.10
(ii)	(A) (B) (C) (D) (E) Figure 4.12	(A) (B) (C) (D) (E) Figure 4.14	(A) (B) (C) (D) (E) Figure 4.16
(iii)	(A) (B) (C) (D) (E) Figure 4.18a	(A) (B) (C) (D) (E) Figure 4.18b	(A) (B) (C) (D) (E) Figure 4.18c
(iv)	(A) (B) (C) (D) (E) Figure 4.7	(A) (B) (C) (D) (E) Figure 4.9	(A) (B) (C) (D) (E) Figure 4.11
(v)	(A) (B) (C) (D) (E) Figure 4.13	(A) (B) (C) (D) (E) Figure 4.15	(A) (B) (C) (D) (E) Figure 4.17
(vi)	(A) (B) (C) (D) (E) Figure 4.19a	(A) (B) (C) (D) (E) Figure 4.19b	(A) (B) (C) (D) (E) Figure 4.19c
(vii)	(A) (B) (C) (D) (E) Figure 4.20a	(A) (B) (C) (D) (E) Figure 4.20b	(A) (B) (C) (D) (E) Figure 4.20c

different modelling and inversion schemes can be employed along with arbitrarily shaped scatterers. It has to be noted that it is cumbersome to derive analytical expressions of the reduced Hessians in the case of arbitrarily shaped scatterers, when full-waveform modelling is employed.

4.6.2. Point-shaped Scatterers

We first consider the inverse problems of reconstructing the point-shaped scatterers in Table 4.4. When scheme I is adopted for modelling and inversion, the error-bowl analysis, identical to the point-scatterer analysis, suggests that $\mathbf{m}_{K,\rho}$- or $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations have the fastest convergence because of their corresponding circular error bowls, as plotted in Figures 4.6c, 4.6e, 4.7c and 4.7. In addition to that, we observe that the bowls for $\mathbf{m}_{c_p,Z_p}$-parameterization are more circular than either the $\mathbf{m}_{c_p,\rho}$- or $\mathbf{m}_{Z_p,\rho}$-parameterization. The least-squares misfit plotted against the iteration count while estimating the c_p-only contrast, in Figure 4.6f, and the ρ-only contrast, in Figure 4.7f, shows that the $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations have faster convergence than the others.

We now employ the modelling and inversion scheme II. Figures 4.8 and 4.9 show that the ellipticity of the error bowls for all the parameterizations, except for the $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization, are different from the previous case. Note that the $\mathbf{m}_{K,\rho}$-parameterization is no longer equivalent to the $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization because of the non-linear re-parameterization. However, similar to the previous case with the scheme I, we observe that the $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations converge faster.

Figures 4.10 and 4.11 display the error bowls when modelling and inversion with

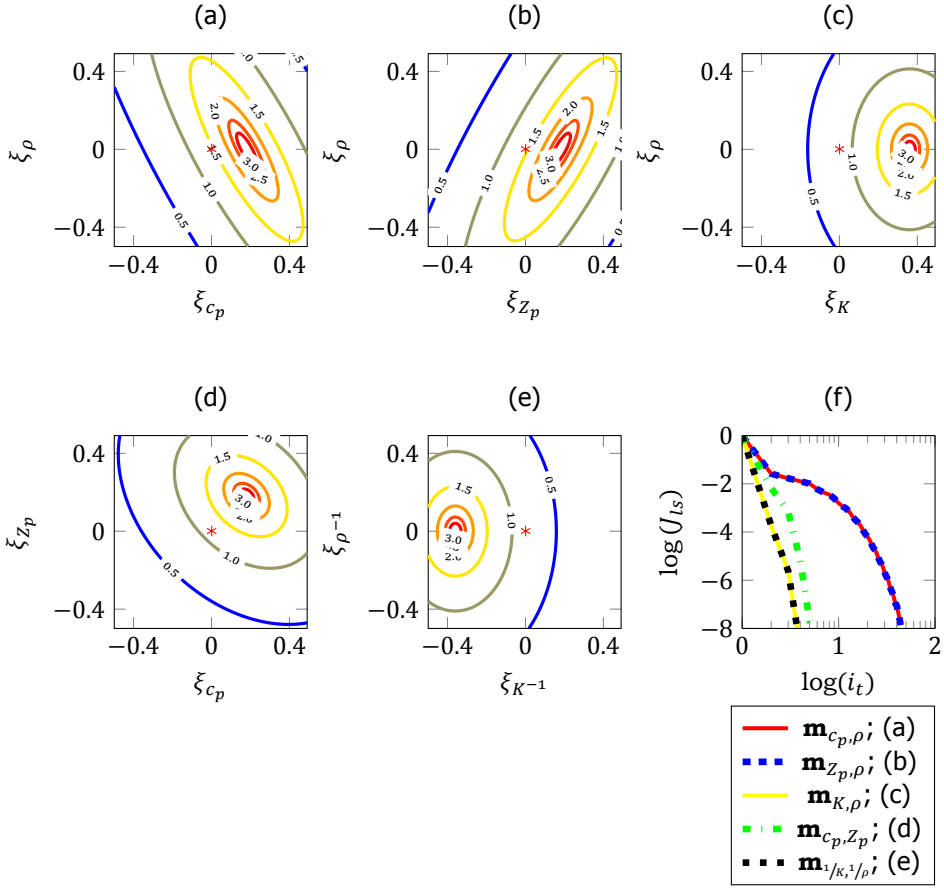


Figure 4.6: Reconstruction of the point-shaped c_p-only scatterer (i) when the modelling and inversion scheme I, with the Born approximation and linear re-parameterization, is adopted. Error-bowl analysis is performed by plotting the logarithmic contours of the least-squares misfit as a function of the two subsurface parameters while using the following parameterization choices: a) $\mathbf{m}_{c_p,\rho}$— slower convergence expected due to high ellipticity; b) $\mathbf{m}_{Z_p,\rho}$— slower convergence expected due to high ellipticity; c) $\mathbf{m}_{K,\rho}$—expect faster convergence due to circular contours; d) $\mathbf{m}_{c_p,Z_p}$—expect faster convergence due to almost circular contours; e) $\mathbf{m}_{1/K,1/\rho}$— expect faster convergence due to circular contours. In all the plots, the starting homogeneous model, $(0,0)^{\mathrm{T}}$, is marked by the red star. f) The least-squares misfit is plotted as a function of the iteration count on a log-log scale. This plot shows that the suggestions of the point-scatterer analysis are valid in this case.

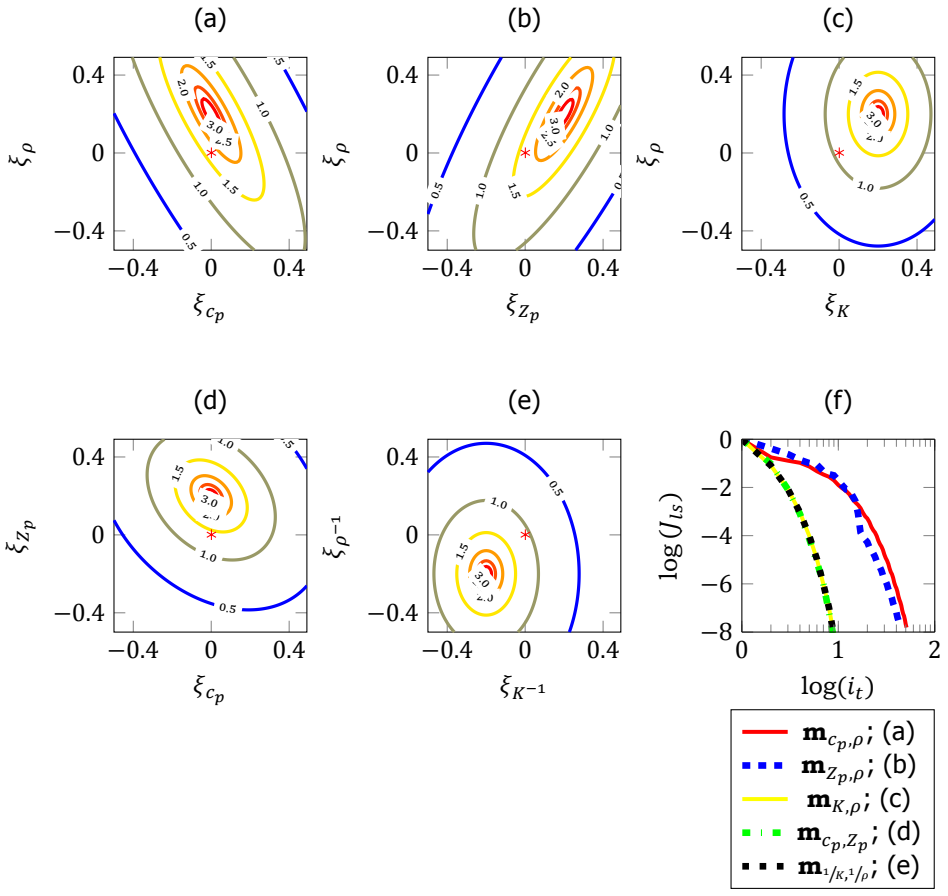


Figure 4.7: Reconstruction of the point-shaped ρ-only scatterer (iv) when the modelling and inversion scheme I, with the Born approximation and linear re-parameterization, is adopted. Error-bowl analysis is performed by plotting the logarithmic contours of the least-squares misfit as a function of the two subsurface parameters while using the following parameterization choices: a) $\mathbf{m}_{c_p,\rho}$— slower convergence expected because of high ellipticity; b) $\mathbf{m}_{Z_p,\rho}$— slower convergence expected because of high ellipticity; c) $\mathbf{m}_{K,\rho}$— faster convergence expected because of circular contours; d) $\mathbf{m}_{c_p,Z_p}$— faster convergence expected because of almost circular contours; e) $\mathbf{m}_{1/K,1/\rho}$— faster convergence expected because of circular contours. In all the plots, the starting homogeneous model, $(0,0)^{\mathrm{T}}$, is marked by the red star. f) The least-squares misfit is plotted as a function of the iteration count on a log-log scale. This plot shows that the suggestions of the point-scatterer analysis are valid in this case.

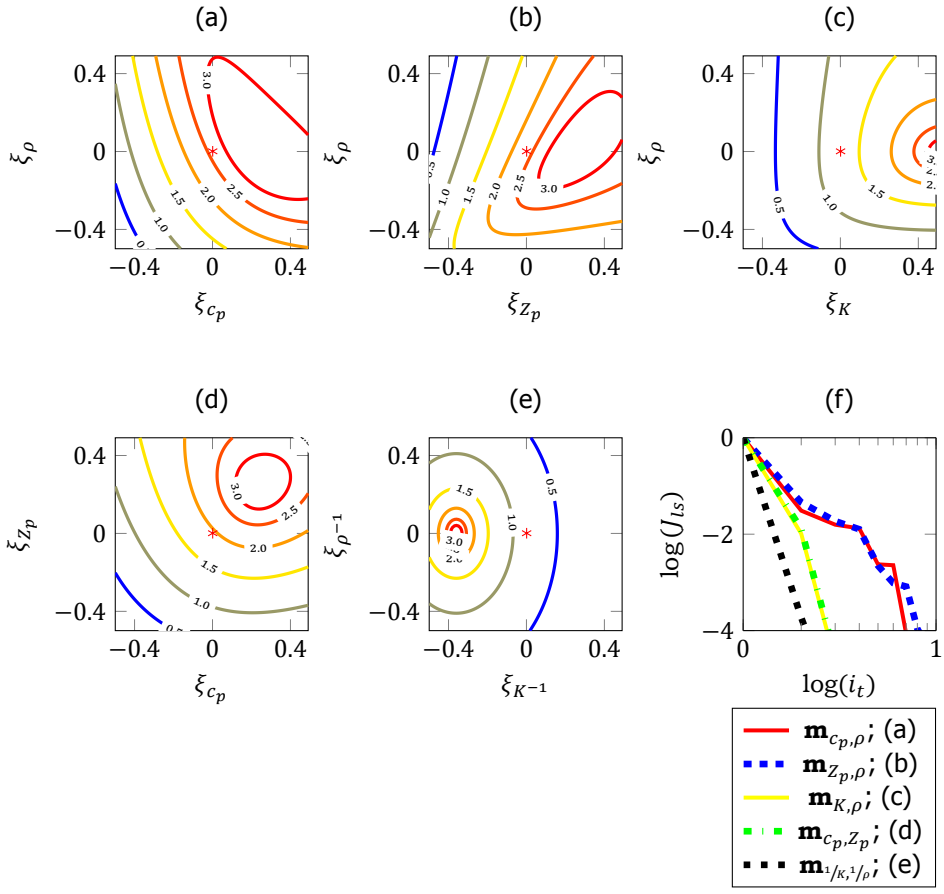


Figure 4.8: Same as Figure 4.6, except for adopting the modelling and inversion scheme II, with the Born approximation and non-linear re-parameterization. It can be seen that the error bowls for the $\mathbf{m}_{K,\rho}$-parameterization are not equivalent to the $\mathbf{m}_{1/K,1/\rho}$-parameterization because of the non-linear re-parameterization. Also, the shapes of the error bowls (a)–(d) are different compared to that of Figure 4.6. It can be noted that the error bowls (c)–(e) are more *circular* than error bowls (a)–(b). The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

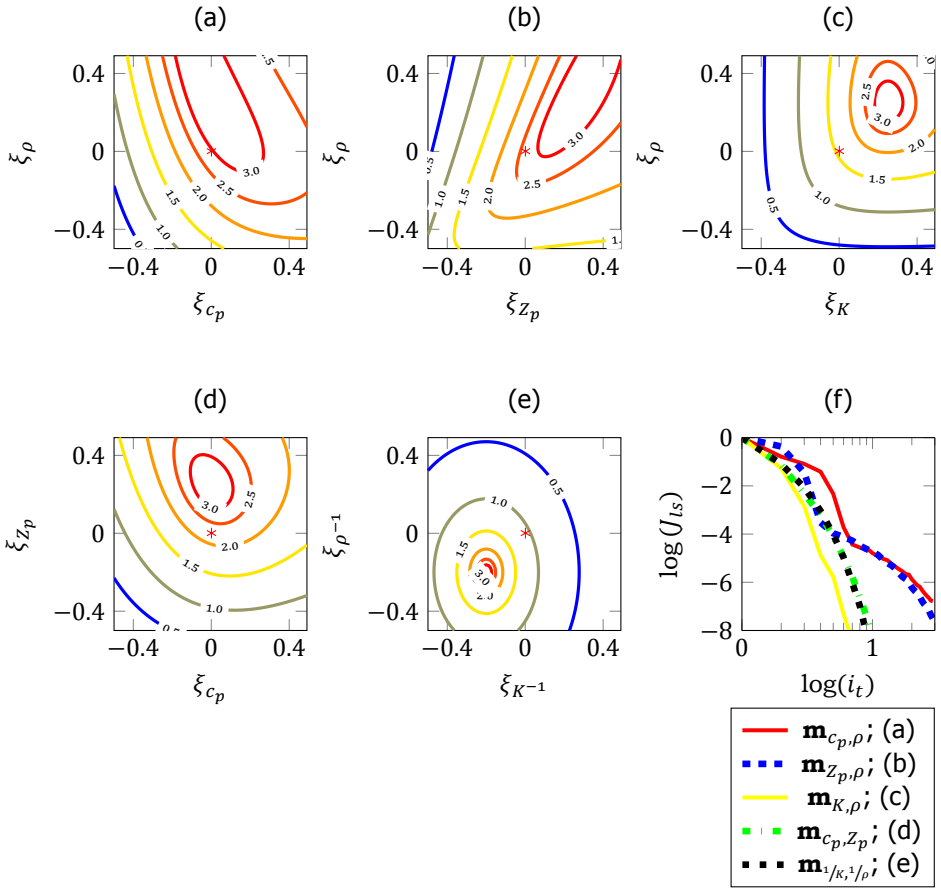


Figure 4.9: Same as Figure 4.7, except for adopting scheme II, with the Born approximation and non-linear re-parameterization. The least-squares misfit plot, similar to that of Figure 4.8, shows that the suggestions of the point-scatterer analysis are valid.

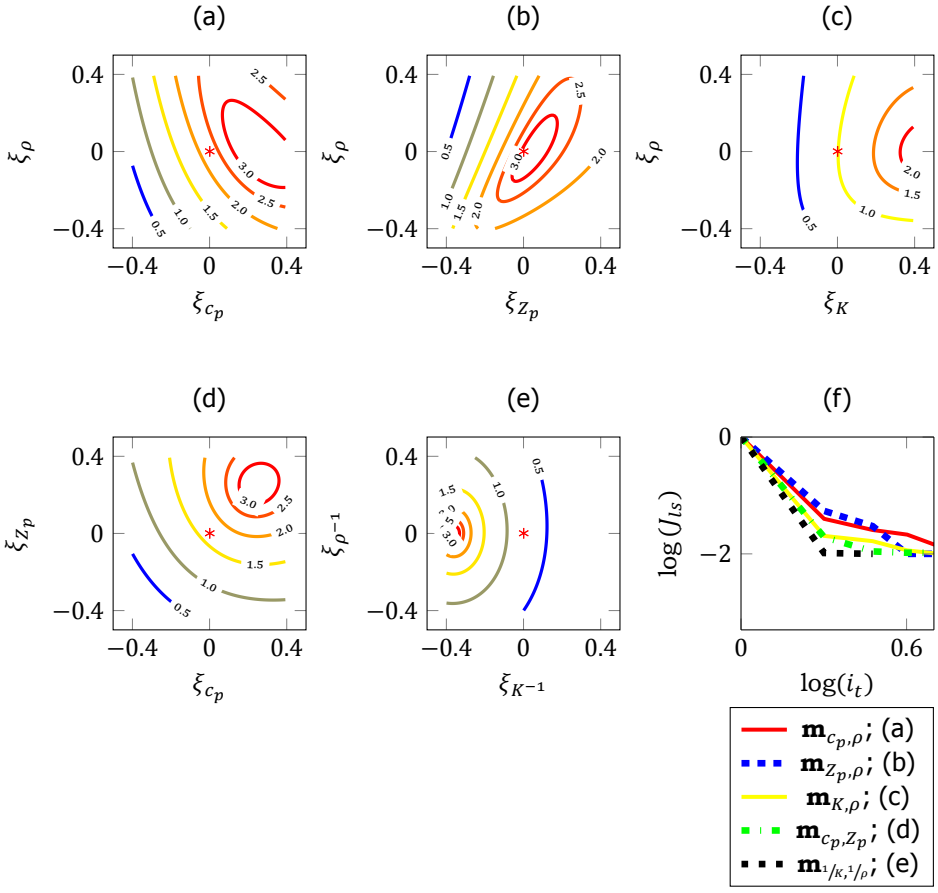


Figure 4.10: Same as Figure 4.6, except for adopting scheme III, with full-waveform modelling and inversion. We see that the shapes of the error bowls and the relative convergence rates of different parameterization choices are similar to that of Figure 4.8. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

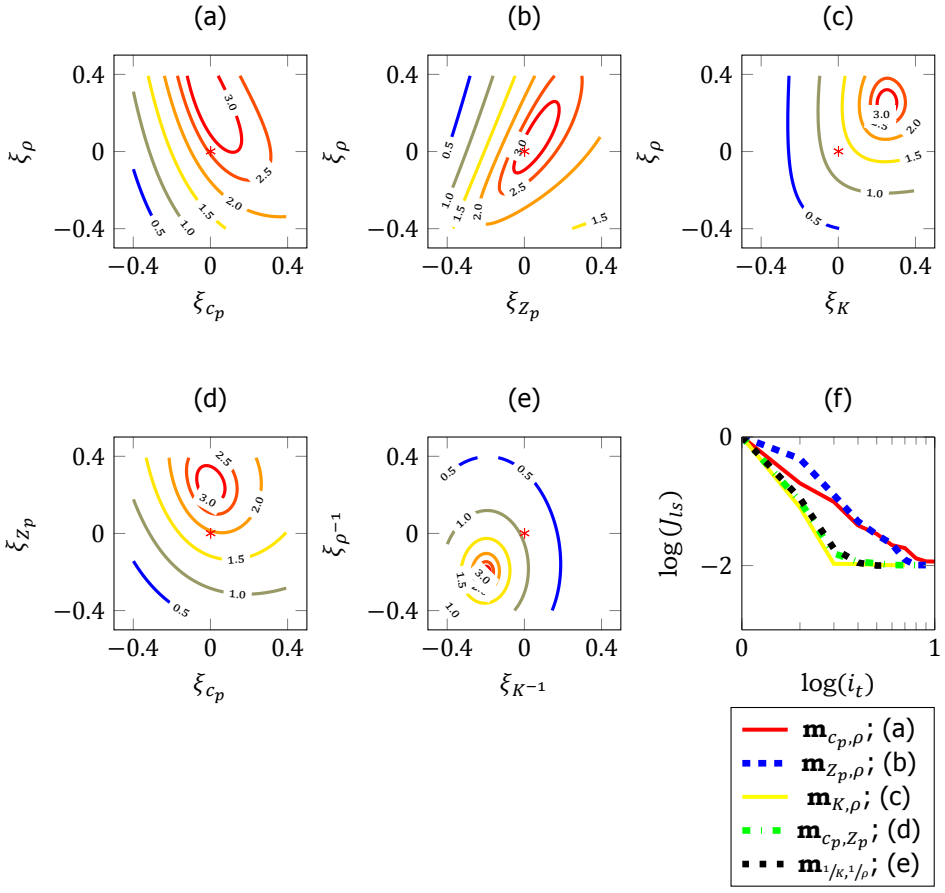


Figure 4.11: Same as Figure 4.7, except for adopting scheme III, with full-waveform modelling and inversion. It can be seen that the shapes of the error bowls and the relative convergence rates of different parameterization choices are similar to that of Figure 4.9. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid.

scheme III is employed. It can be seen that the suggestions of the point-scatterer analysis remain valid even when employing the modelling and inversion scheme II or III. This is because the error bowls (c)–(e) are more *circular* or less *elliptical* than the error bowls (a)–(b) in the Figures 4.8, 4.9, 4.10 and 4.11.

4.6.3. Gaussian-shaped Scatterers

We now want to see if the sub-wavelength Gaussian-shaped scatterers at $\mathbf{x}_0$ will result in error bowls of a different shape compared to the point-shaped scatterers of the previous subsection. When the modelling and inversion scheme I is employed, both while reconstructing the scatterer (ii) with the c_p-only contrast, in Figure 4.12, and the scatterer (v) with the ρ-only contrast, in Figure 4.13, the error bowls are more elliptical for the $\mathbf{m}_{c_p,\rho}$-, $\mathbf{m}_{Z_p,\rho}$- and $\mathbf{m}_{c_p,Z_p}$-parameterizations than those with point-shaped scatterers. Moreover, the error bowls for the $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations are not circular anymore. We still observe that the $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations have a faster convergence because their corresponding error bowls are more circular.

When either modelling and inversion scheme II or III is employed, the error bowls for all parameterizations not only have different ellipticities but also different orientations. This indicates that the suggestions of the point-scatterer analysis are no longer valid. We observe that the $\mathbf{m}_{c_p,\rho}$-parameterization is faster in reconstructing a c_p-only contrast (Figure 4.14f and 4.16f) and that the $\mathbf{m}_{1/\kappa,1/\rho}$-parameterization has the fastest convergence when reconstructing a ρ-only contrast (Figures 4.15f and 4.17f). Hence, when modelling and inversion schemes II and III are employed to reconstruct non-point shaped scatterers, the rate of convergence of a particular parameterization also depends on the type of contrast that has to be reconstructed.

4.6.4. Scatterers with Unknown Shape and Contrast

We now consider the inverse problems of estimating both the shape and contrast of the sub-wavelength Gaussian scatterers (iii), (vi) and (vii). As shown in Figures 4.18, 4.19 and 4.20, the $\mathbf{m}_{c_p,\rho}$-parameterization is the fastest while estimating a c_p-only scatterer, whereas the $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/\kappa,1/\rho}$-parameterizations have a faster convergence while estimating a ρ-only or a non-K scatterer. This observation is similar to that of estimating the contrast of the Gaussian-shaped scatterers in the last subsection.

4.7. Conclusions

We have briefly outlined the conventional point-scatterer and the diffraction-pattern parameterization-analysis methods for the 2-D acoustic inverse problem. Using almost well-posed-numerical examples, we have shown that the suggestions of these conventional methods are valid only when the contrasts of the point-shaped scatterers at a known location are estimated. The numerical examples employ three different modelling and inversion schemes using both Born and full-waveform modelling. As expected, for almost well-posed inverse problems, we observed that a change in parametrization will result in a different convergence rate. Furthermore, the relative rate of convergence for a particular choice of parameterization depends on (a) the modelling and inversion

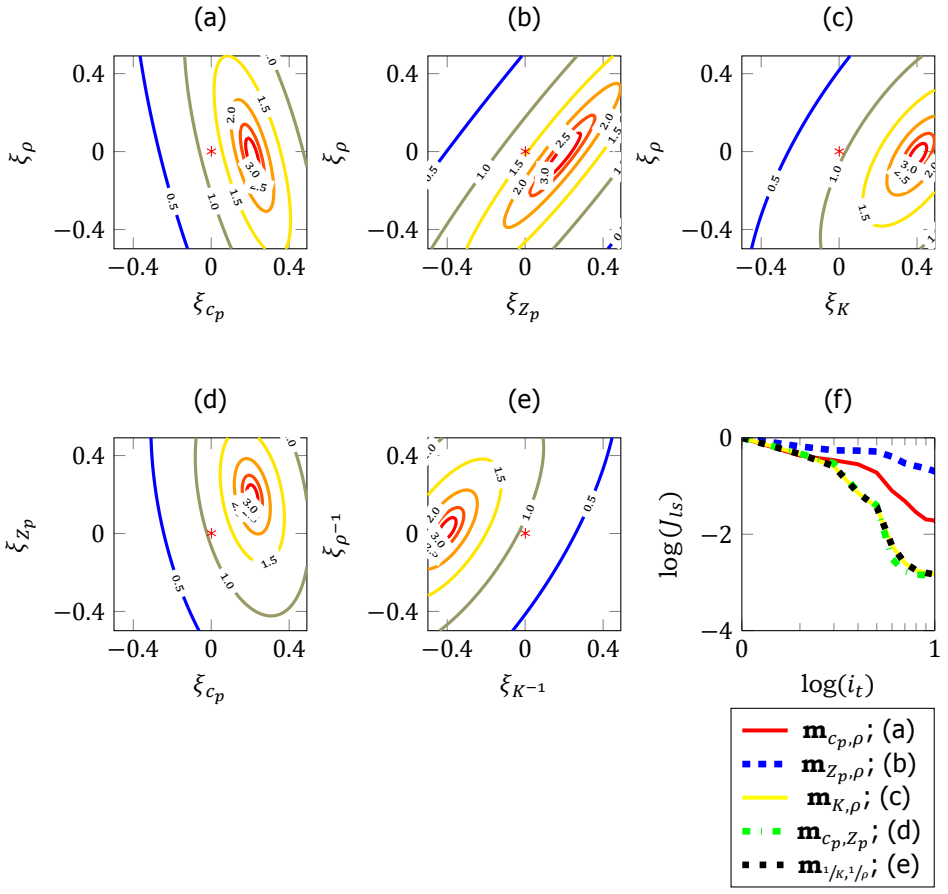


Figure 4.12: Same as Figure 4.6, except for the Gaussian-shaped c_p-only scatterer (ii). We observe that the errors bowls for all the parameterization choices are more elliptical compared to that of Figure 4.6. The least-squares misfit plot shows that the suggestion of the point-scatterer analysis is valid. The contours in the case $\mathbf{m}_{1/K,1/\rho}$- and $\mathbf{m}_{K,\rho}$-parameterizations are not circular. The rate of convergence in the case of $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/K,1/\rho}$-parameterizations is higher as their corresponding error bowls are more circular compared to the rest.

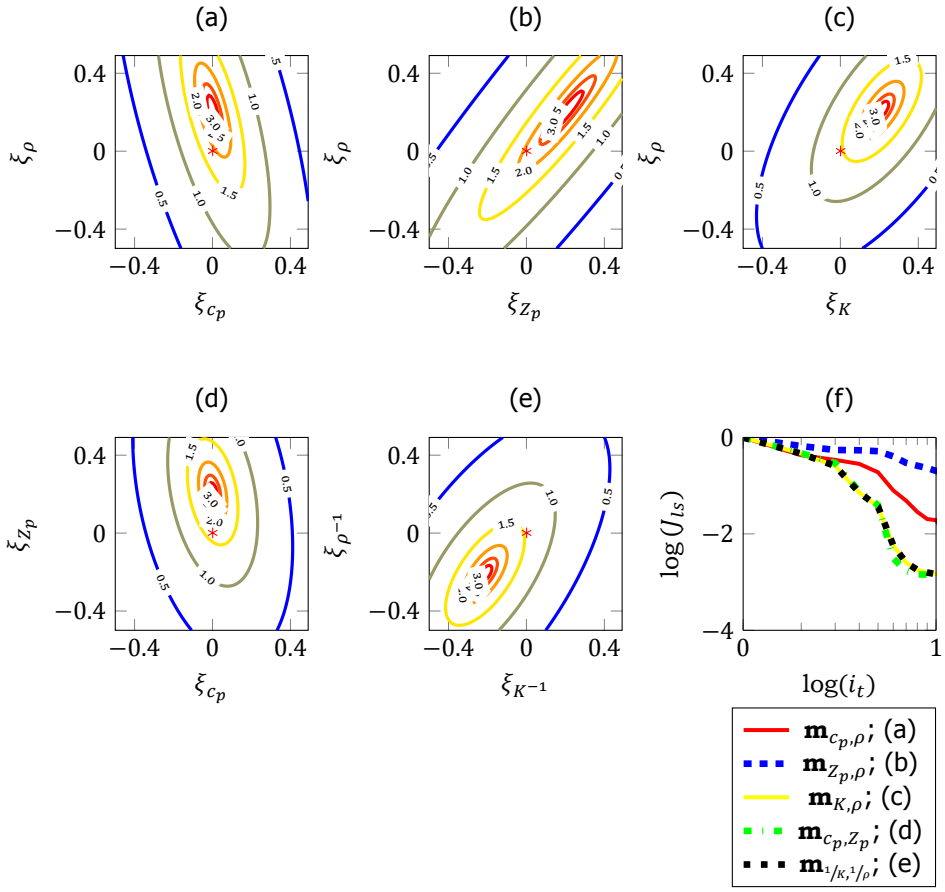


Figure 4.13: Same as Figure 4.7, except for the Gaussian-shaped ρ-only scatterer (v). We observe that the errors bowls for all the parameterization choices are more elliptical compared to that of Figure 4.7. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are valid. The rate of convergence in the case of $\mathbf{m}_{c_p,Z_p}$-, $\mathbf{m}_{K,\rho}$- and $\mathbf{m}_{1/K,1/\rho}$-parameterizations is higher as their corresponding error bowls are more circular compared to the rest.

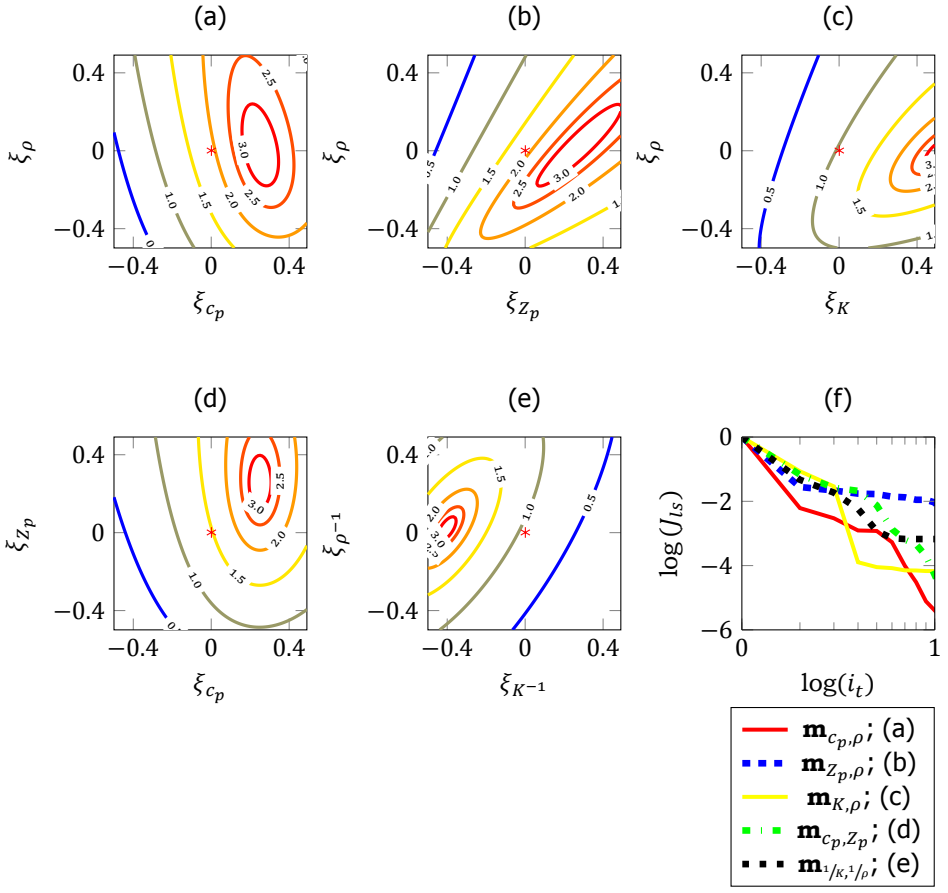


Figure 4.14: Same as Figure 4.6, except for scatterer (ii) and adopting scheme II, with the Born approximation and non-linear reparameterization. Compared to the Figure 4.12, we see that the error bowls, in this case, have much different orientation. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are no longer valid.

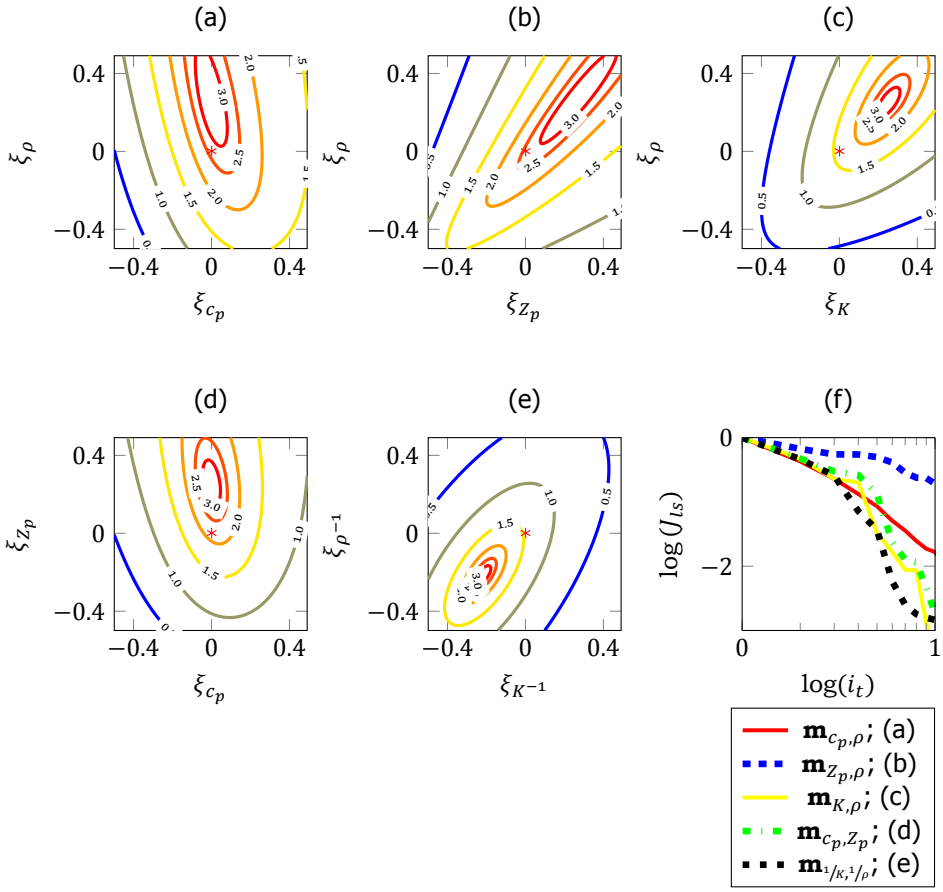


Figure 4.15: Same as Figure 4.7, except for scatterer (v) and adopting scheme II, with the Born approximation and non-linear reparameterization. Compared to the Figure 4.13, we see that the error bowls in this case have different ellipticity and orientation.

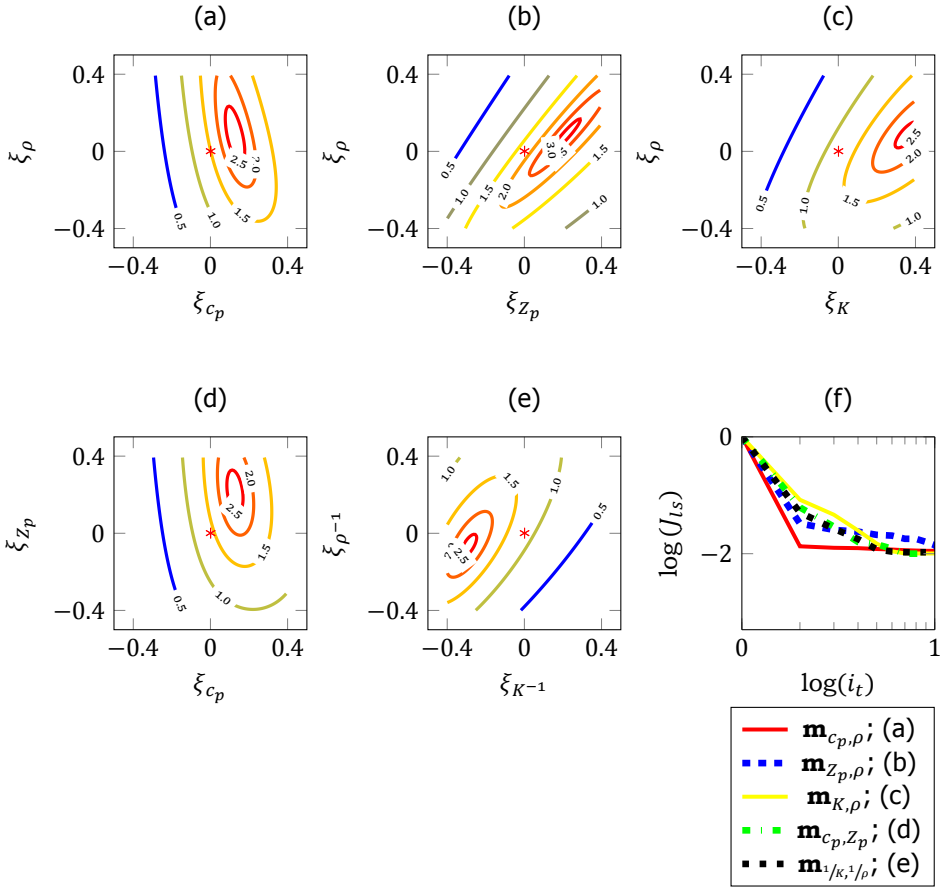


Figure 4.16: Same as Figure 4.6, except for scatterer (ii) and adopting scheme III, with full-wavefrom modelling and inversion. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are no longer valid. Parameterization using $\mathbf{m}_{c_p,\rho}$ has the best convergence rate.

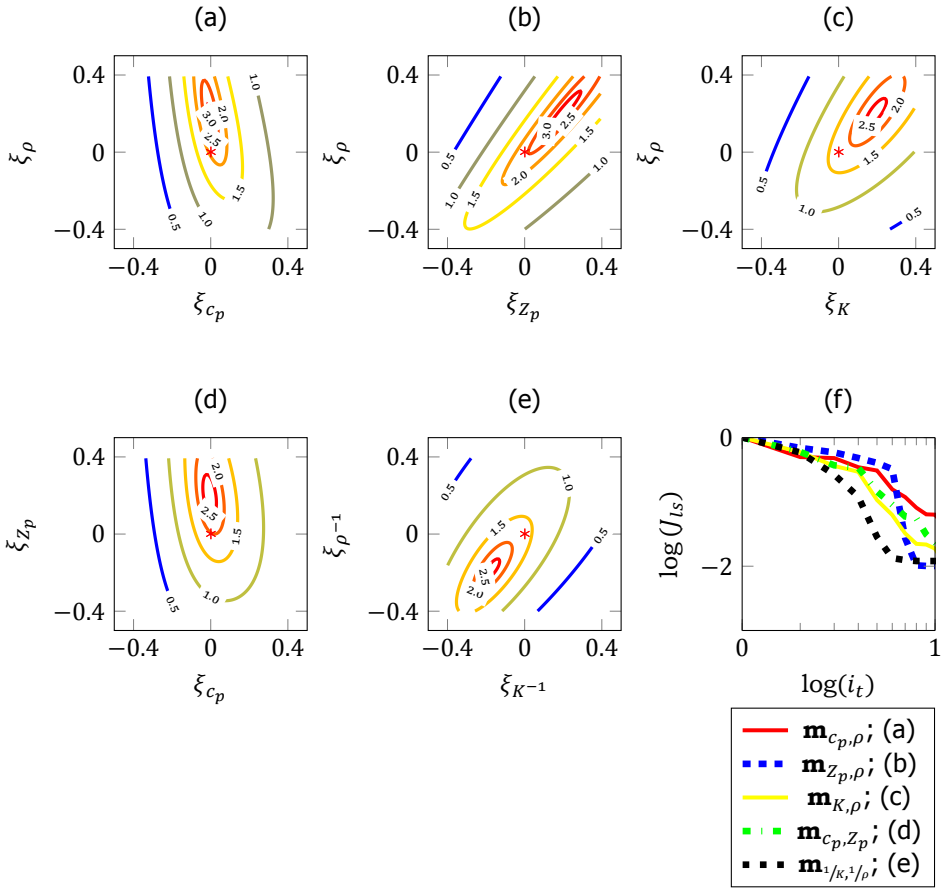


Figure 4.17: Same as Figure 4.7, except for scatterer (v) and adopting scheme III, with full-wavefrom modelling and inversion. The least-squares misfit plot shows that the suggestions of the point-scatterer analysis are no longer valid.

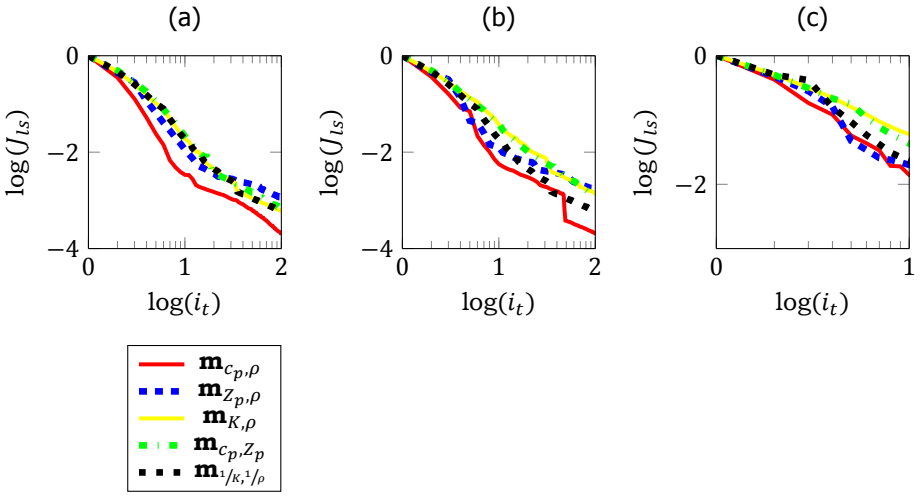


Figure 4.18: The least-squares misfit is plotted as a function of the iteration count on a log-log scale during the reconstruction of scatterer (iii), where both shape and contrast of a c_p-only scatterer are unknown. The modelling and inversion schemes (a) I, (b) II and (c) III are adopted. It can be observed that the parameterization using $\mathbf{m}_{K,\rho}$ has the worst convergence rate in all the cases.

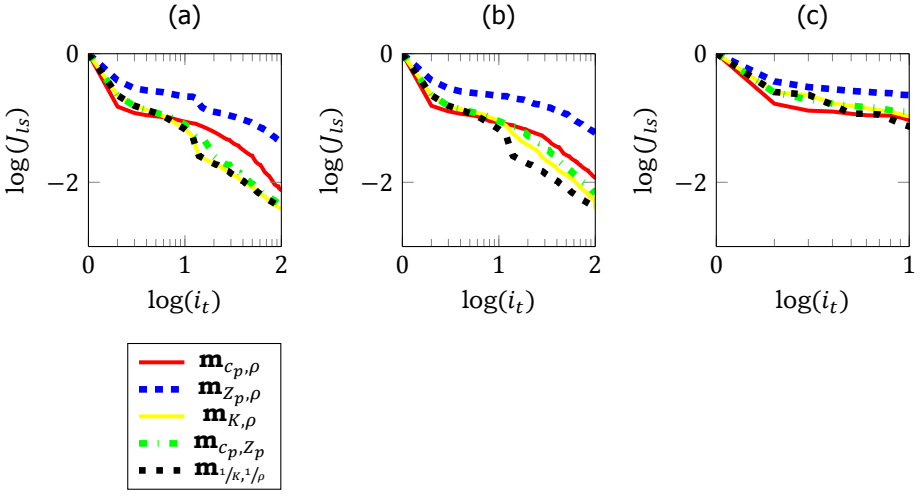


Figure 4.19: Same as Figure 4.18, except for scatterer (vi), with ρ-only contrast. Note that the relative rates of convergence of various parameterization choices are different compared to that of the Figure 4.18. It can be observed that the parameterization using $\mathbf{m}_{Z_p,\rho}$ has the worst convergence rate in all the cases.

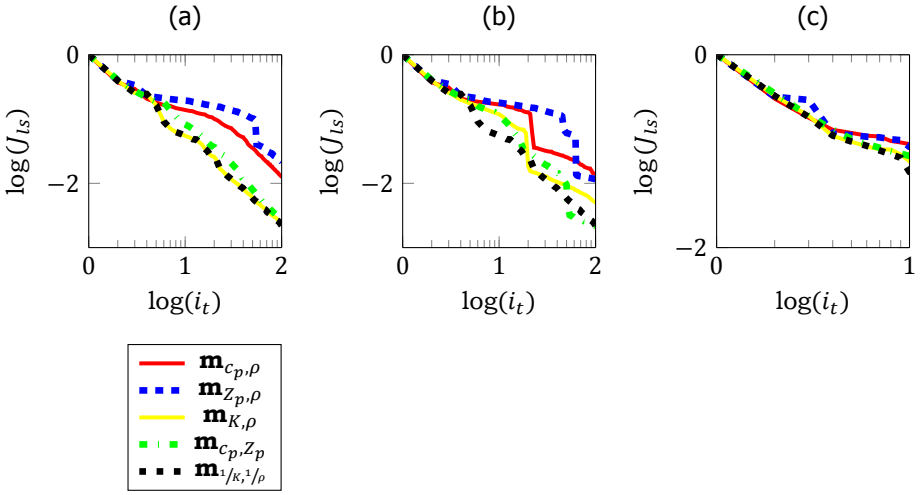


Figure 4.20: Same as Figure 4.18, except for scatterer (vii), with a non-K contrast. Note that relative rates of convergence of various parameterization choices are different compared to that of the Figures 4.18 and 4.19. It can be observed that the parameterization using $\mathbf{m}_{Z_p,\rho}$ has the worst convergence rate in all the cases.

scheme employed; (b) the contrast of subsurface scatterer that has to be reconstructed; (c) the shape of the sub-wavelength scatterer. We observed that in most of the cases the $\mathbf{m}_{Z_p,\rho}$-parameterization did not have the fastest convergence. Finally, our numerical examples show that, in general, there is no such thing as the best parametrization choice that provides the fastest convergence for acoustic inversion.

References

A. Tarantola, *strategy for nonlinear elastic inversion of seismic reflection data,* Geophysics **51**, 1893 (1986).

J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics,* Geophysics **74**, WCC1 (2009).

R. G. Pratt, C.Shin, and G. Hicks, *Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion,* Geophysical Journal International **133**, 341 (1998).

S. Operto, Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, *A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice,* The Leading Edge **32**, 1040 (2013).

V. Prieux, R. Brossier, S. Operto, and J. Virieux, *Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: imaging compressional wave speed, density and attenuation,* Geophysical Journal International **194**, 1640 (2013).

R. H. Byrd, P. Lu, J. Nocedal, and C. Y. Zhu, *A limited memory algorithm for bound constrained optimization,* SIAM Journal on Scientific Computing **16**, 1190 (1995).

R. Modrak, J. Tromp, and Y. Yuan, *On the choice of material parameters for elastic waveform inversion,* in *SEG Technical Program Expanded Abstracts 2016* (Society of Exploration Geophysicists, 2016) pp. 1115–1119.

A. Fichtner and J. Trampert, *Hessian kernels of seismic data functionals based upon adjoint techniques,* Geophysical Journal International **185**, 775 (2011).

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

R. Wu and K. Aki, *Scattering characteristics of elastic waves by an elastic heterogeneity,* Geophysics **50**, 582 (1985), http://library.seg.org/doi/pdf/10.1190/1.1441934 .

M. Malinowski, S. Operto, and A. Ribodetti, *High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion,* Geophysical Journal International **186**, 1179 (2011).

Y. Gholami, R. Brossier, S. Operto, A. Ribodetti, and J. Virieux, *Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? Part 1: Sensitivity and trade-off analysis,* Geophysics **78**, R81 (2013), http://library.seg.org/doi/pdf/10.1190/geo2012-0204.1 .

W. He and R.-É. Plessix, *Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic Earth with a vertical axis of symmetry,* Geophysical Prospecting , 1 (2016).

E. Forgues and G. Lambaré, *Parameterization study for acoustic and elastic ray plus born inversion,* Journal of Seismic Exploration **6**, 253 (1997).

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *Parametrization for 2-D SH full waveform inversion,* in *76th EAGE Conference and Exhibition 2014* (2014).

5

Near-surface Application of 2-D SH Full Waveform Inversion

We are what our thoughts have made us;
so take care about what you think. Words are secondary.
Thoughts live; they travel far.

Swami Vivekananda

In the near-surface with unconsolidated soils, shear-wave properties can often be imaged better and with a higher resolution than P-wave properties. To enable imaging ahead of a tunnel-boring machine (TBM), we developed a seismic prediction system with a few shear-wave vibrators and horizontal receivers. The boring process is interrupted at regular intervals to carry out active surveys. The vibrators are then pushed against the rock or soil in front of the machine's cutting wheel. Their design is based on linear synchronous motor technology that can generate very low frequencies. They inject a force in a direction perpendicular to the tunnel axis. Horizontal receivers measure the particle velocity, mainly due to the horizontal polarized shear (SH) waves. Because imaging with conventional migration methods suffers from artefacts, caused by the incomplete aperture and inaccuracies in the assumed velocity model, we use 2-D SH full waveform inversion (FWI) to image the subsurface shear properties. The classic cycle-skipping problem, which can make the application of FWI cumbersome, is avoided by the capability to generate low frequencies. In this paper, we demonstrate the capabilities of the proposed seismic system by a number of synthetic and field experiments.

Parts of this chapter have been accepted for publication in Near Surface Geophysics.

5.1. Introduction

While excavating a tunnel with a tunnel-boring machine (TBM), the geology and the ground conditions along the planned-tunnel trajectory need to be investigated in order to safely and efficiently carry out underground operations. This entails detecting the occurrence of faults, boulders, foundations, pipes, etc., necessary to avoid hazards that can cause time-consuming delays in the tunnel-boring operations. In order to predict ground conditions ahead of a TBM, seismic exploration techniques such as data acquisition, processing and inversion can be deployed. We will discuss how each of those can be applied to tunnel excavation with a TBM.

Seismic-ground-prediction systems for TBM record waves generated by either controlled sources or its rotating cutter wheel. Hauser [2001] and Petronio and Poletto [2002] obtained interpretable seismic data by cross-correlating signals generated by the cutter wheel with pilot signals recorded at reference receivers. Alternatively, seismic interferometry uses cross-correlation to turn the noise generated by the cutter wheel into useful virtual source records [Poletto and Petronio, 2006; Harmankaya *et al.*, 2016]. However, artefacts are present in the virtual source records since the location of the sources generating the noise is limited to the cutter wheel. Because of this, ground-prediction systems that employ active sources are more successful. In these systems, acquisition is usually carried out with a focus on a particular propagation mode of the seismic waves. The use of Rayleigh waves was proposed by Bohlen *et al.* [2007] and Jetschny [2010], considering a system that excites and records tunnel surface waves at the tunnel wall behind the cutter head of the TBM. With 3-D elastic modelling, they showed that the high-amplitude Rayleigh waves are converted into high-amplitude shear waves at the front face of the tunnel and vice-versa. P waves are commonly used in hydrocarbon exploration. Kneib *et al.* [2000] describes a seismic system for use in soft soil, which uses P waves from 1.8 to 6 kHz. These are higher compared to the 10- to 120-Hz frequencies commonly used in surface-seismic techniques. The advantage of using shear or S body waves has been demonstrated by several authors in the case of soft-soil near-surface applications [Omnes, 1978; Helbig and Mesdag, 1982; Stümpel *et al.*, 1984; Guy *et al.*, 2003; Haines and Ellefsen, 2010]. S waves turn out to be very suitable for soft soils since shear waves are not sensitive to the type of fluid or gas in the pores. Hence, estimated shear-wave properties using shear waves correlate well with subsurface lithology. In these soils, propagating shear waves often have a shorter wavelength than P waves [Ghose *et al.*, 1998; Haines and Ellefsen, 2010; Miller *et al.*, 2001], resulting in a better resolution while imaging. Also, in the near surface, where the soft-soil TBM usually operates, relative shear-wave variations are much larger than relative P-wave variations.

Recorded data are processed with the aim to obtain the subsurface parameters that control the seismic wave-propagation. A *reflectivity* image of the subsurface, which depicts the interfaces between different soil types, can be produced by using a subsurface wave-speed or velocity model. The conventional methods of estimating the velocity model directly from seismic data are not fully automatic and require time-consuming human interaction, commonly taking several days to obtain the final images. In tunnel-boring operations, this time is not available: results need to be available within an hour

or even minutes, to allow for preventive action when obstacles or potentially dangerous situations ahead of the TBM show up.

Most of the current systems for seismic exploration produce reflectivity images using an assumed velocity model instead of an estimated one. Swinnen *et al.* [2007] discusses an imaging technique based on focusing operators in an assumed model. Tzavaras [2010] applies Kirchhoff pre-stack depth migration and Fresnel-volume migration to produce 3-D reflectivity images in the case of hard-rock tunnelling. Ashida [2001] describes a method to detect the interfaces using data from multi-component receivers. These conventional near-surface imaging techniques all use an assumed velocity model and suffer from various pitfalls [Steeples and Miller, 1998]. When using shear waves, Miller *et al.* [2001] has shown that Love waves can stack coherently in a common midpoint gather, leading to a wrong interpretation. Incomplete acquisition, due to the limited space available on the TBM, causes recording footprint noise in conventional images. Inaccuracies in the assumed velocity model will result in migration images that are not properly focused. Therefore, in tunnelling applications, there is a need for a seismic system that can automatically estimate the wave-velocity prior to imaging. Bellino *et al.* [2013] proposes such a fully automatic method, which can estimate the average wave-velocity as well as the distance to an interface.

Recently, an approach called full-waveform inversion (FWI) [Tarantola, 1986; Virieux and Operto, 2009] has been used to automatically produce a subsurface velocity model for tunnel exploration [Musayev *et al.*, 2013; Bharadwaj *et al.*, 2015]. FWI is a non-linear data fitting procedure that minimises the misfit between the recorded and the modelled seismic data, in a least-squares sense, to estimate the subsurface parameters. It requires low frequencies in the data to avoid convergence to local minima caused by the notorious cycle-skipping problem [Virieux and Operto, 2009]. The advantage of using FWI over conventional imaging techniques is that the least-squares imaging condition used in FWI will suppress some of the acquisition-related artefacts [Nemeth *et al.*, 1999].

The application of waveform inversion to near-surface land data is even more challenging than the marine case because of strong elastic effects such as ground-roll and near-surface attenuation. It is difficult to fit the surface waves due to a heterogeneous near-surface during inversion. In the tunnel environment, the surface waves propagate along the tunnel wall. In order to properly account for the surface waves, elastic modelling is commonly used for near-surface FWI [Bretaudeau *et al.*, 2013]. Even when using elastic modelling, Brossier *et al.* [2009] illustrated that near-surface FWI generates accurate results via judicious data pre-conditioning and/or muting rather than using surface waves. Moreover, using elastic modelling makes the inverse problem computationally more expensive.

In this chapter, we focus on a system that uses horizontally polarised shear (SH) waves [Bharadwaj *et al.*, 2015], because shear waves are better suited for exploration in soft soils than P waves, as discussed earlier. Our objective is to investigate and demonstrate the feasibility of using SH waves in unconsolidated soils for TBM-like situations and geometries. For data acquisition in this system, shear vibrators and receivers are placed on the soil in front of the cutter head to generate and record mainly the SH wavefield. Acquisition is carried out when the tunnel-boring machine (TBM) is not in

operation. The design of the vibrator is based on linear synchronous motors technology [Noorlandt *et al.*, 2015], which can expand the source frequency band to frequencies as low as 5 Hz. For inverting the data, we use full-waveform inversion to estimate the subsurface shear-wave speed and reflectivity. The capabality of the seismic vibrator to generate low frequencies allows us to circumvent the classic problem of cycle-skipping. Of course, this assumes that these frequencies can actually be injected into the subsurface, which may not always be the case. The inability to inject low frequencies is illustrated by an acoustic model with a linear vertical velocity gradient in which waves below a certain frequency do not propagate [Kuvshinov and Mulder, 2006].

Since the acquired data need to be processed in near real time with current computing technology, we simplified the SH full-waveform inversion problem to 2D. We applied a crude but simple correction to the measured 3-D data to make them resemble 2-D data. The 2-D approach implicitly assumes invariance in the out-of-plane direction. In that case, the SH waves are decoupled from P, SV, and Rayleigh waves and we can simplify the elastic wave equation to a 2-D SH wave-equation. Due to the absence of Rayleigh waves, the observed data are easier to fit compared to P-wave land datasets. However, Love waves that might be present in the observed data are also modelled by solving the 2-D SH wave equation [Luo *et al.*, 2010]. It should be noted that Love waves are guided waves that are only generated in the presence of a low S-wave velocity layer close to the sources and receivers. The occurrence of such a layer will be rare as subsurface layering is often nearly horizontal and perpendicular to cutter-head front.

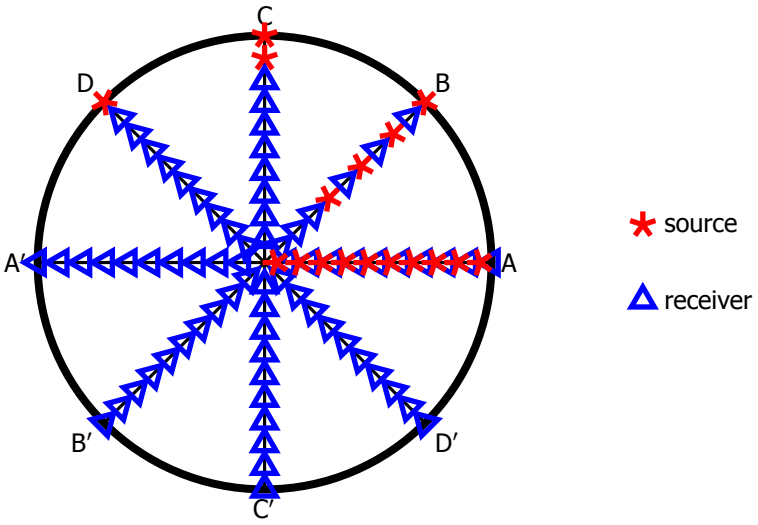
The remainder of the chapter is organised into five sections. The next section describes the data acquisition and pre-processing of our system in more detail. After that, we elaborate on the FWI algorithm used to process the data. In the fourth and fifth section, we demonstrate the application of the seismic system using synthetic scenarios as well as field data, respectively. The last section summarises the chapter.

5.2. Data Acquisition and Pre-processing

Our system places sources and receivers on the soil along a diameter of the cutter head to generate and record seismic shear waves. Figure 5.1 shows four possible source-receiver geometries along diameters AA′, BB′, CC′ and DD′, respectively. Geometry A is impractical but best illuminates the target. The other geometries use only a few source and receiver positions because of the limited space on a TBM. In the case of a 10-m TBM, geometries B, C and D also take the following practical considerations into account:

1. sources and receivers cannot be placed at the same position;
2. sources and receivers cannot be placed within a 1-m radius around the centre of the cutter head;
3. the minimum distance between two positions, each with either a source or receiver, is 0.5 m.

The sources and receivers are pushed against the tunnel face to improve the coupling with the soil. The acquisition takes the rotation and advance of the TBM into account.



Geometry	Number of Sources	Number of Receivers	Practical?
A	20	20	no
B	4	16	yes
C	2	18	yes
D	1	19	yes

Figure 5.1: Cutter head of the tunnel-boring machine showing source-receiver acquisition geometries along different diameters.

Making use of the rotation of the TBM, different measurements are combined to obtain data along a particular transect, oriented along one of the diameters of the cutter head. As an example, we can combine the measurements at 45 and 225 degrees to obtain data along the transect BB′ in Figure 5.1. The new dataset will have eight source positions and we refer it as a *combined* dataset. Depending on the rotation speed of the cutter head, combined datasets along 3 to 4 transects can be obtained before the TBM advances ahead, closer to a possible target reflector that needs to be imaged.

In the rest of the chapter, we consider pre-processing and inversion for a combined dataset along one particular transect. We choose a coordinate system for this transect such that the x-axis is always along the transect. The TBM's direction of advance corresponds to the z-axis. The vibrator source primarily injects a known sweep as a ground force in the y-direction, perpendicular to the transect. As a simplified but useful model, we consider 2-D SH waves in the $[x, z]$-plane, with the particle velocity in the y-direction. Shear vibrators and receivers generate and record only the SH wavefield. This is under the assumption that the medium properties should be invariant in the y-direction. Our

vibrator source can excite signals down to 5 Hz, a low enough frequency for shallow shear-wave surveying.

We denote the sweep signal in the frequency domain by Φ_s. The uncorrelated recorded data at $\mathbf{x}_r = [x_r, z_r]$ due to a source at $\mathbf{x}_s = [x_s, z_s]$ are given in the frequency domain by

$$Q_u(\mathbf{x}_r, f;\ \mathbf{x}_s) = \Phi_s(f)\Gamma_s^{(0)}(\mathbf{x}_s, f)\Gamma_r^{(0)}(\mathbf{x}_r)I_r(f)G_{3D}(\mathbf{x}_r, f;\ \mathbf{x}_s). \tag{5.1}$$

In the above equation, $\Gamma_s^{(0)}(\mathbf{x}_s, f)$ is a factor that takes frequency-dependent ground coupling at the source position $\mathbf{x}_s$ into account. Similarly, $\Gamma_r^{(0)}(\mathbf{x}_r)$ denotes coupling at the receiver position $\mathbf{x}_r$ and $I_r(f)$ describes the frequency-dependent instrument response of the receivers used in the survey, which is assumed to be known. G_{3D} is the 3-D Green's functions or impulse response of the Earth. After acquiring the data, a cross-correlation with the source-sweep signal ϕ_s in the time domain or multiplication with the complex-conjugate spectrum Φ_s^* in the frequency domain is performed. Furthermore, to broaden the bandwidth of the data, a spectral division with the amplitude spectrum of the sweep and the amplitude spectrum of the instrument response is carried out. This results in correlated data $Q^{(3D)}$, given in the frequency domain by

$$Q^{(3D)}(\mathbf{x}_s, \mathbf{x}_r, f) \quad = \quad \frac{\Phi_s^*(f)\,\Phi_s(f)\Gamma_s^{(0)}(\mathbf{x}_s, f)\Gamma_r^{(0)}(\mathbf{x}_r)I_r(f)G_{3D}(\mathbf{x}_s, \mathbf{x}_r, f)}{(|\Phi_s(f)|^2 + \epsilon_1)(|I_r(f)| + \epsilon_2)}, \tag{5.2}$$

where ϵ_1 and ϵ_2 are small stabilisation factors.

We invert the combined data along each transect individually as if line sources, which are oriented along the direction perpendicular to the transect, were used. This is a consequence of the 2-D assumption. Since, in reality, we use nearly point-like 3-D sources, we apply an approximate correction to the combined data, $q^{(3D)}$, in the time domain to make them resemble 2-D data. We let

$$q^{(2D)}(\mathbf{x}_s, \mathbf{x}_r, t) = \sqrt{t}\, q^{(3D)}(\mathbf{x}_s, \mathbf{x}_r, t), \tag{5.3}$$

where the correlated data before and after applying correction are denoted by $q^{(3D)}$ and $q^{(2D)}$, respectively. This correction mainly compensates for the difference in geometrical spreading of the waves [Wapenaar *et al.*, 1992]; we skipped the customary phase correction by $\sqrt{\imath f}$, which is handled instead by including a source-related filter as unknown during the inversion. The correction replaces the 3-D Green's functions G_{3D} in the equation (5.2) with the 2-D Green's functions G_{2D}. After this correction, additional pre-processing such as band-pass filtering, shot-gather normalisation and time-domain tapering, etc., are performed on $q^{(2D)}$ to obtain the pre-processed observed data q_p, used for inversion.

Finally, the inversion results from datasets along different transects can be combined into a single 3-D image.

5.3. 2-D SH Full-Waveform Inversion

An important characteristic of our ground prediction seismic system is its ability to *automatically* produce subsurface maps of the shear-wave velocity, c_s, and/or mass-density,

ρ. This is mainly achieved by the combination of a specially designed seismic vibrator, based on linear synchronous motors and capable of generating low frequencies, and FWI. Once these maps are produced, the *reflectivity* image of the subsurface can be obtained by taking the derivative of the estimated S-wave impedance in either x- or z-direction or by taking the Laplacian. Applying the derivative boosts the reflectors, the interfaces between different materials where seismic waves are reflected.

To compute the modelled data, we solve the 2-D SH wave equation in terms of the particle velocity in the out-of-plane or y-direction. We use a time-domain staggered-grid finite-difference solver [Virieux, 1984] for the forward modelling. To properly account for the effect of the tunnel wall on the wave-propagation, we impose the appropriate Neumann boundary conditions.

Next, we will discuss the objective function to be minimised during full-waveform inversion.

5.3.1. Least-squares Functional With Source Filters and Receiver-coupling Factors

The classic least-squares inversion [Tarantola, 1984; Pratt *et al.*, 1998; Virieux and Operto, 2009; Fichtner, 2010; **?**] minimises the difference between the pre-processed observed data and the modelled data iteratively. The objective function with unknown source-related filters and receiver-coupling factors along with the modelled data is given by

$$J_{ls} = \frac{1}{2} \sum_{s} \sum_{r} \sum_{t} \left[p(\mathbf{x}_r, t;\ \mathbf{x}_s) *_t \gamma_s(\mathbf{x}_s,\ t)\, \gamma_r(\mathbf{x}_r) - q_p(\mathbf{x}_r, t;\ \mathbf{x}_s) \right]^2 . \tag{5.4}$$

Here, $*_t$ denotes convolution in time. We denote the source filters in the time domain and receiver-coupling factors by $\gamma_s(\mathbf{x}_s, t)$ and $\gamma_r(\mathbf{x}_r)$, respectively. Note that the receiver-coupling factors are chosen to be time-independent scalars. The pre-processed observed data and the modelled data are denoted by q_p and p, respectively. For further analysis, we consider the data in the frequency domain where convolution in time corresponds to a simple product operation. Using Parseval's theorem, we rewrite the objective function in equation (5.4) by Fourier transforming the data from time t to frequency f as

$$J_{ls} = \frac{1}{4\pi} \sum_{s} \sum_{r} \sum_{f \geq 0} \left| P(\mathbf{x}_r, f;\ \mathbf{x}_s) \Gamma_s(\mathbf{x}_s,\ f)\, \Gamma_r(\mathbf{x}_r) - Q_p(\mathbf{x}_r, f;\ \mathbf{x}_s) \right|^2 , \tag{5.5}$$

where we used the fact that p, q_p, γ_s and γ_r are real valued. Here, P, Q_p, Γ_s and Γ_r denote the frequency domain representations of p, q_p, γ_s and γ_r, respectively. Note that $\Gamma_r = \gamma_r$, as the receiver-coupling factors are frequency independent.

During the inversion, Γ_s and Γ_r are to be estimated in addition to the medium parameters. The source filters and the receiver-coupling factors compensate for additional unknowns such as source signature and ground coupling at source and receiver locations in the seismic experiment. In order to understand their significance, we consider a case in which the modelled data are generated for the correct medium parameters

with a source wavelet Φ. For that case, we express the pre-processed observed data discussed in the previous section as

$$Q_p(\mathbf{x}_s, \mathbf{x}_r, f) \approx \frac{\Phi_s^*(f)\,\Phi_s(f)\Gamma_s^{(0)}(\mathbf{x}_s, f)\Gamma_r^{(0)}(\mathbf{x}_r) I_r(f) G_{2D}(\mathbf{x}_s, \mathbf{x}_r, f)}{(|\Phi_s|^2 + \epsilon_1)(|I_r| + \epsilon_2)}, \tag{5.6}$$

$$\approx \underbrace{\frac{\Phi_s^*(f)\,\Phi_s(f)\Gamma_s^{(0)}(\mathbf{x}_s, f) I_r(f)}{\Phi(f)(|\Phi_s|^2 + \epsilon_1)(|I_r| + \epsilon_2)}}_{\Gamma_s(\mathbf{x}_s, f)} \underbrace{\Gamma_r^{(0)}(\mathbf{x}_r)}_{\Gamma_r(\mathbf{x}_r)} P(\mathbf{x}_s, \mathbf{x}_r, f). \tag{5.7}$$

It is obvious from the above equation that the factors Γ_s and Γ_r are required to match the modelled data to the observed data even when the correct medium parameters are used.

We use gradient-based techniques to minimise the functional J_{ls}. Hence, we first derive the expressions for the gradient of J_{ls} with respect to the modelled data, source filters and receiver-coupling factors. These are denoted by $\nabla_p J_{ls}$, $\nabla_{\gamma_s} J_{ls}$ and $\nabla_{\gamma_r} J_{ls}$, respectively. Then, we will explain our optimisation strategy.

Gradient With Respect To Medium Parameters

The gradient of J_{ls} with respect to the medium parameters, c_s and ρ, is required to update the subsurface maps. It is computed by correlating the forward-propagated source wavefield with the adjoint wavefield at each point in the subsurface. The adjoint wavefield is generated by injecting the adjoint source functions $\nabla_p J_{ls}$ from the receiver positions. In order to compute the adjoint source functions for the least-squares functional with source filters and receiver-coupling factors, the following steps are performed sequentially:

1. source filters and receiver-coupling factors are applied to the modelled data,
2. the difference between the data after applying filters and the observed data is calculated,
3. the difference is cross-correlated with the source filters and multiplied with receiver-coupling factors.

In order to derive an expression for $\nabla_p J_{ls}$, we rewrite equation (5.5) using the real and imaginary parts of the absolute-valued data as

$$J_{ls} = \tfrac{1}{4\pi} \sum_s \sum_r \sum_{f \geq 0} \{[\Re(P)\Re(\Gamma_s)\Gamma_r - \Im(P)\Im(\Gamma_s)\Gamma_r - \Re(Q_p)]^2 + [\Im(P)\Re(\Gamma_s)\Gamma_r + \Im(\Gamma_s)\Re(P)\Gamma_r - \Im(Q_p)]^2\}. \tag{5.8}$$

We now differentiate the above equation with respect to the real and imaginary parts of the modelled data P to obtain

$$\begin{aligned} \nabla_P J_{ls} &= \nabla_{\Re(P)} J_{ls} + \imath \nabla_{\Im(P)} J_{ls}, \\ &= \tfrac{1}{2\pi} \left[P(\mathbf{x}_r, f;\, \mathbf{x}_s)\Gamma_s(\mathbf{x}_s,\, f)\,\Gamma_r(\mathbf{x}_r) - Q_p(\mathbf{x}_r, f;\, \mathbf{x}_s)\right] \\ &\qquad \Gamma_s^*(\mathbf{x}_s,\, f)\,\Gamma_r(\mathbf{x}_r), \end{aligned} \tag{5.9}$$

where the superscript $*$ denotes complex conjugation. Alternatively, the Wirtinger calculus can be used to derive an expression for $\nabla_P J_{ls}$.

Using the chain rule, we can write the derivative of the functional with respect to the modelled data in the time domain as

$$\nabla_p J_{ls} = \left[p(\mathbf{x}_r, t;\ \mathbf{x}_s) *_t \gamma_s(\mathbf{x}_s,\ t)\gamma_r(\mathbf{x}_r) - q_p(\mathbf{x}_r, t;\ \mathbf{x}_s)\right] \otimes_t \gamma_s(\mathbf{x}_s,\ t)\gamma_r(\mathbf{x}_r), \quad (5.10)$$

where $\otimes_t$ denote cross-correlation.

Updating Source and Receiver Filters

In order to compute the gradient of the objective function in the equation (5.8) with respect to the source filters, the same steps as those for the gradient in the preceding subsection are followed except for the step 3, where the difference is cross-correlated with the modelled data and multiplied with the receiver-coupling factors. Now, an additional summation over the receiver coordinate is also performed. This leads to

$$\nabla_{\gamma_s} J_{ls} = \sum_r \left[p(\mathbf{x}_r, t;\ \mathbf{x}_s) *_t \gamma_s(\mathbf{x}_s,\ t)\gamma_r(\mathbf{x}_r) - q_p(\mathbf{x}_r, t;\ \mathbf{x}_s)\right] \otimes_t p(\mathbf{x}_r,\ t;\ \mathbf{x}_s)\gamma_r(\mathbf{x}_r). \quad (5.11)$$

Similarly, the gradient with respect to the receiver-coupling factors can be computed as

$$\nabla_{\gamma_r} J_{ls} = \sum_s \sum_t \left[p(\mathbf{x}_r, t;\ \mathbf{x}_s) *_t \gamma_s(\mathbf{x}_s,\ t)\gamma_r(\mathbf{x}_r) - q_p(\mathbf{x}_r, t;\ \mathbf{x}_s)\right] \otimes_t p(\mathbf{x}_r, t;\ \mathbf{x}_s) \otimes_t \gamma_s(\mathbf{x}_s, t). \quad (5.12)$$

5.3.2. Optimisation Strategy

We estimate the unknown medium parameters, source filters and receiver-coupling factors that affect equation (5.4) by using an optimisation strategy outlined by the flowchart of Figure 5.2. We use the conjugate-gradient minimisation scheme that uses the gradients of the objective function with respect to the unknowns. The inputs of the optimisation strategy are the pre-processed observed data, the initial values of the source filters, the receiver-coupling factors and the medium parameters. Initially, we choose the receiver-coupling factors to be equal for all the receivers, for instance $\gamma_r = 1$, and the source filters all zero. The initial shear-wave velocity model is a homogeneous one and the velocity is chosen based on the move-out of the direct arrivals in the observed data.

The outermost loop in our strategy runs over the frequency band (i_b) of the observed data selected for inversion. This approach corresponds to multi-scale full-waveform inversion [Bunks *et al.*, 1995; Boonyasiriwat *et al.*, 2009], where we first invert the low-frequency data and then gradually include higher frequencies. The innermost loop consists of two round-trips (i_r) for the observed data in a given frequency band. The output of the first round trip is used as input to the second. In each round-trip, the unknowns that should minimize J_{ls} are estimated in the following order:

1. source-related filters, γ_s, having both positive and negative lags, where a maximum lag time is chosen to prevent error-leakage from medium parameters;
2. receiver-coupling factors, γ_r, which should be positive;
3. the velocity, c_s, everywhere ahead of the TBM, by fitting both the direct arrivals and the reflections in the observed data;
4. the velocity, c_s, in a region with roughly one dominant wavelength away from the TBM, so that the minimization mainly fits the arrivals reflected off the scatterers in the region. This is accomplished by muting the gradient close to the TBM.

While updating one of the unknowns, the other unknowns are kept constant. During the step 4, we focus on fitting only the reflections in the data because the direct arrivals are often stronger and dominate the inversion during step 3. Note that the limited maximum offset causes refractions and diving waves to be absent. The motivation behind performing more than one round trip is that the estimate of the source and receiver filters is improved during the second round-trip when an updated velocity model from the first round-trip is available, assuming that each step reduces the error between modelled and observed data. Similarly, the velocity model is better estimated in the second round-trip since improved filters are applied to the modelled data.

5.4. Synthetic Scenarios

In order to evaluate the performance of 2-D SH FWI in a TBM-like setting, we present results from three typical hazardous scenarios from the tunnelling industry. Table 5.1 lists the values for the seismic shear-wave properties of various subsurface materials, used to generate 2-D synthetic c_s and ρ models. Random velocity and density perturbations (approximately 10% of the background) are added to the synthetic models to make them more realistic. For each scenario, synthetic models serve as input to the 2-D SH finite-difference wave-equation solver to generate synthetic 'observed' data. Table 5.2 lists the parameters chosen for the forward modelling. The records corresponding to each receiver are then multiplied with a random number between 0.2 to 1 to introduce receiver-coupling factors $\Gamma_r^{(0)}(\mathbf{x}_r)$, as in Equation (5.1). As specified in Table 5.2, we deliberately chose different source wavelets for the observed data and for the initially modelled data so that the estimation of the source filter is necessary during the inversion.

We invert only for the shear-wave speed, c_s, while the source filter is assumed to be independent of the source location. Table 5.3 summarizes the inversion parameters. The mass-density, ρ, is taken as a constant during the inversion. The tunnel axis is assumed to be at a depth of 10 m below the surface at $z = 0$. The sources and receivers are constrained to depths between 5 m to 15 m below the surface, thereby assuming a TBM diameter of 10 m (see, e.g., Figure 5.1). Absorbing boundary conditions are applied at the surface. For quantitative evaluation of the output model vectors, say $\mathbf{m}$, we used a correlation measure with respect to a reference model vector, $\mathbf{m}_0$, given by

$$C_{\mathbf{m},\mathbf{m}_0} = \frac{\langle \mathbf{m}, \mathbf{m}_0 \rangle}{\langle \mathbf{m}_0, \mathbf{m}_0 \rangle}, \tag{5.13}$$

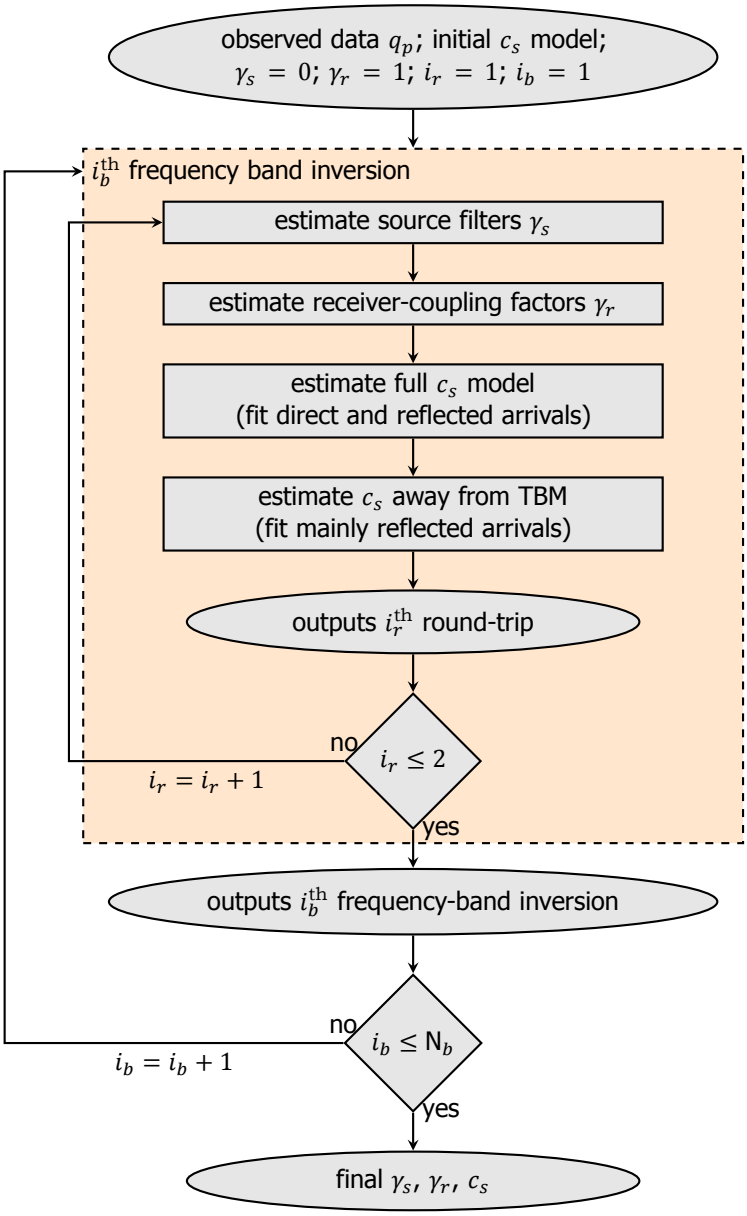


Figure 5.2: Flowchart depicting the optimisation strategy used to estimate source, receivers filters and shear-wave velocity in the medium. The subscript b in i_b stands for frequency band and the subscript r in i_r stands for round-trip. N_b denotes the total number of frequency bands inverted.

Table 5.1: Shear properties of some materials as used in synthetic scenarios, given as shear-wave velocity c_s and mass density ρ.

Synthetic Scenario(s)		Background Material	Anomaly
Abrupt Change & Fault Region		Sand	Clay
	c_s (m s^{-1})	500	160
	ρ (g cm^{-3})	2.2	1.8
Inclusion		Clay	Limestone
	c_s (m s^{-1})	160	1300
	ρ (g cm^{-3})	1.8	2.4

Table 5.2: Forward modelling parameters used for different examples.

Example	Source wavelet for q	Source wavelet for p 4th order minimum-phase Butterworth	Record time (s)
Abrupt Change (synthetic)	40–80–300–400 Hz Ormsby	40–400 Hz	0.1
Inclusion (synthetic)	40 Hz Ricker	1–150 Hz	0.15
Fault Region (synthetic)	40–80–300–400 Hz Ormsby	40–400 Hz	0.1
Inclusion (field)	unknown	10–120 Hz	0.35

where $\langle .,. \rangle$ denotes zero-lag correlation.

5.4.1. Scenario A: Abrupt Change

This scenario defines a sudden change in geology, for instance, when a compacted sand layer lies next to clay. Figures 5.3e and 5.3j show 2-D cross-sections of the shear-wave velocity and mass-density models for such a scenario. Here, the sources and receivers are positioned in the sand and we consider all the four acquisition geometries shown in Figure 5.1. As listed in Table 5.2, we used a 40–80–300–400 Hz Ormsby source wavelet, plotted in Figure 5.4, to generate the observed seismic data. To generate the modelled data, we used a 40–400 Hz fourth-order minimum-phase Butterworth wavelet. Maximum recording time for the observed and the modelled data is 0.1 s.

The source filter, γ_s, has both positive and negative lags with a maximum lag time of 0.01 s. We applied multi-scale full-waveform inversion by first inverting the data between 40–70 Hz, followed by three bands: 40–100 Hz, 40–200 Hz, and 40–400 Hz, as

Table 5.3: Inversion parameters used in examples.

Scenario	Band 1 (Hz)	Band 2 (Hz)	Band 3 (Hz)	Band 4 (Hz)	Length of γ_s (s)	Initial c_s (m s^{-1})
Abrupt change (synthetic)	40–70	40–100	40–200	40–400	0.01	400
Inclusion (synthetic)	full				0.01	170
Fault Region (synthetic)	40–70	40–100	40–200	40–400	0.01	400
Inclusion (field)	full				0.04	110

shown in the Table 5.3. We start the inversion from a homogeneous velocity model with $c_s = 400\,\text{m/s}$. As already mentioned, the mass-density is kept constant during the inversion.

The updated velocity models after inversion with the acquisition geometries A, B, C and D are displayed in Figures 5.3a–d. The cross-hatched pattern indicates the location of the TBM. The horizontal and vertical reflectivity models, shown in Figures 5.3f–i and Figures 5.3k–n, are obtained by differentiating the output velocity models with respect to x and z, respectively. The abrupt change in geology is well imaged with acquisition geometries A and B, although better with A than with B. This is caused by the fact that more sources and receivers are used in acquisition geometry A. In both cases, the horizontal reflectivity models depict a reflector that is positioned correctly. In addition to the first reflector, the inversion has also imaged a ghost reflector that can be misinterpreted as a second reflector. Acquisition geometries C and D were unsuccessful in illuminating the target reflector, particularly in the centre, because they use a smaller number of sources. As shown in the Figure 5.3, the correlation measure (Equation 5.13) decreases (by a factor of 2) for these geometries compared to geometry B, where the outputs using geometry A are chosen as reference. This indicates that one should aim for an acquisition with properly distributed sources and receivers. In all cases, the estimated source filters and receiver-coupling factors after inversion are close to their true values. In the case of geometry B, the source wavelet used to generate the observed data and the estimated source filter γ_s applied to the modelled-data wavelet ϕ are plotted in Figure 5.4. Figure 5.5 shows the estimated and true receiver-coupling factors.

Advance of the TBM

As the TBM advances, it approaches the subsurface target that we would like to image. Our system performs inversion at each stage of advance separately. Often the target is better illuminated when the TBM comes closer, resulting in a better image. To demonstrate this fact, we consider the current scenario with three stages of advance and with acquisition geometry B. During the first stage, the TBM is far away from the target reflector as in Figure 5.6a. The position of the target reflector cannot be determined accurately because of the lack of offset-dependent information, which determines the background velocity. The image of the target reflector in the output reflectivity model of Figures 5.6d appears closer to the TBM than in the actual synthetic model in Figure 5.6g. It can be noticed that, in the output velocity model (Figure 5.6a), the velocity between the reflector and the TBM is lower than the actual value. The inversion results from the second stage, Figures 5.6b and 5.6e, show that the target is better illuminated and slightly better positioned. In the third stage, the inversion results in Figures 5.6c and 5.6f show that the target is correctly imaged.

In practice, it is possible to use the inversion outputs from a particular stage as initial models for the next stage. Also, the datasets from two or three consecutive stages can be combined to perform joint inversion.

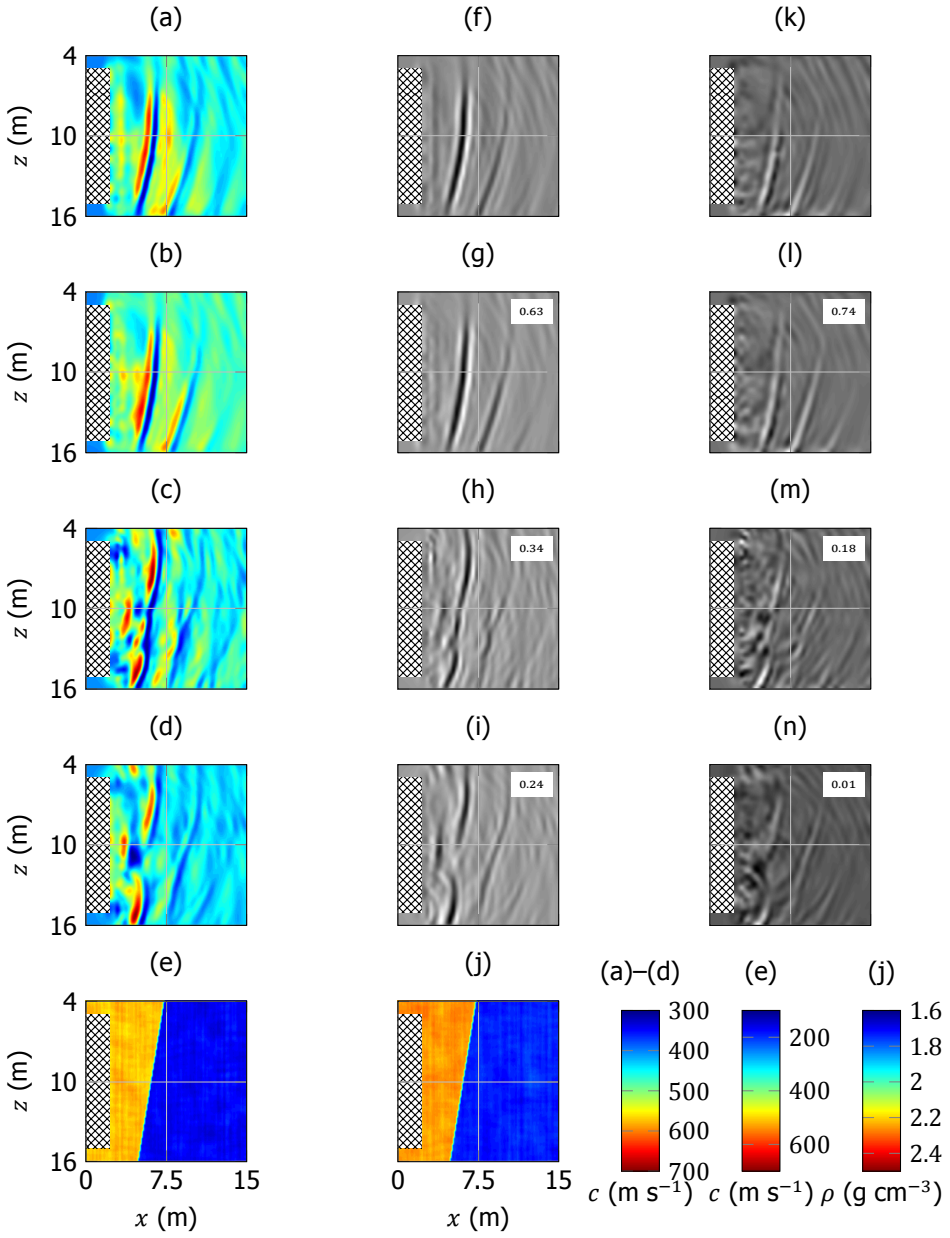


Figure 5.3: Scenario A: Abrupt Change. The output c_s models using geometries A, B, C and D are plotted in (a), (b), (c) and (d), respectively. The corresponding horizontal and vertical derivatives of the c_s models are plotted in (f)–(i) and (k)–(n), respectively. The c_s (e) and ρ (j) models used to generate the observed data are shown as well. Crosshatch pattern indicates the location of the TBM. The correlation measures of the outputs using geometries B, C and D are also given, where the outputs using geometry A are chosen for reference.

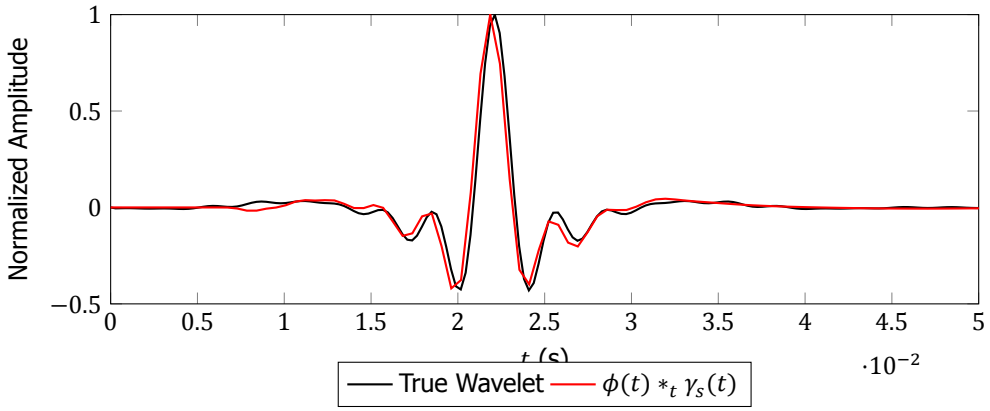


Figure 5.4: For the abrupt-change scenario, the source wavelet used to generate the synthetic *observed* data is plotted in black. The estimated source filter γ_s applied to the modelled-data wavelet ϕ is plotted in red, when using acquisition geometry B.

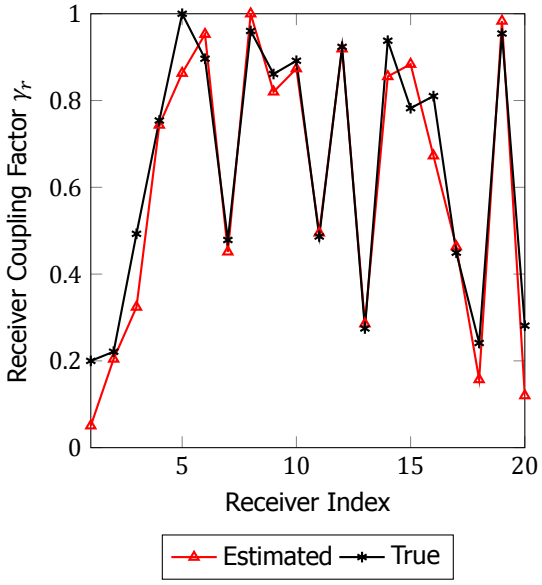


Figure 5.5: For the abrupt-change scenario, the true and estimated receiver-coupling factors are plotted in black and red, respectively. Acquisition geometry B is used.

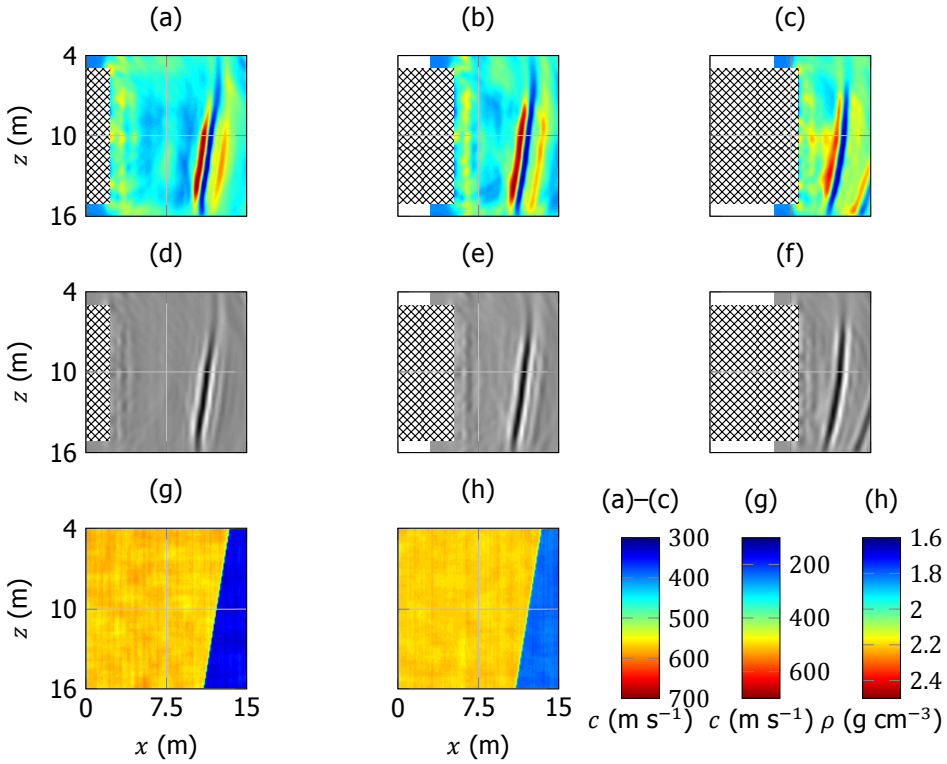


Figure 5.6: Inversion results during three stages of advance of the TBM for Scenario A using acquisition geometry B. The output velocity models from inversion are plotted in (a)–(c). The horizontal derivatives of the output velocity models are plotted in (d)–(f), which depict the target reflector. The cross-hatched pattern indicates the location of the TBM. The c_s (g) and ρ (h) models used to generate the observed data are also plotted.

5.4.2. Scenario B: Hard-rock Inclusion

In this scenario, we used the c_s and ρ models shown in Figures 5.7e and 5.7j. The models represent a rock-type inclusion in a soft-soil environment. Often, this type of inclusion is more or less horizontal in young sediments. We use clay as the background medium and limestone for the hard inclusion. The shear-wave velocity of the propagating waves in clay is lower than in sand or limestone and high frequencies are attenuated due to losses. Based on our experience with field experiments, we used a 40 -Hz Ricker wavelet to generate the observed seismic data. The modelling and inversion parameters for this scenario are summarized in Tables 5.2 and 5.3, respectively. In this example, we invert the observed data only in a single frequency band because the initial homogeneous velocity model has roughly the same background velocity as the actual model. The output velocity models, horizontal-reflectivity models and vertical-reflectivity models after inversion using different acquisition geometries are plotted in Figure 5.7. The correlation measure is only slightly lower for geometries C and D compared to geometry B, where the outputs using geometry A are chosen as reference. In all cases, we were able to image only the tip of the inclusion in front of the TBM. This was to be expected since the source cannot illuminate the sides of the inclusion, in a way that would cause waves to be reflected back to the receivers. All the acquisition geometries image the tip of the inclusion well. However, it can be seen that using less source positions, as in the case of geometries C and D, causes more artefacts in the output velocity models.

5.4.3. Scenario C: Fault Region

Figures 5.8e and 5.8j show the velocity and density models corresponding to this scenario. The aim is to predict the characteristics of the fault region, in particular its width and filling material. We consider the case where the geology is the same on both sides of the fault region. The medium inside the fault region is considered to be clay and the background medium is considered to be sand. Since the sources and receivers are in the sand area, we used the same modelling and inversion parameters (Tables 5.2 and 5.3) as those for the abrupt-change scenario. The output velocity models, horizontal-reflectivity models and vertical-reflectivity models after inversion with the different acquisition geometries are displayed in Figure 5.8. The fault region model is only imaged properly in the case of geometries A and B. Again, when compared to geometry B, the correlation measure is lower by almost a factor two when using geometries C and D, because of the smaller number of sources and receivers. The width of the reconstructed fault structure exceeds the actual width. The exact velocity of the filling material, clay, cannot be determined by the inversion because of the limited aperture. However, the horizontal-reflectivity images depict the boundaries of the fault region. Note that the imaged reflector corresponding to the outer boundary of the fault region might also be misinterpreted as a ghost reflector in Figure 5.3.

5.5. Field Test: Inclusion

So far, we have considered synthetic scenarios. We will contnue with a scenario that was built in the field. This took place at a site near Eindhoven Airport in the Netherlands, where a number of scenarios were built. Here, we will discuss only one of them, namely

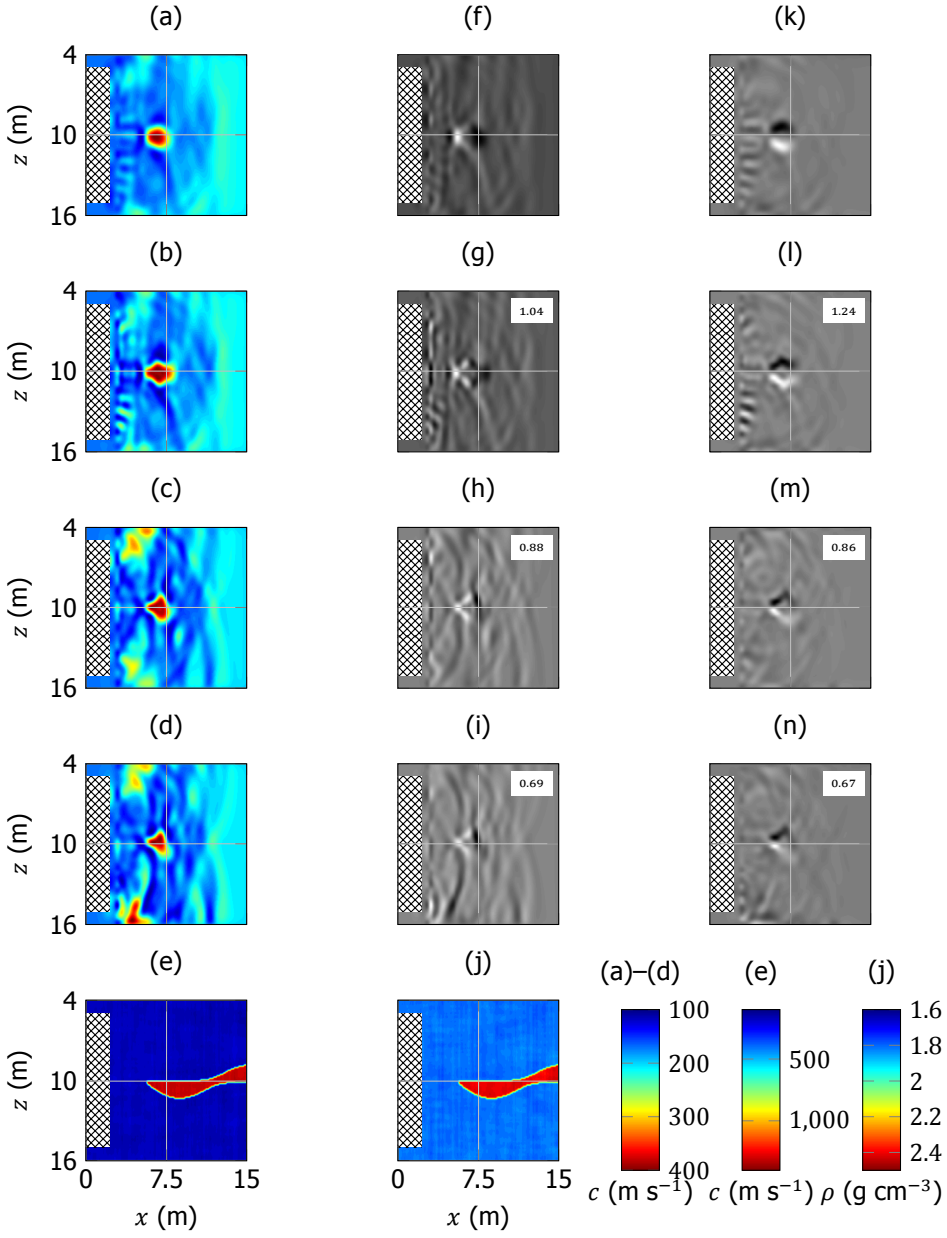


Figure 5.7: Same as Figure 5.3, except for Scenario B: Hard-rock Inclusion.

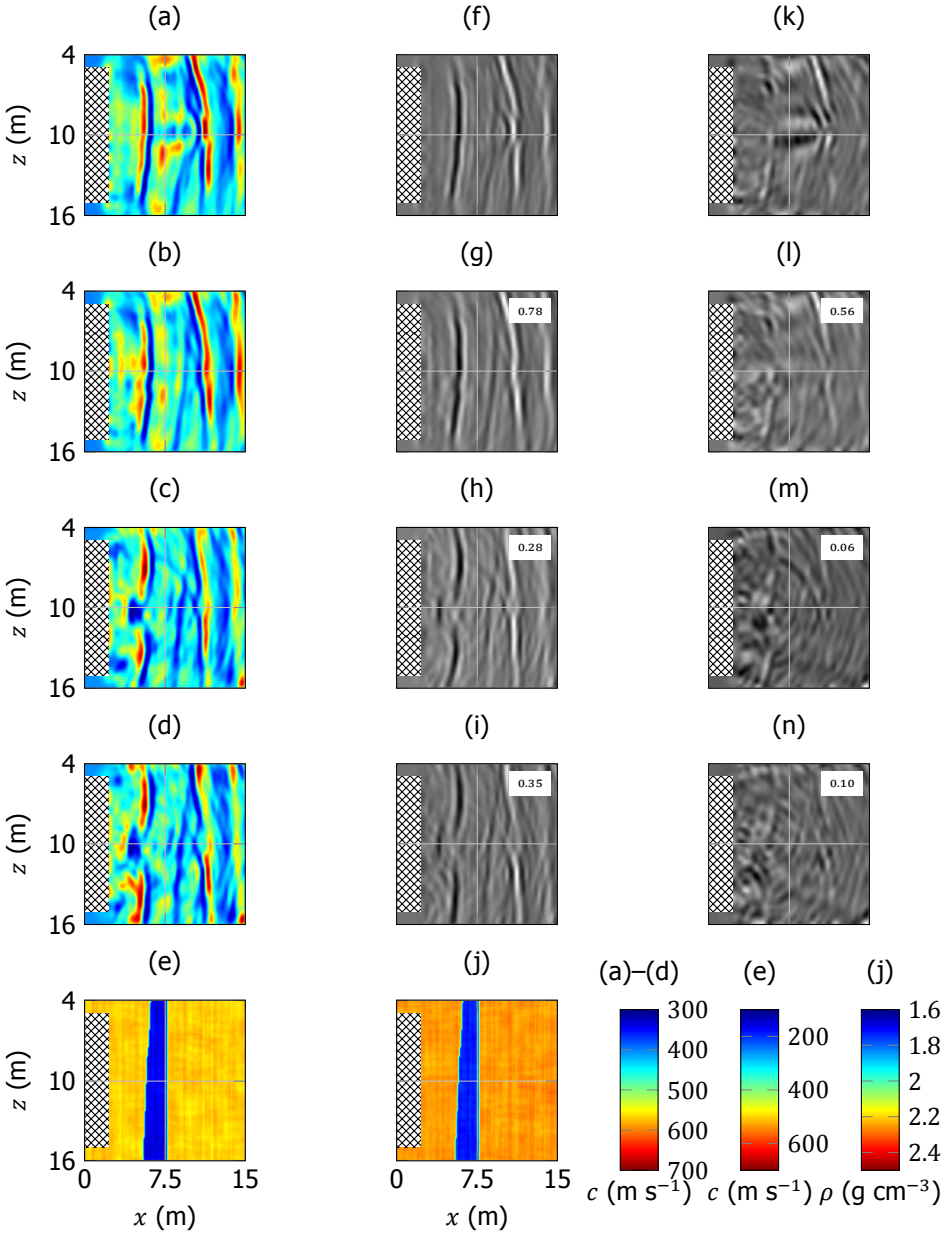


Figure 5.8: Same as Figure 5.3, except for Scenario C: Fault Region.

an inclusion. A TBM setting was simulated by having a fixed spread at the surface as if that spread would represent the source(s) and receivers on the cutter head of a TBM. Therefore, this mimicks a TBM measuring in the vertical rather than the horizontal direction, as considered in the preceding section. The rotation of the cutter head was simulated by rotating the spread by an angle of 60 degrees compared to the previous one, keeping the central point of the spread fixed.

In order to simulate an inclusion, a vertical concrete tube of 1.2 m height and 0.6 m diameter was filled with gravel and placed at a depth of 6 m, exactly below the middle point of the spread. The tube was buried underneath a thick cover of clay. The clay cover was subsequently peeled off, mimicking the advancement of the TBM. Here we will show the results after removing 2 m of the clay cover, when the concrete tube was 4 m, below the acquisition surface; in Figures 5.10 and 5.11, the tube is situated at $z = 6$ m, while the acquisition surface was at $z = 2$ m.

As for the receivers positions: the experiment was conducted with seventeen 10-Hz horizontal SM9© geophones from the company ION, evenly spaced at 0.5 m from $x = -4$ m to 4 m, mimicking the case of a TBM with a diameter of 8 m. And for the type of seismograph: we used a 24-bit Sigma Delta system Geode© from the company Geometrics. The geophones measure the out-of-plane component of the particle velocity, in this case the SH waves. As for the source, we generated an out-of-plane component of the force, i.e., generating SH waves. The shear-wave vibrator was placed at the positions of the 1st, 3rd, 5th, 13th, 15th and 17th receivers, so at $x = -4$ m, -3 m, -2 m, 2 m, 3 m, and 4 m. With the source placed there, the geophone was removed from that particular position and put back when the source was moved to the next position. Figure 5.9, taken during the acquisition, shows a picture of the whole set up.

We used the newly developed shear-wave vibrator that allowed us to input a broadband signal into the ground, namely a linear sweep from 5 to 120 Hz, including the low frequencies required by FWI to perform the processing in a fully automated way.

Observed shot gathers for a source at $x = 4$ m after pre-processing are plotted in Figures 5.10a and 5.11a. The pre-processing consisted in applying the correlation with the input sweep, the correction for the amplitude of the receiver's transfer function and the correction for 3D-to-2D amplitudes, all according to Equation (5.3). We fitted the recorded shot gathers starting from a homogeneous Earth model with $c_s = 110$ m/s and $\rho = 1$ g/cm^3. The mass-density model is not updated during the inversion and only the data in the bandwidth 50–120 Hz are used. Unfortunately, frequencies below 50 Hz had to be muted to suppress the noise from nearby vehicles. We noticed that such a muting operation did not cause severe cycle-skipping problems in this test.

We use the optimisation strategy shown in Figure 5.2. After inversion, the modelled data, plotted in Figures 5.10b and 5.11b, match the observed data quite well. Figures 5.10c and 5.11c show the final output velocity model for both transects. Figures 5.10d and 5.11d depict the vertical derivatives of the output velocity models. Note that in Figure 5.11d, there is a slight mismatch with the true position of the buried target. In Figure 5.11b, there is a data mismatch around an offset of 4 to 5 m. For each transect, the inversion of pre-processed data takes about half an hour on six compute cores. The results were used for interpretation and we were able to detect a high-velocity anomaly,

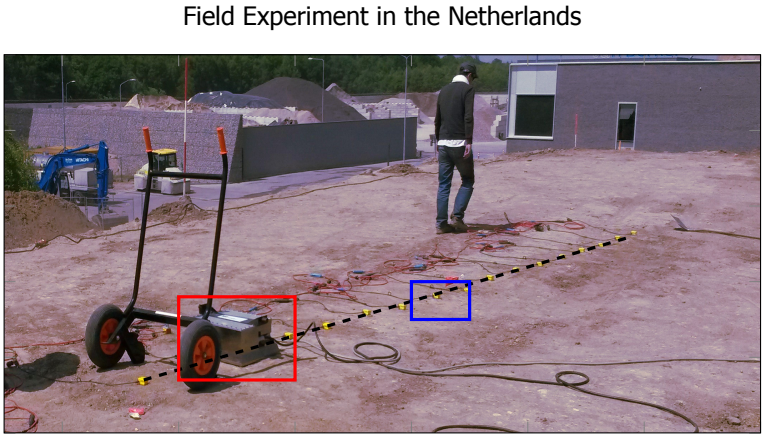


Figure 5.9: Field acquisition with the shear-wave vibrator (marked by the red box), especially designed for this application, and the receivers (blue box) along a transect (dashed line) similar to what is typical for a TBM situation.

corresponding to the location of the concrete-tube inclusion. This demonstrates the success of our SH-wave seismic system to produce subsurface images, including the possibility to have a fully automatic system.

5.6. Discussion and Conclusions

During the field experiment, the recorded signals below 50 Hz are masked by the noise due to nearby vehicles. In these situations with unreliable low frequencies in the data, the inversion is prone to suffer from the cycle-skipping problem [Mulder and Plessix, 2008]. The cycle-skipping problem is severe only when recovering low-wavenumber anomalies, which is not the case in our field example. Alternative inversion algorithms are proposed by many authors [van Leeuwen and Mulder, 2010; Bozdağ *et al.*, 2011; Bharadwaj *et al.*, 2016; Li and Demanet, 2016] to reduce the severity of this problem.

In the examples of this chapter, we only estimated the subsurface shear-wave velocity. In addition to that, full-waveform inversion can also estimate the mass-density of the subsurface provided if the inverse problem is properly regularized with necessary constraints.

We developed a ground prediction system that uses horizontally polarised shear-waves for imaging in front of a TBM, in the case of unconsolidated soils. Compared to the conventional systems, this system produces subsurface images with less artefacts, in an automated way. Seismic data are acquired by the receivers on the cutter head of the TBM, which also hosts the vibrators. The design of the vibrator is based on linear synchronous motors technology that is capable of generating low frequencies. With these low frequencies, the seismic system can use full-waveform inversion as an imaging engine and estimate the shear-wave velocity model required for subsurface

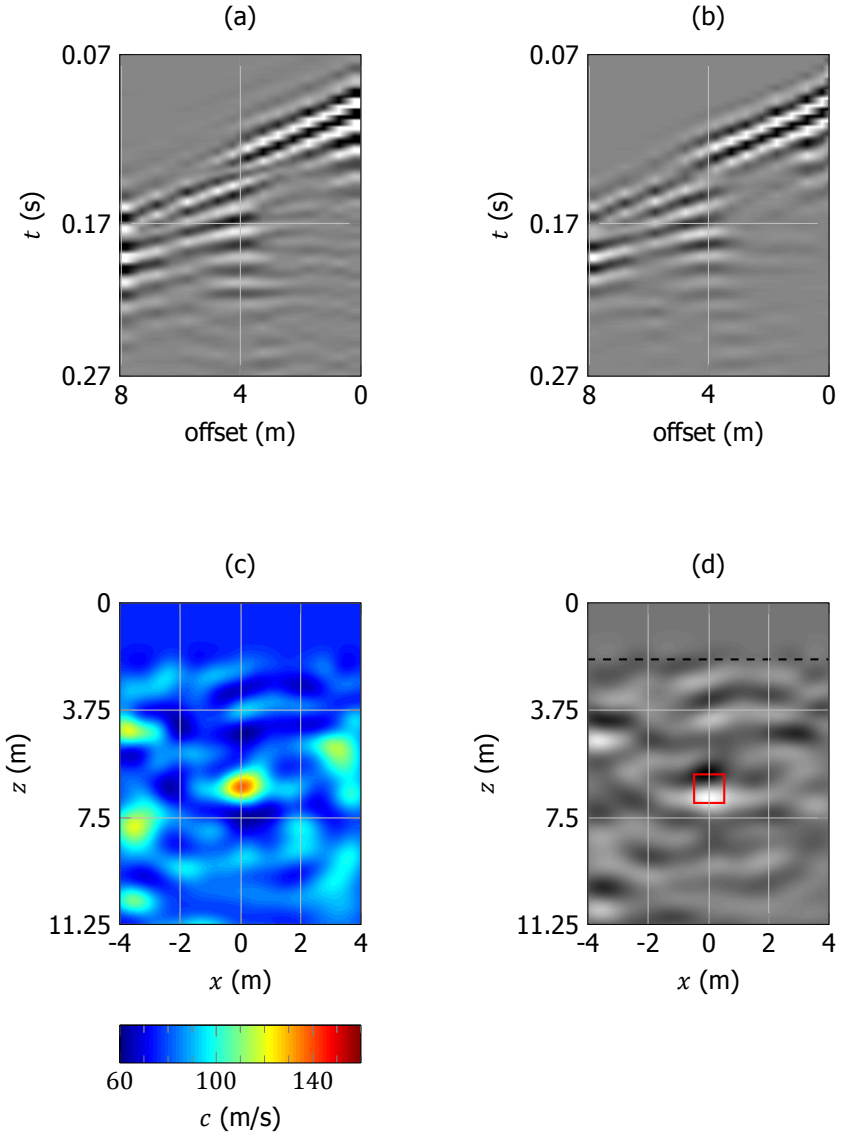


Figure 5.10: The pre-processed data and inversion results of the first transect. a) Observed shot gather for a source at $x = 4\,\mathrm{m}$. b) Modelled shot gather after inversion. c) Estimated shear-wave velocity model of subsurface. d) Image of subsurface depicting the inclusion. The actual location of the inclusion is marked by the red box.

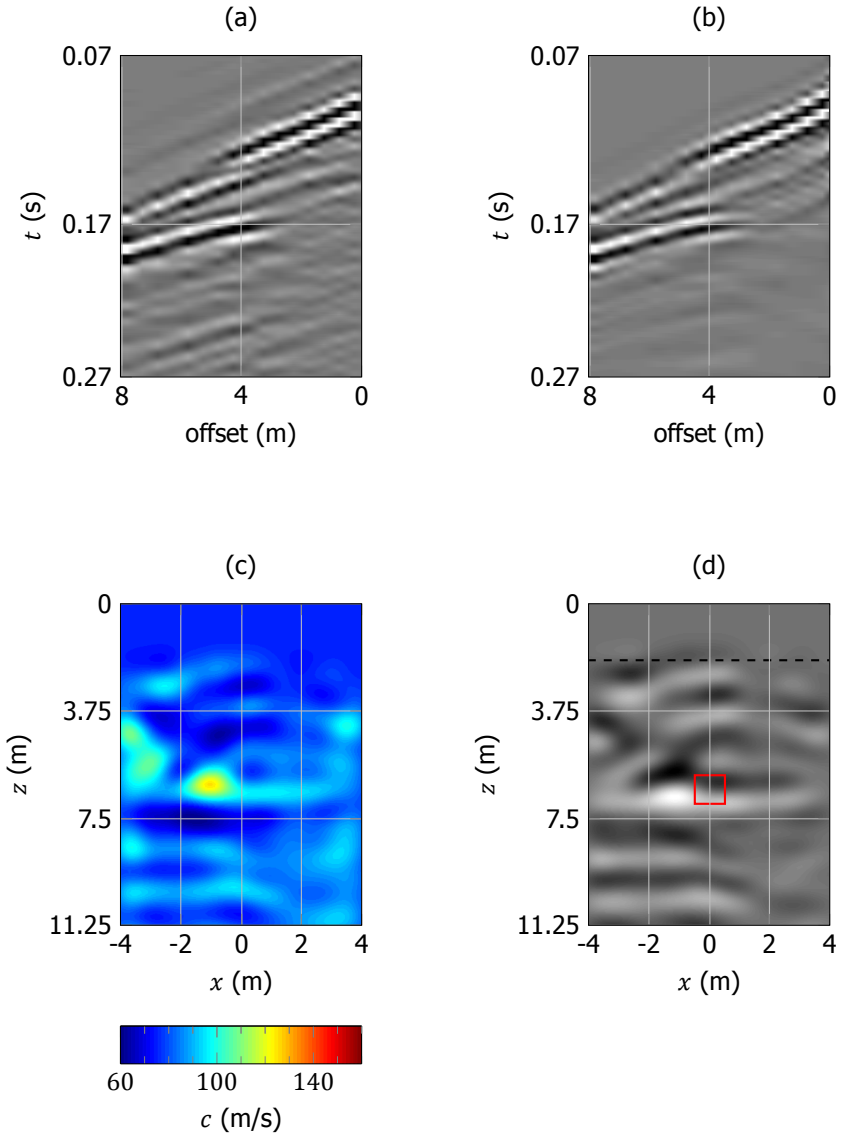


Figure 5.11: Same as Figure 5.10, except for the data corresponding to the second transect.

imaging in a fully automatic way, without human intervention. In addition to shear-wave velocity, our full-waveform inversion algorithm estimates the source-related filters and the receiver-coupling factors.

We investigated the potential of the seismic system using both synthetic and field experiments with TBM configurations. The synthetic examples use different acquisition geometries that take practical constraints into account. We showed that the output subsurface image quality goes up almost by a factor of two when an acquisition with properly distributed sources and receivers is used. In the case of the field experiments, the system was able to detect a buried object in the subsurface. Hence, our ground prediction system can be used for hazard assessment during TBM operation in soft soils.

References

E. C. Hauser, *Detection and location of obstructions ahead of a tunnel boring machine using the tunneling vibrations as a seismic source: The first successful example,* in *2001 Symposium on Application of Geophysics to Environmental and Engineering Problems (SAGEEP)* (2001).

L. Petronio and F. Poletto, *Seismic-while-drilling by using tunnel boring machine noise,* Geophysics **67**, 1798 (2002).

F. Poletto and L. Petronio, *Seismic interferometry with a TBM source of transmitted and reflected waves,* Geophysics **71**, SI85 (2006).

U. Harmankaya, A. Kaslilar, K. Wapenaar, and D. Draganov, *Locating scatterers ahead of a tunnel boring machine using noise correlation,* in *78th EAGE Conference and Exhibition 2016* (2016).

T. Bohlen, U. Lorang, W. Rabbel, C. Müller, R. Giese, S. Lüth, and S. Jetschny, *Rayleigh-to-shear wave conversion at the tunnel face: From 3D-FD modeling to ahead-of-drill exploration,* Geophysics **72**, T67 (2007).

S. Jetschny, *Seismic prediction and imaging of geological structures ahead of a tunnel using surface waves*, Ph.D. thesis, Karlsruher Inst. für Technologie, 2010 (2010).

G. Kneib, A. Kassel, and K. Lorenz, *Automatic seismic prediction ahead of the tunnel boring machine,* First Break **18** (2000).

G. Omnes, *Exploring with SH-waves,* Canadian Journal of Exploration Geophysics **14**, 40 (1978).

K. Helbig and C. Mesdag, *The potential of shear-wave observations,* Geophysical Prospecting **30**, 413 (1982).

H. Stümpel, S. Kähler, R. Meissner, and B. Milkereit, *The use of seismic shear waves and compressional waves for lithological problems of shallow sediments,* Geophysical Prospecting **32**, 662 (1984).

E. D. Guy, R. C. Nolen-Hoeksema, J. J. Daniels, and T. Lefchik, *High-resolution SH-wave seismic reflection investigations near a coal mine-related roadway collapse feature,* Journal of Applied Geophysics **54**, 51 (2003).

S. S. Haines and K. J. Ellefsen, *Shear-wave seismic reflection studies of unconsolidated sediments in the near surface,* Geophysics **75**, B59 (2010).

R. Ghose, V. Nijhof, J. Brouwer, Y. Matsubara, Y. Kaida, and T. Takahashi, *Shallow to very shallow, high-resolution reflection seismic using a portable vibrator system,* Geophysics **63**, 1295 (1998).

R. D. Miller, J. Xia, and C. B. Park, *Love waves: A menace to shallow shear wave reflection surveying,* in *71st Annual Meeting Society of Exploration Geophysicists Expanded Abstracts* (2001) pp. 1377–1380.

G. Swinnen, J. Thorbecke, and G. Drijkoningen, *Seismic imaging from a TBM,* Rock Mechanics and Rock Engineering **40**, 577 (2007).

J. Tzavaras, *3D Tunnel Seismic Imaging*, Ph.D. thesis, Freie Universität Berlin (2010).

Y. Ashida, *Seismic imaging ahead of a tunnel face with three-component geophones,* International Journal of Rock Mechanics and Mining Sciences **38**, 823 (2001).

D. W. Steeples and R. D. Miller, *Avoiding pitfalls in shallow seismic reflection surveys,* Geophysics **63**, 1213 (1998).

A. Bellino, L. Garibaldi, and A. Godio, *An automatic method for data processing of seismic data in tunneling,* Journal of Applied Geophysics **98**, 243 (2013).

A. Tarantola, *strategy for nonlinear elastic inversion of seismic reflection data,* Geophysics **51**, 1893 (1986).

J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics,* Geophysics **74**, WCC1 (2009).

K. Musayev, K. Hackl, and M. Baitsch, *Frequency domain waveform inversion in a tunnel environment,* Proceedings of Applied Mathematics and Mechanics **13**, 323 (2013).

P. Bharadwaj, W. Mulder, G. Drijkoningen, and R. Reijnen, *Looking ahead of a tunnel boring machine with 2-D SH full waveform inversion,* in *77th EAGE Conference and Exhibition 2015* (2015).

T. Nemeth, C. Wu, and G. T. Schuster, *Least-squares migration of incomplete reflection data,* Geophysics **64**, 208 (1999).

F. Bretaudeau, R. Brossier, D. Leparoux, O. Abraham, and J. Virieux, *2D elastic full-waveform imaging of the near-surface: Application to synthetic and physical modelling data sets,* Near Surface Geophysics **11**, 307 (2013).

R. Brossier, S. Operto, and J. Virieux, *Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion,* Geophysics **74**, WCC105 (2009).

R. Noorlandt, G. Drijkoningen, J. Dams, and R. Jenneskens, *A seismic vertical vibrator driven by linear synchronous motors,* Geophysics **80**, EN57 (2015).

B. N. Kuvshinov and W. A. Mulder, *The exact solution of the time-harmonic wave equation for a linear velocity profile,* Geophysical Journal International **167**, 659 (2006).

Y. Luo, J. Xia, Y. Xu, C. Zeng, and J. Liu, *Finite-Difference Modeling and Dispersion Analysis of High-Frequency Love Waves for Near-Surface Applications,* Pure and Applied Geophysics **167**, 1525 (2010).

C. Wapenaar, D. J. Verschuur, and P. Herrmann, *Amplitude preprocessing of single and multicomponent seismic data,* Geophysics **57**, 1178 (1992).

J. Virieux, *SH-wave propagation in heterogeneous media: velocity-stress finite-difference method,* Geophysics **49**, 1933 (1984).

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

R. G. Pratt, C.Shin, and G. Hicks, *Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion,* Geophysical Journal International **133**, 341 (1998).

A. Fichtner, *Full seismic waveform modelling and inversion* (Springer, 2010).

C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, *Multiscale seismic waveform inversion,* Geophysics **60**, 1457 (1995).

C. Boonyasiriwat, P. Valasek, P. Routh, W. Cao, G. T. Schuster, and B. Macy, *An efficient multiscale method for time-domain waveform tomography,* Geophysics **74**, WCC59 (2009).

W. Mulder and R.-E. Plessix, *Exploring some issues in acoustic full waveform inversion,* Geophysical Prospecting **56**, 827 (2008).

T. van Leeuwen and W. A. Mulder, *A correlation-based misfit criterion for wave-equation traveltime tomography,* Geophysical Journal International **182**, 1383 (2010).

E. Bozdağ, J. Trampert, and J. Tromp, *Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements,* Geophysical Journal International **185**, 845 (2011).

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *Full waveform inversion with an auxiliary bump functional,* Geophysical Journal International **206**, 1076 (2016).

Y. E. Li and L. Demanet, *Full waveform inversion with extrapolated low frequency data,* arXiv preprint arXiv:1601.05232 (2016).

6

Conclusions and Further Research

I would rather have questions
that can't be answered than answers that can't be questioned.

Richard Feynman

In the chapter 2 of this thesis, we have outlined the theory of full waveform inversion (FWI). Using a simple numerical example, we have demonstrated that cycle-skipping is the main cause of the local-minima problem, which limits the success of FWI when reliable low-frequency observed data are absent.

To overcome this problem, in chapter 3, we formulated a data-domain functional that matches the observed and the modelled data in a simplified form. It can be viewed as a generalized envelope-based misfit. The simplification results in bumpy data, obtained by taking the absolute value of the data and subsequent smoothing or blurring with a Gaussian. Using numerical examples involving either transmission or reflection data, we illustrated the following characteristics of the bump functional:

- the functional is less sensitive to cycle skipping and does not rely on the low frequencies present in the data;
- blurring increases the size of the basin of attraction that corresponds to the functional and hence its global-convergence robustness;
- the bump-functional inversion suffers from the fact that the model that matches given bumpy data tends to be highly non-unique.

In the same chapter, we observed that the single-objective bump-functional inversion can produce acceptable results while fitting transmitted arrivals. While in the case of reflected arrivals, the non-uniqueness prevents the bump functional to update the background velocity of the model. Therefore, in order to reach the global minimum corresponding to the least-squares objective, we proposed a multi-objective inversion scheme that uses the bump functional as an auxiliary functional. We demonstrated the potential of the bump functional to pull the trapped solution out of the least-squares local minimum whenever necessary. Finally, we have tested the applicability of the multi-objective inversion scheme using realistic numerical examples as well as cross-well field data. In all the cases, the scheme found the model that well explains the observed data and hence corresponds to the global minimum of the least-squares functional, even in the absence of low frequencies in the data. Some immediate research oppurtunities regarding the multi-objective inversion scheme using the bump functional are given below.

- More physical and/or mathematical appreciation of the scheme is necessary to show that its achievements are not limited to the numerical examples chosen in the chapter.
- The applicability of the scheme has to be tested with a real data scenario, where the background velocity has to be reconstructed mainly using the reflected arrivals. This is to investigate its sensitivity to the noise and the unaccountability of the wave equation to the recorded amplitudes.

In chapter 4, we have briefly outlined the conventional point-scatterer and the diffraction-pattern parameterization-analysis methods for the 2-D acoustic inverse problem. Using almost well-posed numerical examples, we have shown that the suggestions

of the conventional analyses methods are valid only when the contrasts of the point-shaped scatterers at a known location are estimated. The numerical examples employ three different modelling and inversion schemes using both Born and full waveform modelling. As expected, for almost well-posed inverse problems, we observed that the change in parametrization will result in a different convergence rate. Furthermore, the relative rate of convergence for a particular choice of parameterization depends on

- the modelling and inversion schemes employed;
- the contrast of the subsurface scatterer that has to be reconstructed;
- the shape of the sub-wavelength scatterer.

We observe that in most of the cases a parameterization with compressional-wave impedance and mass-density did not have the fastest convergence. Finally, our numerical example shows that, in general, there is no such thing as the best parametrization choice that provides the fastest convergence for acoustic inversion.

In chapter 5, we developed a ground prediction system that uses horizontally polarised shear-waves for imaging in front of a TBM, in the case of unconsolidated soils. Seismic data are acquired by the receivers on the cutter head of the TBM, which also hosts the vibrators. The design of the vibrator is based on linear synchronous motors technology that is capable of generating low frequencies. With these low frequencies, the seismic system can use full waveform inversion as an imaging engine and estimate the shear-wave velocity model required for subsurface imaging in a fully automatic way, without human intervention. In addition to shear-wave velocity, our full waveform inversion algorithm estimates the source-related filters and the receiver-coupling factors. We investigated the potential of the seismic system using both synthetic and field experiments with TBM configurations. The synthetic examples use different acquisition geometries that take practical constraints into account. In the case of the field experiments, the system was able to detect a buried object in the subsurface. We now give some suggestions for further research below.

- Our inversion algorithm considers the the recorded data only at a particular stage of the TBM advance. The benefits of the following multiple-stage algorithms are to be investigated: 1. the inversion outputs from a particular stage can be used as initial models for the next stage; 2. the data from two or three consecutive stages can be combined to perform joint inversion.
- Love waves are not seen as a problem in our numerical examples. They are known to be dominant in the presence of a reflector close to the source-recevier array. The performance of the inversion should be tested in the presence of Love waves because: 1. they are dispersive; 2. they mask the useful reflections in the data.

Acknowledgements

This work is a part of the NeTTUN project (nettun.org), funded from the European Commissions Seventh Framework Programme for Research, Technological Development and Demonstration (FP7 2007-2013) under Grant Agreement 280712. This work was sponsored by NWO Exacte Wetenschappen (Physical Sciences) for the use of supercomputer facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO). The computations required in this thesis were carried out on the Dutch national e-infrastructure with the support of the SURF Foundation (www.surfsara.nl).

During the last four years, it has been a pleasure to work on this project and I would like to express gratitude to my promotor Wim Mulder and co-promotor Guy Drijkoningen for their invaluable guidance and support. They made sure that I have a rewarding graduate school experience. At the start of the PhD, my thinking was mostly driven by physical intuition. I remember those Thursdays when I walked up to Wim's office with crazy ideas that made total sense in my mind, but were mathematically inexact. Wim calmly listened, taught me how to be precise, and helped me gain a better mathematical understanding of full-waveform inversion. He engaged me in new ideas and I would have struggled at least two times more without his attention. I also express my warm thanks to him for translating the propositions and summary of this thesis to Dutch.

A very special gratitude goes to Guy for his efficient project planning, accepting me to work with him in Delft. His expertise was very precious during field data acquisition and pre-processing. Thanks are also due to Thomas Tscharner (Geo2X), Rob Jenneskens (MI Partners) and Ralph Reijnen (MI Partners), who have spent time in the field to acquire shear-wave data.

It was a great motivation talking to Kees Wapenaar, and at the same time, attending his wave-theory classes helped me learn a lot. I would like to express my gratitude to Jan Thorbecke for sharing his knowledge about high-performance computing. He helped me write better software with capable modelling and inversion algorithms. I am also really thankful to Eric Verschuur for listening to me now and then and giving his opinion on my research.

Before I started my PhD, I had the pleasure of working as an intern with Gerard Schuster in KAUST. Those interns are just unforgettable and have a huge impact on both my personal and professional spheres of life. I would like to thank Jerry for straightening me out and showing a direction. His book on geophysical inverse theory is one of the best out there. I would also like to thank Chaiwoot, Sherif, Yunsong, Abdullah, Ian, Gaurav, Xin Wang and Wei Dai for a lot of fruitful discussions in KAUST. During my masters, I also received a lot of support from Kirk McIntosh in UT Austin and Tarje Nissen-Meyer in Oxford University.

Now its time for me to thank my cheerful group of colleagues in Delft. Amarjeet

was my first friend in a new continent and cheers to all the skating trips we had! I had a lot of peaceful late-night discussions in an empty office with Carlos. Joost answered my numerous questions via his white-board sketches. I would like to thank Jürg, Elmer, Deyan and Kees Weemstra for educational scientific discussions. Max Holicki introduced me to this amazing thing called PGFplots. If you are impressed by the figures in my thesis, the credit goes to him. Iris was a good friend, we had a lot of nice biking trips to Scheveningen, sometimes even when it was raining. Sixue makes me feel at home by sending her birthday wishes every year in my mother tongue. Ranjani and Syam share lip-smacking food with me all the time; they are the best hosts ever. Siddarth and Yohei: I enjoyed a lot of our chats during coffee breaks every now and then. Niels, thank you for taking me along for King's Day celebrations and our coffee-room fun. Boris, I enjoyed our kayaking trip in Delft-Middenland with you a lot! Matteo, your jokes made me laugh throughout and thanks for that awesome Italian dinner. Casagrande and Diego Barrera, we had a good time during your short and sweet visits to our department! Thanks chaps, keep up the good work. I would also like to thank Tomohide, Asiya, Andreas, Cees, Liang, Marko for being good friends at work. I appreciate the unfailing assistance of Lydia, Marlijn, Hannie and Ralf in our department. Thanks are due to Guus for his support when applying for NWO-SURFSara computational grant.

A very special gratitude goes to my badminton buddies in United Shuttles Smashing Right (USSR), Delft: Queenie, Bjorn, Nicky, Xia, Mohit, Reshu and Jiaguang for fun during both practice and competitions. With a special mention, I really enjoyed the visits of my close friends, Udai Bhan and Akshat Pandey, from India to Delft. I owe a debt of gratitude to Sumedha for a wonderful cover design. I want to also thank Aayush, Deepak and Sana for being my close friends in Delft.

At last but not the least, everything was possible because of an awesome family I have. I am grateful to my siblings Rahul and Swetha, and parents Shobha Rani and Raghava Rao, who have provided me with a through moral and emotional support, and unconditional love in life.

Thanks for all your encouragement!

– Pawan Bharadwaj

Curriculum Vitæ

Pawan Bharadwaj

03-11-1989	Born in Hyderabad, India.

Education

2012–2016	Ph. D. in Geophysics Delft University of Technology, Delft, the Netherlands
2007–2012	Master of Science and Technology in Applied Geophysics Indian Institute of Technology (Indian School of Mines), Dhanbad

Employment

2017–present	Post-doctoral Associate, Massachusetts Institute of Technology
2012–2016	Ph. D. Candidate, Delft University of Technology
2011	Researcher, King Abdullah University of Science and Technology, Saudi Arabia

Grants

2013–2015	Computational Resources (Netherlands Organisation for Scientific Research, NWO)

Other Activities

- Reviewer: Geophysical Journal International, Geophysics, National Science Foundation Proposals Earth Sciences Division (USA).
- Presentations: Earth Resources Laboratory in Massachusetts Institute of Technology [Cambridge, 2016], Institute of Geophysics and Planetary Physics in University of California [Santa Cruz, 2016], Department of Earth Sciences in Utrecht University [Utrecht, 2016], King Abdullah University of Science and Technology [Thuwal, 2013], Institute for Geophysics in University of Texas [Austin, 2011].

List of Publications

Journal Papers

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *A parameterization analysis for acoustic full waveform inversion,* Submitted to Geophysical Journal International (2017a).

P. Bharadwaj, W. Mulder, G. Drijkoningen, J. Thorbecke, B. Neducza, and R. Jenneskens, *A shear-wave seismic system using full waveform inversion to look ahead of a tunnel-boring machine,* Accepted for Publication in Near Surface Geophysics (2017b).

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *Full waveform inversion with an auxiliary bump functional,* Geophysical Journal International **206 (2)**, 1076 (2016).

P. Bharadwaj, T. Nissen-Meyer, G. Schuster, and P. M. Mai, *Enhancing core-diffracted arrivals by supervirtual interferometry,* Geophysical Journal International **196**, 1177 (2014).

P. Bharadwaj, X. Wang, G. Schuster, and K. McIntosh, *Increasing the number and signal-to-noise ratio of OBS traces with supervirtual refraction interferometry and free-surface multiples,* Geophysical Journal International **192**, 1070 (2013).

P. Bharadwaj, G. Schuster, I. Mallinson, and W. Dai, *Theory of supervirtual refraction interferometry,* Geophysical Journal International **188**, 263 (2012).

V. Srivastava, D. N. Giri, and P. Bharadwaj, *Study and mapping of ground water prospect using remote sensing, GIS and geoelectrical resistivity techniques — a case study of Dhanbad district, Jharkhand, India,* The Journal of Indian Geophysical Union **16**, 55 (2012).

I. Mallinson, P. Bharadwaj, G. Schuster, and H. Jakubowicz, *Enhanced refractor imaging by supervirtual interferometry,* The Leading Edge **30**, 546 (2011).

Conference Proceedings

P. Bharadwaj, G. Drijkoningen, W. Mulder, T. Tscharner, and R. Jenneskens, *A shear-wave seismic system to look ahead of a tunnel boring machine,* in *Proceedings of the World Tunneling Congress, Society for Mining, Metallurgy & Exploration* (2016).

P. Bharadwaj, W. Mulder, G. Drijkoningen, and R. Reijnen, *Looking ahead of a tunnel boring machine with 2-D SH full waveform inversion,* in *77th European Association of Geoscientists & Engineers Conference and Exhibition 2015* (2015).

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *An auxiliary bump functional to overcome cycle skipping in waveform inversion,* in *77th European Association of Geoscientists & Engineers Conference and Exhibition-Workshops* (2015).

P. Bharadwaj, W. Mulder, and G. Drijkoningen, *Parametrization for 2-D SH full waveform inversion,* in *76th European Association of Geoscientists & Engineers Conference and Exhibition* (2014).

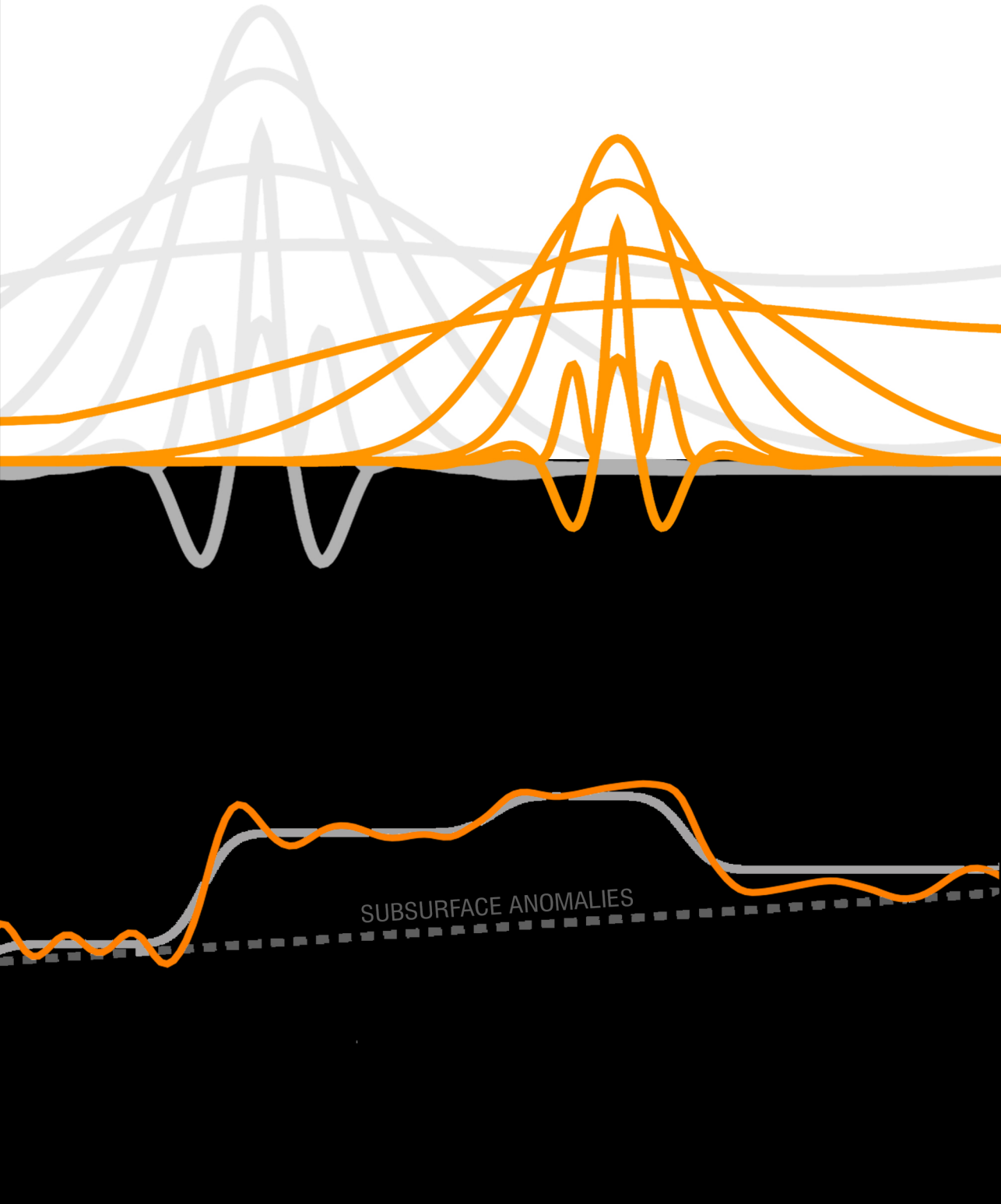
P. Bharadwaj, W. A. Mulder, and G. Drijkoningen, *Multi-objective full waveform inversion in the absence of low frequencies,* in *Society of Exploration Geophysicists Annual Meeting* (Society of Exploration Geophysicists, 2013).

G. Schuster, P. Bharadwaj, T. Nissen-Meyer, and P. Mai, *Extracting CMB diffractions by super-virtual interferometry,* in *American Geophysical Union Fall Meeting Abstracts*, Vol. 1 (2012) p. 2414.

P. Bharadwaj, G. T. Schuster, and I. Mallinson, *Super-virtual refraction interferometry: theory,* in *Society of Exploration Geophysicists Technical Program Expanded Abstracts 2011*, Vol. 30 (2011) pp. 2011–3809.

G. Schuster, P. Bharadwaj, and K. McIntosh, *Refraction interferometry and free-surface multiples for large-offset OBS data,* in *American Geophysical Union Fall Meeting Abstracts*, Vol. 1 (2011) p. 05.

G. Schuster and P. Bharadwaj, *Extending the aperture and enhancing the S/N ratio of refraction imaging by super-virtual interferometry,* in *American Geophysical Union Fall Meeting Abstracts*, Vol. 1 (2010) p. 04.


SUBSURFACE ANOMALIES