2020-10-01

Viscoelastic full-waveform inversion: treating attenuation uncertainty, characterizing cross-talk, and quantifying confidence in inversion results

Keating, Scott Douglas

UNIVERSITY OF CALGARY

Viscoelastic full-waveform inversion: treating attenuation uncertainty, characterizing cross-talk, and quantifying confidence in inversion results

by

Scott Douglas Keating

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

OCTOBER, 2020

Abstract

Full waveform inversion (FWI) is a powerful technique for estimating the subsurface properties which affect seismic wave propagation but it is often limited by simplistic treatment of wave propagation physics. In particular, attenuation and dispersion, which play major roles in most seismic experiments, are commonly neglected. This omission often occurs because of the many complicating factors a treatment of attenuation introduces. Particularly notable challenges include a strong degree of cross-talk, in which attenuation and elastic property estimates are incorrectly influenced by one another, and the need to decide which of several plausible models of seismic attenuation to include in the inversion. In this thesis, I propose approaches to mitigate these challenges and make the inclusion of attenuation in FWI more practical.

While it is necessary to assume a specific relation between attenuation, dispersion, and frequency in FWI, many such relations exist, and these can differ significantly in some of their specific predictions. In general, subsurface attenuation can differ substantially from the attenuation-frequency relation assumed. In this thesis, I propose a flexible inversion approach which mitigates the errors caused by this type of discrepancy by relaxing the assumption that the assumed relation holds over all frequencies, and instead requiring consistency only over smaller bands of frequencies. This approach can be effective in mitigating the impact of errors in the assumed attenuation physics.

Understanding the behaviour of cross-talk as a function of acquisition geometry is often considered essential when using FWI to estimate more than one physical property. Cross-talk with attenuation variables has not previously been deeply investigated in this way, because the conventional approaches used for assessing cross-talk are poorly suited for the modes of cross-talk that are important when considering attenuation. I propose an alternative approach for assessing cross-talk in this thesis. This approach is better able to treat attenuation and also allows for consideration of both the optimization strategy used and cross-talk between spatially separated variables, impor-

tant features of cross-talk not treated in conventional approaches. I use this approach to identify important cross-talk behaviours in viscoelastic FWI.

Finally, I propose an approach for targeted uncertainty quantification in FWI. This approach does not account for all potential uncertainties in the FWI problem, but does quantify a major part of the uncertainty associated with cross-talk. I present numerical examples to illustrate how this approach can be used to assess confidence in viscoelastic FWI results.

Preface

This thesis is written in a manuscript-style format, based on two published papers and one submitted, accepted paper. I was the lead investigator and manuscript composer for each of these papers. These papers have been republished in this thesis with permission from the co-authors and the Society of Exploration Geophysicists.

A version of Chapter 3 has been published as: Keating, S. and K. A. Innanen, 2019, Parameter cross-talk and modelling errors in viscoacoustic seismic full waveform inversion, *Geophysics*, **84**, 4, R641-R653.

A version of Chapter 4 has been published as: Keating, S. and K. A. Innanen, 2020, Parameter cross-talk and leakage between spatially-separated unknowns in viscoelastic full waveform inversion, *Geophysics*, **85**, 4, R397-R408.

A version of Chapter 5 has been accepted for publication in *Geophysics* as: Keating, S. and K. A. Innanen, 2020, Targeted uncertainty analysis in full-waveform inversion using null-space shuttles.

Acknowledgements

I would first like to thank my supervisor, Kris Innanen, for making my PhD studies both interesting and enjoyable. Kris has insightful suggestions for every project, and his guidance was crucial to my research. He was always ready to point out the value of research projects that didn't reach the expected or desired results, which I appreciated immensely and helped to keep me positive.

I am grateful to Daniel Trad for many conversations. His viewpoint helped me to better understand both least-squares migration and full waveform inversion.

Kevin Hall was extremely helpful in everything computer related. Several projects wouldn't have gotten off the ground without help from Kevin. I especially appreciated the time Kevin took to explain computer concepts. Conversations with Kevin were very helpful, and especially let me learn more about seismic fieldwork and distributed acoustic sensing.

Laura Baird made my life immeasurably simpler by taking care of administration in CREWES. She was always very friendly, and kept the candy bowl stocked with gummy tarantulas.

I would like to thak Kevin Bertram for his technical support, especially with posters. He also had valuable advice for giving conference presentations.

Wenyong Pan, Jian Sun, Penny Pan and Junxiao Li each provided me with codes which were key for starting new projects. Without their help, these projects would have taken much more time. Wenyong Pan in particular has my gratitude for taking the time to discuss full waveform inversion with me at length on multiple occasions.

I would like to thank all CREWES staff and students for making my studies enjoyable, especially Eric Rops, Bobby Gunning, Adrian Smith, Babatunde Arenrin, Penny Pan, Raul Cova, Andrew Mills, Adriana Gordon, Khalid Almuteri, Yu Geng, Andy Iverson, Evan Mutual, Tyler Spackman, Matt Eaid and Tim Cary.

I would like to thank the Government of Alberta, the CSEG and the SEG for financial support through scholarships. I also thank NSERC and CREWES sponsors for making my CREWES

support possible.

Finally, I would like to thank Emily, Kayla, and my parents, John and Barb, for their support.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
BFGS	Broyden-Fletcher-Goldfarb-Shanno method
FWI	Full waveform inversion
KF	Kolsky-Futterman
L-BFGS	Limited memory Broyden-Fletcher-Goldfarb-Shanno method
PML	Perfectly matched layers
QFWI	Multi-parameter FWI including attenuation
SD	Steepest descent
SLS	Standard linear solid
TGN	Truncated Gauss-Newton
TN	Truncated Newton
a_1, a_2, a_3, a_4	Coefficients used in regularization
B_k	The BFGS approximation of the inverse Hessian at the kth iterate
B_k^0	The initial estimate of the inverse Hessian used in L-BFGS at the kth iterate
b_1, b_2, b_3	Coefficients used to define a hypothesis function
C	A function giving 0 if and only if the wave equation assumed is satisfied
c	A phase velocity
c_0	Phase velocity at ω_0
c_{0_u}	An unperturbed c_0
$\mathbf{d}$	Measured data
E	Peak strain energy stored
$\mathbf{f}$	A set of seismic source terms
f	A single seismic source term
f_1, f_2	Horizontal force source term, vertical force source term

f_N	A frequency defining the lower limit of the Nth frequency band
$\mathbf{g}$	The gradient (first derivatives of ϕ with respect to $\mathbf{m}$)
$\mathbf{g}_0$	The objective function gradient at the inversion result
H	The Hessian (second derivatives of ϕ with respect to $\mathbf{m}$)
$\mathscr{H}$	The Heaviside step function
H_{GN}	The Gauss-Newton approximation of the Hessian
$\mathbf{h}$	The product of a wavefield with an arbitrary vector
i	The unit imaginary number
J	The Jacobian matrix
L	The Lagrangian of an optimization problem
$\mathbf{M}$	The set of acceptable inversion models
$\mathbf{M}'$	The set of models with objective function equal to ϕ_{max}
$\mathbf{m}$	A subsurface model
$\hat{\mathbf{m}}$	A model with a residual removed
$\mathbf{m}'$	A model with an objective function equal to ϕ_{max}
$\mathbf{m}^*$	A model with an objective function below ϕ_{max}
$\mathbf{m}'^*$	An approximate minimizer of ψ
$\mathbf{m}^0$	An initial model
$\mathbf{m}_{FWI}$	An FWI result
$\hat{\mathbf{m}}_{FWI}$	An inversion result based on data from $\hat{\mathbf{m}}$
$\mathbf{m}_k$	The kth iterate of a model
$\mathbf{m}_{min}$	An ideal FWI result
$\mathbf{m}_N$	A model for the Nth frequency band
$\mathbf{m}_{vp}, \mathbf{m}_{vs}, \mathbf{m}_{qp}, \mathbf{m}_{qs}, \mathbf{m}_\rho$	Model subsets defining v_P, v_S, Q_P, Q_S, and ρ
$\mathbf{m}^0_{vp}, \mathbf{m}^0_{vs}, \mathbf{m}^0_{qp}, \mathbf{m}^0_{qs}, \mathbf{m}^0_\rho$	Initial values of these model subsets
N_F	Number of discrete frequencies considered

$N_{\mathbf{m}}$	The number of variables in $\mathbf{m}$
N_S	Number of sources considered
$\mathbf{p}$	A model update direction
$\mathbf{p}_k$	The model update direction used at the kth inversion iteration
Q	Quality factor
Q_P	P-wave quality factor
Q_S	S-wave quality factor
Q'_P, Q'_S	An outer ring of Q_P or Q_S
$Q_{P_{crit}}, Q_{S_{crit}}$	Threshold values used in regularization
$\bar{Q}_P^{-1}$	The average value of Q_P^{-1} in the model
Q_u	An unperturbed Q
$\mathbf{R}$	A matrix applying receiver sampling
$\mathbf{r}$	A vector of spatial positions
$\mathbf{r}_{A_n}$	The nth element of the model residual subset for parameter A
$\mathbf{r}_{\mathbf{m}}$	A model residual
$\mathbf{r}_N$	Newton system residual
$\mathbf{S}$	Helmholtz operator
$\mathbf{S}_1, \mathbf{S}_2$	Helmholtz operators invoking different attenuation physics
s	A term representing a complex, dispersive phase velocity
$\mathbf{u}$	A set of wavefields
u	A wavefield
$\bar{u}$	A wavefield satisfying the wave equation used
u_x	Displacement in x
u_z	Displacement in z
$\mathscr{V}$	A scattering potential
V_{c_0}, V_Q	Terms in $\mathscr{V}$ representing contributions of c_0 and Q

v_i	The ith element of $\mathbf{v}$
v_P	P-wave velocity
$\bar{v}_P^{-2}$	The average value of v_P^{-2} in the model
v_S	S-wave velocity
$\mathbf{v}, \mathbf{w}$	Arbitrary vectors
x	Position coordinate orthogonal to z
$\mathbf{x}^*$	A vector used to approximate x^*
z	Depth coordinate
α	A step length
β	A term defining the imaginary and frequency dependent part of s
γ	The data residual
Δc_0	A change in c_0
ΔE	Energy lost
$\Delta \mathbf{g}_k$	Gradient change at kth iteration
ΔQ	A change in Q
$\Delta \mathbf{m}$	The maximum viable model-space step
$\Delta \mathbf{m}'$	A null-space shuttle from $\mathbf{m}_{FWI}$
$\Delta \mathbf{m}'^*$	The null-space shuttle corresponding to $\mathbf{m}'^*$
$\delta \hat{\mathbf{m}}$	The unit vector pointing in the direction $\Delta \mathbf{m}$
$\Delta \mathbf{m}_0$	An initial estimate of a model-space direction minimizing ψ
$\Delta \mathbf{m}_{FWI}$	A difference between inversions with and without a model residual
$\Delta \mathbf{m}_k$	Model change at kth iteration
$\varepsilon_1, \varepsilon_2$	Regularization weighting terms
$\zeta_1, \zeta_2, \ldots$	Unitless scaling terms
η	A forcing term in TN
κ	A Lagrange multiplier

$\bar{\kappa}$	A special value of κ
$\tilde{\lambda}, \tilde{\mu}$	Complex Lamé parameters
$\xi, \bar{\xi}$	A Lagrange multiplier and a special value of that multiplier
ρ	Density
$\bar{\rho}$	The average value of ρ in the model
$\tilde{\sigma}$	The average v_P to v_S ratio in the model
$\tau, \tau_\varepsilon, \tau_\sigma$	Relaxation times in the SLS model
ϕ	An FWI objective function
ϕ_0	The objective function value of the starting model
ϕ_D	Data-fitting part of the objective function
ϕ_{max}	A maximum acceptable objective function value
ϕ_P	Prior-fitting part of the objective function
ϕ_{P1}, ϕ_{P2}	Prior-fitting objective function terms
ϕ_{WE}	A wave equation enforcing part of an objective function
ψ	A hypothesis function
ω_0	A reference angular frequency
ω	Angular frequency
†	The adjoint operator
^{T}	The transpose operator
$\|\| \cdot \|\|$	The L_2 norm operator
$< \cdot, \cdot >$	Inner product
∇	The derivative operator

Chapter 1

Introduction

1.1 Full waveform inversion

Inversion is the process of using measurements of physical systems to estimate the properties of those systems (Tarantola, 2005). In this sense, inversion is ubiquitous throughout science, though it can often be a trivial procedure. Inversion assumes a special importance when available measurements relate to the desired properties in complex, indirect ways. In geophysics, where direct measurements are often impractical or impossible, inversion is a crucial tool for determining the properties of the subsurface. Geophysical measurements used to investigate the subsurface include gravity (e.g. Wessel and Watts, 1988; Chen et al., 2006; Van Camp et al., 2017) , magnetic (e.g. Kuang and Bloxham, 1997; Sabaka et al., 2002; Parker, 2003) and electric (e.g. Daily et al., 1992; Zhou et al., 2002; Sudha et al., 2009) field measurements, as well as measurements of electromagnetic (e.g. Ward and Hohmann, 1988; Jiracek, 1990; Annan, 2005) and seismic (e.g. Van Gestel et al., 2008; Tape et al., 2009; Maxwell et al., 2010) waves. Seismic inversion in particular is a powerful tool for subsurface characterization, with the ability to characterize the subsurface with fine resolution over large volumes. Seismic inversion can be used for very large-scale investigation, such as determining the properties of the Earth's mantle and core, or continental-scale characterization of the crust; for study on the scale of tens of kilometers, such as characterization of oil and gas reservoirs or monitoring CO_2 sequestration; and for scales on the order of hundreds of meters, such as microseismic monitoring of small fracturing events. These wide-spread applications make seismic inversion an important tool.

Well established inversion techniques in exploration seismology make use of portions of the measured data to recover specific characteristics of the subsurface. Traveltime tomography (e.g. Bording et al., 1987; Nolet, 1987; Bregman et al., 1989) utilizes picked travel times from early

arrivals to obtain a long wavelength estimate of the seismic velocities in the subsurface. Migration (e.g. Stolt, 1978; Baysal et al., 1983; Gray and May, 1994) and least-squares migration (e.g. Nemeth et al., 1999; Kühl and Sacchi, 2003) use measurements of reflection events to generate images of seismic reflectivity. Amplitude variation with offset (AVO) analysis (e.g. Smith and Gidlow, 1987; Fatti et al., 1994; Goodway et al., 1997; Downton and Ursenbach, 2006) makes use of changes in measured reflection amplitudes from a chosen reflector to generate an estimate of the contrast in elastic parameters at the reflector. While these methods have been very successful in recovering subsurface properties, their focused nature results in assumptions and incomplete usage of the measured data, both of which can limit their effectiveness. The aim of full-waveform inversion (FWI) is to instead recover the physical parameters which predict the full information content of a set of seismic recordings. The goal in FWI, then, is to encompass traveltime tomography, migration, AVO, and other seismic inversion techniques and to further utilize information employed by none of these. While FWI is not currently a replacement for these methods, it offers significant utility and considerable potential by making use of data other inversion approaches neglect. In particular, FWI has been consistently and very effectively used as a tool for recovering velocity structures on wavelengths intermediate between the very long ones contributed by traveltime tomography, and the very short ones reconstructed by migration (Virieux and Operto, 2009).

Tarantola (1984) and Lailly (1983) are widely credited with the development of the full-waveform inversion approach. Though these authors were not the first to propose treating the seismic inversion problem holistically, they were the first to propose a method with potentially feasible computational cost. They achieved this by 1) framing the inversion as a local optimization problem, 2) considering a simplified model of wave propagation physics, and 3) recognizing that the gradient in the optimization problem could be evaluated through a process highly similar to reverse-time migration (an approach already being efficiently implemented and optimized at that time). While the gradient formulation presented by these authors was the key development making the FWI problem tractable, the first and second points represents simplifications which have

proved to be both necessary and problematic. Each of these simplifications can have significant impacts on inversion results and the challenges associated with mitigating these impacts have been the focus of much of the research in FWI from its initial proposal to today.

1.2 Local optimization

Full waveform inversion is generally formulated as an optimization problem in which an objective function, a measure of the discrepancy between measured data and data simulated from a model estimate, is minimized. Ideally, the objective function could be assessed over a representative sampling of feasible model space, allowing for the likelihood of any particular model given the measured data to be assessed in a Bayesian framework. Unfortunately, such comprehensive exploration of the objective function is infeasible in practice; even one simulation of the data which would be recorded with a given model is computationally expensive. Because of this large computational expense, FWI is generally framed as a local optimization problem (Virieux and Operto, 2009), in which only model changes that decrease the objective function are allowed and the minimum of the objective function is considered to be the best representation of the subsurface. This approach is much less computationally intensive than stochastic methods; ideally the objective function will be assessed at only a few intermediate points in model space between the starting model and the minimizing model. It also has drawbacks however, especially relating to local minima and uncertainty quantification.

Local optimization of a non quadratic function (like the FWI objective function) bears an inherent risk of converging to a model which is locally, but not globally, optimal. Local minima are problematic because they represent non-optimal solutions to which our algorithm may converge. Substantial research in FWI has focused on mitigating the threat posed by local minima (e.g., Bozdağ et al., 2011; Van Leeuwen and Herrmann, 2013; Warner and Guasch, 2014; Engquist and Froese, 2014). Perhaps the most intuitive type of local minimum in FWI is caused by what is referred to as cycle skipping. Cycle skipping occurs when the synthetic data corresponding to

a candidate FWI model matches peaks and troughs of the measured data at a different cycle from the correct match. This type of model represent a local minimum in the optimization. If the initial model estimate is more than a half wavelength from the true model, cycle skipping minima are serious hazards, lying between the current model and the global minimum. An example of the cycle-skipping problem is illustrated in figure 1.1. Bunks et al. (1995) proposes avoidance of these local minima by beginning the inversion using only low frequencies (where a half wavelength represents a relatively large distance), and introducing higher frequencies (representing shorter wavelengths) gradually as iterations proceed. In this fashion, local minima corresponding to high frequency components of the data, which lie nearer to the global minimum, are bypassed before these frequencies are considered. This strategy has been demonstrated to be effective in avoiding local minima, provided sufficiently low frequency data is available, but introduces a constraint on how the available data frequencies are employed in FWI. More recently, significant research has focused on re-formulating the objective function to eliminate local minima, through filter-based methods (e.g. Luo and Sava, 2011; Warner and Guasch, 2014), relaxation of partial differential equation constraints (e.g. Van Leeuwen and Herrmann, 2013; da Silva and Yao, 2017), optimal transport metrics (e.g. Engquist and Froese, 2014; Métivier et al., 2016) and many other approaches (e.g. Bozdağ et al., 2011; Huang and Symes, 2015). These methods have been shown to allow for effective FWI, even when important low frequency data are missing, but these approaches still require some low frequencies in order to function, and are often recommended in conjunction with a multi-scaling approach.

The use of local optimization in FWI also introduces challenges for uncertainty quantification. In a Bayesian approach, an approximation of the probability density function can be made by assessing the objective function throughout feasible model space (Tarantola, 2005). Unfortunately, local optimization methods, by design, evaluate the objective function at only a few points in model space and provide insufficient information for an accurate estimation of uncertainty. While a posteriori methods exist for uncertainty estimation, these too are limited by computation considerations,

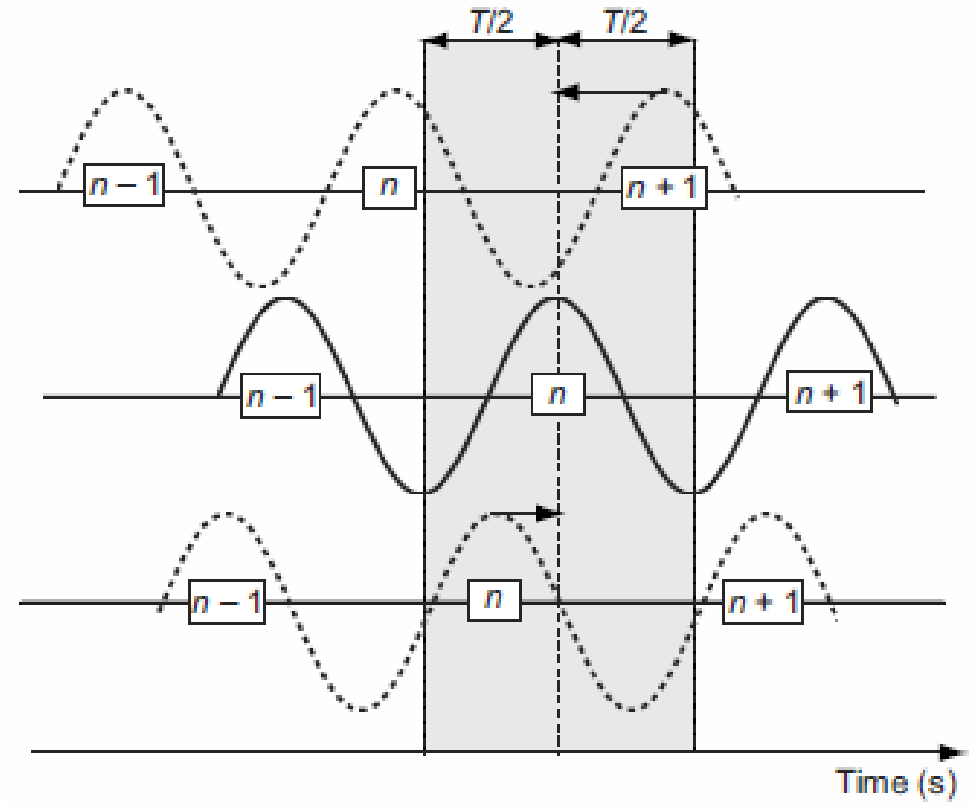

Figure 1.1: Illustration of cycle skipping. In the top example cycle skipping occurs due to the starting error exceeding a half period. In the bottom example, the initial guess is close enough to the true model to avoid the cycle-skipped minimum. From Virieux and Operto (2009).

and typically struggle to fully consider the dimensionality of the inversion problem. These methods provide, of necessity, only local estimates of inversion uncertainty and are unable to quantify the confidence that the inversion result represents a global minimum (Bui-Thanh et al., 2013; Zhu et al., 2016; Eliasson and Romdhane, 2017; Liu and Peter, 2019). Because of these challenges, efficient estimation of confidence in inversion results is an ongoing research topic with significant implications for the effective use of FWI.

1.3 Multiparameter FWI

The first proposals of FWI neglected elastic, anisotropic, and attenuative wave propagation effects, instead assuming an acoustic model of wave propagation (Lailly, 1983; Tarantola, 1984). The assumption of acoustic wave propagation was important for making the problem tractable, reducing the computational cost of wavefield modeling and reducing the dimensionality of the inverse problem, but compromised the ability of the inversion to make use of all the measured data, as seismic wave-propagation involves many significant non-acoustic effects. Early successful versions of FWI used the acoustic model and made the further assumption of constant density. While this formulation of FWI has been very successful in recovering the P-wave velocity of the subsurface, effective implementations often discard or minimize the consideration of portions of the measured data, including surface waves, shear arrivals, and to the extent possible, amplitude variations (e.g Shin and Min, 2006; Luo et al., 2016; Fu et al., 2018). This type of approach treats data more holistically than traveltime tomography or migration, but it neglects data and physics in a way that causes several important shortcomings. First, when the representation of seismic wave propagation is very simplified and not all data can be predicted, it becomes uncertain whether the model best reproducing the data also represents a best estimate of subsurface properties; there may be data introduced by neglected physics effects being erroneously matched in the inversion. Second, by neglecting much of the information content in the data, this approach fails to discern as much information about the subsurface as the data could, in principle, provide. For both of these reasons,

multi-parameter FWI formulations have been proposed.

A multi-parameter FWI inverts for more than one physical property. In its very simplest implementation, multiparameter FWI considers acoustic wave propagation, but recovers both *P*-wave velocity and density in the inversion, as in the original formulation. More complex formulations consider better representations of seismic wave propagations, employing elastic (e.g. Innanen, 2014; Wang et al., 2016; Pan et al., 2018), viscoacoustic (e.g. Kamei and Pratt, 2013; Métivier et al., 2015; Plessix et al., 2016; Yang et al., 2016), anisotropic (e.g. Warner et al., 2013; Alkhalifah and Plessix, 2014; Kamath and Tsvankin, 2016), or viscoelastic (e.g. Yang et al., 2016; Fabien-Ouellet et al., 2017) physics among others. While these approaches improve the capacity of the inversion to match the observed data, they also have several drawbacks which make their implementation challenging. The number of variables in these inversion problems can be very large, both because a larger number of physical properties (parameters) are considered and because more complete physics often requires finer spatial sampling in modeling. The computational costs of wavefield modeling can likewise be much larger than in the single-parameter case. Multi-parameter inversion also makes explicit the problem of cross-talk. Cross-talk occurs when differences between measured and modeled data (data residuals) are mistakenly attributed to model errors in the wrong parameters. While cross-talk occurs even in single parameter inversion (wherein all data residuals are attributed to one parameter), effective multi-parameter inversion requires that cross-talk be largely mitigated in order to have confidence in the estimates of different parameters.

Strategies for mitigating cross-talk have been extensively investigated by the FWI community. One successful approach for acoustic and elastic inversion has been to re-parameterize the inversion in terms of properties that each contribute to unique portions of the data, allowing for recovery of each by focusing the inversion on different data sub-sets (e.g. Sears et al., 2008; Gholami et al., 2013; Kamath et al., 2017). While this approach has been shown to be effective in some problems, it is inadequate where the necessary data for discrimination are unavailable, or where the physics considered do not support a parameterization that assigns different properties to different sub-sets

of data. Where these approaches fail, the main strategy that has been proposed is based on using more effective numerical optimization approaches to reduce cross-talk (e.g. Operto et al., 2013; Innanen, 2014; Métivier et al., 2015; Pan et al., 2016).

While FWI is a nonlinear problem, the Newton method of optimization (which assumes near-linearity) has been demonstrated to very effectively suppress cross-talk for a variety of different FWI formulations. Unfortunately, the computational cost of Newton optimization is extremely high in the FWI problem, preventing its use in practical inversion. Instead, approximations to Newton optimization are used, typically either quasi-Newton approaches, especially the limited memory Broyden, Fletcher, Goldfarb and Shanno (L-BFGS) algorithm (e.g. Vigh et al., 2014), or truncated Newton approaches (e.g. Métivier et al., 2013; Pan et al., 2016). Quasi-Newton methods work by solving an approximation of the Newton system exactly, while truncated Newton methods provide an approximate solution to the exact Newton system. Both these approaches attempt to recover the important information in the Newton update at substantially reduced cost, but it can be challenging to mitigate cross-talk with these strategies at manageable computation levels.

The choice of wave physics and parameterization to be considered in an inversion will typically be based on the subsurface properties we hope to recover and the physical behaviours which must be included in order to reproduce the parts of the data we hope to use. For instance, if we hope to recover shear-wave velocity or make use of the information content in mode conversions and AVO effects in the data, an acoustic FWI formulation will be insufficient; isotropic elastic physics, at minimum, must be considered. The second consideration (adequate physics for data-matching) has prompted substantial research in elastic (e.g., Tarantola, 1986; Choi et al., 2008) and anisotropic (e.g., Barnes et al., 2008; Alkhalifah and Plessix, 2014) formulations of FWI, as these effects typically play a major role in seismic data. Attenuation effects are also important in seismic data, having a dominant role in determining measured signal amplitudes, but while viscoacoustic and viscoelastic FWI formulations have been investigated (e.g., Hicks and Pratt, 2001; Kamei and Pratt, 2013; Fabien-Ouellet et al., 2017), these have not been as well developed as elastic or anistotropic

formulations due to the unique inversion challenges associated with attenuation.

1.4 Attenuation

All seismic waves experience attenuation, where the waves lose energy as they propagate, and this has a major effect on seismic measurements. An associated phenomenon in seismic wave propagation is dispersion (where wave-speed changes with frequency), which is necessary for the preservation of causality in the presence of attenuation (Aki and Richards, 2002). Seismic attenuation is often described in terms of the quality factor Q, which can be defined as

$$\frac{1}{Q(\omega)} = \frac{\Delta E}{2\pi E}, \tag{1.1}$$

where E and ΔE are the peak strain energy stored and energy lost during a given cycle (Aki and Richards, 2002). While many possible mechanisms for attenuation have been proposed, no one mechanism does a satisfactory job of describing observed seismic behaviour. For example, Biot (1956) proposes a theory for poroelastic wave propagation: wave propagation in a fluid saturated porous solid. This theory predicts, among other effects, attenuation and dispersion behaviours, however these are not found to correspond to attenuation of the magnitude commonly observed in seismic experiments. An extension of Biot's theory, the squirt-flow model (Mavko and Nur, 1975) was later proposed, and provided an additional source of attenuation. Squirt-flow mechanisms provide realistic attenuation at ultrasonic frequencies, but are again insufficient to describe seismic attenuation. White (1975) proposes attenuation due to patchy saturated poroelasticity, where each of two immiscible fluids exclusively saturates a porous medium over characteristic, meso-scale patches. This mechanism has been shown to produce attenuation of realistic magnitude at seismic frequencies, however a strong frequency dependence in Q is predicted, in disagreement with observed behaviour.

In addition to intrinsic attenuation, where energy is dissipated by the medium through which the wave propagates, there exists also an apparent attenuation in the subsurface, known as stratigraphic

attenuation. In stratigraphic attenuation, energy is multiply scattered by small scale heterogeneities in the medium and redistributed from the primary wavefront to a complicated coda (O'Doherty and Anstey, 1971). Stratigraphic attenuation does not truly dissipate energy, but for a subsurface of non-white reflectivity it generates an effective dispersion and a frequency dependent decay in amplitudes for the primary wavefield. Stratigraphic attenuation is caused by heterogeneities on scales much smaller than the resolution of seismic, and is thus very difficult to distinguish from intrinsic attenuation.

It has been proposed that no single attenuation mechanism is dominant over the range of seismic frequencies, and that instead the observed seismic attenuation is the combined result of many distinct dissipative mechanisms (Liu et al., 1976). Such a scenario provides a framework within which observed seismic behaviour can be plausibly explained by theoretical attenuation mechanisms. While useful conceptually, this model leads to an unmanageably complex characterization of a dissipative medium, and is of little help in modelling wave propagation or inverting seismic data. Consequently, it is difficult to use theoretical attenuation mechanisms to successfully describe seismic measurements. Instead, attenuation is commonly modelled based on the empirical observation that Q is often nearly independent of frequency, called the nearly constant Q approximation. Causality and linearity cannot be achieved without a frequency dependence in Q and a dispersion term (Aki and Richards, 2002), so these are introduced in nearly constant Q models, with the objective of creating a Q which is close to constant on the frequency band of interest. The details of the frequency dependence chosen give rise to the different types of nearly constant Q models. While these can be largely successful in reproducing similar behaviour over a narrow frequency band, on broader ranges of frequencies they display increasingly dissimilar characteristics (Ursin and Toverud, 2002).

1.5 Attenuation in FWI

The Q values observed in the sedimentary environments where FWI is typically employed can range from about 200 to 10 (e.g. Johnston et al., 1979; Quan and Harris, 1997; Sams et al., 1997). With Q in this range, attenuation is typically a dominant or at least major factor in determining measured wave amplitudes in seismic experiments, especially at high frequencies. Conversely, if nearly-constant Q behaviour is expected, the low frequencies of seismic waves must be subject to substantial dispersion. Because of these significant data effects, it is important to consider attenuation and dispersion when trying to make full use of the information in measured data. Failure to account for these effects will significantly harm our ability to recover acoustic or elastic information from the data. Consideration of attenuation in FWI also offers an additional source of information for use in characterizing the subsurface which can be effectively used for challenging problems such as fluid characterization (Best et al., 1994; Pride et al., 2004; Odebeatu et al., 2006; Ren et al., 2009). Whether to better recover other parameters or to learn about Q itself, it is important to include attenuation in an FWI problem.

Despite the major role that attenuation and dispersion play in seismic wave propagation, there are several substantial challenges with including these effects in FWI. Chief among these are that 1) attenuation effects can be highly ambiguous with other amplitude-changing effects, introducing severe cross-talk, 2) the frequency-dependent effects which distinguish Q can complicate other aspects of the inversion (for instance multi-scaling) and 3) uncertainty about the physics of attenuation is generally much greater than uncertainty in other aspects of wave propagation.

Formulations of FWI which include attenuation (QFWI for short), have been investigated (e.g., Hicks and Pratt, 2001; Hak and Mulder, 2011; Malinowski et al., 2011; Kamei and Pratt, 2013; Métivier et al., 2015), but have not developed to the point that QFWI is preferred to more conventional FWI. A common finding in these investigations is the considerable number of significant challenges present in QFWI which do not appear in single parameter FWI. One of the most prominent challenges observed is a strong tendency to cross-talk between Q and velocity variables. Two

broad approaches to cross-talk reduction have emerged: sequential and simultaneous inversion strategies. Sequential inversion, wherein different subsets of parameters are estimated in succession, has been used for viscoacoustic inversion, typically by first estimating the P-wave velocity, then recovering Q at a later stage (e.g., Malinowski et al., 2011; Kamei and Pratt, 2013). This style of approach has the disadvantage of guaranteeing total cross-talk from the parameters not considered at each stage into those which are considered. This cross-talk can be difficult to remove later in the inversion, given the local nature of the optimization methods considered. The advantage of this approach is the ability to prescribe which modes of cross-talk are active, and authors have investigated strategies to ensure the minimally cross-talking modes are active at the initial stage of the inversion. Other authors have focused on simultaneous inversion, in which all parameters considered are determined in the same process (e.g. Métivier et al., 2015). While this approach offers the possibility for cross-talk to be avoided, in practice it is very difficult to prevent severe cross-talk in this type of formulation. Mulder and Hak (2009) demonstrate that without consideration of dispersion, the viscoacoustic cross-talk problem cannot be resolved with a surface seismic acquisition. Resolution of the cross-talk problem via changes in acquisition geometry requires a seismic experiment with data measurements at all scattering angles (Hak and Mulder, 2010); even very comprehensive acquisition, as in a cross-hole survey, cannot resolve the ambiguity between inverted parameters. This stands in sharp contrast with inversion for simultaneous variations in acoustic and/or elastic properties, which can be separately determined primarily because of differences in the angle-dependence of scattering from one parameter to another. With P-wave velocity and Q it is instead the differences in the frequency-dependence of scattering amplitudes which permits them to be distinguished (Innanen and Weglein, 2007; Hak and Mulder, 2011). This means that established, geometry-based methods for elastic or anisotropic cross-talk prevention will generally not be directly applicable to QFWI.

An additional complication in QFWI is the large degree of uncertainty around the physics of attenuation. Crucially, it is assumed in FWI that these and all physics governing wave propagation

are known, and accurately reproduced in the simulation component of the FWI procedure. If the wave propagation equations miss, or incorrectly model, important features of the data, FWI will seek to match those data features through often dramatically un-physical spatial arrangements of the available model parameters. QFWI is especially prone to modelling errors, because (1) even small changes in the Q model-type can lead to large differences in, for instance, wave velocities at low frequencies, and (2) many model-types exist, and which is suitable in any given instance may not be clear.

1.6 Thesis objectives

Attenuation plays a major role in seismic wave propagation, but is frequently neglected in FWI due to the complexities and challenges arising from its treatment. In this thesis, I seek to develop effective strategies for coping with some of the special challenges of simultaneous inversion QFWI. In particular, I develop an approach for coping with the physics uncertainty which is faced in a QFWI problem, introduce an approach for assessing expected cross-talk modes in the QFWI problem, which may also be useful for survey and inversion design for cross-talk prevention, and I present a strategy for targeted uncertainty quantification in the FWI problem, allowing for possible cross-talk in a QFWI result to be assessed. The structure of the thesis is as follows:

In Chapter 2, I introduce the wave propagation and numerical optimization strategies used in inversions throughout the thesis. I present both viscoacoustic and viscoelastic finite-difference wave propagation formulations. The truncated Gauss Newton (TGN) optimization strategy, which is used extensively throughout the thesis, is outlined in depth and the LBFGS optimization strategy, often used in FWI and implemented within the TGN approach I use, is also explained. I also present a general adjoint-state approach for evaluating the gradient of the FWI objective function considered, as well as for evaluating the product of the Hessian of the objective function with an arbitrary vector, as both these terms play a major role in using the TGN approach in FWI.

In Chapter 3, I demonstrate with synthetic tests the effects of assuming one attenuation physics

model-type when this model-type misrepresents the true attenuation behavior. I go on to outline a frequency updating strategy which can reduce these negative effects while still allowing for the advantages of a multiscale FWI implementation.

In Chapter 4, I present an investigation of cross-talk in the QFWI problem. I begin by demonstrating the challenges associated with using scattering radiation pattern analysis, which is typically used to investigate the cross-talk behaviour in FWI, in the QFWI problem and proceed to develop a numerical approach for cross-talk characterization. Using this tool to investigate, I highlight the modes of cross-talk expected in viscoelastic FWI given different data acquisition geometries, source types and optimization strategies. This analysis helps to identify which data and optimization techniques are effective in suppressing cross-talk, as well as which modes of cross-talk are expected in a QFWI result when effective suppression is impractical.

In Chapter 5, I develop a strategy for targeted uncertainty quantification in FWI and apply this approach to numerical viscoelastic FWI examples. This approach estimates a lower bound on the maximum uncertainty of the inversion in a given hypothesis about the inversion result. The type of uncertainty quantified includes the uncertainty introduced by cross-talk, so this approach offers an effective way to assess whether cross-talk present in QFWI makes the important features of an inversion result ambiguous.

In Chapter 6, I provide a summary of the novel contributions of the thesis. I go on to explain the major outstanding problems of QFWI as I see them, and suggest the future work that may be helpful in resolving these obstacles.

Chapter 2

Optimization and wave modeling tools for attenuative full waveform inversion

Before presenting the novel contributions of this thesis, I will introduce in this chapter some of the important tools used extensively throughout. The key concepts I will introduce here are the formulation of FWI as an optimization problem, the finite-difference strategies I use for wave propagation, the objective function I minimize, the way I parameterize the model, and the important derivatives of the objective function used in the optimization.

2.1 Full waveform inversion as an optimization problem

Full waveform inversion is generally treated as an optimization problem (e.g. Tarantola, 1984; Virieux and Operto, 2009). In this problem, the subsurface model that produces data in closest agreement with the observed seismic data is sought, as this model is expected to best approximate the true subsurface. This optimization problem can be stated in a fairly general form as

$$\mathbf{m}_{min} = \min_{\mathbf{m}} \phi(\mathbf{u}, \mathbf{m}, \mathbf{d}) \text{ subject to } C(\mathbf{u}, \mathbf{m}) = 0, \tag{2.1}$$

where $\mathbf{m}$ is a subsurface model, $\mathbf{m}_{min}$ is the inversion result, $\mathbf{u}$ is a simulated wavefield, ϕ is an objective function, which is small when $\mathbf{u}$ is consistent with measured data ,$\mathbf{d}$, and $\mathbf{m}$ is consistent with prior information and larger otherwise, and $C = 0$ holds only when the wave equation assumed in the inversion is satisfied. Careful examination of equation 2.1 shows that a number of important decisions need to be made before FWI can be implemented. Specifically, we must decide which model of wave propagation to enforce with the $C = 0$ constraint, which objective function, ϕ, to use as measure of data mismatch, which optimization strategy we use to obtain a solution or approximation to the solution of equation 2.1, and how to parameterize the subsurface in models

like **m**. For each of these decisions, there are a broad variety of different alternatives which have been explored in the FWI literature. In this thesis, I will be making these decisions with a first priority of facilitating our goal of investigating effective viscoacoustic and viscoelastic FWI and a second priority of choosing a relatively simple, common approach where there is little direct impact of a decision on QFWI. While the inversion strategies I propose in this thesis should be amenable to a wide range of specific FWI formulations, in the following sections I will describe some key features of the QFWI algorithms I implement here.

2.2 Forward modeling

The process of generating an estimate of the data corresponding to a given subsurface model is referred to as 'forward modeling', or 'the forward problem' (Tarantola, 2005), and is represented by the $C = 0$ constraint in equation 2.1. No model of seismic wave propagation is exact in representing the physics of seismic waves, so any forward modeling is, necessarily, only an approximation. The goal of this thesis is to consider the treatment of attenuation in FWI, so any representation of seismic wave propagation I consider should include attenuation and dispersion. In this thesis I consider both a constant-density, viscoacoustic model of wave propagation (Chapter 3) and a viscoelastic model of wave propagation (Chapters 4 and 5). Constant-density, viscoacoustic wave propagation is significant because it represents the addition of attenuation and dispersion to the constant-density acoustic model of wave propagation, which is the most extensively used model in FWI literature (e.g. Pratt et al., 1998; Virieux and Operto, 2009; Barnes and Charara, 2009). In this sense, it represents a minimum complexity model of wave propagation for QFWI. This approach shares many of the drawbacks of a constant-density acoustic model, however, modeling only P-waves and predicting wave amplitudes poorly. Viscoelastic wave propagation is a more complex model, requiring more physical properties to be defined in the subsurface, but it does a much better job of representing the true behaviour of seismic waves. More accurate models of wave propagation can be considered, but I choose the models above because they capture important features of

seismic waves without requiring unmanageable numbers of physical properties to be defined. The decision to consider these models is not meant to be restrictive; it should be possible to extend the approaches I propose in this thesis to FWI implementations with more complete models of seismic wave propagation.

When the type of physics to consider in the forward modeling is decided, the question of how to frame the partial differential equations representing those physics becomes important. The specific decision which typically plays a major role in FWI is whether to frame our partial differential equations in terms of time- or frequency-dependence. The analytical expressions for the two choices are equivalent, but the decision will have major effects on the computational cost of forward modeling and the specific approximations it will be feasible to use in solving them. Time domain modeling has the advantage of being more computationally efficient for the very large scale problems which are typically considered in FWI (e.g. Vigh et al., 2009), while frequency domain modeling can be more efficient for smaller problems (e.g. Pratt et al., 1998) and provides a much simpler environment for the implementation of frequency-dependent effects (notably attenuation and dispersion, e.g. Aki and Richards (2002)). Because I do not treat the large-scale three-dimensional type of models which necessitate time-domain modeling and I focus on the treatment of attenuation, I will consider frequency-domain forward modeling throughout this thesis. The inversion approaches I introduce later should be equally applicable to either formulation; the frequency-domain modeling I use is not integral to the inversions I will discuss.

2.2.1 Viscoacoustic modeling

Acoustic wave propagation can be represented in the frequency domain by the equation

$$\left[\frac{\omega^2}{c(\mathbf{r})^2}+\nabla^2\right]u(\mathbf{r},\omega)=f(\mathbf{r},\omega), \tag{2.2}$$

where u is the pressure field, f is a source term, c is the phase velocity, $\mathbf{r}$ is a spatial location, and ω is an angular frequency (e.g. Song and Williamson, 1995). This model of wave propagation can

be altered to a viscoacoustic model by replacing $c(\mathbf{r})$ with the variable $s(\mathbf{r},\omega)$:

$$\left[\omega^2 s(\mathbf{r},\omega)+\nabla^2\right] u(\mathbf{r},\omega) = f(\mathbf{r},\omega), \tag{2.3}$$

where s combines the effects of dispersive velocity and attenuation and is given by

$$s(\mathbf{r},\omega) = \left[c(\mathbf{r},\omega)\left(1-\frac{i}{2Q(\mathbf{r},\omega)}\right)\right]^{-2}, \tag{2.4}$$

where c is now a frequency dependent phase velocity, and Q is a quality factor. The viscoacoustic FWI problem is to determine unknown spatial and frequency distributions of two parameters: c and Q. In general, c and Q in this form as arbitrary functions of frequency and space are not possible to constrain given typical seismic data. The freedom of these variables with respect to frequency can be curbed by assuming a particular model of attenuation in the forward modeling. Some of the consequences of assuming a particular model when lacking certainty about the behaviour of attenuation in the subsurface are investigated in Chapter 3, but for the majority of this thesis, I assume a Kolsky-Futterman nearly-constant Q model of attenuation (Kolsky, 1956; Futterman, 1962), which prescribes a nearly-constant Q, and a c proportional to the logarithm of frequency. With this approximation, equation 3.2 is replaced with

$$s(\mathbf{r},\omega) = \left[c_0(\mathbf{r})\left(1+\frac{1}{\pi Q(\mathbf{r})}\log\frac{\omega}{\omega_0}-\frac{i}{2Q(\mathbf{r})}\right)\right]^{-2}, \tag{2.5}$$

where ω_0 is a reference frequency, and $c(\omega_0) = c_0$.

Unfortunately, there exists no analytic solution to equation 2.3 for an arbitrary medium. To obtain an approximate solution, I adopt a finite-difference approach, wherein the partial derivatives in equation 2.3 are replaced with their finite-difference approximations and discretize our solution in space and frequency. I consider a second-order, centered finite difference approximation of the ∇ operator, and consider wave propagation in only two dimensions, resulting, for instance, in the differencing equation

$$\left[\omega^2 s_{h,j}(\omega)u_{h,j}(\omega)+\frac{u_{h-1,j}(\omega)-2u_{h,j}(\omega)+u_{h+1,j}(\omega)}{\Delta x^2}+\frac{u_{h,j-1}(\omega)-2u_{h,j}(\omega)+u_{h,j+1}(\omega)}{\Delta z^2}\right] = f_{h,j}(\omega) \quad , \tag{2.6}$$

for the approximation of equation 2.3 at the location centered at cell h in the x direction and j in the z direction, where subscripts denote the index of the spatial grid cell considered in the x and z directions , Δx is the space increment considered in the x direction, and Δz is the equivalent in the z direction. The implementation of this difference at all locations in the spatial grid considered can be expressed as

$$\mathbf{S}(\mathbf{m})\mathbf{u} - \mathbf{f} = \mathbf{0}, \tag{2.7}$$

where $\mathbf{u}$ is a matrix representing pressure wavefields at the frequency considered, whose columns represent the wavefield for different seismic sources, and whose rows represent different spatial locations, $\mathbf{f}$ is a matrix representing the sources with the same indexing, and $\mathbf{S}$ is a matrix which applies the finite difference scheme in equation 2.6. This form allows for direct solution for $\mathbf{u}$ and, while it represents a very large system of equations for most problems of interest, the very sparse structure of S allows for moderately-sized two-dimensional problems to be solved at manageable computational cost. At the spatial edges of the region considered, equation 2.6 cannot be implemented in S, as locations outside of the study area would be needed. To deal with the boundaries, I implement a perfectly matched layer (PML) region (Berenger, 1994) at the edge of the model to simulate the effect of an infinite extension of the model, preventing reflections from the model edges. These conditions were also applied at the top of the model, so reflections from the surface are not modeled. The left hand side of equation 2.7 can be chosen as the C in equation 2.1 and enforces the constraint that the assumed wave equation must be satisfied when viscoacoustic wave propagation is assumed.

2.2.2 Viscoelastic modeling

While a viscoacoustic representation of wave propagation is closer to reality than an acoustic representation, both fail to represent some of the key features of observed seismic wave propagation. This can lead to a situation where inversions using viscoacoustic wave simulations can better reproduce measured data than acoustic inversions, but cannot do so accurately enough to merit the

use of additional aspects of the data. This type of effect can be seen in the treatment of seismic amplitudes. Acoustic modeling largely captures the kinematics of body P-waves, accurately predicting correct travel times, but often fails to reproduce accurate dynamics for these waves, incorrectly modeling wave amplitudes and leading to inversion strategies which de-emphasize the information contained in seismic amplitudes. Because attenuation has a major effect on measured amplitudes, viscoacoustic modeling can do a better job of correctly modeling the dynamics of wave propagation, but important elastic effects which have substantial impacts on wave amplitudes are still neglected. The result can be a situation where seismic amplitude information is unsuitable for detailed analysis in viscoacoustic inversion, despite the improvement over acoustic inversion. This type of concern leads to consideration of viscoelastic modeling, which accounts for most of the major effects on seismic amplitudes.

A system of equations representing isotropic viscoelastic wave propagation in two dimensions is described by Pratt (1990) as:

$$\begin{aligned}\omega^2 \rho u_x + \frac{\partial}{\partial x}\left[\tilde{\lambda}\left(\frac{\partial u_x}{\partial x}+\frac{\partial u_z}{\partial z}\right)+2\tilde{\mu}\frac{\partial u_x}{\partial x}\right] \\ + \frac{\partial}{\partial z}\left[\tilde{\mu}\left(\frac{\partial u_z}{\partial x}+\frac{\partial u_x}{\partial z}\right)\right]+f_1=0\end{aligned} \tag{2.8}$$

and

$$\begin{aligned}\omega^2 \rho u_z + \frac{\partial}{\partial z}\left[\tilde{\lambda}\left(\frac{\partial u_x}{\partial x}+\frac{\partial u_z}{\partial z}\right)+2\tilde{\mu}\frac{\partial u_z}{\partial z}\right] \\ + \frac{\partial}{\partial x}\left[\tilde{\mu}\left(\frac{\partial u_z}{\partial x}+\frac{\partial u_x}{\partial z}\right)\right]+f_2=0,\end{aligned} \tag{2.9}$$

where ω is the angular frequency, ρ is the density, u_x and u_z are the horizontal and vertical displacements, f_1 and f_2 are their respective source terms, and $\tilde{\lambda}$ and $\tilde{\mu}$ are the complex, frequency-dependent Lamé parameters. Assuming a Kolsky-Futterman model of attenuation (Kolsky, 1956; Futterman, 1962), these are defined in terms of ρ, the P- and S- wave speeds, v_P and v_S, as well as the quality factors Q_P and Q_S, by

$$\tilde{\mu} = \left\{ v_S \left[1 + \frac{1}{Q_S}\left(\frac{1}{\pi}\log\frac{\omega}{\omega_0}+\frac{i}{2}\right)\right]\right\}^2 \rho \tag{2.10}$$

and

$$\tilde{\lambda} = \left\{ v_P \left[1 + \frac{1}{Q_P} \left(\frac{1}{\pi} \log \frac{\omega}{\omega_0} + \frac{i}{2} \right) \right] \right\}^2 \rho - 2\tilde{\mu}, \tag{2.11}$$

where ω_0 is a reference frequency. As in the viscoacoustic case, we solve these equations using a finite-difference approach with second-order centered finite difference approximations of the spatial derivatives. The finite-difference equations I use are described in detail by Pratt (1990). For the purposes of our discussion here, it is only important to note that these equations can be structured to take a similar form to the viscoacoustic case:

$$\mathbf{S}(\mathbf{m})\mathbf{u} - \mathbf{f} = 0, \tag{2.12}$$

where $\mathbf{u}$ is a vector containing u_x and u_z, $\mathbf{f}$ is a vector containing f_1 and f_2, and $\mathbf{S}$ is a Helmholtz matrix containing the finite-difference coefficients. Once again, I use perfectly matched layers (Berenger, 1994) to prevent reflections from the boundaries of the model. The left hand side of equation 2.12 can be chosen as the C in equation 2.1 and enforces the constraint that the assumed wave equation must be satisfied when viscoelastic wave propagation is assumed. The only difference between equations 2.7 and 2.12 is in the definitions of $\mathbf{S}$, $\mathbf{u}$, and $\mathbf{f}$. By expressing the forward modeling equations in this common form, I simplify the treatment of objective function derivatives later in this chapter.

2.2.3 Major forward modeling assumptions

Here I briefly summarize the major assumptions I make in the forward modeling approaches used for inversion. A key assumption is that the equations of isotropic viscoacoustic or viscoelastic wave propagation are accurate representations of seismic wave propagation. Anisotropy is neglected not on the basis of minimal impact on measured data (in fact, anisotropy typically plays a key role in determining the kinematics of seismic wave propagation (e.g. Prieux et al., 2011)), but rather to simplify the modeling and because of the relatively minor interaction expected between anistotropy- and attenuation-defining parameters. I assume that the direction of maximal attenuation is parallel to that of increasing phase (homogeneous wave propagation), again for simplicity,

but this type of behaviour is not generally necessary in viscoelastic wave propagation (Borcherdt, 1973). Often, I adopt a Kolsky-Futterman model of attenuation. Surface reflections and surface waves are neglected in the modeling - these are often removed from data as a matter of course in seismic processing and inversion, and lead to significant complications which are not the focus of my work here. Further, the numerical errors associated with the finite-difference approximation and any reflections from the PML region I assume to be negligible. These assumptions are based on choosing a finite-difference grid size appropriately small for the frequencies I consider and a PML region appropriately thick.

Each of the assumptions above is likely to cause at least some discrepancy between modeled and true wave propagation, complicating comparison of simulations to field data. While I explicitly investigate the implications of one of these assumptions in Chapter 3, I hold the other assumptions without investigating or accounting for their possible failure throughout the thesis. It is worth noting that I do this not because of a belief that they will always (or ever) hold exactly true, but because the problems I wish to investigate exist even when these assumptions are appropriate, and I expect that these problems will play an important role even with more complete forward modeling strategies. With this limitation in mind, I will attempt to present inversion strategies which should be applicable even if more accurate models of wave propagation are considered.

2.3 Objective function

The objective function, or objective, expressed as ϕ in equation 2.1 is a metric which penalizes discrepancies between measured data and data generated from the current subsurface model, and promotes consistency of the subsurface model with prior information about the subsurface. Specific implementations of ϕ, however, are extremely variable and often the specific subject of intense research (e.g. Bozdağ et al., 2011; Luo and Sava, 2011; Engquist and Froese, 2014). One commonality between most objective function implementations is an explicit separation into a wavefield-dependent data-fitting and a model-dependent prior-fitting part, usually additively combined as in:

$$\phi(\mathbf{u},\mathbf{m}) = \phi_D(\mathbf{u}) + \phi_P(\mathbf{m}), \tag{2.13}$$

where ϕ_D is the data-fitting part, and ϕ_P is the prior-fitting part. The only type of objective function not amenable to this type of separation that I am familiar with is used in approaches which do not exactly enforce the constraint $C = 0$ in equation 2.1, but instead include a wave equation enforcing objective term $\phi_{WE}(C(\mathbf{u},\mathbf{m}))$ (Van Leeuwen and Herrmann, 2013; da Silva and Yao, 2017). This type of approach makes up a small minority of proposed FWI strategies.

Many different formulations of ϕ_D have been proposed, most often with the goal of reducing the tendency of FWI to converge to local minima caused by cycle-skipping. Unfortunately, there is no one objective function which is conclusively best at avoiding cycle-skipping, so choosing a generally acceptable cycle-skipping based objective function is difficult. Almost all of these approaches, however, are used only at the early stages of an inversion; it is common when using these methods to switch to a faster-converging ϕ_D when the cycle-skipping minima have been bypassed (e.g. Bozdağ et al., 2011; Van Leeuwen and Herrmann, 2013; Engquist and Froese, 2014). Commonly in these approaches, the later-stage ϕ_D is the one originally proposed by Tarantola (1984):

$$\phi_D = \sum_{n=1}^{N_F} \sum_{m=1}^{N_S} \frac{1}{2} ||\mathbf{R}\mathbf{u}_{n,m} - \mathbf{d}_{n,m}||^2, \tag{2.14}$$

where N_F is the number of discrete frequencies considered, N_S is the number of sources considered, $\mathbf{R}$ is a matrix applying the receiver sampling, $\mathbf{d}$ is the measured data, and subscripts denote the frequency and seismic source considered. As my focus here is not on cycle-skipping issues, this choice of ϕ_D is appropriate for the problems I consider here, and I use it throughout the thesis.

The choice of the prior-fitting objective term ϕ_P is less simple, as the details and type of the prior information to be considered are typically highly case-dependent. Approaches for including prior information in geologic settings including salt bodies, for instance, have been closely studied (e.g. Esser et al., 2018; Kalita et al., 2019), but these approaches are typically not appropriate in other settings. Because I do not want to restrict the approaches discussed to any particular case, I

will largely neglect the ϕ_P term, introducing it primarily to constrain recovered medium properties to physically reasonable values. In each chapter, I will introduce any ϕ_P terms used.

2.4 Model parameterization

The parameterization of the wavefield **u** in equation 2.1 is largely dictated by the relevant differential equations and the finite-difference approach I use to solve them: the grid on which I define the finite-difference equations is necessarily the grid on which **u** is recovered, which represents the appropriate wave feature from equation 2.7 or 2.12. There is substantially greater freedom in FWI when deciding how the subsurface model, **m**, is parameterized. Certainly, each cell of the finite-difference grid must have all the appropriate properties defined to create the Helmholtz matrix **S**, but this need not be the same way that the model is characterized in the optimization problem of equation 2.1. This distinction is not trivial: different model parameterizations change the number and meaning of variables in the optimization procedure in ways that can dramatically change the output of the inversion. Broadly speaking, two types of decisions need to be made when parameterizing an FWI model: which physical properties are represented by each variable, and which spatial region a variable represents those properties in. Both decisions can have large effects on an inversion result, though the former has been studied in much greater detail than the latter.

Property-type re-parameterization has been studied in detail in multi-parameter FWI, with the objective of reducing the extent of cross-talk, wherein different physical properties are confused in the inversion. Parameterizations which are better at separating the effects of different physical properties into different subsets of data can be effective in reducing cross-talk. This type of re-parameterization for cross-talk avoidance is especially well developed for elastic FWI (e.g. Tarantola, 1986; Operto et al., 2013; Vigh et al., 2014). In the elastic case, the variables which are most often considered are P-wave squared slowness $\left(\frac{1}{v_P}^2\right)$, S-wave squared slowness $\left(\frac{1}{v_S}^2\right)$, density ($\rho$), and both P- and S-impedances (ρv_P,ρv_S). Different combinations of these variables have been effective in different settings. In this thesis, the parameterization I consider uses vari-

ables defining squared P-wave slowness at a reference frequency, the reciprocal of P-wave quality factor $\left(\frac{1}{Q_P}\right)$, and, for the anelastic case, squared S-wave slowness at a reference frequency, the reciprocal of S-wave quality factor $\left(\frac{1}{Q_S}\right)$ and density. I choose this parameterization of the elastic variables because 1) I am not primarily interested in cross-talk between the elastic variables, 2) squared-slowness formulations have been shown to have better convergence than velocity formulations (Anagaw, 2014), and 3) there is no clear consensus that the elastic properties I have chosen are more prone to cross-talk than others. I make this decision for the Q variables because $\frac{1}{Q}$ is well defined in the zero-attenuation case, unlike Q, and it appears simply in the forward modeling equations I consider, like the squared slownesses. I do not consider variables which hybridize attenuation and elastic properties because I am not familiar with any rationale supporting the idea that such a combination would be effective.

Another option in FWI parameterization is to consider only certain physical properties as variable in the inversion, while leaving others fixed. This type of approach allows for accurate models of wave-propagation to be considered without introducing the complexity of additional variables to the optimization problem. These advantages come with the drawback of potentially misrepresenting the fixed properties. I am interested in FWI approaches which characterize attenuation, so I do not neglect Q variables in this way here.

Space-type parameterizations have been little investigated relative to their property-type counterparts, but represent an even greater degree of variability. Most often in FWI, the variables in the inversion are considered to represent physical properties on the same spatial grid used for forward modeling. This parameterization likely allows for maximal spatial resolution in results, but also causes the inversion problem to have an extremely large number of variables. This type of parameterization is highly redundant in some respects, as it defines variables on a scale much smaller than the seismic wavelengths considered. Parameterizations which define fewer variables on a larger scale can allow for relatively low dimensional, long-wavelength models to be reconstructed at low computational cost (e.g. Debens et al., 2015). More sophisticated approaches can

use prior information about the subsurface to create sparse, efficient, representations of high resolution models (for instance, by defining the model in terms of layer interface locations) (e.g. Ma et al., 2012). In this thesis, I consider the simple parameterization defining a set of variables at each finite-difference grid cell because 1) this approach is common, 2) I do not want to restrict the spatial resolution of the models recovered, and 3) I do not want to make the type of structural assumptions about the subsurface needed for more efficient representations.

Because the derivatives of ϕ with respect to the model parameters play a major role in FWI, it is generally helpful to consider scaled variables in order to ensure that objective function derivatives have comparable amplitudes. I apply scaling terms in this thesis to avoid the numerical difficulties that can arise when comparing variables with highly dissimilar magnitudes of impact on the objective function. I often refer to the variables used without mention of the scaling terms or the exact formulation of the inversion variables for simplicity, describing, for instance, the inversion variables defining $\zeta\left(\frac{1}{Q_P}\right)$ (where ζ is a scaling term) as 'the Q_P variables'.

2.5 Numerical optimization

In the previous sections, I have detailed the choices made in my implementations of FWI for each of the terms appearing in equation 2.1, specifically the wave-equation constraint, C, the wavefield $\mathbf{u}$, the objective function, ϕ, and the subsurface model, $\mathbf{m}$. The remaining key feature of equation 2.1, and the engine of the FWI problem is the minimization procedure itself. In this section I will briefly explain some of the key optimization strategies used in FWI in general, and in this thesis specifically.

2.5.1 Steepest descent optimization

The steepest descent (SD) or gradient descent optimization strategy was the first proposed driver for the FWI problem (Tarantola, 1984). In this optimization strategy, ϕ is decreased at each iteration by updating the model in the direction anti-parallel to the gradient of the objective function ϕ with

respect to the model parameters, $\mathbf{m}$. This type of approach can be expressed as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \frac{\partial \phi(\mathbf{m}_k)}{\partial \mathbf{m}}, \tag{2.15}$$

where $\mathbf{m}_k$ is the current subsurface model, $\mathbf{m}_{k+1}$ is the updated model, and α is a step-length which approximately minimizes ϕ in the model-space direction considered. I will discuss the strategy I use to determine α and how $\frac{\partial \phi(\mathbf{m}_k)}{\partial \mathbf{m}}$ can be calculated later in this Chapter, here our focus is on the optimization strategy itself. The SD approach has the advantages of being simple to implement and requiring few assessments of ϕ or its derivatives (which are the main computational costs for FWI) in each iteration. Unfortunately, this approach is slow to converge to a solution (e.g. Pratt et al., 1998; Métivier et al., 2013), and extremely prone to cross-talk, as the consideration of only first-derivatives in the approach can cause variable changes with similar effects on the data to be confused with one another; there is no mechanism to account for how changes in one variable modify the suitability of changes in another, except for iteration.

2.5.2 Newton optimization

To deal with the problems which hamper SD FWI, the inversion is often implemented using optimization strategies that incorporate second-derivative information. An effective approach for including this type of information is Newton optimization, with the updating strategy

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \left(\frac{\partial^2 \phi}{\partial \mathbf{m}^2} \right)^{-1} \frac{\partial \phi(\mathbf{m}_k)}{\partial \mathbf{m}}. \tag{2.16}$$

This approach effectively incorporates second-derivative information, which can dramatically improve the ability of the inversion to anticipate how changing one variable will alter the derivative of another. This allows for substantial cross-talk reduction (Innanen, 2014). Unfortunately, the large dimensionality of the FWI problem means that $\frac{\partial^2 \phi}{\partial \mathbf{m}^2}$ (also called the Hessian) has an extremely large associated storage and computation cost due to its $N_{\mathbf{m}}^2$ elements, where $N_{\mathbf{m}}$ is the number of variables in $\mathbf{m}$. Instead, FWI approaches usually make use of optimization strategies that approximate the Newton step. In the following sections I discuss two important approaches of this type: L-BFGS optimization, and truncated Newton optimization.

2.5.3 L-BFGS optimization

In this section, I provide a brief explanation of L-BFGS optimization. My discussion follows that of Nocedal and Wright (2006), where the reader can find a more detailed explanation. The Broyden, Fletcher, Goldfarb and Shanno (BFGS) optimization strategy (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) approximates the Newton update through exact solution of an approximation of the Newton system (equation 2.16), which can be stated as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha B_k \frac{\partial \phi(\mathbf{m}_k)}{\partial \mathbf{m}}, \tag{2.17}$$

where B_k is the BFGS approximation to the inverse Hessian matrix, $\frac{\partial^2 \phi}{\partial \mathbf{m}^2}^{-1}$. In the BFGS approach, the matrix B_k is designed to satisfy two key conditions. The first is that B_k should satisfy the condition

$$B_k^{-1}[\mathbf{m}_k - \mathbf{m}_{k-1}] = \frac{\partial \phi(\mathbf{m}_k)}{\partial \mathbf{m}} - \frac{\partial \phi(\mathbf{m}_{k-1})}{\partial \mathbf{m}}, \tag{2.18}$$

just as the inverse Hessian would for an objective function with a constant second derivative. The second condition is that of all the matrices satisfying equation 2.18, B_k is the closest to B_{k-1} by the metric of a weighted Frobenius norm. While the specific metric used for this closeness criterion is important in distinguishing the BFGS method from similar approaches, the key feature of this requirement for our purposes is that it ensures that B_k will be in as close agreement with B_{k-1} as possible while still satisfying equation 2.18. By combining these two conditions, the BFGS approximation to the inverse Hessian generally improves as iterations are completed, gathering more information about the second derivatives of the objective function at each step.

The BFGS method tends to have very fast, cost-effective convergence, but has the disadvantage of large memory requirements, especially if the elements of B_k are explicitly calculated and stored. In FWI, for instance, a matrix with a size of B_k (or equivalently, the Hessian) is generally not feasible to handle. In response to this issue, a limited memory BFGS (L-BFGS) approach has been developed (Liu and Nocedal, 1989). In this approach, only previous models and gradients are stored, and direct calculation of B_k is avoided; instead only the products of this matrix with vectors are calculated. If all of the previous gradients and models are stored, this approach is just

as effective as BFGS, but with dramatically reduced memory requirements. Often in L-BFGS, memory requirements are further reduced by storing only the last n gradients and models, and using the BFGS approximation as if the $k-n$th iteration were the first iteration. This approach can still be highly effective, even when n is as small as five. In this thesis, I will use the L-BFGS approach with storage of all previous models and gradients as these represent a relatively small memory cost in the models I consider.

In order to initialize BFGS or L-BFGS, it is necessary to define an initial approximation to the inverse of the Hessian as B_0. An appropriate approximation suggested by Nocedal and Wright (2006) is to use the model locations and objective function gradients at the last two iterations to generate an appropriate scaling for the identity matrix I:

$$B_0^k = I\frac{\Delta\mathbf{m}_k^T\Delta\mathbf{g}_k}{\Delta\mathbf{g}_k^T\Delta\mathbf{g}_k}, \tag{2.19}$$

where ^{T} denotes the transpose of a vector or matrix,

$$\Delta\mathbf{m}_k = \mathbf{m}_k - \mathbf{m}_{k-1} \tag{2.20}$$

and

$$\Delta\mathbf{g}_k = \frac{\partial\phi(\mathbf{m}_k)}{\partial\mathbf{m}} - \frac{\partial\phi(\mathbf{m}_{k-1})}{\partial\mathbf{m}}. \tag{2.21}$$

While this approach results in a different B_0^k at each iteration, this doesn't present a problem in the L-BFGS approach, and it allows the L-BFGS approximation to change faster if the curvature of the objective function changes. At the first iteration of the L-BFGS procedure, where such an initial estimate cannot be created, I use a SD step instead of equation 2.17.

The L-BFGS approach is widely used in FWI, both for single parameter and multiparameter inversion. It has been shown to produce significantly faster convergence than SD optimization and, in the FWI problem, incurs little additional computational cost as the vector-products that are assessed to evaluate B_k are computationally negligible in comparison to the evaluation of ϕ and its derivative (which I discuss later). The main disadvantage of L-BFGS as compared to SD in FWI is the larger memory cost, but because this can be varied as needed, this approach is typically

preferred to SD. A disadvantage of L-BFGS in FWI relative to other optimization strategies is that B_k gains second-derivative information iteratively. This means that early steps can be very similar to SD steps, and introduce substantial cross-talk. It has been observed in FWI that second-derivative information can be more effective in preventing cross-talk than in removing it from a model. This may limit the effectiveness of the L-BFGS approach in producing cross-talk free results.

2.5.4 Truncated Newton optimization

The truncated-Newton (TN) approach to optimization uses iterative methods to find a solution to the Newton system (equation 2.16), but stops iteration (or 'truncates') before an exact solution has been determined. This means that truncated Newton optimization contains two nested iterative optimization problems. The first, often referred to as the 'outer loop', applies the truncated Newton updates:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha \mathbf{p}_k. \tag{2.22}$$

The second, referred to as the 'inner loop', determines the model update direction, $\mathbf{p}_k$, through approximate solution of

$$\mathbf{p}_k = \min_{\mathbf{p}} \frac{1}{2}\mathbf{p}^T H(\mathbf{m}_k)\mathbf{p} + \mathbf{p}^T \mathbf{g}, \tag{2.23}$$

where H represents the second derivatives of the objective function (the Hessian), and $\mathbf{g}$ represents the first derivatives (gradient). The inner loop problem has an exact solution given by $\mathbf{p}_k = -H^{-1}\mathbf{g}$: the model update satisfying the Newton system, and, as the minimum in equation 2.23 is approached, the model update approaches the Newton update (Nocedal and Wright, 2006). The inner loop is not solved exactly in TN due to computational constraints, but is instead terminated based on stopping conditions. These stopping conditions are typically set as 1) a maximum on the number of optimization iterations used in the inner loop, and 2) a tolerance for the Newton system residual, $\mathbf{r}_N = ||H\mathbf{p}_k + \mathbf{g}||$, with iteration being terminated when $\mathbf{r}_N$ falls below a threshold amplitude. In my implementation of this approach, I have found that the iteration maximum is a

powerful tool for controlling the accuracy and computational cost of the method, while effective tolerance values are typically so low that they are generally not reached unless a very large iteration maximum is considered. Consequently, I implement the inner loop tolerance only in Chapter 3, where a very large maximum on inner loop iterations is set. In Chapters 4 and 5, I use the iteration maximum as the only stopping criterion. I begin each inner loop with the estimate $\mathbf{p} = 0$.

The conjugate gradient approach is often used as the optimization strategy for the inner-loop problem in equation 2.23 (Nocedal and Wright, 2006; Métivier et al., 2013). In my implemenation of this approach, I instead use the L-BFGS approach. I do this because the L-BFGS approach typically has faster convergence and the memory costs are not too high for the FWI problem (as this approach is often used as the main FWI driver). The L-BFGS approach used for this application is identical to the one explained above, with the exception that the line-search step used to determine α can be replaced with an exact calculation for the linear problem:

$$\alpha = -\frac{(H(\mathbf{m}_k)\mathbf{p} + \mathbf{g}(\mathbf{m}_k))^T \mathbf{p}}{\mathbf{p}^T H(\mathbf{m}_k)\mathbf{p}}. \tag{2.24}$$

The TN approach directly approximates a solution to the Newton sytem at each iteration, as compared to the L-BFGS approach, which exactly solved an approximate version of the Newton sytem. For practical purposes, this means that while the L-BFGS approach relies on an incrementally improving approximation to the Hessian applied during the optimization, the TN approach allows for a chosen level of approximation to the Hessian to be applied at each step of the inversion. This tunability means that the TN approach can work with an effective level of second-derivative information even at the first step of the optimization, something the L-BFGS approach cannot do.

2.5.5 Line-search

An important step in the optimization strategies outlined above is the line-search step, where the step-length α is determined. An effective line-search strategy is important, because ineffective strategies can dramatically increase the cost of optimization by requiring many additional objective function evaluations at each iteration. I use the line-search approach described by Nocedal and

Wright (2006). In this approach, the line-search is terminated when the two Wolfe conditions are satisfied. The first of these conditions requires that the objective function decrease by at least a small fraction (I use 0.001) of the expected decrease for that step, defined as the product $\frac{d\phi}{d\alpha}\alpha$. The second condition requires that the derivative $\frac{d\phi}{d\alpha}$ decrease to a fraction of the original amplitude (I typically use 0.9). The approach defines an upper and lower bound on the step length, with the lower bound starting at 0. If, based on the objective values and derivatives observed, the bounds do not conclusively contain a minimum, and the first Wolfe condition is satisfied, the upper bound is doubled. If the first Wolfe condition is not satisfied, the upper bound is halved. If the bounds do conclusively contain a minimum (negative derivative on one bound and positive on the other) and the first Wolfe condition is satisfied, an expected minimum based on a cubic interpolation between the bounds is used to replace one of the bounds. When one of these bounds satisfies both the Wolfe conditions, the line-search accepts that value. I have found that, with this line-search approach, only one or two objective function evaluations are typically necessary in each line search when using TN or L-BFGS optimization for FWI.

2.6 Objective function derivative evaluation

Each of the optimization techniques I describe above (and each of the optimization strategies used in almost all FWI implementations) is a local optimization strategy, using the derivatives of the objective function to locate better models. In FWI, evaluating these derivatives constitutes the major computational cost. Almost all FWI approaches require the calculation of the gradient of the objective function with respect to the model parameters. Truncated-Newton approaches and several uncertainty quantification techniques also require the calculation of the product of the Hessian matrix with one or more vectors. In this section, I briefly detail an adjoint-state method for the evaluation of these derivatives. In presenting these expressions, I closely follow Métivier et al. (2013). The objective function I will consider here will be the L_2 data-fit objective in equation 2.14. The structure of the derivatives is typically very similar for other data-fit objective function

terms, and prior-fitting objective function terms generally do not require wavefield modeling, and are consequently much simpler to calculate.

2.6.1 Gradient

After replacing the expressions in equation 2.1 with the specific choices I use in this thesis, the FWI problem becomes

$$\mathbf{m}_{min} = \min_{\mathbf{m}} \sum_{j=1}^{N_F} \sum_{k=1}^{N_S} \frac{1}{2} ||\mathbf{R}\mathbf{u}_{j,k} - \mathbf{d}_{j,k}||^2, \text{ subject to } \mathbf{S}(\mathbf{m})\mathbf{u} = \mathbf{f}. \tag{2.25}$$

To simplify the following discussion, I will consider the gradient calculation at a single frequency, and allow for $\mathbf{u}$, $\mathbf{d}$, and $\mathbf{f}$ to stand for matrices representing the wavefields, measured data and sources, with each column representing one source, and each row representing one location in space. The full gradient can be determined through a sum of single-frequency gradients. The Lagrangian of this problem is

$$L = \frac{1}{2}||\mathbf{R}\mathbf{u} - \mathbf{d}||_2^2 + < \mathbf{S}(\mathbf{m})\mathbf{u} - \mathbf{f}, \kappa >, \tag{2.26}$$

where κ is an as yet unconstrained Lagrange multiplier, and $<,>$ represents an inner product, such that $< \mathbf{a}, \mathbf{b} >= \sum_{k=1}^{N_S} \mathbf{a}_k^\dagger \mathbf{b}_k$. If $\mathbf{u} = \bar{\mathbf{u}}$, where $\bar{\mathbf{u}}$ satisfies $\mathbf{S}(\mathbf{m})\bar{\mathbf{u}} = \mathbf{f}$, then $L = \phi_D$. The derivative of ϕ_D with respect to $\mathbf{m}$ is then equal to the derivative of L with respect to $\mathbf{m}$ at $\mathbf{u} = \bar{\mathbf{u}}$. This derivative is

$$\frac{d\phi}{d\mathbf{m}} = \frac{dL(\bar{\mathbf{u}})}{d\mathbf{m}} = \frac{\partial L}{\partial \bar{\mathbf{u}}} \frac{\partial \bar{\mathbf{u}}}{\partial \mathbf{m}} + \frac{\partial L}{\partial \mathbf{m}}. \tag{2.27}$$

Because the term $\frac{\partial \bar{\mathbf{u}}}{\partial \mathbf{m}}$ is prohibitively expensive to calculate, it is desirable to choose $\kappa = \bar{\kappa}$ such that the term $\frac{\partial L}{\partial \bar{\mathbf{u}}}$ becomes zero. This choice will mean that

$$\frac{\partial L}{\partial \bar{\mathbf{u}}} = \mathbf{R}^T(\mathbf{R}\bar{\mathbf{u}} - \mathbf{d}) + \mathbf{S}^\dagger \bar{\kappa} = 0. \tag{2.28}$$

Equation 2.28 shows that $\bar{\kappa}$ can be calculated by forward modeling of the data residuals. The choice of $\bar{\kappa}$ reduces 2.27 to

$$\frac{dL}{d\mathbf{m}} = \frac{\partial L}{\partial \mathbf{m}} =< \frac{\partial \mathbf{S}}{\partial \mathbf{m}} \bar{\mathbf{u}}, \bar{\kappa} > . \tag{2.29}$$

Because $\frac{\partial \mathbf{S}}{\partial \mathbf{m}}$ is typically very sparse, this can be quickly calculated for many variables m, with the main cost being the calculation of $\bar{\mathbf{u}}$ and $\bar{\kappa}$.

2.6.2 Hessian-vector product

While an expression for the product of the Hessian matrix with a vector can be determined using the adjoint state method (Métivier et al., 2013), in this thesis I will instead use the Gauss-Newton approximation to the Hessian when evaluating such products. In the Gauss-Newton approximation, it is assumed that the residual of the objective function, in this case $\gamma = R\mathbf{u} - \mathbf{d}$, is small, or equivalently, the current model is close to the solution. The resulting approximate Hessian is always positive definite, and has been suggested to allow for faster inner-loop convergence in the TN method, leading to faster FWI convergence (e.g. Métivier et al., 2013). I refer to the TN method when using the Gauss-Newton approximation as the truncated Gauss Newton (TGN) method.

The main term of the objective funtion in equation 2.25 is of the form $\phi = \frac{1}{2}\gamma^2$. The second derivative of an objective function like this with respect to model parameter $\mathbf{m}$ is

$$\frac{\partial^2 \phi}{\partial \mathbf{m}^2} = \frac{\partial^2 \gamma}{\partial \mathbf{m}^2}^T \gamma + \frac{\partial \gamma}{\partial \mathbf{m}}^T \frac{\partial \gamma}{\partial \mathbf{m}} \tag{2.30}$$

Under the Gauss-Newton approximation, I neglect the first term on the right-hand side in equation 2.30 due to the assumption that γ is small.

For the FWI objective I define in equation 2.25, the Gauss-Newton Hessian is given by the relation

$$J^\dagger \mathbf{R}^T \mathbf{R} J, \tag{2.31}$$

where J is the Jacobian matrix $\frac{d\mathbf{u}}{d\mathbf{m}}$. The matrix J is too costly to directly calculate in FWI, but we can use the adjoint state method to avoid the need for such a calculation. We can begin by noting that the derivative of the function

$$\mathbf{h} = < \mathbf{u}(\mathbf{m}), \mathbf{w} >, \tag{2.32}$$

where $\mathbf{w}$ is an arbitrary vector, with respect to $\mathbf{m}$ is

$$\nabla \mathbf{h} = J^\dagger \mathbf{w}. \tag{2.33}$$

Consequently, if $\mathbf{w}$ is chosen to be $\mathbf{R}^T\mathbf{R}J\mathbf{v}$, then the Hessian-vector product $H_{GN}\mathbf{v}$ is equal to the derivative of $\mathbf{h}$. This derivative can be calculated in exactly the same way as the gradient was by considering the Lagrangian

$$L =< \mathbf{u}(\mathbf{m}), \mathbf{w} > + < \mathbf{S}(\mathbf{m})\mathbf{u} - \mathbf{f}, \xi > \tag{2.34}$$

instead of equation 2.26. The same procedure follows, but instead of requiring Lagrange multiplier $\bar{\kappa}$ to satisfy equation 2.28, the removal of $\frac{\partial \bar{\mathbf{u}}}{\partial \mathbf{m}}$ requires that the Lagrange multiplier for this problem, ξ, satisfies

$$\mathbf{S}^{\dagger}\bar{\xi} = -\mathbf{w}. \tag{2.35}$$

As before, if the Lagrange multiplier satisfies this condition, then

$$\frac{dL}{d\mathbf{m}} = \frac{\partial L}{\partial \mathbf{m}} =< \frac{\partial S}{\partial \mathbf{m}}\bar{\mathbf{u}}, \bar{\xi} > . \tag{2.36}$$

The calculation of $\bar{\xi}$ does require that $\mathbf{w}$ is known, however, and we cannot directly calculate J. We can calculate the product of J with the vector $\mathbf{v}$ through consideration of the derivative of the forward problem with respect to variables m_i multiplied by vector elements v_i. Using the relation $\mathbf{Su} - \mathbf{f} = 0$, it can be shown that

$$\frac{\partial(\mathbf{Su} - \mathbf{f})v_i}{\partial m_i} = \mathbf{S}\left(\frac{\partial \mathbf{u}}{\partial m_i}v_i\right) + \mathbf{u}\left(\frac{\partial \mathbf{S}}{\partial m_i}v_i\right) = 0. \tag{2.37}$$

A sum over i for the term $S\left(\frac{\partial \mathbf{u}}{\partial m_i}v_i\right)$ is equal to the product $S(J\mathbf{v})$, so equation 2.37 is equivalent to

$$\mathbf{S}(J\mathbf{v}) = -\mathbf{u}\sum\left(\frac{\partial \mathbf{S}}{\partial m_i}v_i\right) \tag{2.38}$$

This equation can be solved for the product $J\mathbf{v}$. Using this term to define $\mathbf{w}$ as $\mathbf{R}^T\mathbf{R}J\mathbf{v}$, we can solve equation 2.35 for ξ. With this choice for ξ, the derivative $\frac{dL}{d\mathbf{m}}$ from equation 2.36 becomes the Gauss-Newton Hessian vector product $H_{GN}\mathbf{v}$.

The major computational costs required in the assessment of the derivatives above comes when solving wave-propagation equations. With the formulation of the FWI problem I present, this is equivalent to solving equations of the form $\mathbf{Sa} = \mathbf{b}$ for $\mathbf{a}$. The large computational costs associated

with wave propagation make other computational steps in the derivative evaluations negligible. In the assessment of the gradient, two such propagation problems must be solved: one for the evaluation of $\bar{\mathbf{u}}$ and one for the evaluation of $\bar{\kappa}$ (equation 2.28). The assessment of the Gauss-Newton Hessian product with a vector likewise requires two wavefield propagation problems to be solved: one for $\bar{\bar{\xi}}$ in equation 2.35 and one for $J\mathbf{v}$ in equation 2.38. Where possible, I will quantify the computational cost of inversions in terms of wavefield propagation problems. Although in the formulation I present here, solving $\mathbf{Sa} = \mathbf{b}$ for for $\mathbf{a}$ actually represents several wave-propagation problems - one for each source considered (equivalently one for each column of $\mathbf{a}$) - I refer to each such solution as one wave-propagation problem. I do this because all the problems of this form required in the derivative calculation have the same number of columns for the $\mathbf{a}$ matrices, so the costs are about equivalent, and because in the frequency-domain formulation I consider here, the cost of solution for additional columns is less than linear (computational cost is approximately proportional to $\log n$, where n is the number of sources).

2.7 Conclusions

In this chapter, I presented a general form of the FWI problem (equation 2.1), that is representative of most, though not all, implementations of FWI. I then explained the major decisions I make in this thesis to obtain specific implementations of the FWI approach. For forward modeling, I use isotropic viscacoustic and viscoelastic frequency-domain finite-difference modeling approaches, allowing for the treatment of attenuation in seismic wave-propagation. I use a conventional objective function, separated into prior-fitting and L_2 data-fitting terms. I parameterize the inversion in terms of two physical properties in the viscoacoustic case and five in the viscoelastic case, considering squared slownesses, reciprocal Q and density, all defined on the same grid as the one used for forward modeling. I use a truncated-Newton optimization strategy with a Gauss-Newton approximation of the Hessian in most cases, though the steepest-descent method is also used for comparison in Chapter 4. The details of the inversion I describe in this section are not original

work of mine (though I am not familiar with another example of L-BFGS optimization being used in the inner loop of TGN for FWI); the later sections of this thesis represent the innovations I have made.

With the exception of the choice to implement forward-modeling strategies which include seismic attenuation, I do not think that any of the conclusions I draw in the following chapters are specific only to the FWI implementation I have described here. While I only describe results based on the above formulation, my experience with FWI is that changes in the parameterization, objective function, and forward modeling do not change the inversion problem so dramatically that I would expect the findings I describe in Chapters 3-5 to be significantly different with another formulation.

Chapter 3

Parameter cross-talk and modeling errors in viscoacoustic seismic full-waveform inversion

3.1 Summary

Simultaneous use of data within relatively broad frequency bands is essential to discriminating between velocity and Q errors in the construction of viscoacoustic full-waveform inversion (QFWI) updates. Individual frequencies or narrow bands in isolation cannot provide sufficient information to resolve parameter cross-talk issues in a surface seismic acquisition geometry. At the same time, too broad a frequency band introduces significant problems in the presence of modeling errors. The risk of modeling errors arising in viscoacoustic full-waveform inversion is high, because of both the range of very different geological contributors to attenuation and dispersion, and the variety of available mathematical descriptions. I show numerical tests which suggest that by relaxing the typical requirement that the frequency dependence of the assumed intrinsic attenuation model be self-consistent across the full spectrum, significant improvement in the fidelity of inversion results can be obtained in cases where the attenuation model assumed in the inversion differs substantially from the true subsurface behaviour. I find that the size of the frequency bands used in this inversion approach is a useful hyperparameter controlling trade-off between cross-talk reduction and flexibility in coping with uncertain physics.

3.2 Introduction

Intrinsic attenuation and dispersion play large and important roles in seismic wave propagation, making their inclusion in FWI a significant priority. Unfortunately, there are few theoretical constraints on the frequency dependence of attenuation and dispersion in the subsurface. This is prob-

lematic, because a seismic experiment does not have sufficient information content to constrain attenuation as an arbitrary function of frequency at each point in space. The requirement of causal seismic wave propagation necessitates that attenuation and phase velocity satisfy the Kramers-Kronig relation (e.g. Aki and Richards, 2002), constituting a required relation between the Hilbert transform of the attenuation and the phase velocity. Because the Hilbert transform involves an integral over all frequencies, however, this requirement provides little constraint in inversion of finite-band seismic data.

Relations between attenuation and phase velocity for specific seismic attenuation mechanisms can be parameterized in relatively small numbers of variables (e.g. Liu et al., 1976). Observed seismic attenuation is affected by many different mechanisms, however, including pore-, wavelength- and meso-scale fluid flow attenuation mechanisms. Furthermore, variations in the sizes of meso-scale heterogeneities in the subsurface cause many different mechanisms at the (likely dominant) meso-scale to be active at once (Pride et al., 2004). Because of the large number of attenuation mechanisms, even with a very good understanding of how attenuation occurs, the dimensionality of the QFWI problem cannot be reduced to manageable levels through theoretical considerations alone.

Because theoretical descriptions of seismic attenuation do not significantly constrain the frequency dependence of attenuation and phase velocity, it is necessary to make assumptions about this frequency dependence in order to frame a tractable inversion problem. Most commonly, the empirical observation of Q which does not consistently increase or decrease over the seismic frequency band (Aki and Richards, 2002) is used to motivate the assumption of a nearly-constant Q. Several nearly-constant Q approximations exist (Ursin and Toverud, 2002); in this thesis I use the Kolsky-Futterman model (Kolsky, 1956; Futterman, 1962), as described in Chapter 2. Most nearly-constant Q models reduce the problem of parameterizing the attenuation and velocity at all frequencies to a problem of parameterizing two frequency-independent properties in the constant-density viscoacoustic case (five for the viscoelastic case).

Crucially, it is assumed in FWI that a complete physics model governing wave propagation is known and accurately reproduced in the forward modeling. If the wave propagation equations miss, or incorrectly model, important features of the data, modeling errors will result. These errors introduce incorrect and often dramatically un-physical spatial arrangements of the available model parameters. In many formulations of FWI where it is well understood which features of the data can be accurately reproduced with the assumed physics, data which is likely to be misleading can be de-emphasized in the objective function in an attempt to avoid this type of behavior. Modeling of attenuation is challenging from this standpoint because, if the assumptions used to obtain manageable parameterizations of attenuation fail, then features of the data which the inversion is explicitly attempting to account for may be mishandled in the inversion. The consequences of such a failure could be significant if the true subsurface attenuation and dispersion differ substantially from the model assumed. A FWI formulation which navigates a route between the extremes of unmanageable parameterization and unwarranted physics assumptions is desirable when confronted with uncertainty in the attenuation physics.

In this chapter, I formulate a frequency-domain QFWI problem in which P-wave velocity and Q are simultaneously updated. Through numerical examples, I demonstrate the importance of a broad range of frequencies in mitigating parameter cross-talk in QFWI, assuming that the observed data are generated following substantially the same mathematical model of attenuation and dispersion as that assumed in the computation of the gradient and residuals. I then demonstrate that such cross-talk mitigation is only effective in the absence of modeling error. In response to this, I adjust the formulation to accommodate uncertainty in the attenuation physics. I investigate the idea of relaxing the constraint that the assumed physics be exactly obeyed by allowing a band-wise frequency dependence in the recovered model. This increased flexibility offers important benefits when the assumed physical model involves different frequency dependence of wave propagation from that of the true (or at least a different, more appropriate) physical model.

3.3 Theory

3.3.1 Viscoacoustic full-waveform inversion formulation

My focus in this chapter is on the effects of errors in the attenuation physics model on QFWI results and on approaches for mitigating these results. To facilitate this investigation, I consider viscoacoustic wave propagation: a relatively simple model of seismic wave propagation which allows for the effects I am interested in to be investigated with minimal complications. The frequency-domain partial differential equation describing this model of wave propagation was described in Chapter 2 as

$$\left[\omega^2 s(\mathbf{r},\omega) + \nabla^2\right] u(\mathbf{r},\omega) = f(\mathbf{r},\omega), \tag{3.1}$$

where u is the pressure field, f is a source term, and the model parameter s depends on the dispersive velocity and attenuation:

$$s(\mathbf{r},\omega) = \left[c(\mathbf{r},\omega)\left(1 - \frac{i}{2Q(\mathbf{r},\omega)}\right)\right]^{-2}, \tag{3.2}$$

where c is the frequency dependent phase velocity and Q is a quality factor. The viscoacoustic FWI problem is to determine the unknown spatial and frequency distributions of c and Q. Because the data available in seismic experiments do not allow for the recovery of c and Q as arbitrary functions of frequency, it is necessary to assume a particular physics model in order to do viscoacoustic inversion.

Many models of attenuation and dispersion exist and are in regular use for processing, imaging, and inverting seismic data (Ursin and Toverud, 2002). They tend to agree in their general reproduction of the amplitude and phase features of dissipating waves, but in their detailed predictions of, e.g., phase velocities at low frequencies they may differ widely. Many of these models require that Q be approximately constant over a given frequency range, based on empirical evidence. A constant Q in a non-dispersive medium violates causality (Aki and Richards, 2002), so in many models a frequency dependent Q, which is nearly constant over the range of seismic frequencies, together with a dispersion term, are adopted. There are many ways to create a function which is

nearly constant over the range of seismic frequencies, so there are many different nearly constant Q model types (Liu et al., 1976; Ursin and Toverud, 2002). In this thesis, I use the nearly constant Q model due to Kolsky and Futterman (Kolsky, 1956; Futterman, 1962), hereafter referred to as the KF model, in which

$$s(\mathbf{r},\omega) = \left[c_0(\mathbf{r})\left(1 + \frac{1}{\pi Q(\mathbf{r})}\log\frac{\omega}{\omega_0} - \frac{i}{2Q(\mathbf{r})}\right)\right]^{-2}, \tag{3.3}$$

where $c(\omega)$ is the wave velocity, ω_0 is a reference frequency and $c_0 = c(\omega_0)$, as described in Chapter 2. Inspection of equation 3.3 identifies a specific challenge that the QFWI problem faces. The size of the frequency dependent term in s, which produces dispersion, is determined by Q. In effect, both c_0 and Q co-determine the wave velocity at a given frequency. This makes considerable parameter cross-talk likely, and is suggestive that variations from one frequency to another will be instrumental in mitigating it.

3.3.2 Parameter cross-talk in QFWI

Parameter cross-talk is the phenomenon wherein data residuals introduced by an error in the estimate of one model parameter are attributed to errors in the estimate of another parameter. Cross-talk is a major concern in FWI, as it can severely harm the accuracy of the recovered model and the convergence of the inversion (e.g., Plessix et al., 2013; Innanen, 2014; Pan et al., 2016). Gradient-based updates are particularly vulnerable to cross-talk. This is because the gradient considers only the derivative of the objective function with respect to each variable parameterizing the model. If changes in several different variables can reduce the same part of the data residual, all will be changed in a gradient update. Methods approximating the Newton step are better able to prevent cross-talk (e.g., Métivier et al., 2013; Innanen, 2014), but their computational cost generally prohibits the calculation of most of the Hessian information, and a comprehensive appraisal step analyzing potential cross-talk is usually warranted.

The extent of expected cross-talk, and data sufficiency in general for FWI, are typically assessed through consideration of radiation patterns (or scattering potentials). Radiation patterns

represent the derivatives of seismic wavefields with respect to inversion variables. The radiation patterns of point perturbations in active FWI parameters, plotted as functions of experimental variables (e.g., angle between incoming and outgoing rays, frequency, etc.), are often used to determine the degree of expected parameter cross-talk in multi-parameter FWI (e.g., Tarantola, 1986; Virieux and Operto, 2009; Operto et al., 2013; Alkhalifah and Plessix, 2014). Parameters which generate potentials with proportional amplitude variations over a given range of these experimental variables are easily confused with one another. Scattering potentials represent the effective point-sources which would generate the corresponding radiation patterns. The scattering potentials I consider here are generated following the approach of Stolt and Weglein (2012). These potentials consider only terms linear in the model perturbations and represent perturbations to a wavefield in a homogeneous medium. The scattering potential $\mathscr{V}$ associated with the KF viscoacoustic wave equation can be determined by considering small perturbations to the model parameters in equation 3.1. For position $\mathbf{r}$ and frequency ω, this scattering potential is given by

$$\mathscr{V}(\mathbf{r},\omega) \approx -\frac{\omega^2}{c_{0_u}(\mathbf{r})^2}\left[V_Q(\mathbf{r},\omega)+V_{c_0}(\mathbf{r},\omega)\right], \tag{3.4}$$

where

$$V_Q(\mathbf{r},\omega)=\frac{\beta(\omega)}{Q_u(\mathbf{r})}\Delta Q(\mathbf{r}),\quad V_{c_0}(\mathbf{r},\omega)=\left(1+\frac{\beta(\omega)}{Q_u(\mathbf{r})}\right)\Delta c_0(\mathbf{r}), \tag{3.5}$$

Q_u and c_{0_u} represent the unperturbed Q and c_0 values, the $\Delta\cdot$ quantities,

$$\Delta Q(\mathbf{r})=1-\frac{Q_u(\mathbf{r})}{Q(\mathbf{r})},\quad \Delta c_0(\mathbf{r})=1-\frac{c_{0_u}(\mathbf{r})^2}{c_0(\mathbf{r})^2}, \tag{3.6}$$

represent localized jumps in their corresponding model parameters Q^{-1} and c_0^{-2}, and

$$\beta = i-\frac{2}{\pi}\log\left(\omega/\omega_0\right). \tag{3.7}$$

Cross-talk often occurs where similarities between two scattering potentials correspond with data measurements, as these represent different parameters having similar effects on the measured data. The key characteristics distinguishing V_Q and V_{c_0} are their frequency dependence (via the second term in equation 3.7) and their phase (via the first term in equation 3.7). In principle, either of

these can be invoked in an effort to combat cross-talk in QFWI. Hak and Mulder (2010) analyzed the phase term, showing that, at a single frequency, the angle-variations in scattering associated with the phase can suppress cross-talk, and distinguish velocity from Q. However, for these subtle phase differences to be used, sources and receivers must surround the region to be determined. This is necessary because, if a target is illuminated from one side only, the same phase perturbation can be caused by a position-velocity perturbation or by a Q perturbation. This ambiguity can be resolved if the target is illuminated from all sides, but is difficult to otherwise remove. Given limited acquisition geometries, consideration of phase differences alone cannot resolve cross-talk, and the only routes to effective cross-talk reduction require consideration of the frequency dependences of s_Q and s_{c_0}. Provided sets of frequencies are treated simultaneously, cross-talk resilient FWI updates from limited source-receiver geometries should be possible. Consequently, although the scattering angle is active in the problem of cross-talk reduction, as shown by Hak and Mulder (2010), *practical* QFWI must be protected from cross-talk using the frequency dependence associated with Q-dispersion. The frequency dependence of the scattering potentials is illustrated in Figure 3.1. By inspection of this plot, we can furthermore predict that if only a narrow range of frequencies is considered, over which the two scattering potentials vary roughly in proportion, c_0 and Q will be extremely difficult to distinguish. Over broader ranges of frequencies, however, significant differences between the two scattering potentials become prevalent, which should enable a QFWI iteration to create meaningful updates in both. This introduces a new feature to multi-scale FWI workflows, within which demands already exist on the bands of frequencies considered.

The approach I discuss here for coping with potential errors in modeling of attenuation physics produces another forcing effect on the size of frequency bands considered. In this approach, it is necessary to treat the width of the frequency bands as a trade-off parameter, balancing the suppression of modeling error (which will be better performed using narrower bands) with the suppression of parameter cross-talk (which will be better performed using broader bands).

It is notable that while the discussion of radiation patterns above suggests which data are nec-

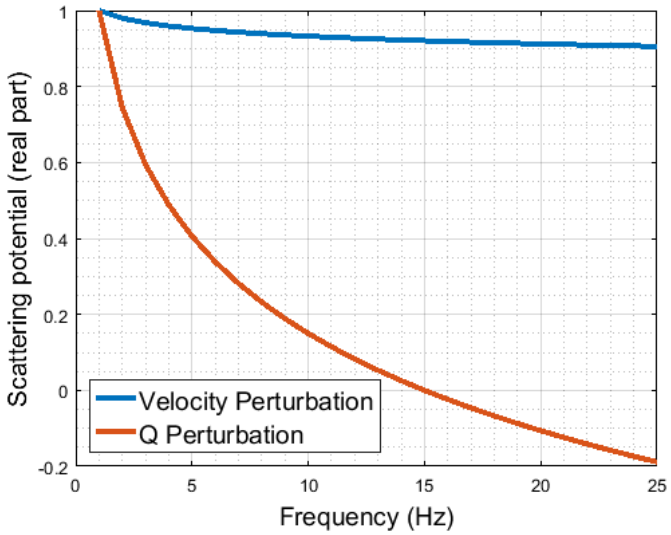

Figure 3.1: Amplitude of the scattering potential as a function of frequency for perturbations in c_0^{-2} (blue) and Q^{-1} (red). Amplitudes have been normalized to 1 at 1 Hz. Background $Q_0 = 20$, and the reference frequency was 15 Hz for this example.

essary for the avoidance of cross-talk, the scattering potential investigated do not present an obvious way of determining which FWI approaches will be effective in suppressing cross-talk. I will present and discuss a more effective approach for characterizing cross-talk in QFWI in Chapter 4.

3.3.3 Simulation of modeling error

A numerical investigation of the effects of errors in assumed attenuation physics is a major topic of this chapter. In order to simulate an error in the attenuation physics model chosen, however, it is necessary to assign both an assumed attenuation physics model, to be used in the inversion, and a true attenuation physics model, which will generate the synthetic data used in the inversion. The attenuation physics model used in this chapter to represent the true subsurface behavior is based on a standard linear solid (SLS) model (e.g. Carcione, 2007). It should be noted that the significance of this choice is not that the SLS is a better or worse match to data actually measured in seismic experiments, but rather that it produces different detailed predictions of wave behaviour than the one underlying the QFWI algorithm.

The SLS model is based on viscoelastic considerations, with a constitutive relation that is linear

in the stress, the strain, and their derivatives (Carcione, 2007). Continua are treated as consisting of a spring and dash-pot in series, in parallel with a second spring. The Q value given by this model is not constant with frequency, but is instead given by

$$Q(\omega) = \frac{1+\omega^2\tau_\varepsilon\tau_\sigma}{\omega(\tau_\varepsilon-\tau_\sigma)}, \tag{3.8}$$

where τ_ε and τ_σ are relaxation times related to the constants of the effective springs and dash-pot of the model (Carcione, 2007). This function is sharply peaked at $\omega = \tau^{-1}$, where $\tau = \sqrt{\tau_\varepsilon\tau_\sigma}$. The P-wave phase velocity for this model in terms of c_0 and ω_0 can be calculated by substituting equation 25 from Casula and Carcione (1992) into equation 8 from the same paper, giving

$$c(\omega) = c_0 \frac{\mathbb{R}\mathrm{e}\sqrt{\frac{1+i\omega_0\tau_\sigma}{1+i\omega_0\tau_\varepsilon}}}{\mathbb{R}\mathrm{e}\sqrt{\frac{1+i\omega\tau_\sigma}{1+i\omega\tau_\varepsilon}}}. \tag{3.9}$$

Many physical processes which could have significant impact on seismic wave attenuation are well modeled by the SLS (Liu et al., 1976). A system of several SLS mechanisms arranged in parallel, called a general standard linear solid, can model the effects of a system with many attenuation mechanisms. This introduces several relaxation mechanisms, and several attenuation peaks. If specific choices for the amplitudes and peak frequencies of these individual SLS components are made, a general SLS with approximately constant Q over a given bandwidth can be constructed. In this case, the dispersive behaviour of the velocity reduces to the KF model, as in equation 3.3, over the nearly constant Q frequency band. If seismic attenuation in the region considered is well represented by such a combination of SLS mechanisms, a KF-based QFWI procedure will suffer from little modeling error. The motivation of this chapter is to develop a methodology which limits modeling errors when the QFWI model (e.g., KF) and the actual model operating in the Earth are dissimilar, so this type of attenuation profile is inappropriate for our purposes here. To establish a model with dissimilarities from KF attenuation, I adopt a model involving only one SLS mechanism, representative of a single, or at least a dominant attenuation mechanism in the subsurface region of interest, in contrast with the constant Q behaviour assumed in the inversion. A comparison of KF and SLS Q and P-wave velocity is shown in Figure 3.2, where the models have the same

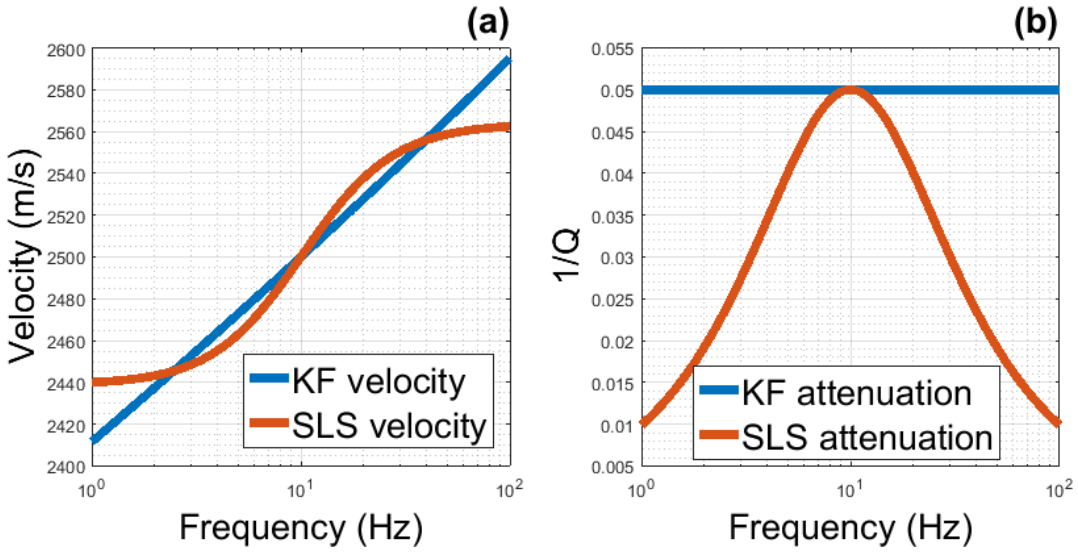

Figure 3.2: Comparison of SLS and KF models for velocity (a) and Q (b). Both models are chosen to have a velocity of 2500m/s and Q=20 at the reference frequency, 15 Hz. Note the semilog scale.

Q and P-wave velocity at 15 Hz.

3.3.4 Adapting QFWI to accommodate modeling errors

In Chapter 2, I introduced a general form of the FWI problem as

$$\mathbf{m}_{min} = \min_{\mathbf{m}} \phi(\mathbf{u},\mathbf{m}) \text{ subject to } C(\mathbf{u},\mathbf{m}) = 0. \tag{3.10}$$

In this equation, the requirement that the assumed wave equation be satisfied is enforced through the $C(\mathbf{u},\mathbf{m}) = 0$ constraint. I also showed that for the frequency-domain finite-difference forward modeling I use in this thesis, this constraint could instead be represented as

$$\mathbf{S}_1(\omega,\mathbf{m})\mathbf{u}(\omega) = \mathbf{f}(\omega) \tag{3.11}$$

for viscoacoustic inversion, where $\mathbf{f}$ is a source term, $\mathbf{u}$ is a wavefield, and $\mathbf{S}_1(\omega,\mathbf{m})$ is a matrix that applies a finite-difference stencil based on, for example, KF attenuation. Equation 3.11 is a discrete instance of equation 3.1, employing an assumed model of attenuation physics. Suppose that the assumed attenuation physics description is inadequate. This implies that the wave propagation in the unknown medium is better represented by

$$\mathbf{S}_2(\omega,\mathbf{m})\mathbf{u}(\omega) = \mathbf{f}(\omega), \tag{3.12}$$

where $\mathbf{S}_2(\omega, \mathbf{m})$ invokes an attenuation model differing from that in $\mathbf{S}_1(\omega, \mathbf{m})$ and is either unknown, or unfeasible to implement. Differences between, for instance, low-frequency velocity dispersion from one Q model to the next can be significant (Ursin and Toverud, 2002), so the two operators cannot be assumed to be similar. Modeling errors are realized when parameter values for a model belonging to $\mathbf{S}_2$ are selected in order to minimize an objective function based on $\mathbf{S}_1$.

We must assume that a given attenuation model, say $\mathbf{S}_1$, holds in order to begin the process of inverting the data. This means adopting equation 3.11 as a constraint. When $\mathbf{S}_2$ is unknown and $\mathbf{S}_1$ uncertain, the approach I propose for managing QFWI modeling errors is to relax this constraint to instead read

$$\mathbf{S}_1(\omega, \mathbf{m}_N)\mathbf{u}(\omega) = \mathbf{f}(\omega), \text{ for } \omega_N < \omega < \omega_{N+1}, \tag{3.13}$$

where $\mathbf{m}_N$ is a subsurface model for the angular frequency range (ω_N, ω_{N+1}). This allows greater freedom in matching the attenuation behaviour of the measured data, because it requires that the assumed physics be satisfied exactly only on a certain frequency band. As the bandwidths ω_N-ω_{N+1} decrease, modeling errors within any given band become less significant. Piecewise application of $\mathbf{S}_1$ can, in other words, closely mimic a model belonging to $\mathbf{S}_2$, without requiring that $\mathbf{S}_2$ be known. The lower limit of this process involves bands containing single frequency components. This is ideal for mitigating error resulting from unknown attenuation physics (as it represents a return to an inversion for c and Q as freely varying functions of frequency), but is impractical due to the considerable potential for harmful parameter cross-talk.

The discrepancies between the KF and SLS models illustrated in Figure 3.2 will have strong negative consequences for a QFWI procedure, if the KF model is assumed and the SLS model (or something like it) actually holds. But, the consequences can be significantly reduced if, in the QFWI procedure, the KF model is not forced to be self-consistent over the full frequency range. The additional flexibility afforded QFWI by imposing the relaxed constraint in equation 3.13 is illustrated in Figure 3.3. An example SLS profile for Q and P-wave velocity is illustrated in this figure as a black dashed line. A KF profile with the same c at 12.5 Hz, the same average Q, and

a reference frequency of 30 Hz is shown in blue. The highly dissimilar frequency dependence of the velocity and attenuation in the different physics models cause the matching here to be very poor. The red line shows the match which can be obtained using a relaxed KF model, also with a reference frequency of 30 Hz, but with mean Q set to match the SLS model at each 1 Hz band, and c set to match at the center frequency of each band. Clearly, this step offers considerable improvement in the ability to match the observed behaviour, despite having physics different from the SLS enforced in each band. Adopting a FWI strategy which allows for this better matching should improve the quality of the results in the case where the true attenuation model is unknown, I will call this approach 'flexible' QFWI. The flexible QFWI algorithm used in this chapter is set out in Algorithm 1. The important feature of this approach is that instead of recovering one model which is assumed to be valid at all frequencies, a different model is recovered on every frequency band, each of which is assumed to be valid only on that band. In consequence, Algorithm 1 presents an approach for recovering a set of models $\mathbf{m}_N$. These are assumed to be similar enough between frequency bands that the final model at one frequency band can be used as a reasonable starting model for a neighboring frequency band, as reflected in second step of the algorithm, where $\mathbf{m}_{N,0}$ is initialized.

Input : Recorded seismic data $d(\mathbf{r}_g, \mathbf{r}_s, \omega)$, initial model $\mathbf{m}_{0,k_{max}}$
Output: Inverted models $\mathbf{m}_{N,k_{max}}$
for $N = 1, \ldots, N_{max}$ **do** ; *// Loop over frequency bands*

  $\mathbf{m}_{N,0} = \mathbf{m}_{N-1,k_{max}}$
  for $k = 1, \ldots, k_{max}$ **do** ; *// Loop over FWI iterations*

    Initialize the descent direction $\Delta s_k = 0$
    for $p = f_{N_{min}}, \ldots, f_{N_{max}}$ **do** ; *// Loop over frequencies*

      Update the descent direction $\Delta \mathbf{m}_k$
    end
    Calculate the step length μ using the line search method Set
     $\mathbf{m}_{N,k} = \mathbf{m}_{N,k-1} + \mu \Delta \mathbf{m}_k$
  end
end

Algorithm 1: Algorithm for flexible frequency domain QFWI

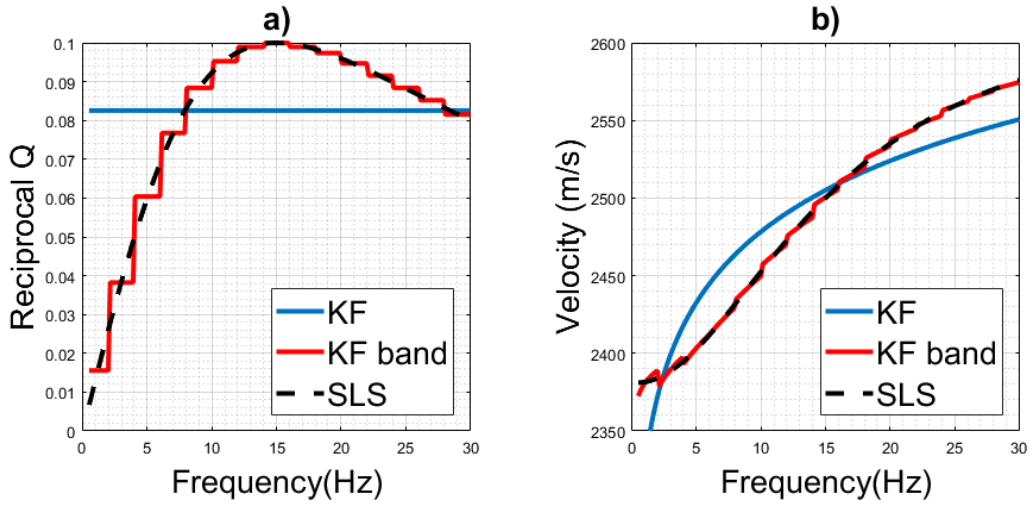

Figure 3.3: Comparison of SLS with best fitting KF, and band-defined KF for reciprocal Q (a), and velocity (b). Due to the highly dissimilar behaviour of the model types, the KF result is a poor approximation of the SLS behaviour. The band-defined KF is capable of matching the SLS behaviour much more closely, though still differs in dispersive behaviour on each band.

While the strategy outlined above has the capacity to match unknown viscoacoustic physics in principle, the question of interest here is whether a QFWI procedure will actually achieve this matching, while constructing interpretable models. Two significant challenges may arise when using this inversion strategy. The first of these is connected to the treatment of dispersion. A band-by-band strategy should be capable of reproducing the dispersive character of the true medium on relatively broad frequency scales, larger than the bands considered, as illustrated in Figure 3.3. The approach discussed here does nothing, however, to match the observed variations in phase velocity associated with dispersion within frequency bands. Errors in the treatment of dispersion may lead the estimated model away from the best approximation at a given frequency band. A second problem with this inversion strategy is associated with the spatial dimension of the inversion. While the approach suggested here should offer one way for the measured data to be reproduced, there remains the possibility that unrealistic spatial variations of KF model parameters will be introduced in the inversion to compensate for the attenuation physics errors as well. It is difficult to address the impact of these concerns analytically, so instead I consider synthetic examples in the next section to gauge the approach.

3.4 Numerical Examples

For the numerical examples presented here, I use the viscoacoustic forward modeling approach described in Chapter 2, using a KF model of attenuation for the forward modeling used in the inversion. Frequencies from 1 Hz to 25 Hz are used, and the sources are considered to have a uniform amplitude spectrum over this range. This extensive frequency coverage was considered in lieu of accurate starting models; the inversions are initialized with constant values of c_0 and Q.

I include regularization terms in the inversions to prevent large spatial variations in the model updates and to limit the extent of Q variations. The objective function I use in this chapter is

$$\phi = \phi_D + \phi_{P_1} + \phi_{P_2}, \tag{3.14}$$

where ϕ_D is the L_2 data-fitting objective described in Chapter 2, and the ϕ_P terms are regularization terms. The first regularization term, ϕ_{P_1}, was designed to stabilize the inversion by discouraging strong spatial variations in the model updates without strongly influencing the inversion results. This term is given by

$$\phi_{P_1} = a_1 ||\nabla(m_{k+1} - m_k)||^2, \tag{3.15}$$

where m_k is the current model, m_{k+1} is the updated model, ∇ is the first order central finite difference operator for the spatial gradient, and a_1 is a weighting term. This term has no gradient contribution when calculating the $k+1$th step, but introduces terms in the Hessian which promote relatively smooth updates. The term penalizing low Q values, ϕ_{R_2}, was given by

$$\phi_{P_2} = a_2 ||Q^{-1}||^2, \tag{3.16}$$

where a_2 is a weighting term. This term was introduced to discourage extreme values of Q in the recovered model.

3.4.1 Frequency bands and cross-talk

The first example I consider explores the QFWI problem where attenuation physics are known, in order to highlight the effects of the frequency bands on the level of inter-parameter cross-talk. The

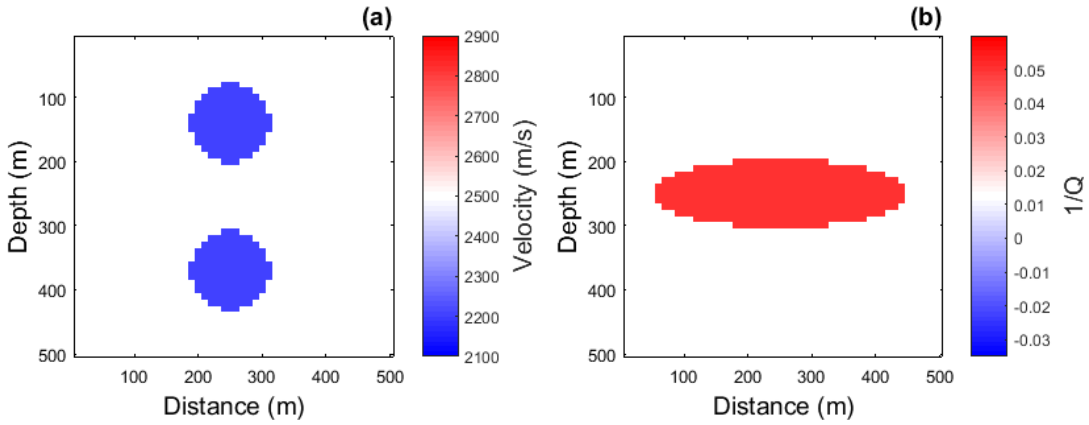

Figure 3.4: Benchmark model c_0 (a) and Q^{-1} (b). The velocity values correspond with reference frequency $\omega_0/2\pi = 30$ Hz.

model used in the inversion is shown in Figure 3.4. The attenuation-physics model used to generate the data in this example is the KF model, which is also assumed in the inversion. To reduce cross-talk as much as possible, I use exact Gauss-Newton optimization in this example. To make this optimization strategy feasible, I consider a small model, defined on a 2D 50×50 grid with 10 m grid cells. Twenty-four explosive sources at 30 m depth were spaced 20 m apart from 10 to 470 m; 48 receivers at 20 m depth were spaced 10 m apart from 10 to 480 m. The inversion was initialized with a constant Q and c_0, equal to the background of the true model. Perfectly matched layers 9 gridpoints thick were applied at every boundary. The reference frequency considered was 30 Hz. No noise was added to the data for this example, and a_2, defining the Q regularization was set to 0. Q was not restricted to physically meaningful values in the inversion, and consequently some inversions resulted in non-physical negative values of Q in parts of the recovered model. Usually, these non-physical values were restricted to poorly constrained regions of the model. I interpret any occurrence of negative Q values in well-constrained regions as a failure of the inversion.

Figure 3.5 illustrates the results of QFWI in the likely problematic setting in which one frequency is inverted at each frequency band. Six iterations were performed at each frequency, beginning at the lowest frequency, 1 Hz, then increasing in 1 Hz increments up to 25 Hz. The problems

in Figure 3.5 highlight important features of QFWI. Cross-talk impairs both the Q estimate and the c_0 estimate, with the Q anomaly not meaningfully recovered. The c_0 estimate is very poor at the lower velocity anomaly (which is obscured from the sources and receivers by the low Q region). Because this cross-talk occurs despite the use of exact Gauss-Newton optimization, and comprehensive simulated acquisition, it is reliably traceable to the use of a single frequency.

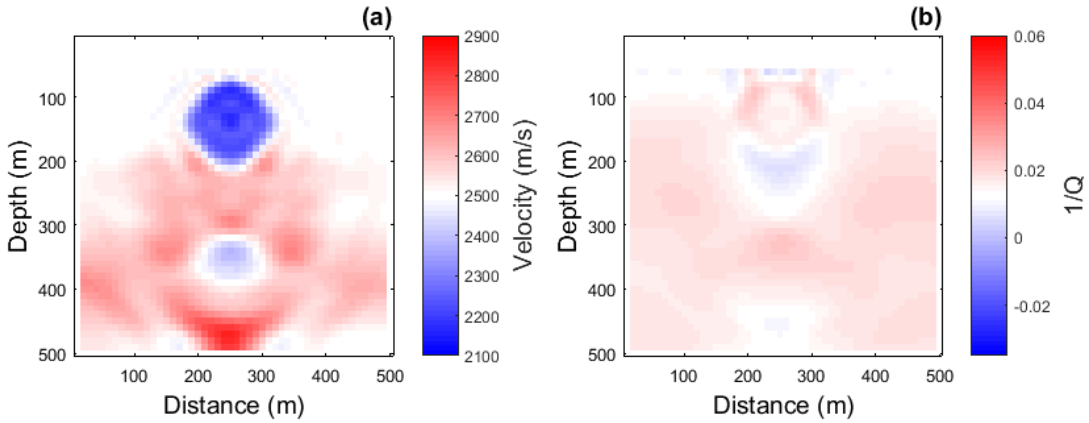

Figure 3.5: Gauss-Newton QFWI c_0 (a) and Q^{-1} (b), inverting only one frequency at each iteration.

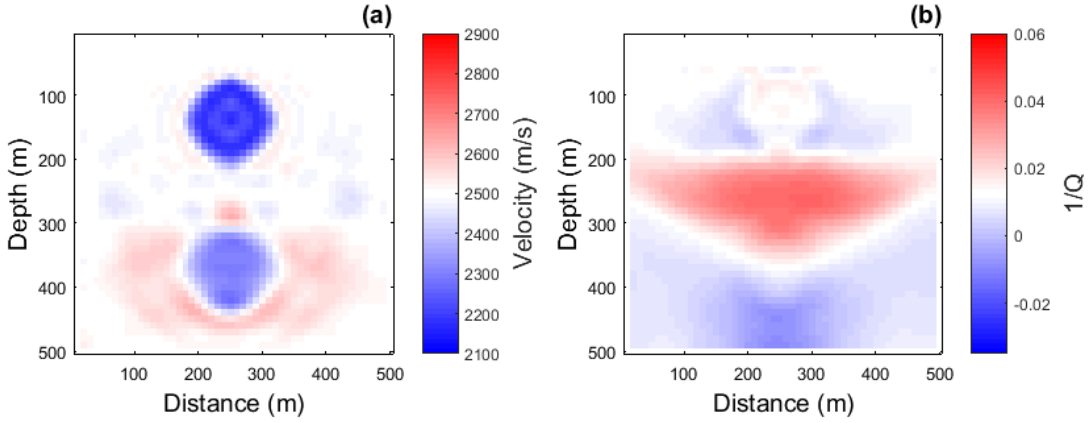

Figure 3.6: Gauss-Newton QFWI c_0 (a) and Q^{-1} (b), inverting a 1 Hz band of frequencies at each iteration. Compare with Figure 3.5.

Figure 3.6 illustrates the results of QFWI in which a narrow range of frequencies (6 evenly-

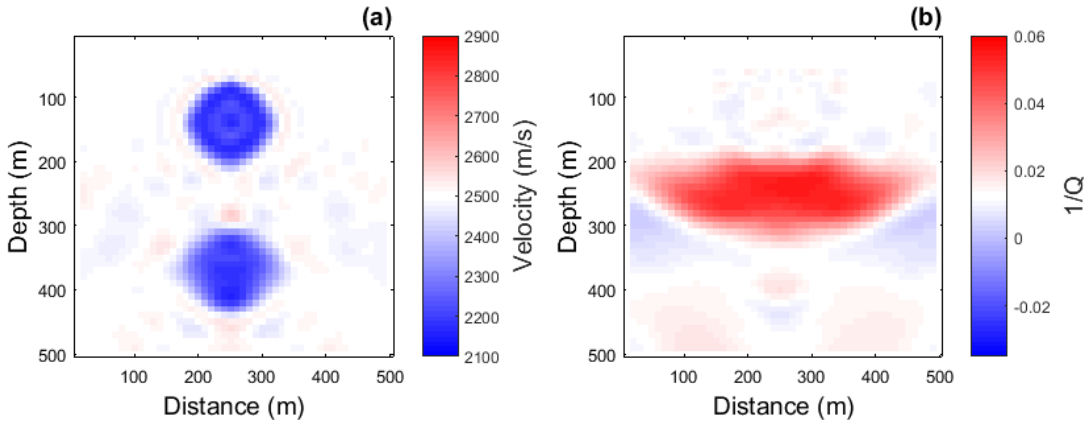

Figure 3.7: Gauss-Newton QFWI c_0 (a) and Q^{-1} (b), using a multiscale approach. Compare with Figures 3.5 and 3.6.

spaced frequencies in a 1 Hz range) were inverted at each frequency band. One iteration was carried out per band, beginning with a band centered at 1.5 Hz, and increasing the center frequency by 1 Hz at each iteration, up to 24.5 Hz. The improved recovery of the deeper velocity anomaly and the Q anomaly are notable (compare with Figure 3.5). This is further evidence supporting the prediction from scattering-potential analysis that groups of frequencies at each iteration offer the only practical means of discriminating between velocity and Q, and mitigating cross-talk.

Figure 3.7 illustrates the results of QFWI using a multiscale approach similar to that proposed by Bunks et al. (1995). In this example, six frequencies were inverted at each iteration, these were distributed from 1 Hz to a maximum frequency which grew as iterations progressed, beginning at 2 Hz, and increasing by 1 Hz per iteration to a maximum of 25 Hz. This approach produces the best recovery of the three strategies. All else having been held fixed in these experiments, I conclude that iterations involving a broad, varied range of frequencies are optimal for suppressing cross-talk. This is in keeping with the general principle that, to separate any two parameters, data must span experimental variables across which the two have different characteristic scattering signatures.

3.4.2 QFWI with attenuation physics uncertainty

While cross-talk suppression is ideally managed by considering a broad range of frequencies in QFWI, such an approach is expected to be error prone where attenuation physics are not well understood. In this subsection, I implement three distinct QFWI approaches numerically and use them to explore their effectiveness in viscoacoustic modeling-error suppression. In the first, which will be referred to as 'low cross-talk' QFWI, equation 3.11 is strictly adhered to, and the updating strategy used is the multiscale approach used to minimize viscoacoustic cross-talk in the previous section. In the second, the flexible QFWI is used with finite frequency bands, obeying instead the constraint in equation 3.13 and using an updating strategy consistent with this constraint, as described in Algorithm 1. In the third case, the flexible approach is again used, but with only one frequency inverted on each band. The QFWI procedure is built around a KF viscoacoustic model in all three cases.

In these examples, the model defined in Figure 3.8 is considered, with SLS attenuation physics used to generate the data. In this model, the maximum Q^{-1} was at 15 Hz. The reference frequency used in the inversion was fixed at 30 Hz. A total of 112 explosive sources at 10 m depth were spaced 20 m apart from 10 to 2250 m; 225 receivers at 0 m depth were spaced 10 m apart from 10 to 2250 m. The initial model for the inversion was a constant, infinite Q, and c_0 of 1700 m/s. Truncated Newton optimization (e.g., Nash, 1985; Métivier et al., 2013) was employed in the inversion. The inner-loop of the truncated Newton optimization was terminated when a maximum of 50 iterations was reached, or when the inequality

$$\eta \leq \frac{||H\mathbf{p}-\mathbf{g}||}{||\mathbf{g}||} \tag{3.17}$$

was satisfied, where H is the Gauss-Newton approximation of the Hessian, $\mathbf{p}$ is the estimated model update, and η is a forcing term, chosen here to be 10^{-5}. Both regularization terms were set to nonzero values in these examples, so $\frac{1}{Q}$ is encouraged to be small, and updates are encouraged to be smooth. Normally distributed noise was added to the modeled data with a signal-to-noise ratio of 10, where signal is defined as the difference between the measured data and the synthetic

data generated from the initial model at each frequency.

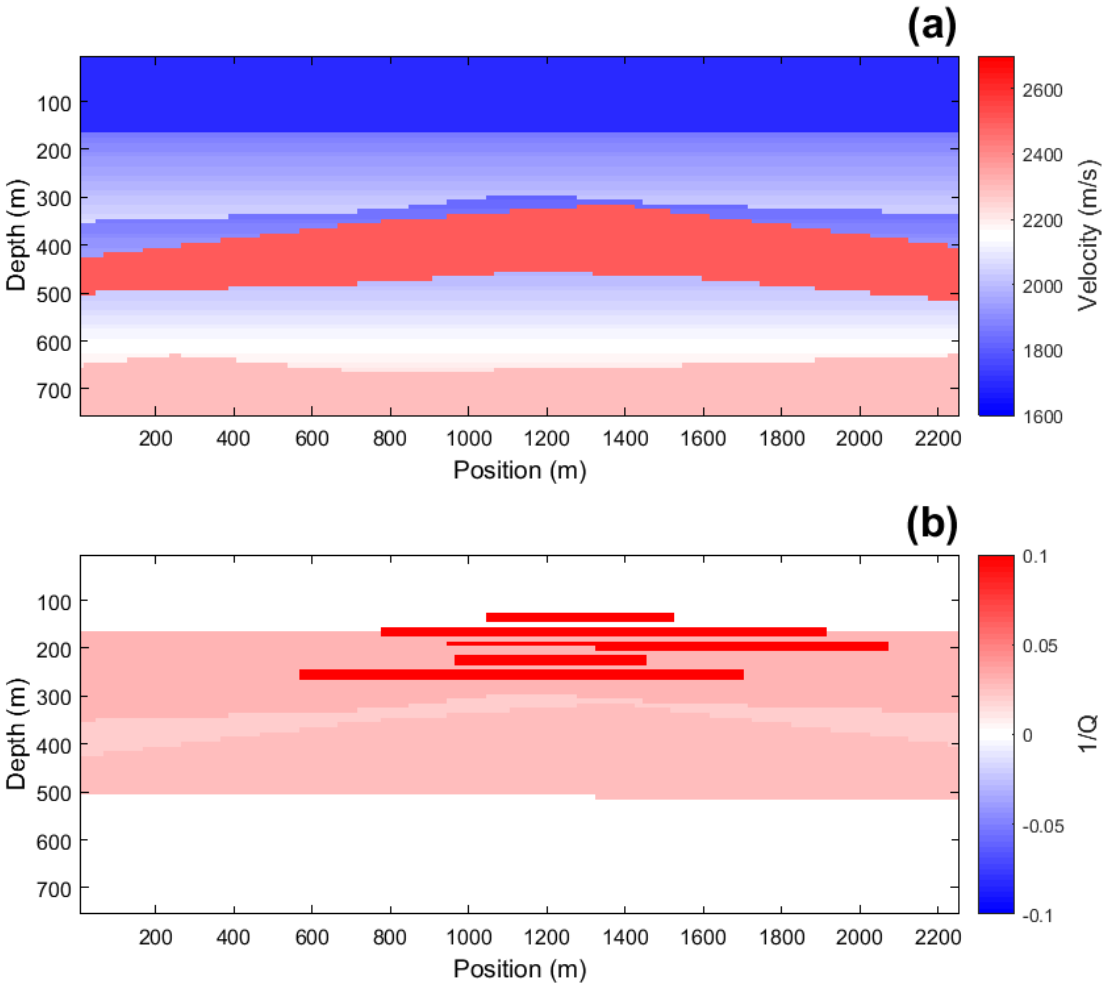

Figure 3.8: Benchmark model c_0 (a) and Q^{-1} (b) for SLS attenuation physics.

In the first approach considered, low cross-talk QFWI was applied, using the multiscale approach of Bunks et al. (1995), as in Figure 3.7. In this example, twelve frequency bands were considered, beginning with 1 Hz to 3 Hz, and increasing the upper limit by 2 Hz at each band, up to the final band of 1 Hz to 25 Hz. At each stage of the inversion, five frequencies spaced evenly from the minimum to the maximum frequency of the band were simultaneously inverted. Four iterations were performed at the lowest frequency band, and two iterations at each higher band. The results of this inversion are illustrated in Figure 3.9. Comparing to the true model in Figure 3.8, it can be seen that the recovered Q model has significant artifacts, including substantial

regions of negative Q. The recovered velocity is also quite poor at the bottom of the model in this inversion, with strong lateral variations not present in the true model appearing. Evidently, QFWI struggles to recover a meaningful velocity or Q model — despite benefiting from dense acquisition and relatively powerful truncated-Newton numerical optimization. This highlights the hazards associated with uncertainty in the QFWI attenuation model, and provides incentive for a more flexible approach.

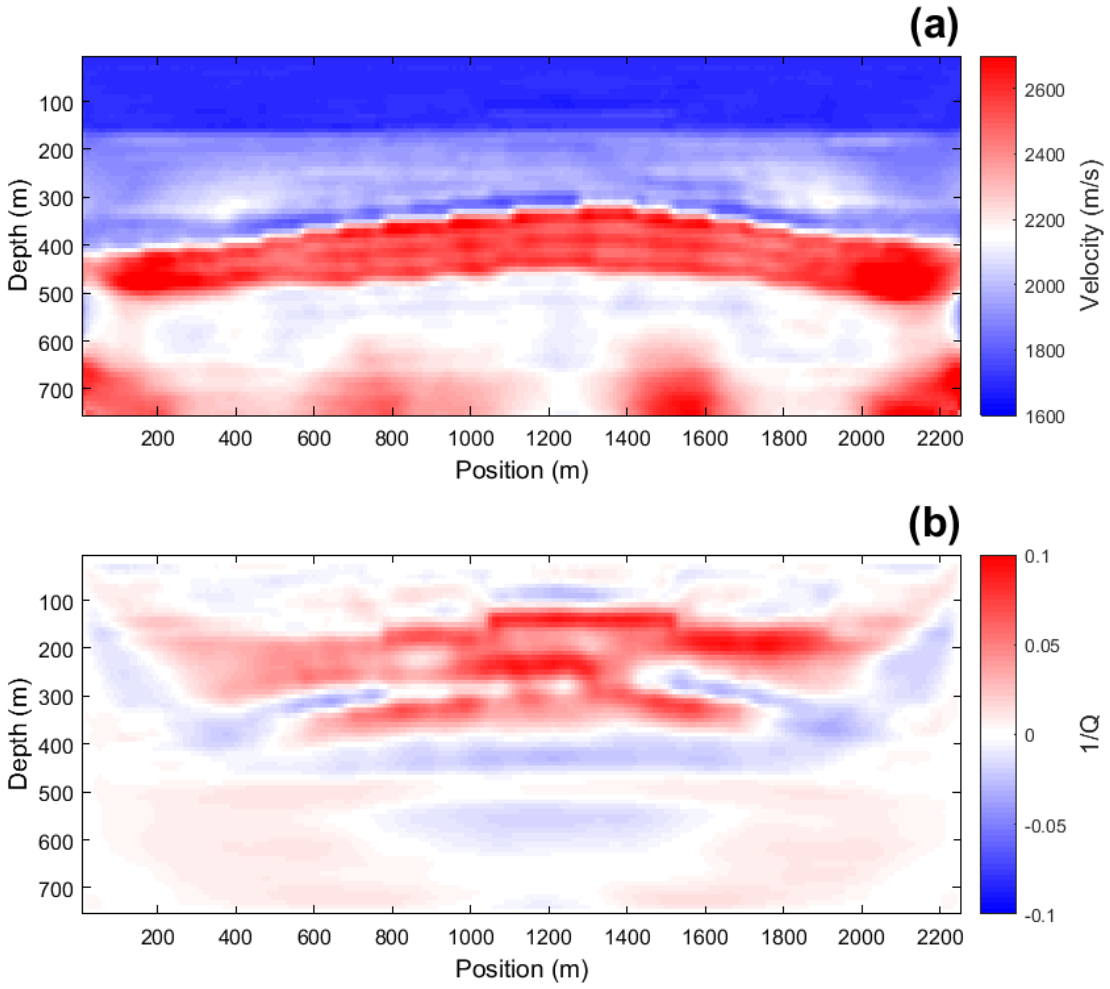

Figure 3.9: Conventional QFWI c_0 (a) and Q^{-1} (b); KF inversion carried out on SLS data..

In the second approach considered, a finite-band flexible approach is used. To determine an appropriate choice of frequency bands for this approach, a number of Gauss-Newton inversions were performed on the model in Figure 3.4, both with KF and SLS physics used. Model residuals

Band type	KF velocity residual	SLS velocity residual	KF Q^{-1} residual	SLS Q^{-1} residual
1 Hz sliding	1.000	0.912	0.870	0.642
2 Hz sliding	0.694	0.741	0.735	0.745
4 Hz sliding	0.594	0.723	0.646	0.852
8 Hz sliding	0.617	0.820	0.625	0.948
Multiscale	0.620	0.913	0.631	1.000

Table 3.1: L_2 model residuals for different frequency band choices. These values are normalized such that the largest residuals in velocity and Q^{-1} are 1.

from these tests were used to assess the effectiveness of each choice of frequency band. A comparison of the relative L_2 model residuals using the flexible approach for several choices of band size as well as the multiscale approach is shown in Table 3.1. The model residual in this table is for the highest frequency band considered in the flexible approach, to limit errors associated with spatial resolution. In this table, the broad frequency band (4 Hz, 8 Hz) flexible and multiscale approaches perform well when the expected KF physics are used in the inversion, as shown in the 'KF residual' columns, giving comparable results in this case. These bands are much less effective when the SLS physics generate the data ('SLS residual' columns). Using a 1 Hz band gives poor results, but these results are less strongly affected by errors in the attenuation physics. Inversion with the 2 Hz band gives the best results when the assumed physics are incorrect. This result is specific to the choices of KF and SLS physics as well as the true model geometry, and is likely to be dependent on other factors, such as noise level. Despite these uncertainties, I interpret this result to mean that a 2 Hz band may be a reasonable choice for the flexible approach, and use this value for the second attempt at inverting for the model in Figure 3.8.

For the flexible inversion, I considered twelve frequency bands of 2 Hz width, with minimum frequencies from 1 Hz to 23 Hz. To recover the model for each frequency band, $\mathbf{m}_N$, five evenly spaced frequencies from f_N to f_{N+1} were inverted simultaneously. Four iterations were performed for $\mathbf{m}_1$, and two iterations at each later $\mathbf{m}_N$, resulting in a total number of iterations equal to the number used in the low cross-talk FWI approach. The $\mathbf{m}_N$ were solved for sequentially, beginning

with the lowest frequency band and then increasing. The initial model used for $\mathbf{m}_1$ was identical to that used in the low cross-talk FWI approach. The initial model for every other $\mathbf{m}_N$ was set equal to the final $\mathbf{m}_{N-1}$, as described in Algorithm 1. In effect, this approach is similar to that used to generate Figure 3.6, but where each intermediate result at a given band can be interpreted as an estimate of the velocity and Q at that band. This approach is summarized in Algorithm 1.

Results produced by applying the finite-band flexible QFWI approach, based on the KF model, on the SLS data, are shown for two example bands in Figures 3.10 and 3.11. In principle, both these results, as well as those at other frequency bands are outputs of the algorithm. The less restrictive constraints used in this approach allow for a significantly improved recovery of the true model behaviour. A substantially more accurate estimate of Q is obtained at both bands, with few of the negative Q artifacts persisting. The velocity estimate is also substantially improved in the deeper regions of the model, likely as a result of the improved Q estimate. A comparison between these results for a profile at an x position of 1500m is shown in figure 3.12, where the flexible approach improves the inversion result considerably.

In the third and last inversion approach I considered, the flexible FWI approach was used with only one frequency inverted at each band. Bands were defined at every 0.5 Hz, beginning at 1 Hz. Four iterations were performed at the lowest band, and two at every higher band. The results of this approach are shown for the highest frequency band in Figure 3.13. In this case, the inversion has struggled to recover an accurate model, giving an inferior result to the finite-band flexible strategy. The relatively poor inversion here is likely a result of the considerable cross-talk expected in the single-frequency case.

Data residuals for the source at 1990 m offset for the initial model and the multiscale and flexible inversions are shown in Figures 3.14, 3.15, 3.16, and 3.17. Differences both in amplitude and in phase are plotted. While the low cross-talk approach predicts much of the measured data (Figure 3.15), the remaining residuals are well above the noise level in the model. By contrast, aside from some evident footprint of the choice of bands, the finite-band flexible approach largely

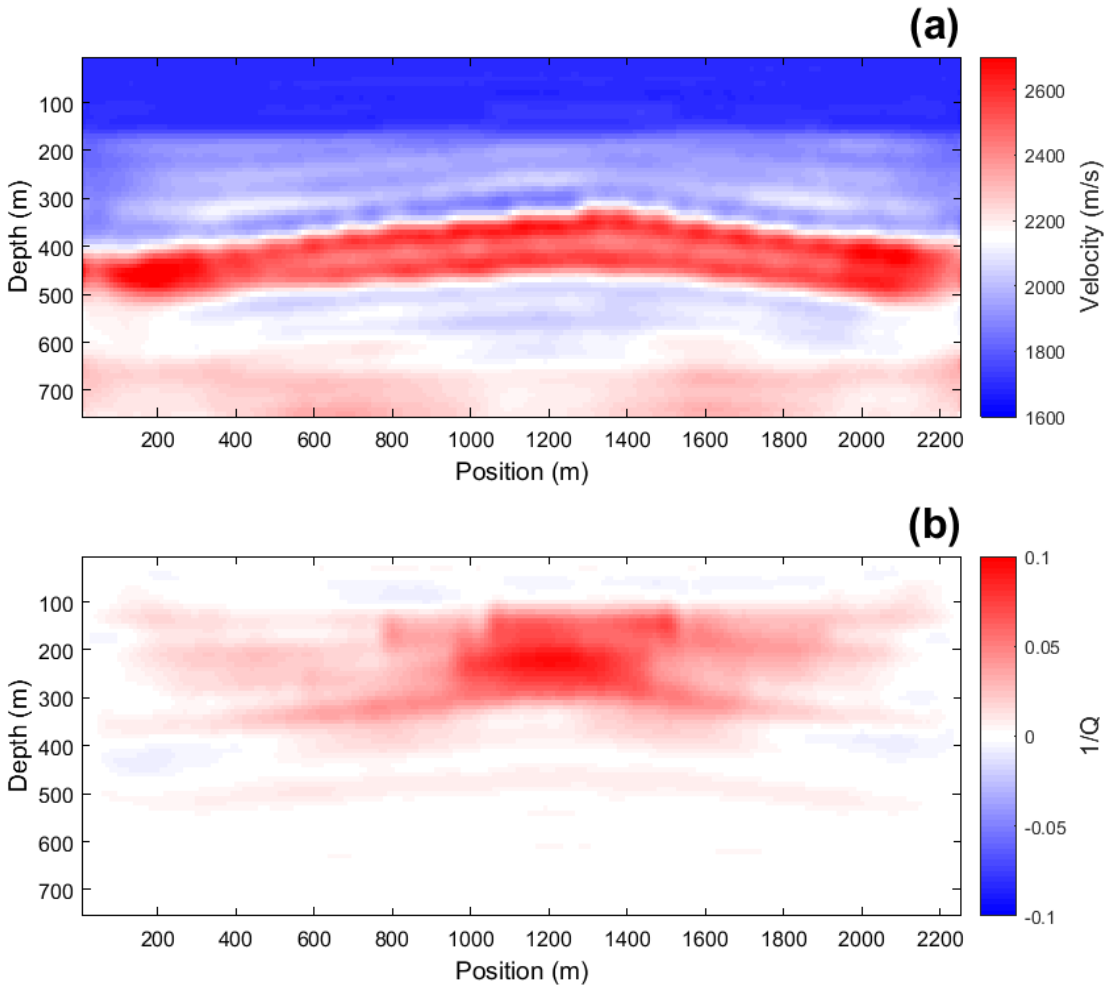

Figure 3.10: Flexible finite-band approach c_0 (a) and Q^{-1} (b) for 13-15 Hz maximum band; KF inversion carried out on SLS data. Velocity plotted is $c(\omega)$ at 15 Hz.

predicts the measured data, in the sense that there is little coherency in the remaining residuals (Figure 3.16). This supports the idea that flexible inversion offers the potential to reproduce the measured data, even with errors in the assumed attenuation physics. The single frequency flexible approach also predicts the measured data well, and removes the footprint (Figure 3.17). This improved data fitting comes at the cost of a poor model reconstruction, however, as shown in Figure 3.13.

The computational costs of the multiscale and finite-band flexible QFWI approaches used here are similar, the greater number of models recovered in the flexible approach being offset by the

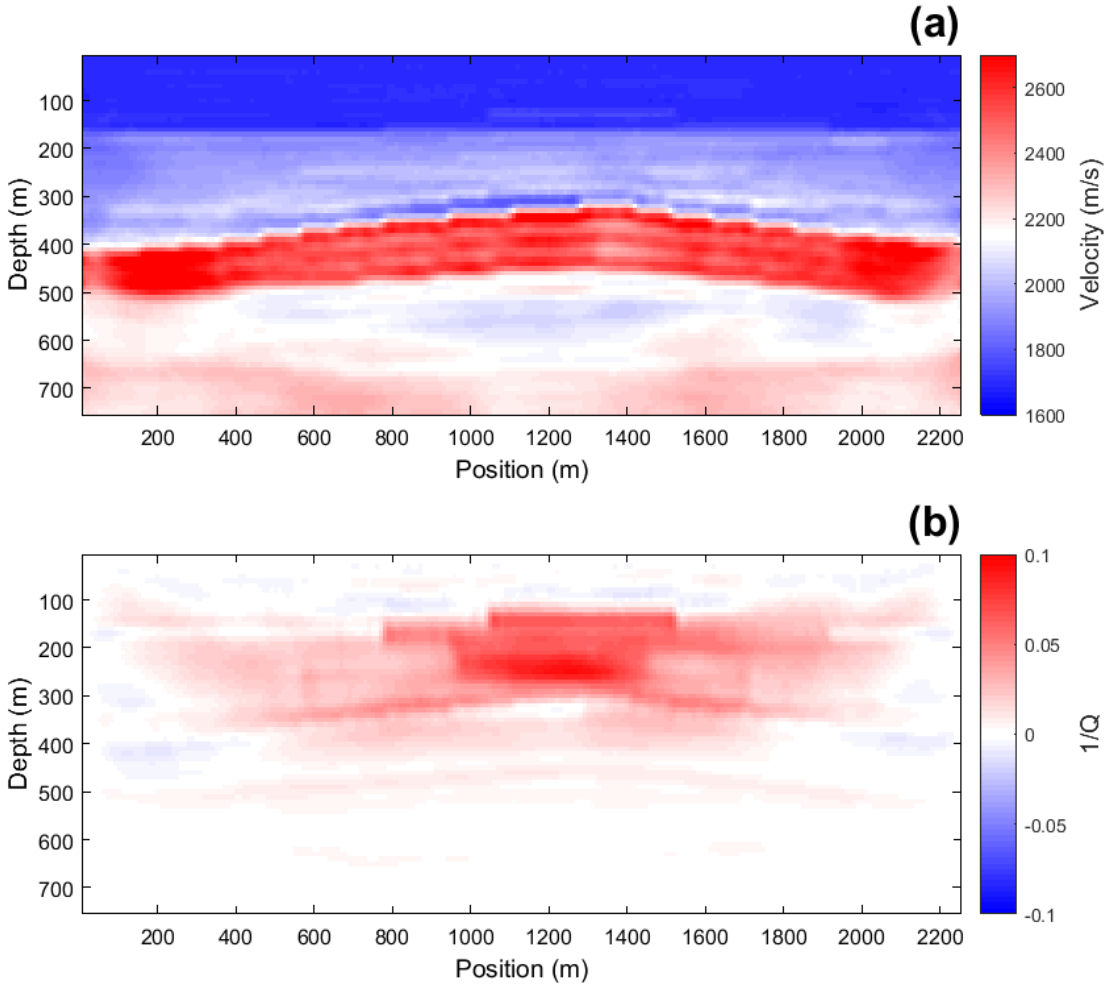

Figure 3.11: Flexible finite-band approach c_0 (a) and Q^{-1} (b) for 23-25 Hz maximum band; KF inversion carried out on SLS data. Velocity plotted is $c(\omega)$ at 25 Hz.

smaller number of iterations used (two) to invert for each. The reason for this similarity is that the flexible approach can be interpreted as an alternative multiscale strategy in conventional FWI, with the caveat that the final result is an approximation of the model behaviour only within the highest frequency band considered, and that the intermediate steps themselves provide an estimate of the model behaviour at their respective frequency ranges.

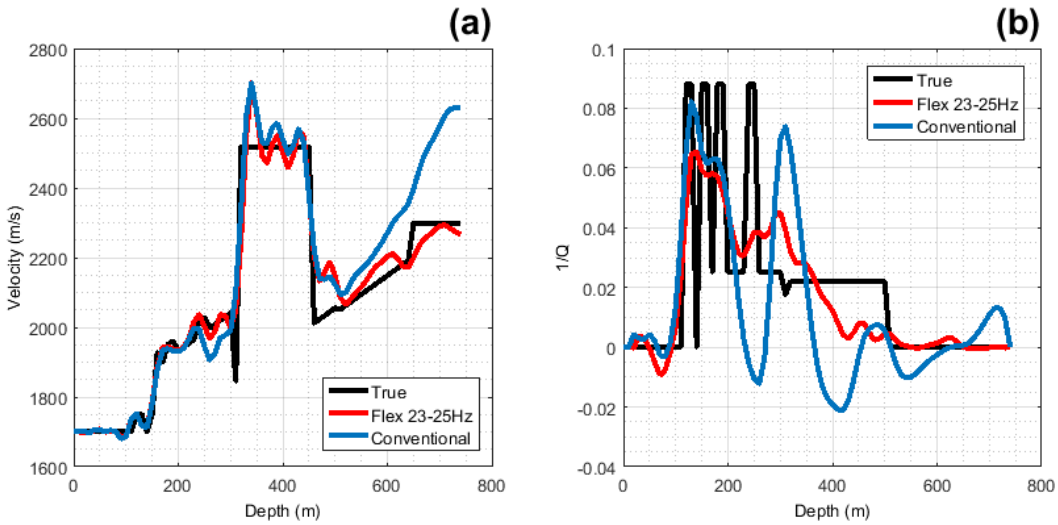

Figure 3.12: A comparison of recovered models for a profile at x = 1500m.

3.5 Discussion

Analysis of the scattering potentials in the QFWI problem, as well as numerical examples, strongly suggest that the choice of frequencies used in a FWI update strongly affects the propensity of QFWI towards cross-talk. Hak and Mulder (2010) demonstrate that an acquisition geometry offering complete angle coverage allows for the discrimination of Q and velocity. This cannot be achieved even with extensive (e.g., cross-hole acquisition) subsets of scattering angles, however, suggesting that the fixed-frequency cross-talk terms experience total cancellation only when all scattering angles are measured. As this is not practical in real seismic surveys, the selection of frequencies used in QFWI plays a major role. The numerical examples used here showed that with clean data, even a very narrow range of frequencies is sufficient to mitigate cross-talk; however noise in the measured data likely introduces stronger demands on the frequencies used in inversion.

Although the flexible QFWI approach outlined here offers a potential route for approaching the QFWI problem in the face of attenuation physics uncertainty, this capability is offset by a sacrifice of other desirable qualities of the inversion. The use of narrower frequency bands in the inversion allows for the observed measurements to be partially explained by physics different from those which generated the data. While this is desirable when the physics are uncertain, it is also

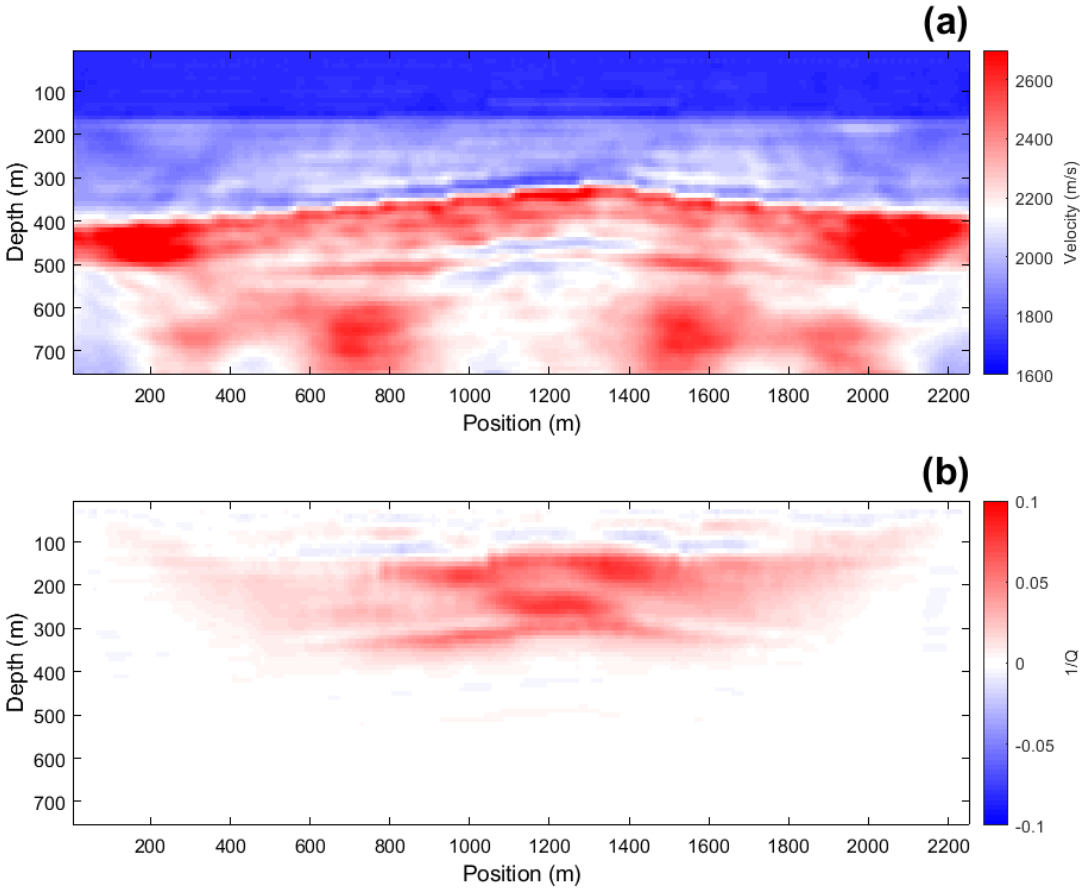

Figure 3.13: Flexible single frequency approach c_0 (a) and Q^{-1} (b) for 25 Hz; KF inversion carried out on SLS data. Velocity plotted is $c(\omega)$ at 25 Hz.

highly conducive to cross-talk, which exploits the same effect. This necessitates a compromise in QFWI, where the most desirable updating strategy is that which most strictly enforces the assumed physics, while still allowing the flexibility to cope with errors in these physics. The ideal size of the frequency bands inverted will depend on the uncertainty in attenuation physics, and is an important topic for further research.

The specific implementation of the flexible QFWI investigated here, where KF physics are assumed, is likely not the ultimate solution. In particular, it may not be desirable to assume a given dispersion relation, even on a narrow frequency band, given the importance of dispersion in solving the viscoacoustic inverse problem (e.g., Innanen and Weglein, 2007; Hak and Mulder, 2011). Introducing, for example, a variable that parameterizes a linear dispersion on each frequency band

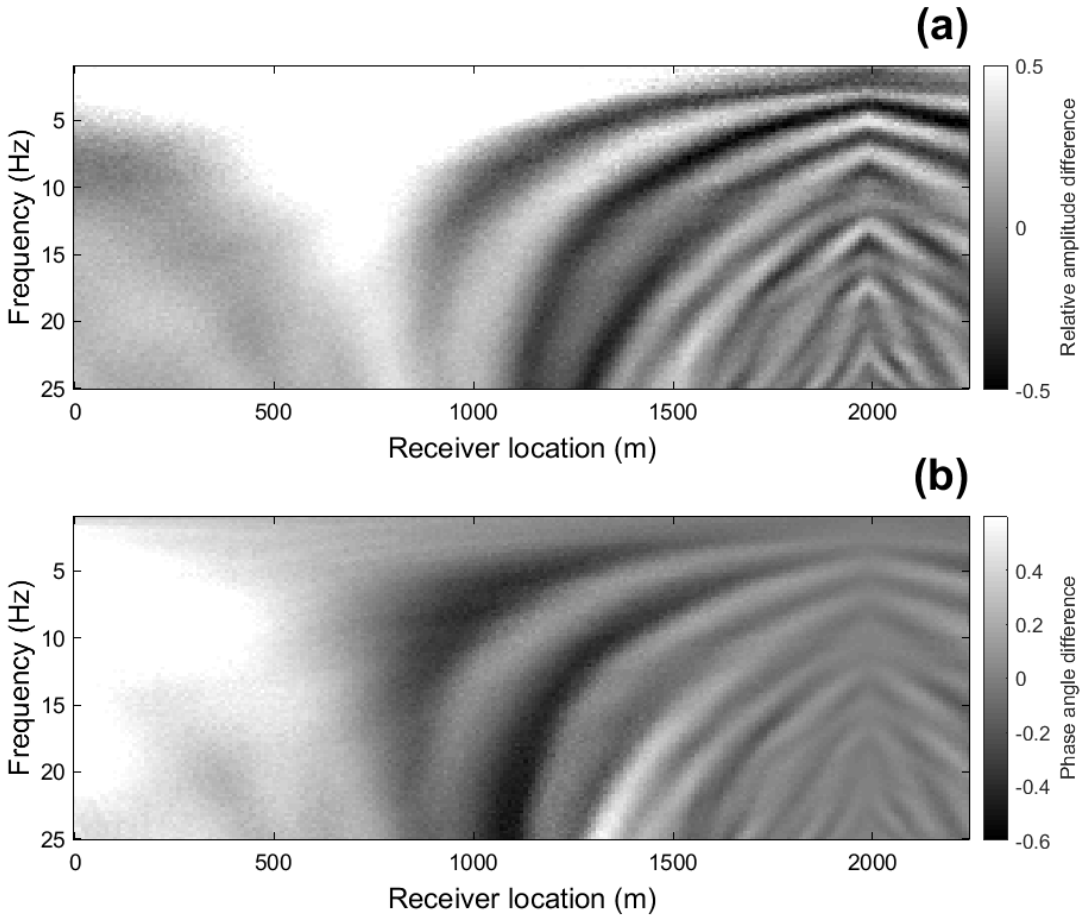

Figure 3.14: Differences in amplitude (top) and phase angle (bottom) between shot records in the true and background models. Results are shown for the shot at 1990 m.

may help to avoid this assumption, but comes at a cost of introducing an additional variable and creating a more difficult, more poorly constrained optimization problem. Similar changes in the degree of flexibility allowed in the recovered physics could offer different benefits and tradeoffs. This appears to be an important topic for further investigation. Another important step is to investigate similar cross-talk mitigation strategies in more complete viscoelastic FWI settings, wherein P-wave and S-wave dispersive velocities and Q-factors leak into one another.

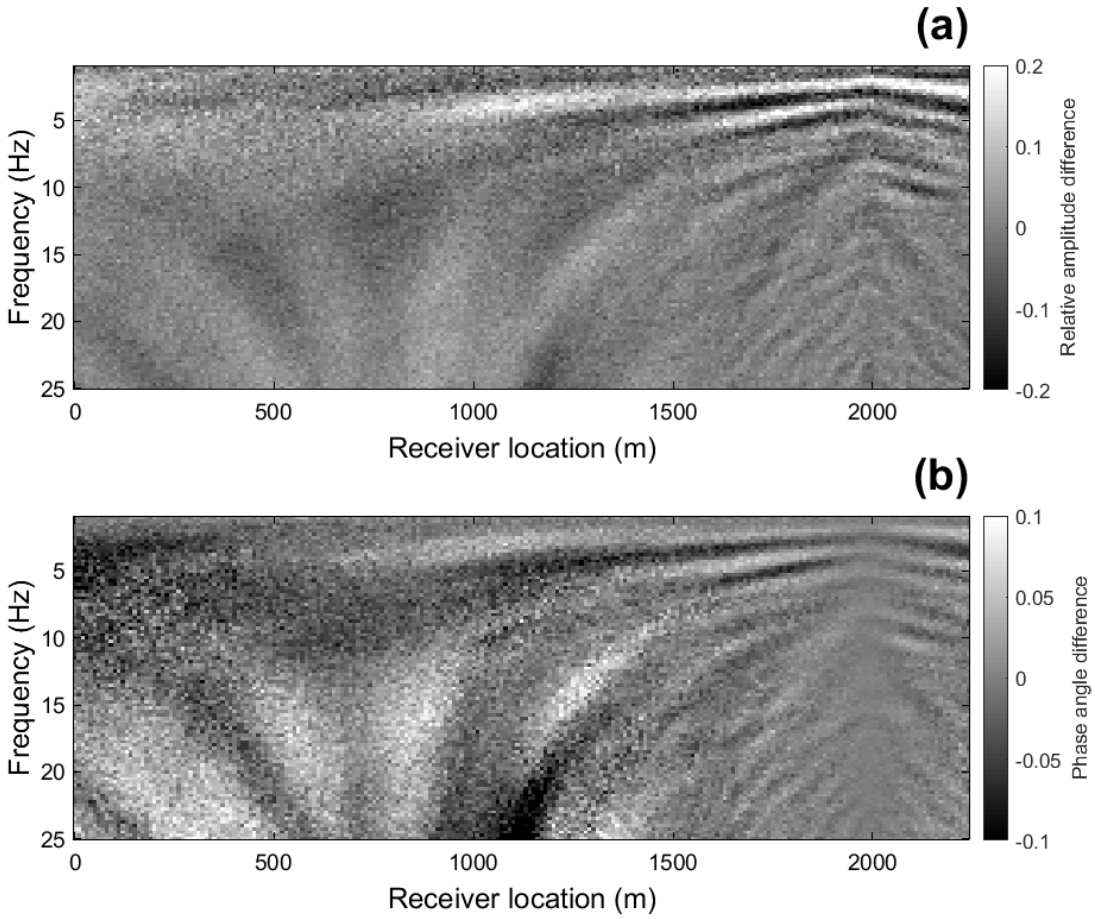

Figure 3.15: Differences in amplitude (top) and phase angle (bottom) between shot records in the true model and in the multiscale inversion result. Results are shown for the shot at 1990 m. Compare to Figure 3.14.

3.6 Conclusions

Cross-talk is a serious concern in viscoacoustic full-waveform inversion, and has a particularly strong impact on the recovered Q model. Frequency dependent effects play a major role in eliminating this cross-talk as suggested by inspection of the relevant scattering potentials. Single-frequency updates are dominated by cross-talk, even when exact Gauss-Newton optimization is used. Inverting even a narrow band of frequencies per iteration offers a notable improvement, and the best results with known attenuation physics were achieved inverting relatively broad frequency bands.

The inclusion of attenuation in QFWI also presents unique risks associated with modeling

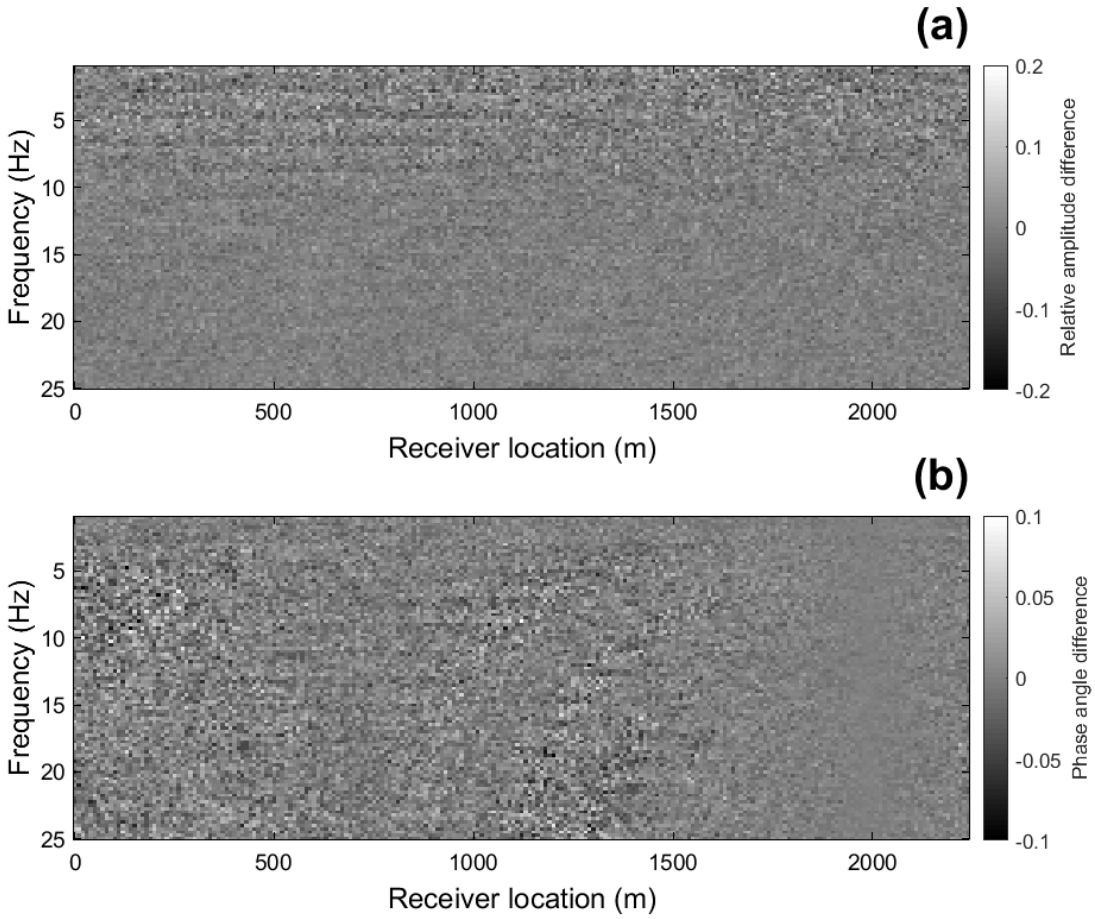

Figure 3.16: Differences in amplitude (top) and phase angle (bottom) between shot records in the true model and in the flexible inversion result. Results are shown for the shot at 1990 m. Compare to Figures 3.14, 3.15.

error. The flexible QFWI approach suggested here relaxes the FWI constraint that the modeled wavefield strictly obeys an assumed physics model across all experimental variables. This allowed for significant improvements over low cross-talk FWI strategies as applied to the QFWI problem where the assumed physics are incorrect. It also introduces a tradeoff between cross-talk reduction and flexibility to cope with unknown physics, in the form of the size of the frequency band used in the inversion. The flexible QFWI approach investigated here is only one possible strategy; possible approaches are far from having been exhaustively considered.

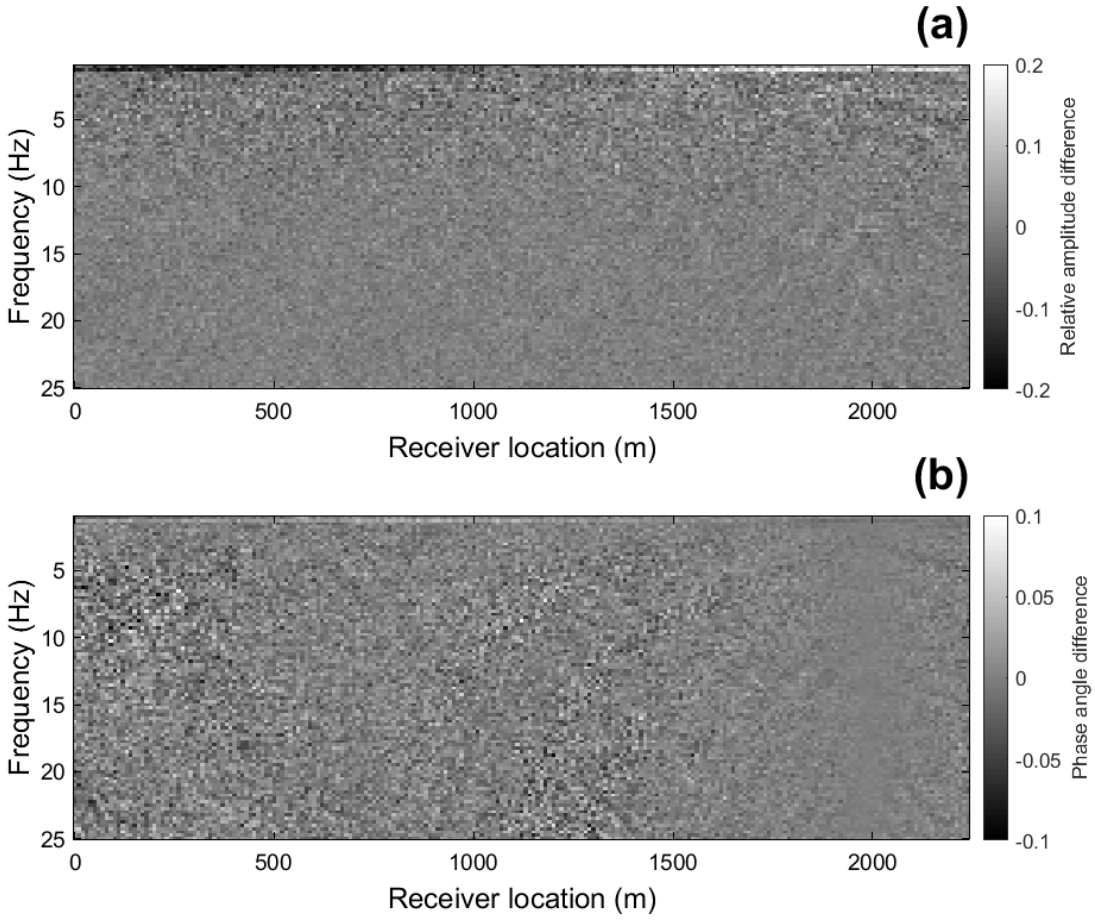

Figure 3.17: Differences in amplitude (top) and phase angle (bottom) between shot records in the true model and the single frequency flexible inversion result. Results are shown for the shot at 1990 m. Compare to figures 3.14, 3.15, 3.16.

Chapter 4

Parameter cross-talk and leakage between spatially separated unknowns in viscoelastic full waveform inversion

4.1 Summary

Elastic and attenuative effects play a major role in the determination of wave amplitudes and phases observed at seismic sensors. Viscoelastic full waveform inversion (FWI) has the potential to recover much of the information content of measured seismic data by simultaneously accounting for these effects. However, the frequency variations and phase information present in viscoelastic FWI introduce additional challenges to the inversion, especially through their impact on inter-parameter cross-talk. Cross-talk is typically characterized through analysis of the radiation patterns of point scatterers; however, the point scatterer model is not well suited to viscoelastic FWI, because (1) attenuation introduces a significant potential for cross-talk between variables distant from one another in space and (2), interpreting the effect of frequency and phase dependence on the radiation patterns of point scatterers is not straightforward. I present and examine a numerical approach for assessing the viscoelastic cross-talk modes expected for a given parameterization, optimization strategy, and acquisition geometry based on differencing various synthetic inversion results. With this approach, I characterize viscoelastic cross-talk for a typical parameterization for several possible acquisition geometries. Of particular note is the strong tendency for Q variables to leak into elastic variables from which they are spatially separated. This type of cross-talk is not easily characterized through the use of radiation patterns.

4.2 Introduction

The goal of full waveform inversion (FWI) is to derive subsurface medium properties by maximizing the use of the information content of measured seismic data (Tarantola, 1984). In reality, harnessing the full information content of measured data is infeasible because approximate models of true wave propagation must be used to relate measurements to physical properties, and these approximations are incapable of reproducing the full complexity of the true measurements. However, the move from simple to more complex model parameterizations, which better approximate true wave propagation, has historically led to improved recovery of useful, accurate information about the subsurface. Much of the information constraining subsurface elastic properties resides in the amplitude and phase of the observed waveforms. Scalar-acoustic FWI cannot use such information effectively, because of its neglect of elastic and attenuative effects. Elastic, viscoacoustic, and, less frequently, viscoelastic FWI approaches have been developed to address this deficiency (e.g., Tarantola, 1986; Hicks and Pratt, 2001; Tromp et al., 2005; Malinowski et al., 2011; Kamei and Pratt, 2013; Métivier et al., 2015; Plessix et al., 2016; Yang et al., 2016; Fabien-Ouellet et al., 2017). As in any multi-parameter FWI problem, inter-parameter trade-off, or 'cross-talk', is a major obstacle to the implementation of these approaches (e.g. Kamei and Pratt, 2013; Alkhalifah and Plessix, 2014; Pan et al., 2016; Fabien-Ouellet et al., 2017). Cross-talk occurs when data residuals caused by an error in the estimate of one physical property are attributed to another, impeding convergence and potentially leading to mis-characterization in the inversion output. Strategies exist for cross-talk reduction, but to design these effectively it is important to understand the cross-talk process, i.e., to determine which properties leak in to one another, and to what extent.

The most commonly used tool for characterizing cross-talk is radiation pattern analysis (e.g. Tarantola, 1986; Moradi and Innanen, 2016; Oh and Alkhalifah, 2016; Kamath et al., 2017; Yang et al., 2019). Radiation patterns express the change in an incident wave-field after interacting with a point scatterer, typically in an otherwise homogeneous medium. These patterns change for different choices of model perturbation. Greater cross-talk is predicted when the radiation patterns

of two variables vary proportionally within the range of scattering angles sensed in an experiment, because this behavior is suggestive that the two variables are not distinguished within the available data. Radiation patterns are not investigated for every variable in the inversion, as the number of these is very large. Instead, a representative radiation pattern for each parameter type is usually investigated, with the scattering point set at a fixed location.

In elastic and anisotropic FWI, radiation pattern analysis has proven effective in predicting cross-talk (e.g., Oh and Alkhalifah, 2016). Guided by analysis of the patterns to select inversion parameters with minimal overlap between different radiation patterns in the data, the extent of cross-talk in the inversion can be reduced. Scattering patterns do not, however, completely characterize cross-talk. They are not well suited to providing information about cross-talk between variables at different spatial locations, as two different scattering angles for variables at different locations may represent the same part of data space. The comparisons they make are between point scatterers; the effects of a sustained model perturbation on cross-talk are difficult to characterize. Further, they are limited to providing information about the gradient and do not naturally allow the effects of iteration to be characterized. Consideration of both the second-order aspects of the objective function (e.g., via the Hessian) and iteration is key to the successful implementation of multi-parameter FWI (e.g. Virieux and Operto, 2009; Operto et al., 2013); hence, scattering patterns do not provide for complete characterization of cross-talk.

Including attenuation in FWI complicates cross-talk in a way that makes radiation pattern analysis even less suitable. Cross-talk involving Q generally involves confusion between different parameters at different points in space. For instance, a density perturbation will exhibit cross-talk with a remote Q region obscuring it from the sources and receivers. Another complication is the frequency and phase dependence of these radiation patterns, which are key to distinguishing Q from elastic properties (Innanen and Weglein, 2007; Hak and Mulder, 2011), but whose role in reducing cross-talk is more difficult to discern from a radiation pattern alone. In the previous Chapter, for instance, I used radiation patterns to support the idea that the frequency- and phase-

dependent behaviour of the data must be key to eliminating cross-talk, but radiation patterns alone do not suggest how these data features should be used to effectively suppress cross-talk.

Cross-talk can also be analyzed by examining numerical FWI examples in which cross-talk occurs (e.g. Köhn et al., 2012; Operto et al., 2013). By comparing inversion results to a known true model, errors in the inversion process can be identified. Some of these errors can be attributed to cross-talk, but because this is not the only source of error in inversion, it can be difficult to accurately and confidently identify it. More accurate approaches to quantifying cross-talk can be designed by comparing inversion results of different 'true' models in synthetic tests, as in Kamei and Pratt (2013), or by investigating the Hessian matrix in a specific case, as in Pan et al. (2018).

Here, I describe and analyze with 2D frequency-domain simulations an alternative approach to characterizing viscoelastic cross-talk, with some specially-designed numerical FWI tests on simple models. This approach allows cross-talk between parameters to be understood as a function of either incidence angle or frequency, or both. It also naturally accounts for the effect of iteration and spatial distribution on parameter resolution. If the models considered are suitably chosen, these results allow general conclusions to be drawn about the nature of cross-talk.

In the following sections, I explain which of the methods described in Chapter 2 are used, I outline the proposed measure of cross-talk, and I present a numerical example of application of this approach to assess cross-talk for anelastic inversion.

4.3 Theory

4.3.1 Forward modeling, optimization and parameterization

In this Chapter, I consider the Kolsky-Futterman model of viscoelastic wave propagation I discussed in Chapter 2. Because my goal in this Chapter is to investigate viscoelastic cross-talk, I will assume that there is high confidence that this model of attenuation physics hold. In relation to the discussion in Chapter 3, this means that I will adopt inversion strategies which recover a single model, consistent across all frequencies, rather than the set of models which is shown in that Chapter to be more effective in mitigating the effects of incorrect attenuation physics. This

allows for the investigation of cross-talk to be simplified, and I expect that the main results should translate to the flexible case.

The objective function I consider in this Chapter is the L_2 data-matching term ϕ_D, discussed in Chapter 2, with no added regularization term. I omit any regularization, because cross-talk is primarily a data-fitting feature; regularization terms are usually not formulated in a way that gives rise to confusion between different parameters. Regularization terms can be expected to reduce cross-talk typically, but only by insisting on a priori information. This type of prior information would be inappropriate in a general investigation of cross-talk. As described in Chapter 2, I consider inversion in five parameters: $\zeta_1\rho$, $\zeta_2 v_P^{-2}$, $\zeta_3 Q_P^{-1}$, $\zeta_4 v_S^{-2}$, and $\zeta_5 Q_S^{-1}$, where the ζ_n are unitless scaling terms introduced to improve conditioning.

4.3.2 A measure of cross-talk

Cross-talk is the attribution of data residuals to the wrong model residuals during inversion (e.g., Innanen, 2014). The result is that one model variable is updated in response to a residual in another model variable. Cross-talk from variable 2 into variable 1 can be defined as *the contribution to the inverted value of variable 1 occurring as the result of model error in variable 2*. Variable 1 can be different from variable 2 because they are distinct in spatial position and/or in type; each leads to behavior that can be classified as cross-talk. Our focus is the case in which the variables are different in their parameter type (e.g., Q_P versus ρ). This definition of cross-talk covers an important class of behaviors that are undesirable in FWI, as they tend to slow convergence and give rise to misleading inversion results. The definition also lends itself to direct calculation in synthetic tests.

Let $\mathrm{FWI}(\cdot)$ be an operator that contains a complete process of full waveform inversion; if $\mathbf{d}$ are simulated data measurements generated from the model $\mathbf{m}$, and an initial model $\mathbf{m}^0$ is given, we reconstruct via $\mathrm{FWI}(\cdot,\cdot)$ the estimate

$$\mathbf{m}_{FWI} = \mathrm{FWI}\big(\mathbf{d}(\mathbf{m}), \mathbf{m}^0\big). \tag{4.1}$$

The overall model residual at the outset of the process is $\mathbf{r}_\mathbf{m} = \mathbf{m} - \mathbf{m}^0$. If the model grid has N rows (or depths) and M columns (or lateral positions), then $\mathbf{m}$ is a $(5MN)$-length vector

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}_{vp} \\ \mathbf{m}_{vs} \\ \mathbf{m}_{qp} \\ \mathbf{m}_{qs} \\ \mathbf{m}_{\rho} \end{bmatrix}, \tag{4.2}$$

where each element (e.g., $\mathbf{m}_{vp}$) is an (MN)-length column vector containing the model parameter values. Organized this way, the model residual is evidently

$$\mathbf{r}_\mathbf{m} = \begin{bmatrix} \mathbf{m}_{vp} - \mathbf{m}^0_{vp} \\ \mathbf{m}_{vs} - \mathbf{m}^0_{vs} \\ \mathbf{m}_{qp} - \mathbf{m}^0_{qp} \\ \mathbf{m}_{qs} - \mathbf{m}^0_{qs} \\ \mathbf{m}_{\rho} - \mathbf{m}^0_{\rho} \end{bmatrix}. \tag{4.3}$$

We can focus on model residuals associated with one parameter class only, for instance Q_P, by constructing model residual vectors where the four other contributions are replaced with zero vectors:

$$\mathbf{r}_{qp} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{m}_{qp} - \mathbf{m}^0_{qp} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{4.4}$$

We can also go further by setting all MN of the remaining non-zero elements, except for the nth,

to zero:

$$\mathbf{r}_{qp_n} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \\ 0 \\ \vdots \\ m_{qp_n} - m^0_{qp_n} \\ \vdots \\ 0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{4.5}$$

Similar constructions, containing four zero vectors and one non-zero vector, allow us to form $\mathbf{r}_A$ for any parameter type A of the five, or further, by setting all but the nth of those vector elements to zero, to form $\mathbf{r}_{A_n}$ for any parameter type A or spatial position n of interest.

If the true subsurface model did not differ from the initial model in the nth element of parameter type A, then there would be no contribution from $\mathbf{r}_{A_n}$ to the inversion. We can consider the case in which a given model residual $\mathbf{r}_{A_n}$ does not exist by defining a model

$$\hat{\mathbf{m}}\left(\mathbf{r}_{A_n}\right) = \mathbf{m} - \mathbf{r}_{A_n}. \tag{4.6}$$

The result of the inversion

$$\hat{\mathbf{m}}_{FWI}(\mathbf{r}_{A_n}) = \mathrm{FWI}\big(\ \mathbf{d}(\hat{\mathbf{m}}(\mathbf{r}_{A_n})), \mathbf{m}^0\ \big) \tag{4.7}$$

is then equal to $\mathbf{m}_{FWI}$ minus the contribution to the inversion from the model residual $\mathbf{r}_{A_n}$, implying that the contribution of $\mathbf{r}_{A_n}$ to the inversion result is

$$\Delta\mathbf{m}_{FWI}(\mathbf{r}_{A_n}) = \mathbf{m}_{FWI} - \hat{\mathbf{m}}_{FWI}(\mathbf{r}_{A_n}). \tag{4.8}$$

Insofar as $\Delta\mathbf{m}_{FWI}(\mathbf{r}_{A_n})$ represents changes in the model for variables other than the single non-zero element of $\mathbf{r}_{A_n}$, it represents cross-talk. Cross-talk contributions from a class of variables,

rather than from a single variable, can be computed as $\Delta\hat{\mathbf{m}}(\mathbf{r}_A)$, where $\mathbf{r}_A$ are the model residuals associated with a full class of parameters (as exemplified for Q_P in equation 4.4). By including all of the variables describing parameter type A, the part of $\Delta\hat{\mathbf{m}}(\mathbf{r}_A)$ representing a change in variable B is a full measure of the cross-talk from A into B. In fact, this framework allows the cross-talk from any set of FWI unknowns into any other set to be discussed quantitatively.

There are several advantages to using this approach to characterize cross-talk. It allows for the determination of cross-talk between any two sets of variables, not just those with the same position in space. This means that cross-talk between spatially separated variables can be identified. It offers the potential to study the effects of iteration on the inversion, a key consideration in FWI. It can also help to characterize the effectiveness of Hessian information in suppressing different modes of cross-talk, which can be very difficult to quantify using other approaches.

One may reasonably ask why I consider the contribution of an unknown parameter A, element n, to the inversion to be the difference between the inversion result with and without $\mathbf{r}_{A_n}$, namely $\Delta\mathbf{m}_{FWI}(\mathbf{r}_{A_n})$, rather than the result of an inversion involving model residual $\mathbf{r}_{A_n}$ only. The key difference here is that the latter does not capture the interaction between $\mathbf{r}_{A_n}$ and other model residuals. This distinction may be minor when analyzing some parameter classes, but when considering cross-talk from Q it can lead to a neglect of changes in the amplitude of other recovered anomalies. The definition of cross-talk I use here includes these effects.

4.4 Numerical analysis of viscoelastic cross-talk

In this section, I will discuss how to use the measure of cross-talk introduced above, to allow for general conclusions to be drawn, analogous to how radiation pattern analysis is used. My main objectives are to (1) characterize acquisition geometry effects on cross-talk, to (2) identify the extent of cross-talk from quality factor variables into elastic variables at other positions, and to (3) investigate these behaviors for several optimization strategies. I calculate the cross-talk for a series of simulated problems, using the same model in each, but each time changing the

acquisition geometry, optimization approach, or choice of $\mathbf{r}$. The examples I investigate provide insights about cross-talk when using the specific model parameterization and objective function we use here. They are also representative of how cross-talk could be investigated using this approach with other FWI formulations.

4.4.1 Model geometry and inversion parameters

To allow for relatively general conclusions about cross-talk to be drawn from our synthetic tests, I consider a simple model geometry. More complex model features would introduce highly model-specific cross-talk features, while my aim is to draw broadly applicable conclusions. Because we do not have to limit the cross-talk investigation to point-scatterers only, some decision about the model geometry must be made. In the examples I consider here I am most interested in the cross-talk terms associated with an obscuring Q region, so I will consider a model with an attenuative region masking the elastic anomalies. Figure 4.1 shows the model I consider. An unknown region of low Q_P and Q_S ($Q \approx$20) obscures from the sources and receivers a smaller region containing unknown changes in the elastic parameters (each a 10% increase over the initial model). The elastic region is a small circle at the center of the model; the attenuative low Q_P and Q_S region surrounds it. For simplicity, the initial model used in the inversion is constant for each parameter, set equal to the background values, and access to sufficiently low frequency data to reconstruct the missing wavelengths in the model is assumed. The model is 1 km in width and depth, and is discretized with a 10 m spacing in the finite-difference calculations. I define the reference frequency, ω_0 as 30 Hz in both forward modeling and inversion.

The choice of perturbation amplitudes used in these examples is, in some sense, arbitrary, due to the lack of a characteristic scale defining how the amplitudes of different model parameters relate to one another. Characterizing the extent of cross-talk from equivalent-amplitude model residuals requires that it be established what an equivalent perturbation might be. There is no direct equivalence between parameters: no density perturbation has a corresponding v_P change. A meaningful measure may relate changes to the expected variance in each parameter, but this will

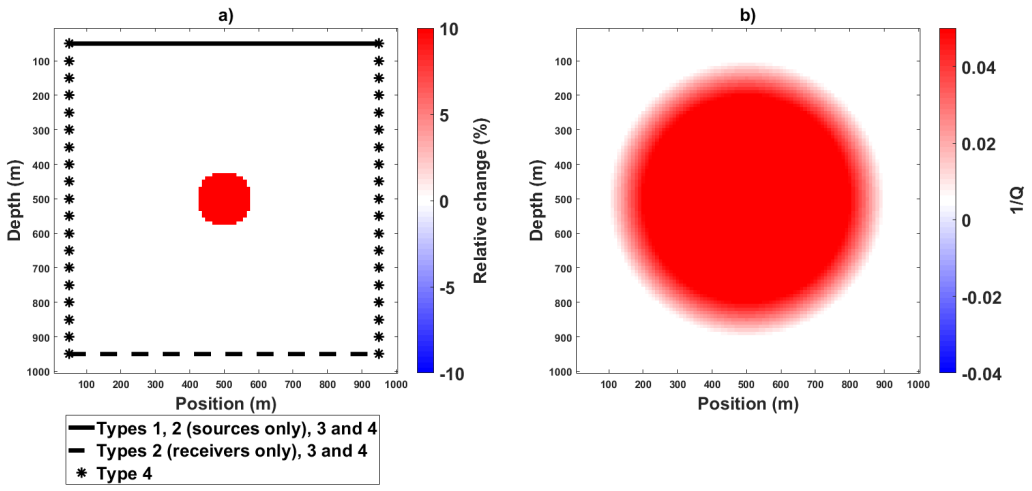

Figure 4.1: The true model involves (a) elastic parameter unknowns confined to a small circular region at the center of the volume, and (b) P- and S-wave mode $1/Q$ unknowns occupying a larger and smooth circular region surrounding the smaller circle.

be highly case-dependent. Here, the only cross-talk amplitudes I seek to define relative to one another are those from one choice of model residual into another model parameter for different acquisition settings. This comparison provides meaningful information about the effect of acquisition geometry on cross-talk amplitudes, and is not as highly dependent on specific model features, like perturbation amplitudes.

In the examples I consider here, I investigate the cross-talk from seven different choices of model residual. In five cases, the residual is chosen to be the full vector of unknowns for each parameter type (i.e., $\mathbf{r}_A$ for parameter A in the terminology of the previous section). Inversion tests on each of these models allow for calculation of the total cross-talk from and to each parameter in the inversion. I refer to the cross-talk from the parameter $\zeta_2 v_P^{-2}$ as 'cross-talk from v_P', and adopt similar terminology for the other parameters. In the other two cases, the residual vector is chosen to include only some of several parameter types, namely, the outer rings of the Q_P and Q_S anomalies, as shown in Figure 4.2 (i.e., $\mathbf{r}_{A_n}$ for values of n corresponding to designated regions of space). These residual vectors are chosen so that we can develop an understanding of cross-talk between the obscuring Q_P and Q_S regions and the elastic anomaly, an important example of

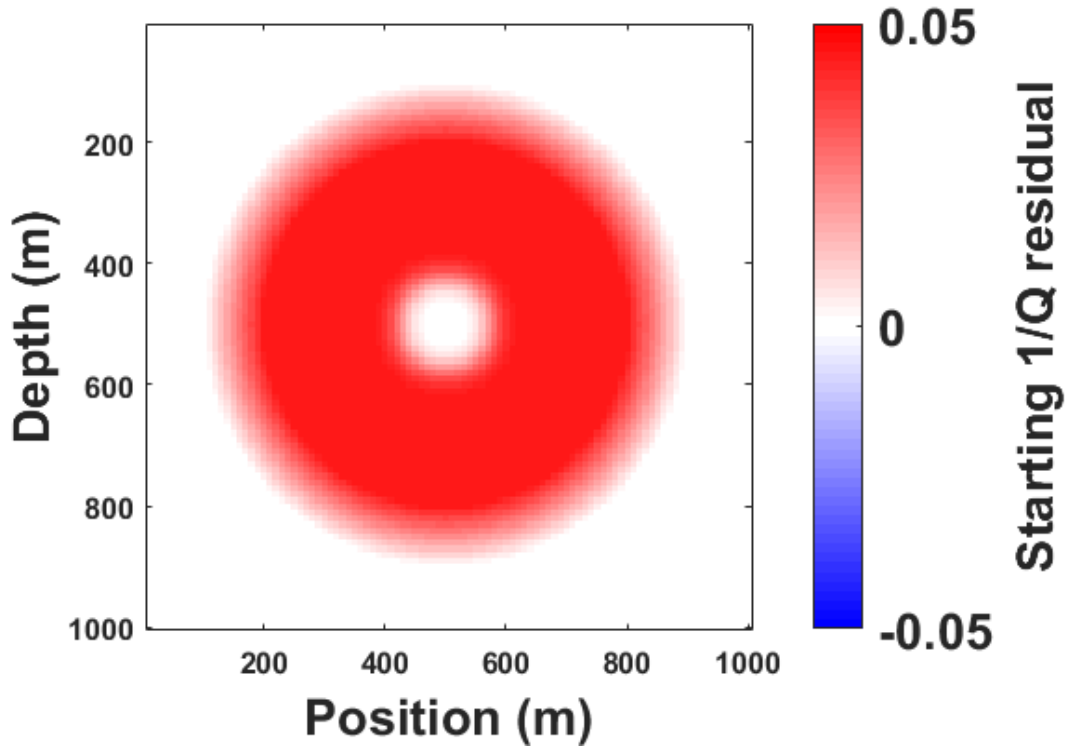

Figure 4.2: Model residual used to calculate cross-talk from Q'_P and Q'_S

cross-talk between spatially distant variables. Instances of cross-talk from the partial (outer ring) model residuals shown in Figure 4.2, for $1/Q_P$ and $1/Q_S$, are referred to as cross-talk from primed variables Q'_P and Q'_S.

I consider four acquisition geometries, each with sources and receivers evenly spaced along one or several edges of the model in Figure 4.1. In Type 1, sources and receivers are placed along the top of the model, and reflections are the main source of information. In Type 2, sources are placed at the top of the model and receivers on the bottom, simulating a transmission or cross-well geometry. In Type 3, sources/receivers are placed on both the top and bottom of the model, providing comprehensive reflection and transmission information. In Type 4, to examine fundamental features of cross-talk, sources/receivers are placed on all four sides. These acquisition geometries were chosen to help develop an understanding of the relative importance of reflection and transmission ray-paths in the resolution of the parameters. They are summarized in Table 4.1. Explosive sources are considered, and both components of displacement are recorded at the receivers. In each

Acquisition	Type 1	Type 2	Type 3	Type 4
Sources top	✓	✓	✓	✓
Receivers top	✓		✓	✓
Sources bottom			✓	✓
Receivers bottom		✓	✓	✓
Sources and receivers sides				✓

Table 4.1: Source and receiver locations for each acquisition geometry type considered.

case, I consider 49 evenly spaced sources and 98 evenly spaced receivers per model edge included.

In each inversion test, ten frequency bands are inverted, each containing five evenly-spaced frequency values. The upper end of the bands is increased with iteration, the first spanning 1-2Hz, and the last 1-20Hz. No regularization is used, except in the examples involving truncated Gauss-Newton optimization, where a small stabilization term is added in the calculation of Hessian-vector products. Four different numerical optimization strategies are considered.

Because the examples we consider here involve different acquisition geometries, the amplitude of the recovered anomalies can vary substantially. In the figures of this section, I plot a normalized 'relative cross-talk' for display purposes. I define this 'relative cross-talk' from model residual r_{A_n} into variable p as $\mathbf{m}_{\mathbf{p}_{\mathbf{FWI}}}(\mathbf{r}_{A_n}) / \max(|\mathbf{m}_{\mathbf{p}_{\mathbf{FWI}}} - \mathbf{m}_{\mathbf{p}}^{\mathbf{0}}|)$. This gives the cross-talk in a given parameter normalized by the total update in that parameter. Using this metric, we can easily see what fraction of the update in a given parameter is cross-talk. This metric does not lend itself to comparison of the amplitude of cross-talk into different parameters, but I try to avoid comparisons of this type regardless, due to a lack of general a priori scale defining amplitudes of one parameter with respect to another.

4.4.2 Optimization

In these examples, I consider four potential optimization strategies, each using the same ten frequency bands defined above. In the first and least intensive approach, I consider one iteration of steepest-descent optimization at each frequency band (ten total iterations). This set-up allows us to examine the cross-talk behavior with limited impact from the iterative nature of the inversion

problem. This is the type of cross-talk that radiation pattern analysis is best suited to describe.

While the approach above is informative about the cross-talk in the gradient, there are limits to the insight it can provide about the actual experience of solving the FWI problem, in which iteration plays an important role. To investigate the effects of iteration on cross-talk, I also consider a more intensive steepest-descent optimization, with five iterations per frequency band (i.e., involving a total of 50 iterations). Contrasting the results obtained using this strategy with those from the single-iteration per band approach should be informative about the effects that iteration has on recovered cross-talk.

The third optimization strategy I consider is a truncated Gauss-Newton approach, wherein the Gauss-Newton step is iteratively estimated (e.g. Métivier et al., 2013). Through consideration of the Hessian, this approach includes second-derivative information that should be useful in reducing cross-talk (Operto et al., 2013; Innanen, 2014). Five inner-loop iterations were used at each FWI iteration to calculate the approximation to the Gauss-Newton step. Each of these inner-loop iterations requires the evaluation of a Hessian-vector product, with a cost of two wavefield evaluations. I consider one outer-loop iteration of this truncated Gauss-Newton optimization per frequency band, resulting in a computational cost similar to that of the five-iteration steepest-descent example. By contrasting the results using this approach with the multi-iteration steepest-descent strategy, we can learn about the effects of the Hessian information considered.

Finally, I consider a computationally intensive approach to allow for large amounts of second-derivative information to be considered. For this approach, I use 30 inner-loop iterations in a truncated Gauss-Newton strategy, with one outer-loop iteration per frequency band. The resulting computational cost is substantially higher than for the other approaches considered, but should allow for an understanding of how cross-talk is affected by more substantial numerical optimization.

4.4.3 Examples

Cross-talk into the inversion parameter $\zeta_2 v_P^{-2}$, in terms of the equivalent relative change in v_P for the first three optimization techniques considered is shown in Figure 4.3. In this plot, each row

corresponds to the cross-talk ($\Delta\mathbf{m}_{FWI}(\mathbf{r}_A)$) for a given optimization strategy and acquisition geometry, and each column corresponds to model residual causing the cross-talk ($\mathbf{r}_A$). The leftmost column shows the inversion result ($\mathbf{m}_{FWI}$) for each acquisition and optimization strategy for comparison, normalized in the same sense as the relative cross-talk terms. Within each panel the x and z dimensions represent x and z position. This plot is continued to the high-cost TGN optimization approach in Figure 4.4.

Figures 4.3 and 4.4 illustrate some important features of cross-talk into v_P. Cross-talk from all residuals considered is substantial in the Type 1 acquisition for each optimization strategy considered. Other acquisition geometries display cross-talk that is much more sensitive to the choice of optimization strategy. The cross-talk from ρ and Q_S is not very persistent relative to that from the other parameters; little cross-talk is observed except in the cases with reflection-type acquisition or single-iteration SD optimization. Transmission-type measurements, and to some extent the more intensive optimization strategies, provide information that allow for the removal of cross-talk from these parameters. Cross-talk from Q_P, in contrast, is much more persistent between acquisition types and optimization strategies, as expected given the mathematically similar roles played by the two parameters in equations 2.10 and 2.28. Leakage from v_S is comparably robust. These modes of cross-talk are severely attenuated only when both transmission type data and intensive geometry are present (Figure 4.4, Type 2-4), or with completely surrounding acquisition geometry (Figure 4.3, Type 4). It is interesting to note that the Type 4, surround acquisition, is much more effective in suppressing these cross-talk modes than the Type 3 acquisition for the low-cost optimization strategies. This suggests that, while comprehensive geometry can help mitigate cross-talk, even significant access to both reflections and transmissions provides insufficient data for removing these modes of cross-talk at low computational cost. Any realistic acquisition will require substantial optimization to eliminate these modes.

Focusing on the effects of optimization strategy in Figure 4.3, it is evident that there is substantial reduction in cross-talk from Q_P when performing more SD iterations, except in the Type

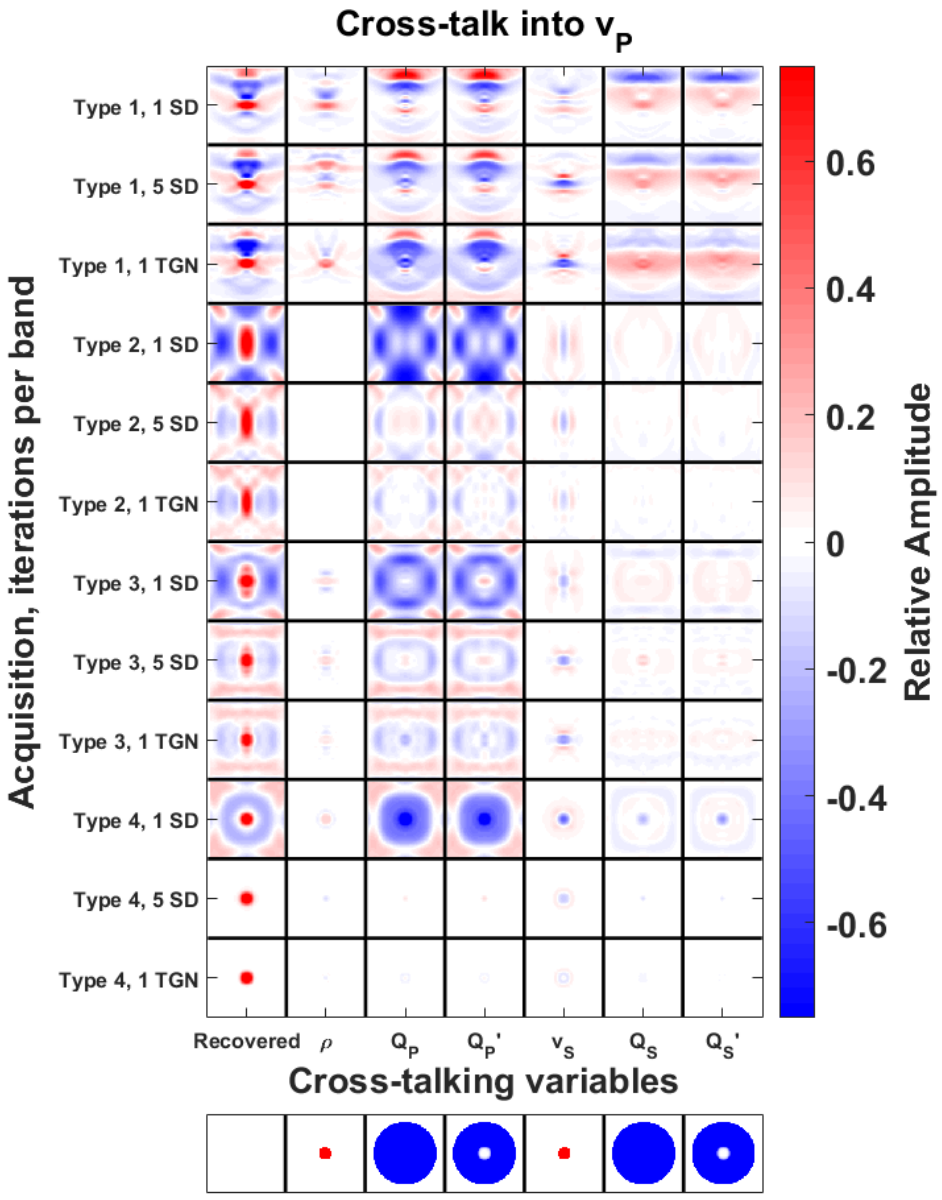

Figure 4.3: Numerically calculated cross-talk into v_P. The left column shows the inversion result for each acquisition geometry and the first three optimization strategies. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

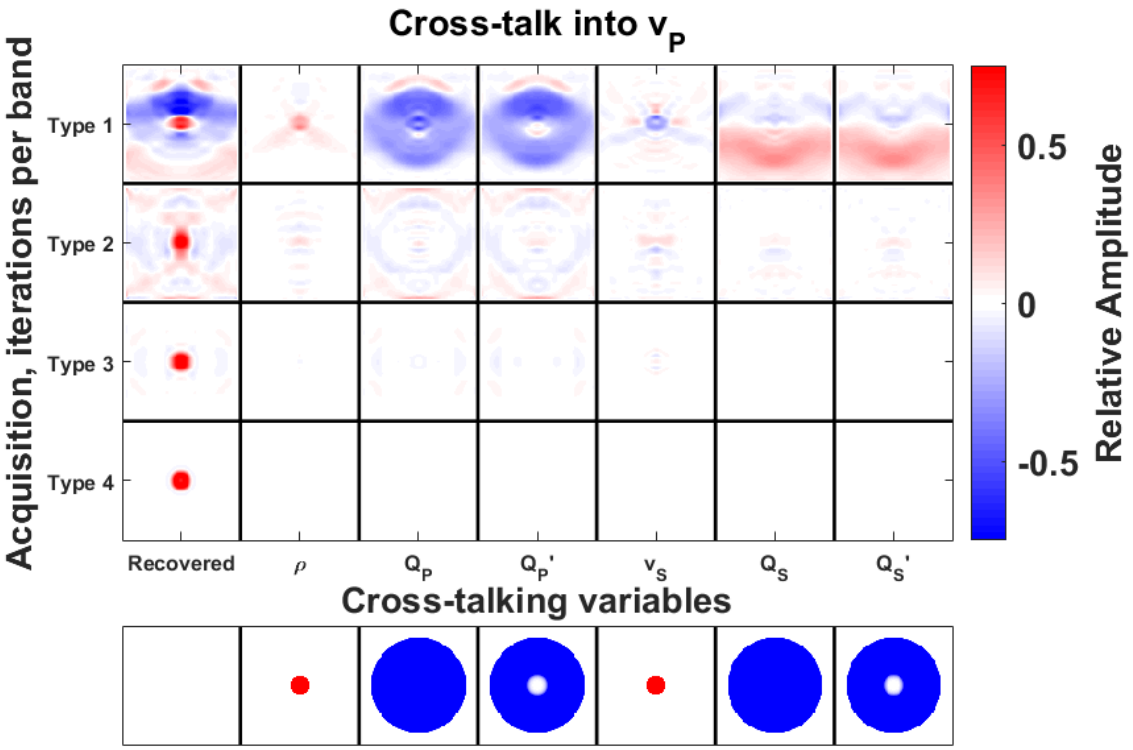

Figure 4.4: Numerically calculated cross-talk into v_P for computationally intensive optimization strategy. The left column shows the inversion result for each acquisition geometry. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

1, reflection dominated acquisition. This can be seen through comparison of the single- and multi-iteration per band SD results. In contrast, there is comparatively little improvement between the multi-iteration SD and the equivalent cost TGN approaches. If these variables are distinguished by the data, the information to discriminate between them should be largely present in the Hessian. The failure of the low-cost TGN approach considered in Figure 4.3 to reduce cross-talk substantially suggests that the update we consider is an inadequate approximation of the Newton step; more inner-loop iterations are likely needed to provide effective cross-talk reducing information in this case. The results for the more intensive TGN optimization strategy are shown in Figure 4.4. Comparison of Figures 4.3 and 4.4 shows substantial reduction in cross-talk from Q_P and v_S in the Type 2 and Type 3 cases. This contrast indicates that there is a strong dependence of these modes of cross-talk on both acquisition type, and on the optimization strategy used; while transmission information doesn't substantially reduce cross-talk with the less intensive optimization strategies (Figure 4.3, Types 2, 3) it becomes important with more intensive optimization (Figure 4.4). In contrast, cross-talk in the reflection-dominated Type 1 acquisition geometry remains fairly strong, even with intensive optimization.

Comparing the cross-talk contributions of Q_P and Q'_P in Figure 4.3, it is clear that substantial cross-talk contributions are made by spatially distant variables. We can observe that cross-talk from Q_P and from Q'_P to v_P is very similar, and that the latter has substantial amplitudes in the region not included in Q'_P (the center). These observations mean that the contribution of the missing part of Q'_P must be small, and that the cross-talk observed at the center region is predominately from outside this region: otherwise it would not be observed in the Q'_P case. This suggests that the cross-talk between spatially separated v_P and Q_P residuals is relatively large for at least some acquisition geometries (Types 1 and 3 here).

The cross-talk into ρ for the same set of examples is shown in Figures 4.5 and 4.6. A prominent difference from the cross-talk into v_P in Figure 4.3 for low-cost optimization is the substantial persistence in the cross-talk from all model residuals considered, even for acquisition geometries

measuring both transmitted and reflected energy. This is likely due to a considerable similarity between the effects of density and other parameters on the data at reflection angles, and the lack of any substantial effect of density features on transmission-type data. Even with Type 4 acquisition, cross-talk remains significant in this case. Unlike the v_P case, cross-talk into ρ is little affected by iteration here. Once again in this case, cross-talk from Q_P and Q_S is little distinguished from that of Q'_P and Q'_S, even in the region where they differ. This supports the idea that it is the obscuring Q region and not the co-located Q features that introduce most of the observed cross-talk at the elastic anomaly.

Figure 4.6 shows the high-computation TGN results for cross-talk into density. As in the v_P case, the changes are most pronounced for the acquisition geometries including transmission information; acquisition Types 2, 3, and 4 all undergo dramatic reductions in cross-talk with the more intensive optimization. For Type 3 and 4 acquisitions, almost all cross-talk into ρ is eliminated in this case, and the heavy cross-talk into density for acquisition Type 2 is substantially reduced. The more intensive strategy has again resulted in only mild improvements for the Type 1 acquisition. This more intensive approach does not seem to preferentially reduce at-a-distance cross-talk: contributions from Q_P and Q'_P remain very similar here.

Figure 4.7 shows the cross-talk into v_S for the low-cost optimization strategies. As with cross-talk into v_P, the reflection-dominated Type 1 acquisition exhibits substantially greater cross-talk than geometries with more transmission information. The parameters contributing the most persistent cross-talk in this case are v_P and Q_S. While iteration continues to have a major effect on cross-talk in acquisition geometries with transmission information, there are also cases in which the low-cost TGN approach outperforms the multi-iteration SD strategy here. For acquisition Types 3 and 4, both cross-talk from v_P and Q_P is substantially reduced in the low-cost TGN case as compared to the multi-iteration SD case. This suggests that the low-cost TGN approximation we consider here is adequate to eliminate cross-talk into some parameters for some acquisitions, but is ineffective in eliminating cross-talk for others. Comparing these results with the high-cost

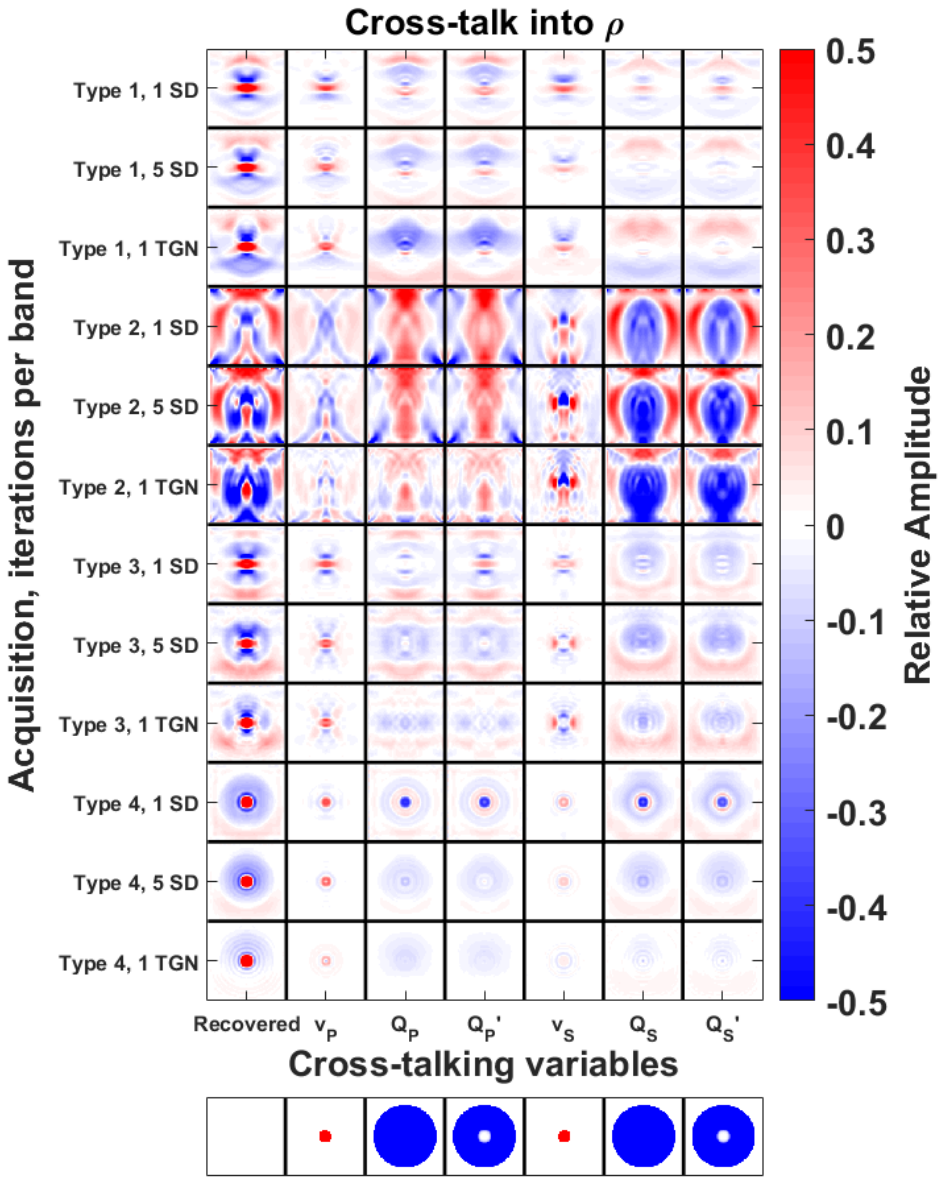

Figure 4.5: Numerically calculated cross-talk into ρ. The left column shows the inversion result for each acquisition geometry and the first three optimization strategies. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

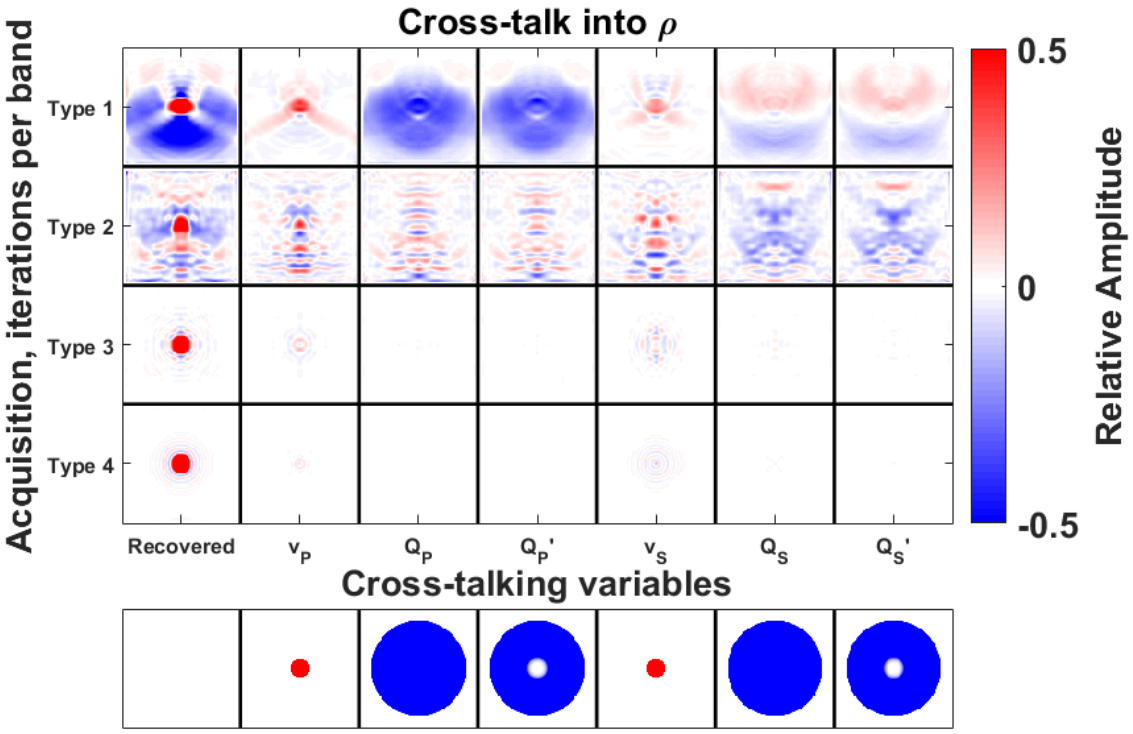

Figure 4.6: Numerically calculated cross-talk into ρ for computationally intensive optimization strategy. The left column shows the inversion result for each acquisition geometry. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

TGN results shown in Figure 4.8, we can see that the more intensive optimization approach is very effective in mitigating cross-talk from v_P, Q_P, and Q_S for the transmission type acquisitions (Types 2-4). Even this more intensive optimization strategy struggles to eliminate cross-talk in the reflection-type acquisition geometry. This suggests that even more Hessian information would be needed to eliminate cross-talk when only reflection-type acquisition is available.

Cross-talk from Q_S and Q'_S into v_S shows strong evidence of leakage between spatially separated variables (Q'_S contributes substantially to cross-talk at the center of the model), but also of cross-talk between co-located variables (the shape of the Q_S model residual has a clear signature). This is similar to the cross-talk from Q variables we observed for most of the v_P and ρ examples considered earlier. Cross-talk from Q_P and Q'_P into v_S, interestingly, shows a different behavior with multi-iteration SD optimization for Types 3 and 4 acquisition. In these cases, there is little evidence of cross-talk between co-located variables; there is no imprint of the Q_P anomaly in the cross-talk. However, there is substantial cross-talk at the center of the model. In the case of the Q'_P residual, this cross-talk has almost no spatial overlap with the model residual, and all cross-talk is clearly occurring between spatially separated variables in this case. It is difficult to see how this mode of cross-talk would be identified with a conventional analysis.

The cross-talk into Q_P, shown in terms of $\frac{1}{Q_P}$ in Figure 4.9, is very strong for the reflection-type data in the Type 1 acquisition as compared with other acquisition geometries. This suggests that transmission-type information plays an important role in generating accurate Q_P estimates. In this case, cross-talk from non-v_P variables becomes relatively small for acquisition other than Type 1, while the v_P cross-talk persists for all acquisitions considered. Strikingly, there is little cross-talk *from* the elastic anomalies *into* an obscuring Q_P region, despite the fact that the reverse is observed for all three elastic parameters. Instead, same-location cross-talk seems to be the main mode from the elastic parameters. This is an interesting asymmetry that should be further investigated. The high-cost optimization results shown in Figure 4.10 show a similar trend to the previous results: cross-talk is substantially reduced when transmission-type data are available, but not for the Type

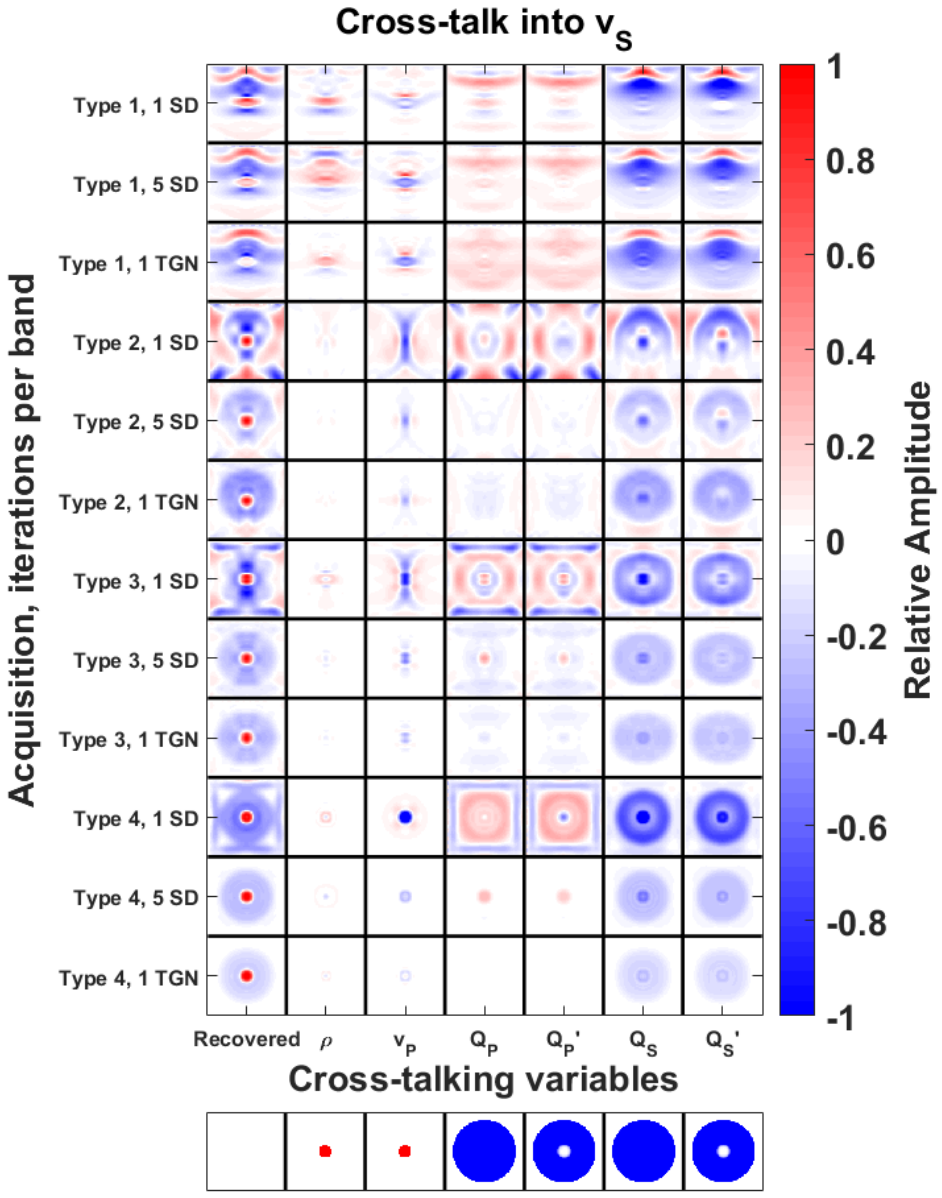

Figure 4.7: Numerically calculated cross-talk into v_S. The left column shows the inversion result for each acquisition geometry and the first three optimization strategies. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

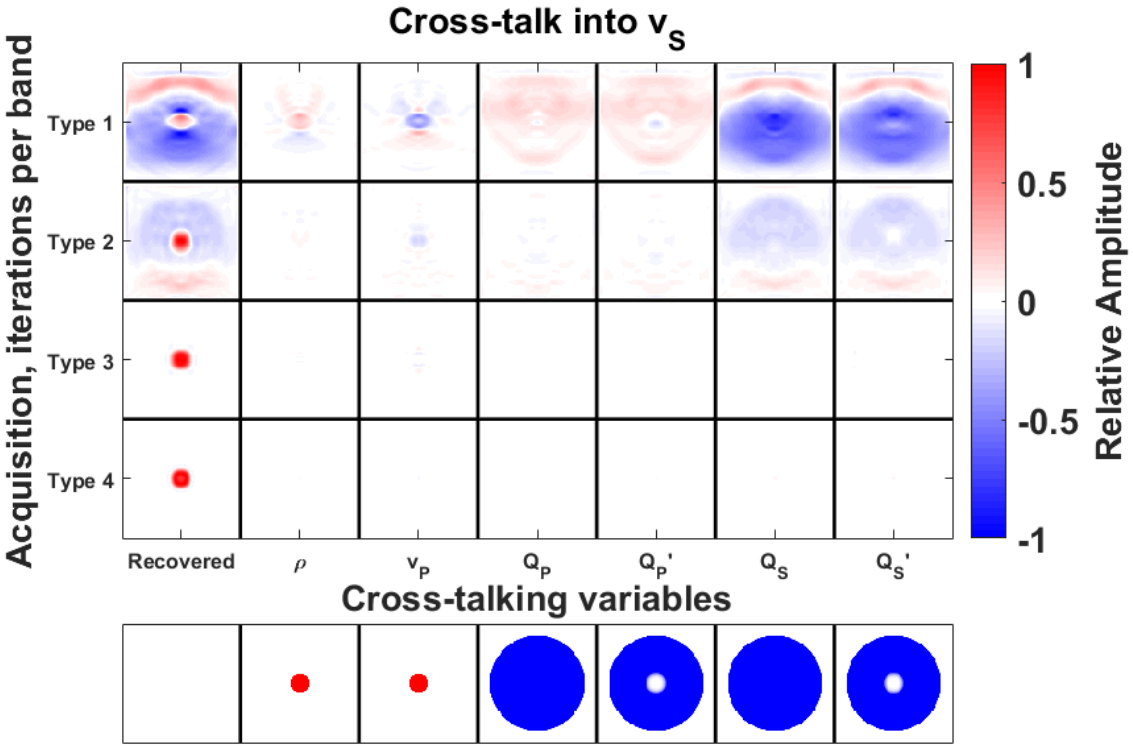

Figure 4.8: Numerically calculated cross-talk into v_S for computationally intensive optimization strategy. The left column shows the inversion result for each acquisition geometry. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

1 acquisition. This reinforces the idea that even the high-cost TGN approximation I consider is inadequate for reflection-type data.

Finally, we can consider cross-talk into Q_S, shown in terms of $\frac{1}{Q_S}$ in Figure 4.11. As expected, cross-talk with v_S is persistent in this case, though leakages from v_P and Q_P are as well. In comparison to cross-talk into Q_P, transmission information seems relatively ineffective in reducing cross-talk. This is likely because I consider explosive sources here, while Q_S will mainly affect shear-waves. This means that mode conversions, which will involve interactions between the model residuals and the P-waves from the source, play a major role, even at transmission angles. This may also explain why v_P and Q_P cross-talk contributions are much more robust in this case than v_S and Q_S leakages into Q_P were (Figure 4.9). As in the previous cases considered, the high-cost TGN approach dramatically reduces cross-talk for acquisition types which include transmission data (Figure 4.12), but is less effective in the Type 1 acquisition. The otherwise persistent cross-talk from v_S is especially attenuated for the transmission-type acquisitions.

4.4.4 The effects of source-type

The acquisition types discussed above differed in the location of the sources and receivers, but considered only explosive sources. While many sources used in seismic surveys are well-approximated by an explosive source, others are better represented by a directed point force. To investigate the effect of source-type, I compare the recovered cross-talk from explosive sources and z-directed point sources in this section. For simplicity, only the multi-iteration SD optimization strategy is considered here.

Figure 4.13 shows a comparison of cross-talk into v_P for explosive and z-oriented point forces. In general, the examples with explosive sources seem less prone to cross-talk than those with directed point forces. In particular, cross-talk from Q_S is substantially greater in the case of directed point forces. This leakage is likely due to the fact that both S- and P-waves originate at the source in this case, and the resulting ambiguities greatly increase the potential for cross-talk.

Cross-talk into ρ for different sources is shown in Figure 4.14. In this case, the relation of

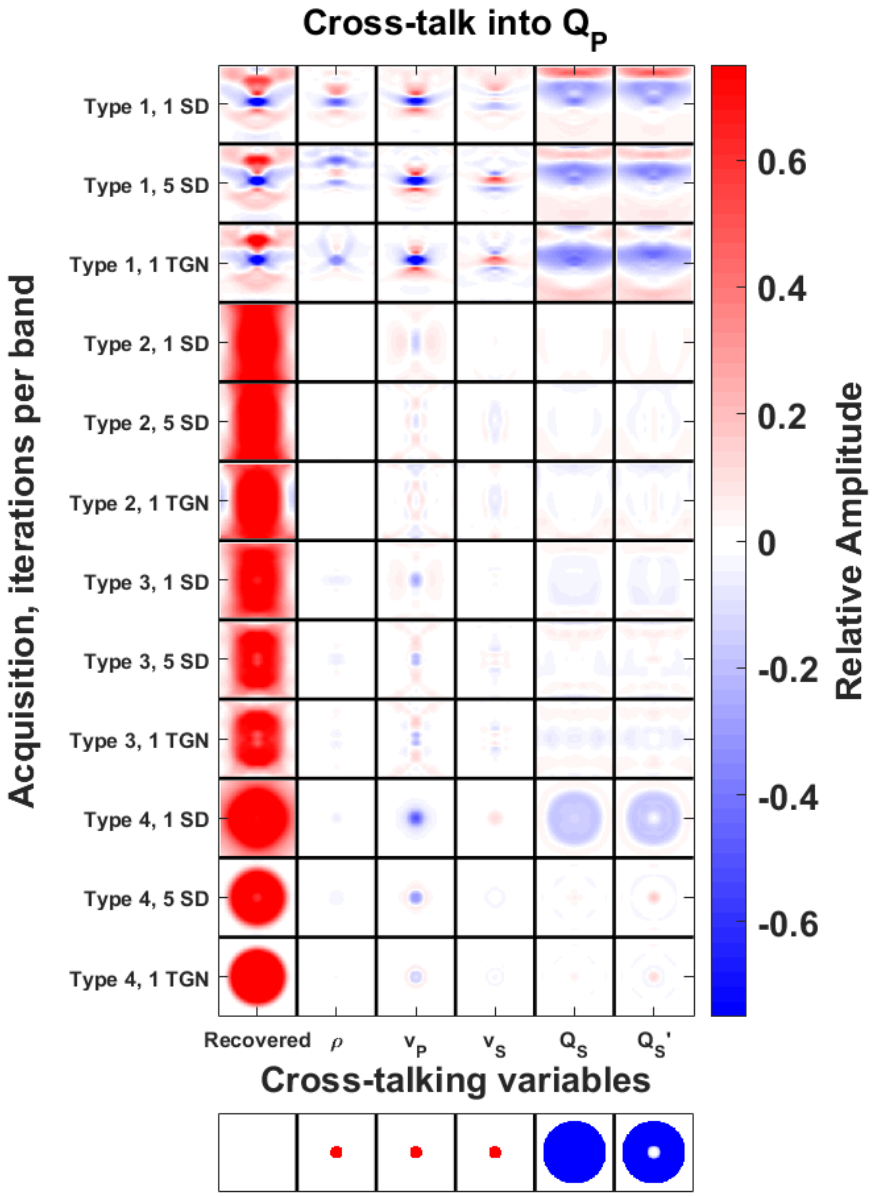

Figure 4.9: Numerically calculated cross-talk into Q_P, shown as relative change in $\frac{1}{Q_P}$. The left column shows the inversion result for each acquisition geometry and the first three optimization strategies. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

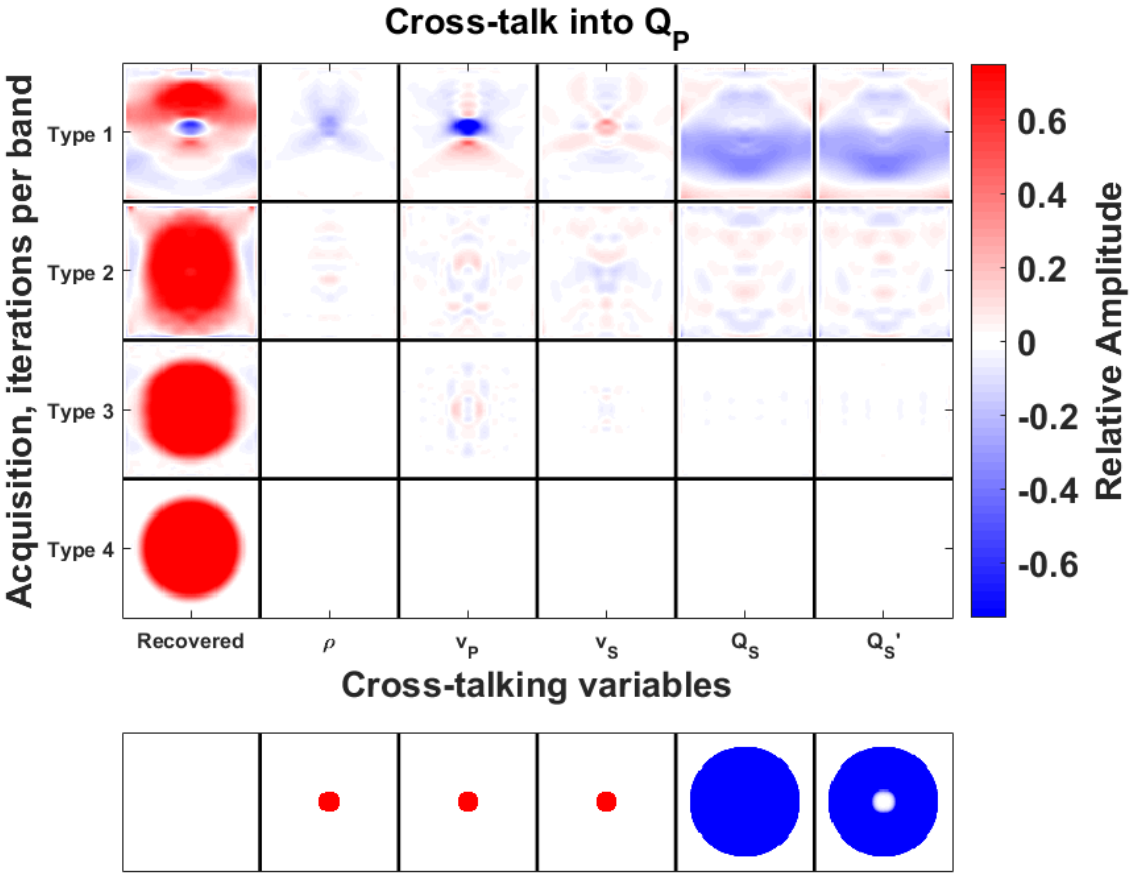

Figure 4.10: Numerically calculated cross-talk into Q_P for computationally intensive optimization strategy, shown as relative change in $\frac{1}{Q_P}$. The left column shows the inversion result for each acquisition geometry. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

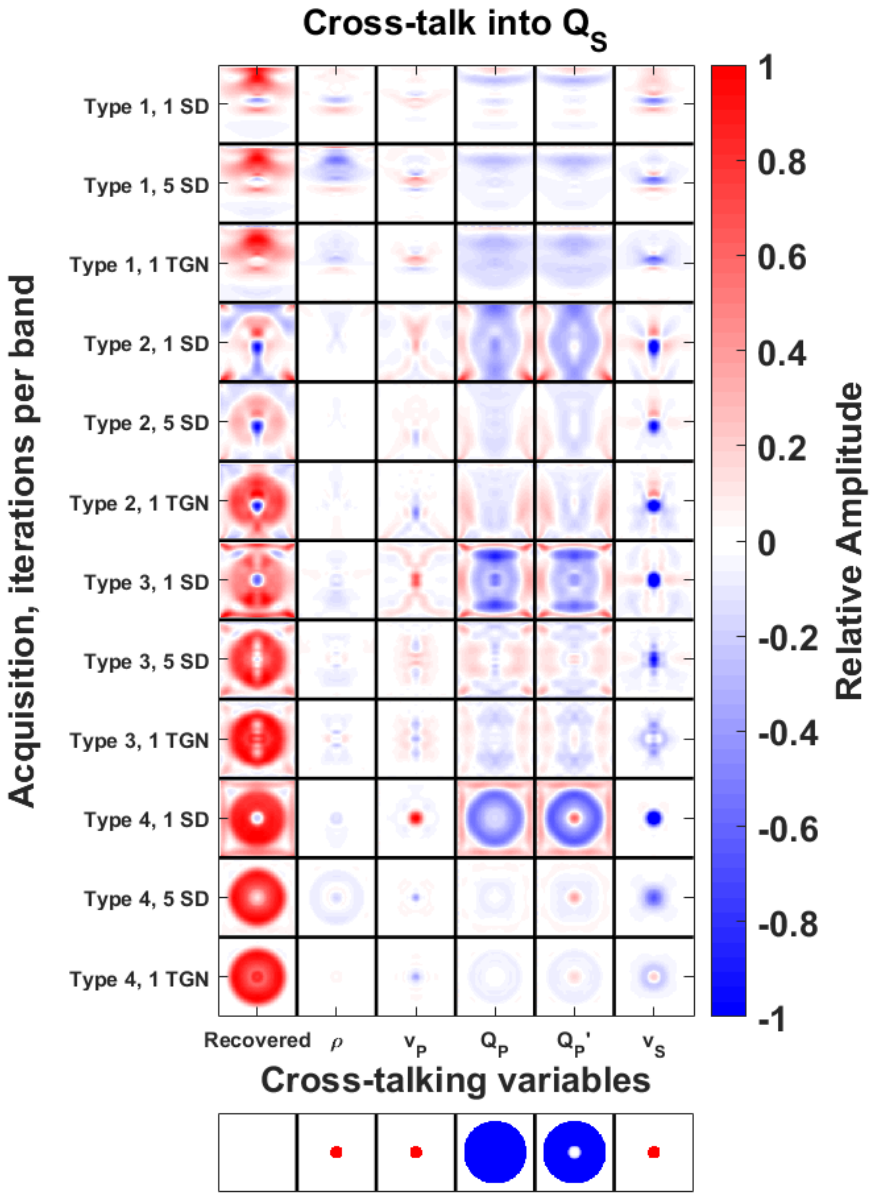

Figure 4.11: Numerically calculated cross-talk into Q_S, shown as relative change in $\frac{1}{Q_S}$. The left column shows the inversion result for each acquisition geometry and the first three optimization strategies. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

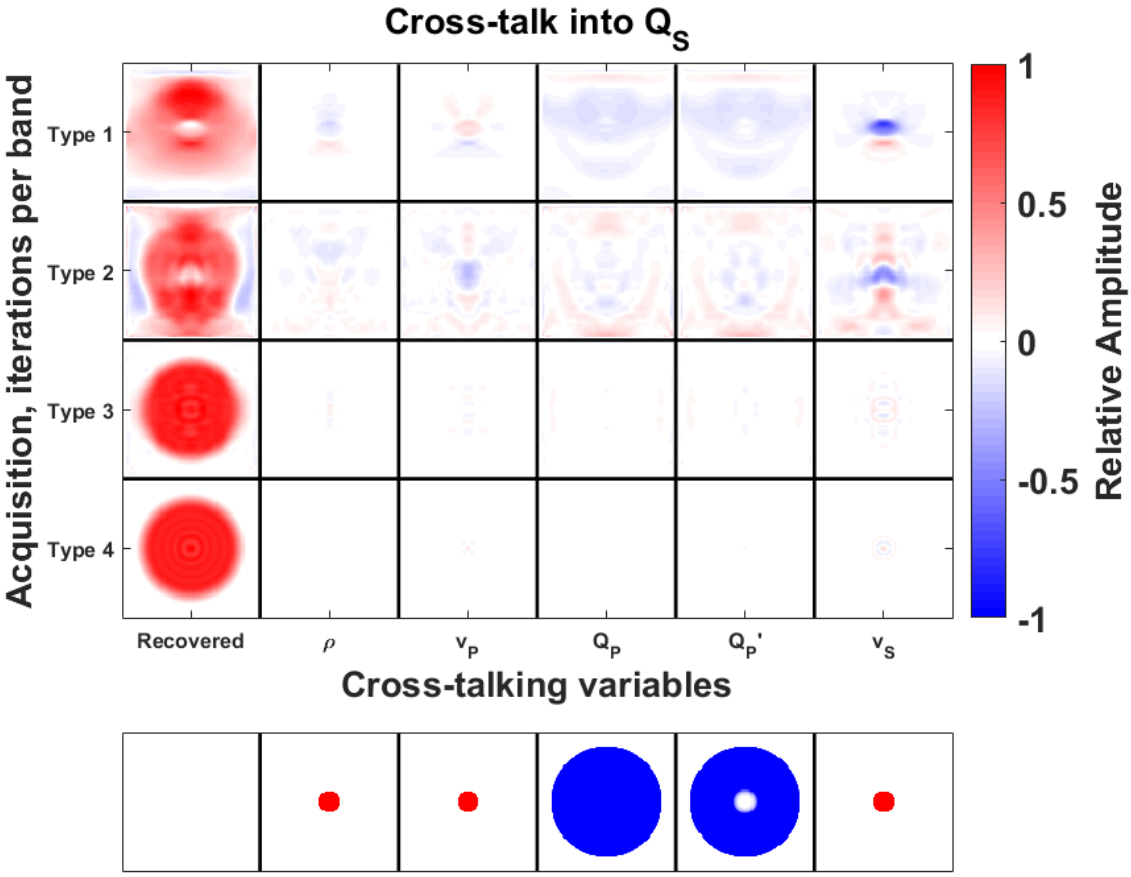

Figure 4.12: Numerically calculated cross-talk into Q_S for computationally intensive optimization strategy, shown as relative change in $\frac{1}{Q_S}$. The left column shows the inversion result for each acquisition geometry. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

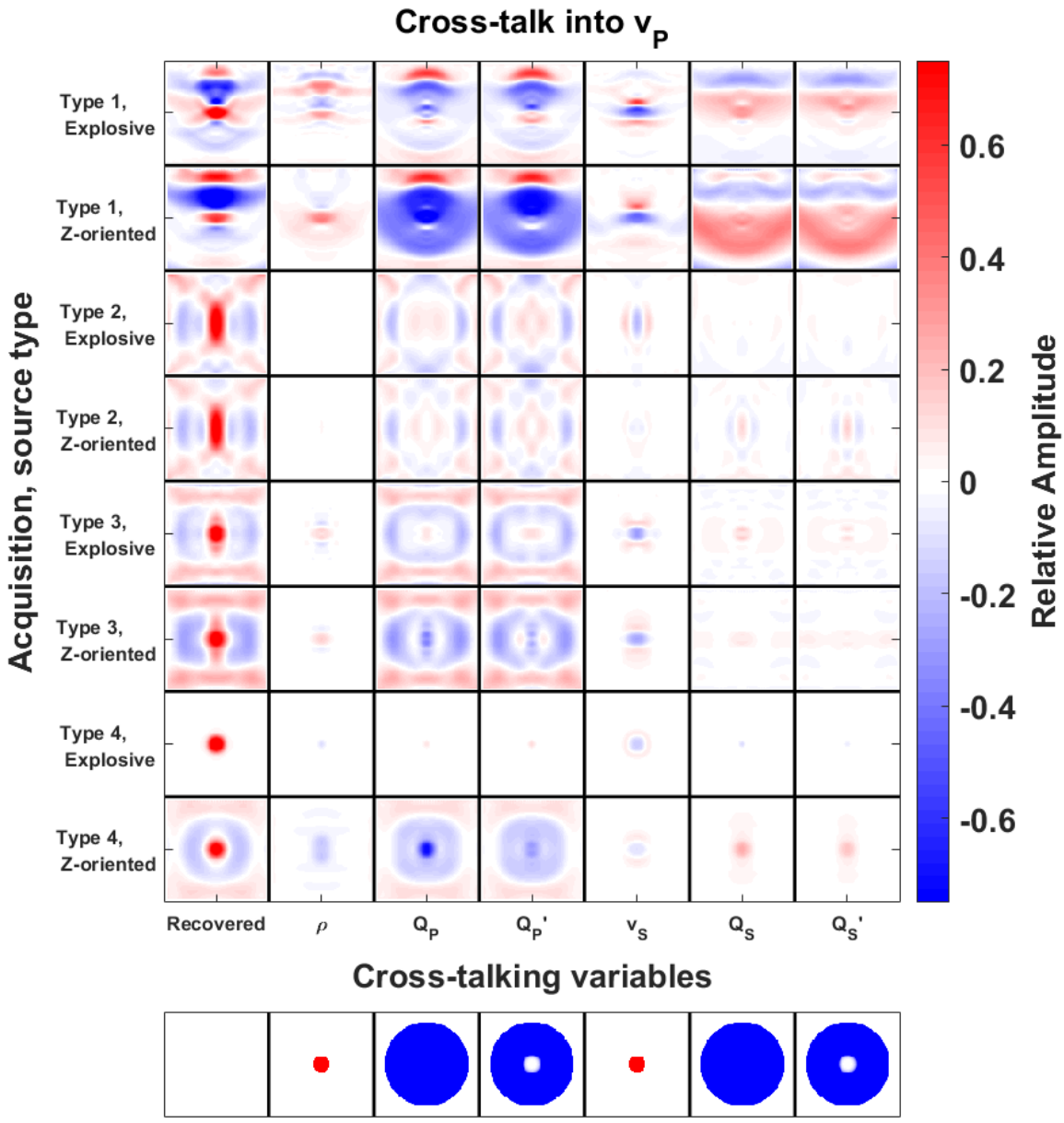

Figure 4.13: Numerically calculated cross-talk into v_P. The left column shows the inversion result for each acquisition geometry and source type considered. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

cross-talk to source-type is more mixed; Type 1 acquisition experiences more cross-talk with a directed point force, while Types 3 and 4 see a reduction in cross-talk in this case. Transmission-only data are very poorly suited to the recovery of density information in either case. In contrast, cross-talk into v_S, (see Figure 4.15) is reduced for all acquisition geometries when considering directed point forces. In particular, acquisitions with transmission information (types 2-4) see substantial improvements. This is presumably due to the S-waves generated at the source providing better constraint on the v_S features. In the case of directed point forces, the cross-talk from Q_S is effectively reduced when transmission information is available, similar to the cross-talk from Q_P to v_P for explosive sources. This suggests that it is much more difficult to distinguish between velocities and the corresponding Q parameters when there is insufficient source energy in the corresponding wave mode (P or S).

The leakage into the Q variables is shown in Figures 4.16 and 4.17. The differences between these sources for cross-talk into Q_P is relatively minor, with a small increase in cross-talk from v_P observed when using directed point forces. The effects of source-type on cross-talk into Q_S are much more pronounced. Generally, the extent of cross-talk is much smaller when directed point forces are used, and only cross-talk from v_S persists for acquisitions with transmission-type information in this case. This again suggests that sources exciting S-waves allow for much better discrimination of S-wave parameters.

4.5 Discussion

Comparison of the numerically calculated cross-talk quantities introduced in this Chapter reveal key aspects of the cross-talk modes present in viscoelastic FWI. While comparison of the amplitudes of different modes of cross-talk is difficult without some reference scale to compare expected parameter changes, these results are indicative of how each mode of cross-talk changes with acquisition geometry, much like scattering radiation pattern analysis. This approach has the added advantage of identifying the effects of optimization strategy on different modes of cross-talk. With this strategy, I was able to reproduce some of the results which radiation pattern analysis can

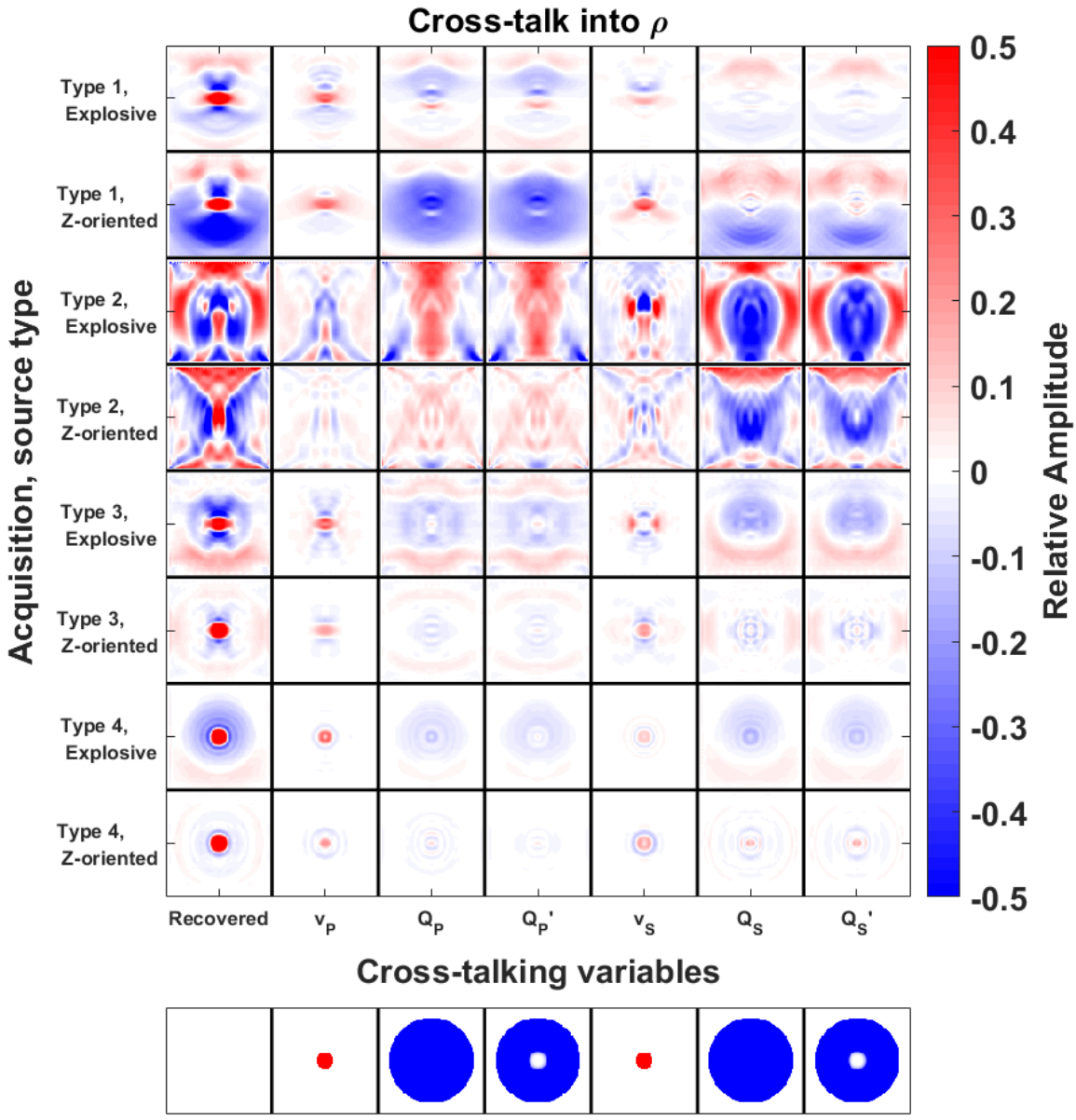

Figure 4.14: Numerically calculated cross-talk into ρ. The left column shows the inversion result for each acquisition geometry and source type considered. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

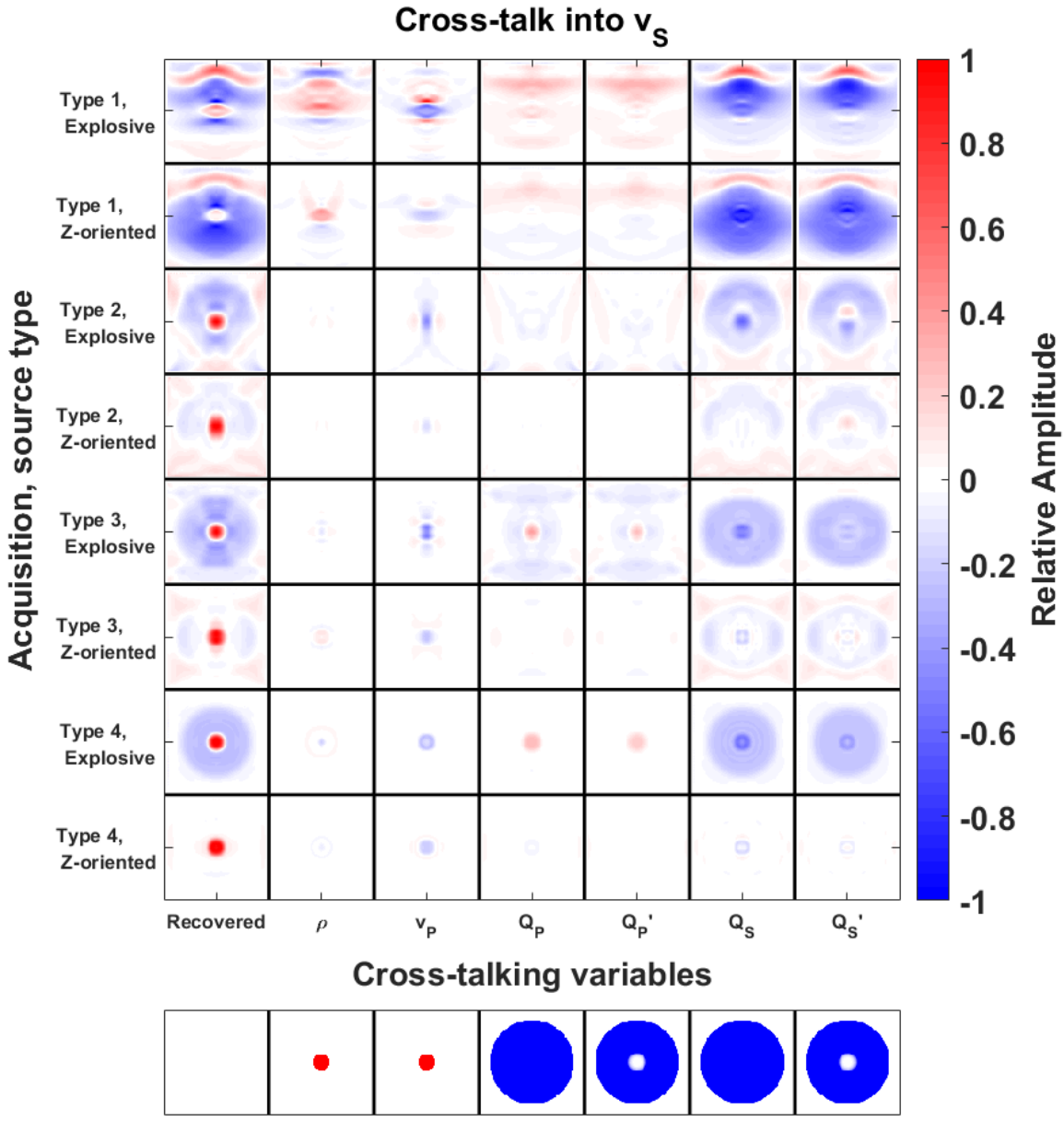

Figure 4.15: Numerically calculated cross-talk into v_S. The left column shows the inversion result for each acquisition geometry and source type considered. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

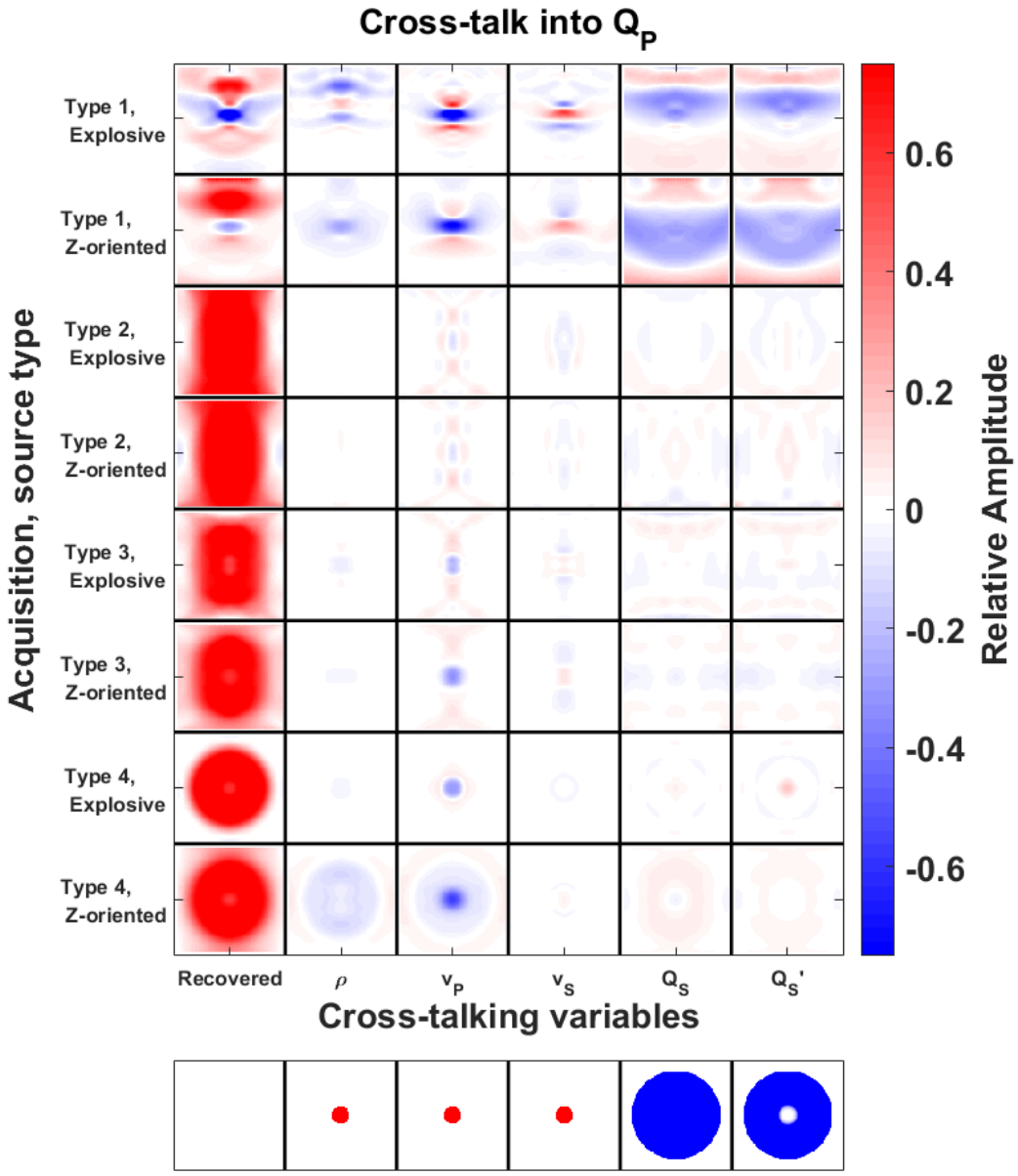

Figure 4.16: Numerically calculated cross-talk into Q_P, shown as relative change in $\frac{1}{Q_P}$. The left column shows the inversion result for each acquisition geometry and source type considered. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

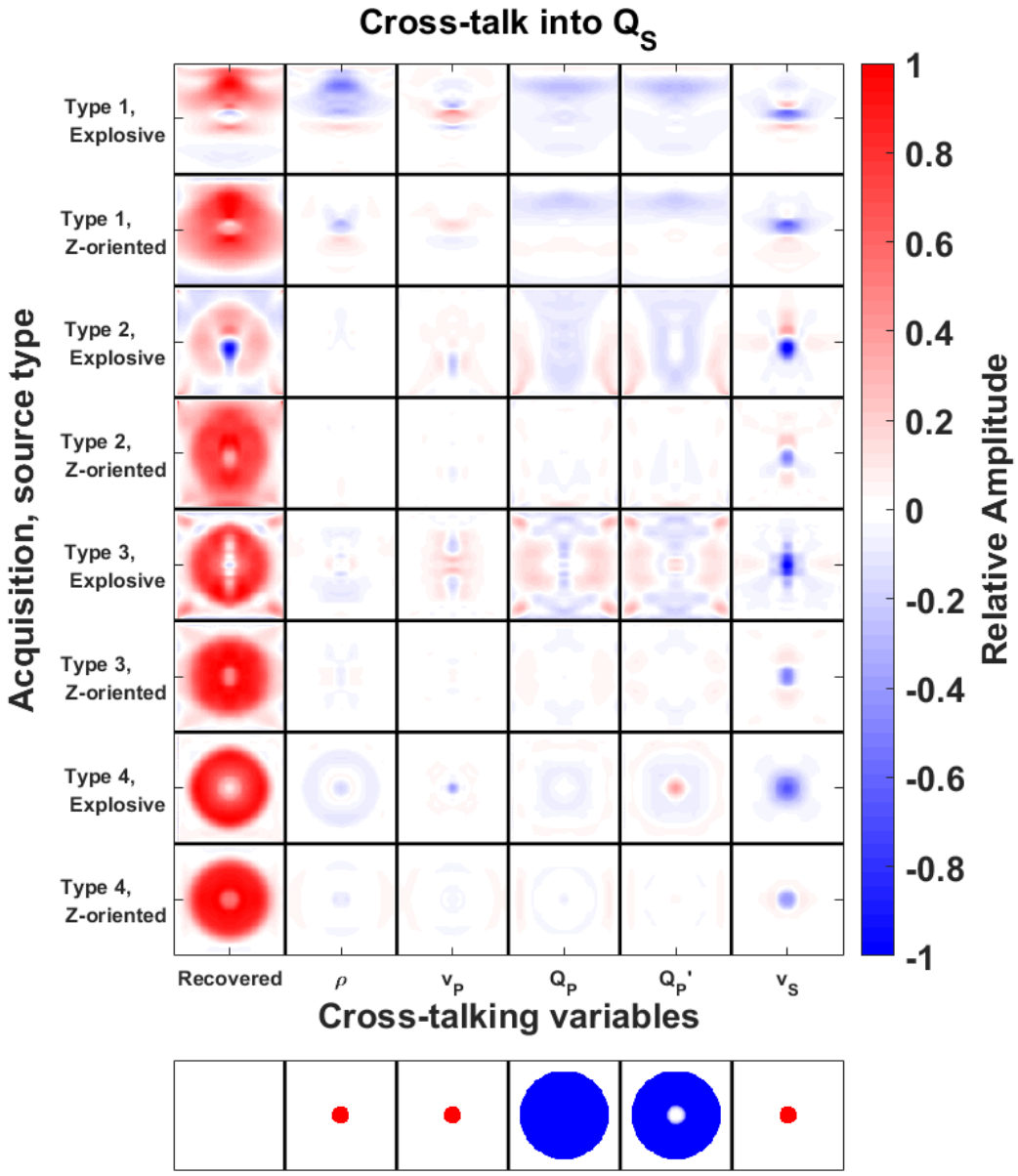

Figure 4.17: Numerically calculated cross-talk into Q_S, shown as relative change in $\frac{1}{Q_S}$. The left column shows the inversion result for each acquisition geometry and source type considered. Below the x-axis the model residuals introducing the cross-talk are shown. Within each panel, x and z position correspond to location in 2D space.

achieve in the elastic case. For instance, my approach highlights the importance of transmission-type data for reducing cross-talk from ρ into v_P, and likewise demonstrates the importance of reflection-type data for reducing cross-talk into ρ from other elastic parameters. This approach identifies cross-talk features in the anelastic problem that conventional radiation pattern analysis would struggle with, however. Cross-talk from an obscuring Q cloud into elastic parameters has a considerable at-a-distance contribution. This mode of cross-talk is difficult to remove, even with more comprehensive acquisition and optimization strategies. Cross-talk into an obscuring Q region conversely, has little contribution between spatially separated parameters and seems to be effectively attenuated when transmission data are available. Insights like these are difficult to glean from conventional radiation pattern analysis, where only co-located, point-like scatterers can be considered and no clear connection between phase- and frequency-dependence with acquisition geometry can be established. The examples here also demonstrate a significant geometry dependence in the cross-talk between v_P and Q_P, and likewise v_S and Q_S, with transmission data being important for reducing these cross-talk modes. This behavior is difficult to motivate from radiation pattern analysis alone, as the corresponding v and Q terms have the same angle-dependence in their radiation patterns (as illustrated in Chapter 3).

While tools for identifying likely modes of cross-talk can be important for preventing misinterpretation of inversion results, it is usually desirable to use these tools to identify ways to avoid cross-talk. Radiation pattern analysis, for instance, is often used in elastic FWI to guide reparameterization strategies for cross-talk reduction based on scattering angles. Often in these approaches, data from angle ranges within which only one parameter has significant radiation energy are used to update just these parameters. For instance, if only v_P substantially alters the data at transmission angles, the data from these angles can be safely used to update v_P. If such ranges do not exist, alternate parameterizations may be sought. While the approach I present here is useful for identifying acquisition geometries which help to prevent cross-talk, a reparameterization strategy is difficult to apply when considering attenuation, because velocity and Q radiation patterns are

not made distinct by scattering angle information (as shown in Chapter 3). This makes it difficult to identify which reparameterizations may help with the problem short of trial and error.

If cross-talk cannot be limited simply through reparameterization strategies, the second derivative information contained in the Hessian must play a major role in a successful cross-talk reduction scheme. The technique I have described here is expected to be a useful means of comparing the efficacy of cross-talk reduction between optimization strategies using different approaches to approximating the effect of the Hessian matrix. While this approach was used to characterize viscoelastic cross-talk here, it could also be used to effectively characterize cross-talk in other FWI formulations, taking into account cross-talk between spatially separated variables and the effects of optimization strategy which are often neglected in other approaches.

4.6 Conclusions

Conventional approaches to characterizing inter-parameter cross-talk in full waveform inversion are poorly suited to the viscoelastic problem, generally lacking simple approaches for characterizing the effects of frequency dependence, phase changes and contributions from spatially distant variables. The cross-talk characterization approach investigated here differenced numerical simulations to directly investigate cross-talk, allowing for simple treatment of these effects. This approach also allowed for the effects of iteration, optimization strategy, acquisition geometry and source type to be observed. With this approach, I verified well-known cross-talk behaviors between elastic parameters and established important cross-talk behaviors for the Q variables. Cross-talk between a v_P anomaly and an obscuring Q_P region was observed to be significantly reduced when transmission-type data were available and computationally intensive optimization strategies were used. The cross-talk mode from Q_P into v_P is dominated by contributions from spatially distant variables in several cases. Similar behaviors were observed in the cross-talk between v_S and Q_S. This approach can be used for other FWI parameterizations, and may be especially useful where the cross-talk contribution of spatially separated parameters is expected to be large. It could also be

used as a tool to assess which optimization strategies are appropriate, given the acquisition geometry and target location for field data. The approach I discuss in this Chapter is an effective tool for assessing the cross-talk expected for a given acquisition geometry, but is not a tool for determining which features of an inversion result are likely attributable to cross-talk. Such a posterior investigation of the effects of cross-talk on an inversion result is more appropriately treated through an uncertainty quantification procedure. I propose one such procedure in the next Chapter.

Chapter 5

Null-space shuttles for targeted uncertainty analysis in full waveform inversion

5.1 Summary

Full-waveform inversion is an effective tool for recovering subsurface information, but many factors make this recovery subject to uncertainty. In particular, unwanted noise in measurements can bias results toward models that are not representative of the true subsurface, and numerical optimization techniques used in the inversion allow for only approximate minimization of the objective function. Both factors contribute to the non-uniqueness of full waveform inversion solutions. Assessing the uncertainty this non-uniqueness introduces can be difficult, due to the large dimensionality of the inversion problem. Fortunately, complete characterization of inversion uncertainty is seldom necessary for applications employing an inversion result, meaning the entire dimensionality of the problem may not be relevant for practical uncertainty quantification. Typically, it is only the uncertainty in a few specific aspects of the inversion that is important (for instance, confidence in a recovered anomaly). A targeted uncertainty quantification, characterizing only the confidence in a specific feature of the subsurface model, can greatly reduce the dimensionality of the uncertainty characterization problem, potentially making it tractable. I propose an approach for quantifying the confidence of inversions in chosen hypotheses about the recovered subsurface model. I test each hypothesis through numerical optimization on the set of equal-objective model-space steps, called null-space shuttles. By approximating the null-space shuttle which maximally violates a given hypothesis about the inversion, this method establishes an effective approximation of the uncertainty in that hypothesis. I demonstrate the use of this technique on several numerical examples for the case of viscoelastic inversion. These examples demonstrate that, at reasonable

computational cost, this method can generate estimates of the lower bound on the maximal uncertainty associated with incomplete numerical optimization. In the viscoelastic examples considered, the velocity variables are much better constrained than the Q and density ones according to this metric.

5.2 Introduction

While FWI has developed to be a powerful inversion approach, the development of tools for uncertainty assessment in FWI has lagged considerably behind. This makes it difficult to know how much an inversion result can be trusted, presenting an obstacle to the effective use of FWI. This problem is further compounded by the presence of known sources of significant uncertainty, notably including cross-talk for QFWI and other multiparameter FWI formulations. While the previous chapter of this thesis provides tools to help characterize when cross-talk might be expected, there are few existing tools in FWI for assessing how much uncertainty these cross-talk modes introduce into an estimate of the subsurface.

In principle, when quantifying uncertainty a user seeks to determine the confidence with which the properties defined in the inversion output have been determined. In a given inversion problem, this implies quantifying the range of possible solutions which could have satisfied the data- and prior-fitting requirements of the inversion. Ideally, such an uncertainty estimate would also identify the probabilities of these solutions. In reality, we are limited by the computational expense of this problem and, as a result, most approaches for assessing confidence in FWI results have been quite simple. Comparison of inversion results to well-log data, for instance (e.g., Bleibinhaus et al., 2007; Owusu et al., 2015; Pan et al., 2018), can provide a measure of the confidence in the inversion result, but this measure only assesses accuracy in one model-space direction for only a small region of model space, and so provides a very rough metric for confidence in the inversion. This strategy is widely used to gauge inversion accuracy, but is lacking as tool for uncertainty quantification.

For a large-scale, fully resolved FWI problem, global optimization techniques are often com-

putationally infeasible, so most proposed uncertainty estimation approaches focus on assessing the local uncertainty, usually through investigation of the Hessian of the inversion objective function. Even in this case, the large memory size and computational cost associated with evaluating the Hessian mean that approximate or matrix-free approaches are needed. Fichtner and Trampert (2011) and Fichtner and Leeuwen (2015), for instance, use Hessian probing to generate point-spread functions and estimate spatial resolution and parameter trade-off from them. Other authors have focused on creating uncertainty estimates based on evaluation of the posterior covariance matrix, which is approximately the inverse of the Hessian near the minimum of the FWI objective function (Tarantola, 2005). Because exact evaluation of the posterior covariance requires the storage and inversion of Hessian-sized matrices, some form of rank-reduction is necessary to generate tractable posterior covariance estimates. Bui-Thanh et al. (2013), Zhu et al. (2016), Eliasson and Romdhane (2017), and Liu and Peter (2019) consider a truncated singular-value decomposition (SVD) of the Hessian to achieve this rank reduction, limiting the uncertainty characterization problem to the best-constrained directions of model space. Thurin et al. (2019) instead uses an ensemble Kalman filter method to obtain a low-rank estimate of the posterior covariance, avoiding the calculation of Hessian-vector products. Fang et al. (2018) does not directly estimate a covariance matrix, but relaxes the PDE constraint of the inversion and works with an approximation of the Hessian to sample elements of a posterior probability density function. In this approach an implicit rank-reduction is present when sampling elements due to the use of an iterative approximate solver instead of the inverse Hessian.

Uncertainty quantification strategies in FWI based on estimation of the posterior covariance matrix have allowed for useful estimates of uncertainty to be generated, but invoke several fundamental assumptions, some of which can be problematic in general. First, these approaches assume the subset of model-space directions implicit in the low-rank estimates of the covariance are sufficient to accurately represent the uncertainty. This assumption is difficult to justify without knowing which uncertainties are important in a given problem. Second, these approaches generate posterior

probability distributions under the assumption that the the inversion output is the model of maximum likelihood. This assumption is valid only if the inversion output minimizes the FWI objective function. However, in FWI we only ever approach the minimum, and never reach it. This means we are unable to precisely interpret the covariance estimates as representative of uncertainty of inversion results. The likely violation of these two assumptions in most FWI applications suggests that other kinds of uncertainty quantification be considered. In fact, a further type of uncertainty within FWI model construction, not addressed by posterior covariance estimates, is introduced by the non-uniqueness of the FWI solution (or the non-uniqueness of the model that covariance estimates center on. This uncertainty reflects the fact that an inversion output which nearly minimizes the objective function is in some sense not to be preferred over any other model which does an equal or better job of this minimization. In this chapter I consider an approach for targeted quantification of this aspect of FWI uncertainty.

One approach to characterizing model uncertainty in specific directions was introduced by Deal and Nolet (1996) in the context of seismic tomography. In this approach, the authors sought to determine how an inverted subsurface model could be modified without changing the data misfit, exploring the uncertainty between models equally acceptable by the metric of the objective function. By projecting a specially designed model-space change onto the null-space of the forward modeling operator, the model update closest to this design which causes no variation in the objective function could be determined. This approach was referred to as null-space shuttling, and was applied as an alternative to regularization. More recently, Fichtner and Zunino (2019) and Liu and Peter (2020) use null-space shuttling approaches to explore the uncertainties associated with the non-uniqueness of solutions to nonlinear inverse problems. Here, I examine the potential of null-space shuttling approaches for explicitly quantifying inversion uncertainty associated with the non-uniqueness of the FWI solution.

Accurate estimation of general uncertainty in FWI is difficult, but in reality the uncertainties in most directions of model space are not particularly informative (in the sense that these uncertainties

relate to confusion between subsurface models we will interpret in the same way). The features of an inversion result which substantially impact our understanding of the subsurface are generally small in number and highly specific. By focusing the quantification of uncertainty on the subset of model-space directions which have an important impact on decision-making, we can reduce the computational demands of uncertainty characterization. This allows for more accurate methods, which would be too costly for the full uncertainty characterization problem to be employed. In this way, the concept of null-space shuttling could be used in FWI to achieve a targeted uncertainty characterization.

Here, I consider a null-space shuttling method for uncertainty quantification with respect to a user-defined hypothesis about the true subsurface. The hypothesis makes a statement about a key feature of the inversion result and asserts that it is true of the subsurface. I characterize the maximal uncertainty in this hypothesis by determining the subsurface model which maximally violates the hypothesis while remaining in the region of model space the inversion considers to be acceptable solutions. This region should include, at minimum, all points which achieve an equivalent or lower objective function value than the inversion output. Quantifying the uncertainty arising from the non-uniqueness of the FWI problem then requires that we explore a subset of the points of model space with the same objective function value as the inversion result. By the definition of Deal and Nolet (1996), the model-space steps relating these points are null-space shuttles, and the problem of finding the maximal uncertainty becomes an optimization problem with respect to the possible shuttles.

The remainder of this Chapter is organized as follows. First, I introduce null-space shuttling in the context of FWI and present my strategy for targeted uncertainty quantification. Next, I provide a brief description of the viscoelastic inversion approach I use for the numerical examples considered here. Finally, I provide numerical examples of the proposed approach.

5.3 Theory

The FWI procedure is generally treated as an optimization problem (e.g., Tarantola, 1984; Virieux and Operto, 2009). In this problem, the subsurface model that produces data in closest agreement with the observed seismic data is sought, as this model should best approximate the true subsurface. The general formulation I introduced in Chapter 2 for a general FWI objective function, ϕ, was

$$\mathbf{m}_{min} = \underset{\mathbf{m}}{\operatorname{argmin}}\ \phi(\mathbf{u},\mathbf{m},\mathbf{d}) \text{ subject to } C(\mathbf{u},\mathbf{m}) = 0, \tag{5.1}$$

where $\mathbf{m}$ is a subsurface model, $\mathbf{m}_{min}$ is the inversion result, $\mathbf{u}$ is a simulated wavefield, $\mathbf{d}$ represents the measured data, and $C = 0$ holds only when the wave equation assumed in the inversion is satisfied. While the specific choices of ϕ, C, and the optimization strategy can greatly affect the behavior of the inversion, a general challenge is the large dimensionality of $\mathbf{m}$ typically required to achieve an acceptably resolved inversion result. This large dimensionality means that the range of possible solutions is large, while the substantial computational cost of seismic wave modeling means that only a few points in model-space can typically be assessed during the inversion. Because of these features, FWI usually makes use of local optimization techniques, where model-space steps are designed based on the derivatives of the objective function at the current model-space location and only steps decreasing the objective function are allowed. This type of optimization is generally not capable of guaranteeing a global minimum, and, importantly for our current purposes, is always terminated before obtaining an exact local minimum, so the inversion result $\mathbf{m}_{FWI}$ is an approximation of the ideal output $\mathbf{m}_{min}$.

When attempting to assess our confidence in $\mathbf{m}_{FWI}$ as an accurate representation of the true subsurface, there are several types of uncertainty to consider. First, there is the consideration that $\mathbf{m}_{min}$ may not be an accurate representation of the subsurface. This could occur either because (1) C does not accurately represent important features of true subsurface wave propagation, (2) the measured data contain noise which would not be reproduced by an accurate subsurface model, or (3) the measured data are insufficient to constrain $\mathbf{m}_{min}$ to a single point in model space. Next, there is consideration that the model recovered by the inversion, $\mathbf{m}_{FWI}$, may not be a sufficiently accurate

estimate of $\mathbf{m}_{min}$. This inaccuracy could manifest in a local sense ($\mathbf{m}_{FWI}$ is not at the minimum of the objective function basin it lies in) and in a global sense (the objective function basin $\mathbf{m}_{FWI}$ lies in is not the one $\mathbf{m}_{min}$ lies in). Some of these modes of uncertainty are more difficult to quantify than others, and treating all simultaneously would be extremely demanding. Here, I focus on developing a technique for quantifying uncertainty due to noise and local optimization inaccuracy; I assume that C is an accurate representation of wave propagation and that $\mathbf{m}_{FWI}$ and $\mathbf{m}_{min}$ lie in the same basin of the objective function. Both of the uncertainty features I consider relate to the non-uniqueness of FWI solutions, and are highly dependent on the local behavior of the objective function near a minimum.

In general, an FWI result can be acceptable without exactly minimizing the objective function; minima are too expensive to exactly compute, and, due to the presence of noise, achieving the best fit to the data is not necessarily desirable. Instead, there is a set of acceptable models, $\mathbf{M}$, satisfying

$$\phi(\mathbf{m}^*) \leq \phi_{max}, \quad \mathbf{m}^* \in \mathbf{M}, \tag{5.2}$$

where ϕ_{max} is the largest acceptable value for the objective function. The set $\mathbf{M}$ contains the models which could be acceptable outputs from the inversion, given the level of noise in the data and the extent of available computation. The existence of multiple acceptable models gives rise to an uncertainty in the inversion between $\mathbf{m}_{FWI}$ and $\mathbf{M}$. In order to quantify this uncertainty, it is necessary to characterize $\mathbf{M}$.

Complete characterization of $\mathbf{M}$ is computationally infeasible for much the same reason that an exact minimum cannot be obtained in the inversion: the dimensionality of the problem is large and evaluating the objective function is expensive. Calculation of specific elements of $\mathbf{M}$, however, is feasible. Any sufficiently short step in a model-space direction with a negative projection onto the gradient from $\mathbf{m}_{FWI}$ should result in a model within $\mathbf{M}$; the definition of the gradient requires that the objective function must go down at least slightly for some step-length. Very short steps are uninteresting from the standpoint of uncertainty characterization, however, as these represent

very small changes in the model. What we are most interested in is the maximum uncertainty in a given model-space direction, that is, how large a model change in a given direction can be without producing an unacceptable value of the objective function.

While the maximum acceptable objective function value, ϕ_{max}, may not be explicitly defined before the inversion, all FWI procedures implicitly involve one, because all FWI procedures end once an acceptable model is produced by the inversion; this is the first model for which $\phi(\mathbf{m})$ is smaller than or equal to the implicit ϕ_{max}. Suppose, then, that we define $\phi_{max} = \phi_{FWI}$. In this case, the maximum model-space step $\Delta\mathbf{m}$ that can be taken in a given direction while remaining in $\mathbf{M}$ is the step which produces the same objective function:

$$\phi(\mathbf{m}_{FWI} + \Delta\mathbf{m}) = \phi(\mathbf{m}_{FWI}). \tag{5.3}$$

The step $\Delta\mathbf{m}$ navigates model-space without changing the objective function. Using the terminology of Deal and Nolet (1996), this type of step is called a null-space shuttle. Provided that $\mathbf{m}_{FWI}$ lies near a minimum, the objective function near $\mathbf{m}_{FWI}$ can be characterized as a function of model change $\Delta\mathbf{m}$ as

$$\phi(\mathbf{m}_{FWI} + \Delta\mathbf{m}) \approx \phi_{FWI} + \mathbf{g}_0\Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^T H_{GN}\Delta\mathbf{m}, \tag{5.4}$$

where $\mathbf{g}_0$ is the gradient at $\mathbf{m}_{FWI}$, ϕ_{FWI} is the objective function at $\mathbf{m}_{FWI}$, and H_{GN} is the Gauss-Newton approximation of the Hessian at $\mathbf{m}_{FWI}$.

I refer to the set of models with an objective function ϕ_{max} as $\mathbf{M}'$. Insofar as equation 5.4 is an accurate representation of the objective function, the model updates $\Delta\mathbf{m}'$ corresponding to element models of $\mathbf{M}'$ are those for which $\phi(\mathbf{m}_{FWI} + \Delta\mathbf{m}) = \phi_{max}$, satisfying

$$\mathbf{g}_0\Delta\mathbf{m}' = -\frac{1}{2}\Delta\mathbf{m}'^T H_{GN}\Delta\mathbf{m}'. \tag{5.5}$$

By enforcing this constraint, we can calculate the maximum step in a given model-space direction $\Delta\mathbf{m}$ expected to lie in $\mathbf{M}$:

$$\Delta\mathbf{m}' = \alpha\delta\hat{\mathbf{m}}, \tag{5.6}$$

where

$$\delta\hat{\mathbf{m}} = \frac{\Delta\mathbf{m}}{||\Delta\mathbf{m}||_2}, \tag{5.7}$$

and

$$\alpha = \frac{-2\mathbf{g}_0\delta\hat{\mathbf{m}}}{\delta\hat{\mathbf{m}}^T H_{GN}\delta\hat{\mathbf{m}}}, \tag{5.8}$$

where $||\ ||_2$ denotes the L_2 norm. In this way, given the vector $\Delta\mathbf{m}$, we can approximate the inversion uncertainty in that direction at the computational cost of one Hessian-vector product evaluation: $H_{GN}\delta\hat{\mathbf{m}}$. A Gauss-Newton Hessian-vector product can be evaluated at the cost of two wavefield propagations, as described in Chapter 2. This computation requirement is small relative to the inversion procedure, but if many directions are considered, more Hessian-vector products are required, and the cost can become substantial. Fortunately, the uncertainty of the model in a random model-space direction is not typically of great importance for applications using an inversion result. In a practical instance of FWI, the key uncertainties in an inversion result (those which affect the understanding of a geophysicist or other interpreter) will typically relate to important, specific features of the model. These specific features will be described by relatively few model-space directions, and so only a small number of null-space shuttles.

The targeted uncertainty quantification strategy I investigate here involves using null-space shuttles to test specific hypotheses about features of the inversion result. Suppose we frame a hypothesis about the subsurface which asserts that a feature of the inversion output is necessary, in the sense that no models exist which both achieve an acceptable level of prior- and data-fitting (as measured by the objective function) and lack this feature. Based on the hypothesis we wish to investigate, we define a scalar function ψ. This hypothesis function is designed such that models which significantly violate the hypothesis have small ψ values, while those which are consistent with the hypothesis have large ψ values. We can quantify the confidence of our inversion in the hypothesis described by ψ by finding the model acceptable to the inversion which maximally violates the hypothesis. If no acceptable model substantially violates the hypothesis, the inversion has high confidence in the claim of the hypothesis. If, on the other hand, there are acceptable

models which do violate the hypothesis, the inversion does not have confidence in the hypothesis. Mathematically, I treat the problem of finding the viable model in $\mathbf{M}'$ which maximally violates the stated hypothesis by solving the problem

$$\mathbf{m}_\psi = \underset{\mathbf{m}'}{\operatorname{argmin}} \ \psi(\mathbf{m}'), \text{ subject to } \mathbf{m}' \in \mathbf{M}'. \tag{5.9}$$

Using the relation defined in equation 5.6 to connect a given direction $\Delta\mathbf{m}$ with the corresponding model in $\mathbf{M}'$, we can re-state the optimization problem as

$$\mathbf{m}_\psi = \mathbf{m} + \underset{\Delta\mathbf{m}}{\operatorname{argmin}} \ \psi(\mathbf{m} + \alpha\delta\hat{\mathbf{m}}). \tag{5.10}$$

An exact solution of equation 5.9 would represent a maximum of the type of uncertainty we consider in the hypothesis. Because the optimization methods we consider will only approximate a solution to equation 5.9, we will be considering a lower bound on this maximum uncertainty (equivalently, an upper bound on the value of ψ achieved by m_ψ). Solving equation5.9 through gradient-based optimization techniques requires the calculation of

$$\frac{\partial\psi}{\partial\Delta\mathbf{m}} = \frac{\partial\psi}{\partial\Delta\mathbf{m}'}\frac{\partial\Delta\mathbf{m}'}{\partial\Delta\mathbf{m}} = \frac{\partial\psi}{\partial\Delta\mathbf{m}'}\left(\frac{\partial\Delta\mathbf{m}'}{\partial\alpha}\frac{\partial\alpha}{\partial\Delta\mathbf{m}} + \frac{\partial\Delta\mathbf{m}'}{\partial\delta\hat{\mathbf{m}}}\frac{\partial\delta\hat{\mathbf{m}}}{\partial\Delta\mathbf{m}}\right). \tag{5.11}$$

Using equations 5.6-5.8, this expression can be expanded to

$$\frac{\partial\psi}{\partial\Delta\mathbf{m}} = \frac{\partial\alpha}{\partial\Delta\mathbf{m}}\left(\frac{\partial\psi}{\partial\Delta\mathbf{m}'}^T\delta\hat{\mathbf{m}}\right) + \alpha\left(\frac{1}{||\Delta\mathbf{m}||_2}\frac{\partial\psi}{\partial\Delta\mathbf{m}'} - \frac{\Delta\mathbf{m}}{||\Delta\mathbf{m}||_2^3}\frac{\partial\psi}{\partial\Delta\mathbf{m}'}^T\Delta\mathbf{m}\right), \tag{5.12}$$

where

$$\frac{\partial\alpha}{\partial\Delta\mathbf{m}} = \frac{1}{||\Delta\mathbf{m}||_2}\frac{\partial\alpha}{\partial\delta\hat{\mathbf{m}}} - \frac{\Delta\mathbf{m}}{||\Delta\mathbf{m}||_2^3}\frac{\partial\alpha}{\partial\delta\hat{\mathbf{m}}}^T\Delta\mathbf{m} \tag{5.13}$$

and

$$\frac{\partial\alpha}{\partial\delta\hat{\mathbf{m}}} = -\frac{2\mathbf{g}_0}{\delta\hat{\mathbf{m}}^T H_{GN}\delta\hat{\mathbf{m}}} + 4H_{GN}\delta\hat{\mathbf{m}}\frac{\mathbf{g}_0\delta\hat{\mathbf{m}}}{\delta\hat{\mathbf{m}}^T H_{GN}\delta\hat{\mathbf{m}}}. \tag{5.14}$$

The somewhat cumbersome choice of $\Delta\mathbf{m}$ as the optimization variable rather than the unit vector $\delta\hat{\mathbf{m}}$ is made here to enable the use of unconstrained optimization techniques in the solution of equation 5.10.

An approximate solution to equation 5.10 can be found through nonlinear optimization. Here, I use the L-BFGS approach (Liu and Nocedal, 1989) for the minimization, though other optimization strategies could be employed instead. The main computational costs of this approach are evaluating the product $H_{GN}\delta\hat{\mathbf{m}}$, which must be done once per iteration to evaluate ψ and $\frac{\partial\psi}{\partial\Delta\mathbf{m}}$, and of the product $H_{GN}\mathbf{p}_{BFGS}$ (where $\mathbf{p}_{BFGS}$ is the calculated L-BFGS descent direction) once per iteration for the line search. While the line-search step of the L-BFGS procedure may require many additional evaluations of ψ and $\frac{\partial\psi}{\partial\Delta\mathbf{m}}$, these will all be evaluated at locations in model space that are linear combinations of $\delta\hat{\mathbf{m}}$ and $\mathbf{p}_{BFGS}$, so the required Hessian-vector products can be determined from those already calculated. As such, the main computational cost of each L-BFGS iteration is the evaluation of two Hessian-vector products.

The initial estimate of $\Delta\mathbf{m}$ can have a significant impact on how quickly a solution to equation 5.10 can be approximated. Better initial estimates should represent directions in which the hypothesis function ψ decreases quickly and the FWI objective ϕ changes slowly, as these are desired features of the optimization output. Accordingly, for the initial estimate, I wish to consider the projection of $-\frac{d\psi}{d\mathbf{m}}$ onto a vector orthogonal to $\frac{d\phi}{d\mathbf{m}}$. As many such vectors exist, I consider the orthogonal vector which maximizes the projection. To approximate this vector, I consider the optimization problem

$$\mathbf{x}^* = \min_{\mathbf{x}} \left(\frac{A\mathbf{x}}{||A\mathbf{x}||_2} \cdot \frac{d\psi}{d\mathbf{m}} \right), \tag{5.15}$$

where the columns of A form an orthogonal basis spanning the space orthogonal to $\frac{d\phi}{d\mathbf{m}}$, and $\cdot$ denotes a dot product. When $\mathbf{x}^*$ has been approximated, the initial estimate of $\Delta\mathbf{m}$ can be defined as

$$\Delta\mathbf{m}_0 = A\mathbf{x}^*. \tag{5.16}$$

Because the evaluation of $\mathbf{x}^*$ in equation 5.15 does not require any wave-propagation problems to be solved, it can be assessed at a low cost relative to the rest of the procedure. Any orthogonal basis spanning the space orthogonal to $\frac{d\phi}{d\mathbf{m}}$ can be used for A, but many of these matrices are memory

intensive to use. I create a sparse A by defining

$$A = \begin{bmatrix} g_0(2) & g_0(3) & g_0(4) & \dots & g_0(n) \\ -g_0(1) & 0 & 0 & \dots & 0 \\ 0 & -g_0(1) & 0 & \dots & 0 \\ 0 & 0 & -g_0(1) & \dots & \vdots \\ \vdots & \vdots & 0 & \ddots & \\ 0 & 0 & 0 & \dots & -g_0(1) \end{bmatrix} \tag{5.17}$$

This matrix requires storage of only $2(n-1)$ elements, and its vectors span the space orthogonal to $\frac{d\phi}{d\mathbf{m}}$ trivially. It is very likely, however, that other formulations of A may be more efficient, the formulation I use here is just one possibility.

Suppose an approximate minimizer of ψ, $\mathbf{m}'^*$, and the corresponding step, $\Delta\mathbf{m}'^*$, have been determined. The step $\Delta\mathbf{m}'^*$ represents a best estimate of the model update which maximally violates our hypothesis without changing the FWI objective function, provided that equation 5.4 holds exactly (i.e. the objective is exactly linear). In reality, the objective function will not be exactly linear, even close to the minimum. This has two major implications for the uncertainty estimate. Firstly, it means changing the model by $\Delta\mathbf{m}'$ will not leave the objective function exactly unchanged. To fix this problem, I suggest using a line-search approach to determine the step length in the direction of $\Delta\mathbf{m}'^*$ corresponding to a negligible change in the objective. The second implication of non-linearity is that H_{GN} is a function of $\mathbf{m}$. This means that even an exact solution of the optimization problem in equation 5.10 may not be the solution to the true problem of interest, equation 5.9.

To obtain a solution of equation 5.9 we can treat the step calculated in equation 5.10 as the first step in an iterative optimization procedure. The procedure I consider is set out in Algorithm 1. In this algorithm, the inner loop corresponds to the iterative approximation of the step, $\Delta\mathbf{m}$, which solves equation 5.10, while the outer loop attempts to solve equation 5.9 by updating the model in the direction of $\Delta\mathbf{m}$, while keeping the FWI objective function constant. At each outer loop iteration, a line-search is used to find the step length corresponding to a true null-space shuttle, which allows for differences between the linear approximation of the objective used in the inner

loop and the true objective function. This line search finds the model in the direction of $\Delta\mathbf{m}$ which achieves a value of ϕ within a chosen tolerance of ϕ_{max}. If the problem is close to linear, we can limit the procedure to only one outer loop iteration. In the examples considered here, I use an iteration maximum in both the inner and outer loops to define a stopping condition for the procedure. Convergence conditions based on the observed change in ψ could be used as well if they are found to be useful.

Input : FWI output: $\mathbf{m}$
Output: Approximate minimizer of ψ: $\mathbf{m}'_{i_{max}+1}$
Initialize model estimate
$\mathbf{m}'_1 = \mathbf{m}$
for $i = 1, \ldots, i_{max}$ **do** ; *// Loop over outer iterations*

  Initialize the descent direction $\Delta\mathbf{m}_1$
  Initialize the L-BFGS inverse Hessian approximation Q
  for $j = 1, \ldots, j_{max}$ **do** ; *// Loop over inner iterations*

    Set $H_{v_1} = H_{GN}(\mathbf{m}'_i)\Delta\mathbf{m}_j$; *// Calculate Hessian - update product*
    Set $\mathbf{g}_\psi = \frac{\partial\psi}{\partial\Delta\mathbf{m}_1}$; *// Calculate gradient using* H_{v_1}
    Set $\mathbf{d} = Q\mathbf{g}_\psi$; *// Calculate L-BFGS descent direction*
    Set $H_{v_2} = H_{GN}(\mathbf{m}'_i)\mathbf{d}$; *// Calculate Hessian - d product*
    Calculate the step length μ to minimize ψ using a line-search ; *// Use* H_{v_1} *and* H_{v_2} *for* ψ *evaluations in line-search*
    Set $\Delta\mathbf{m}_{j+1} = \Delta\mathbf{m}_j + \mu\mathbf{d}$; *// Update the step direction*
    Update Q
  end
  Calculate $\alpha, \delta\hat{\mathbf{m}}$ from $\Delta\mathbf{m}_{j_{max}+1}$
  while $\phi(\alpha\delta\hat{\boldsymbol{m}}) - \phi_{max} \geq \varepsilon * \phi_{max}$ **do** ; *// Line search with tolerance* ε

    Set $\alpha = \alpha + (\frac{d\phi}{d\alpha})^{-1}(\phi_{max} - \phi(\alpha\delta\hat{\mathbf{m}}))$
  end
  Set $\mathbf{m}'_{i+1} = \mathbf{m}'_i + \alpha\delta\hat{\mathbf{m}}$
end

Algorithm 2: Algorithm for ψ minimization through null-space shuttling.

5.4 Numerical Examples

In this section I develop the null-space shuttle procedure with some simple numerical examples. The purpose here is, in part, to demonstrate that the procedure works, but more importantly, to show in greater detail what kind of hypothesis-testing may lead to useful appraisals. My goal is to provide a tool for interpretation, allowing a geoscientist to delineate a structure or region of interest, within a model inferred from FWI, and quantify the confidence with which it was placed there.

The forward modeling approach I consider in this chapter, is the Kolsky-Futterman model of viscoelastic wave propagation I discuss in Chapter 2. Because my goal in this chapter is to investigate uncertainty associated with optimization specifically, I will assume that there is high confidence that this model of attenuation physics hold. In relation to the discussion in Chapter 3, this means that I will adopt inversion strategies which recover a single model, consistent across all frequencies, rather than the set of models which is shown in that Chapter to be more effective in mitigating the effects of incorrect attenuation physics.

5.4.1 Optimization and parameterization

The objective function for the FWI problem I consider here is the one introduced in Chapter 2:

$$\Phi = \phi_D + \phi_P, \tag{5.18}$$

where

$$\phi_D = \sum_{r_s,\ \omega} \frac{1}{2} ||\mathbf{d} - \mathbf{Ru}(\mathbf{m})||_2^2, \tag{5.19}$$

$\mathbf{d}$ is a vector containing the measured data, $\mathbf{R}$ is a matrix representing the receiver sampling of the wave-field, r_s represents the surface location of the source considered, ω is an angular frequency, $\mathbf{m}$ is a vector containing subsurface model parameters, and ϕ_P is a regularization term. As described in Chapter 2, I consider inversion in five parameters: $\alpha_1\rho$, $\alpha_2 v_P^{-2}$, $\alpha_3 Q_P^{-1}$, $\alpha_4 v_S^{-2}$, and $\alpha_5 Q_S^{-1}$, where the α_n are unitless scaling terms introduced to improve conditioning.

In these examples, I use the regularization term to ensure that recovered models have positive Q values. To this end, I define a regularization term which applies a heavy, quadratic penalty as Q^{-1} values approach zero, and a milder penalty away from this region:

$$\begin{aligned}\phi_P = \varepsilon_1 \sum_{j=1}^{N} &\begin{cases} \ln\left(\frac{Q_{P_0}}{Q_{P_j}}\right)^2 : & Q_{P_j}^{-1} \geq Q_{P_{crit}}^{-1} \\ \ln\left(\frac{Q_{P_0}}{Q_{P_{crit}}}\right)^2 + a_{1_j}(Q_{P_{crit}}^{-1} - Q_{P_j}^{-1}) + a_{2_j}(Q_{P_{crit}}^{-1} - Q_{P_j}^{-1})^2 : & Q_{P_j}^{-1} < Q_{P_{crit}}^{-1} \end{cases} \\ + \varepsilon_2 \sum_{j=1}^{N} &\begin{cases} \ln\left(\frac{Q_{S_0}}{Q_{S_j}}\right)^2 : & Q_{S_j}^{-1} \geq Q_{S_{crit}}^{-1} \\ \ln\left(\frac{Q_{S_0}}{Q_{S_{crit}}}\right)^2 + a_{3_j}(Q_{S_{crit}}^{-1} - Q_{S_j}^{-1}) + a_{4_j}(Q_{S_{crit}}^{-1} - Q_{S_j}^{-1})^2 : & Q_{S_j}^{-1} < Q_{S_{crit}}^{-1}, \end{cases}\end{aligned} \tag{5.20}$$

where j represents the model cell index, Q_{P_0} and Q_{S_0} are the average values of Q_P and Q_S in the starting model, ε_1 and ε_2 are weighting terms, and the letter coefficients represent the derivatives of ϕ_P at the critical values of $Q_{P_{crit}}$ and $Q_{S_{crit}}$:

$$\begin{aligned} a_1 &= \frac{d\phi_P(Q_{P_{crit}}^{-1})}{dQ_{P_j}^{-1}}, \\ a_2 &= \frac{d^2\phi_P(Q_{P_{crit}}^{-1})}{d(Q_{P_j}^{-1})^2}, \\ a_3 &= \frac{d\phi_P(Q_{S_{crit}}^{-1})}{dQ_{S_j}^{-1}}, \\ a_4 &= \frac{d^2\phi_P(Q_{S_{crit}}^{-1})}{d(Q_{S_j}^{-1})^2}, \end{aligned} \tag{5.21}$$

which are used to ensure continuity in the transition from the heavy, linear penalty near the constraint to the milder penalty near the starting model.

In the inversions I consider here, ten frequency bands of five evenly spaced frequencies are used. Each band has the same lowest frequency, and other frequencies are linearly spaced to a band-dependent maximum frequency, which increases in later bands. At each band, truncated-Gauss-Newton (TGN) optimization is used to reduced the objective function.

During the uncertainty characterization stage, I consider the objective function at the final band only, which I expect to be representative of the entire objective function. This simplification is

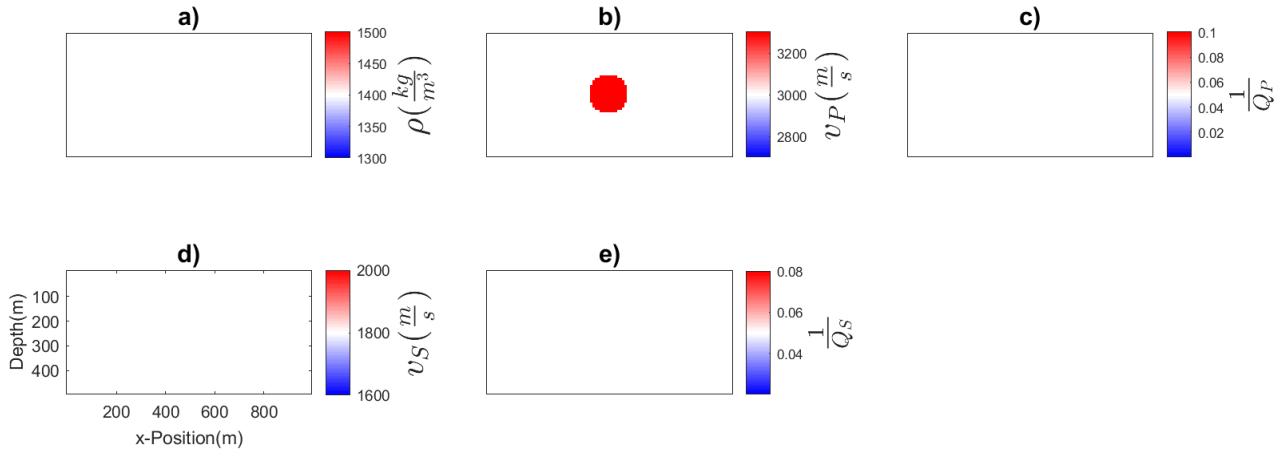

Figure 5.1: Ball model used for synthetic tests, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

used to reduce the computational cost incurred.

5.4.2 Single-target model example

For the first set of synthetic examples I consider, I use the model shown in Figure 5.1. This model is constant in each parameter, with the exception of a circular region at the center of the model, where v_P is higher than its background value. This model is simple, and is characterized by a single target or anomaly, our confidence in which can be quantified as an initial test of the approach. The initial model I use in the inversions is equal to the constant background for each parameter.

For this example, 49 explosive sources and 98 multicomponent receivers are distributed along the top of the model. I consider data from 1 Hz to 20 Hz. In the inversion of the data from this model, five frequency bands are considered, with five frequencies per band. For each band, one iteration of truncated Gauss-Newton optimization is applied, with 20 inner loop iterations per FWI iteration. The inversion result obtained with this approach is shown in Figure 5.2. This result largely captures the geometry of the v_P anomaly, but is noticeably affected by cross-talk, with substantial changes in ρ and Q_P also being recovered.

When investigating an inversion result like the one shown in Figure 5.2, questions about our confidence in the different parameters recovered naturally arise. Three different parameters are

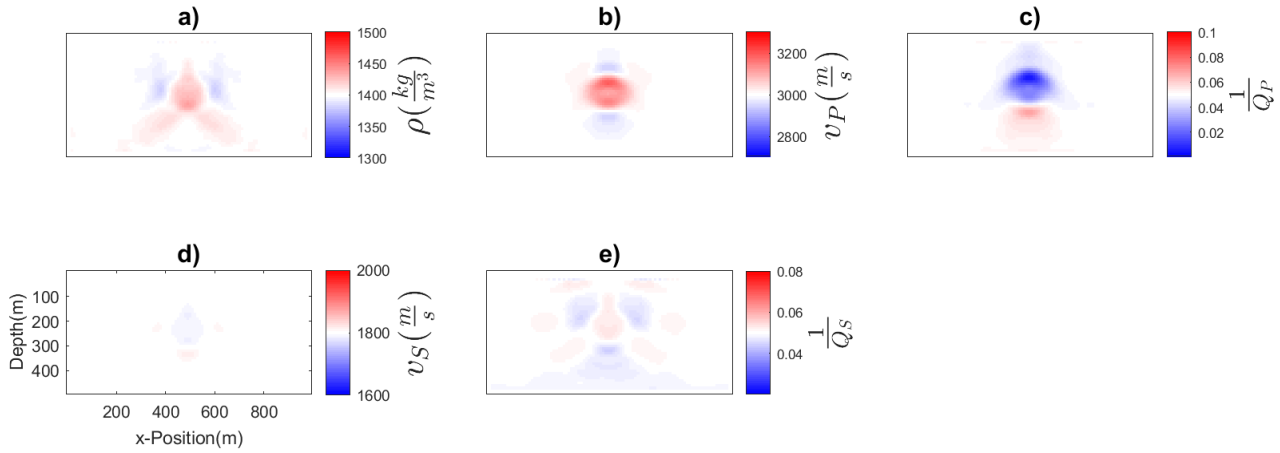

Figure 5.2: FWI output for ball model tests, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

recovered with substantial changes at approximately the same location: ρ, v_P, and Q_P. While changes in rock type will generally alter all the viscoelastic properties, a change in one model parameter will also often introduce cross-talk, causing erroneous model changes in other parameters. This may introduce doubt in one or several of the recovered anomalies. If we suppose that the presence or absence of changes in the v_P model makes a difference to the way this result is used, we may frame the hypothesis here that v_P in the subsurface differs from the initial model used in the inversion. The representative hypothesis function I consider is

$$\psi = \sum_{j=1}^{N} \left(v_{P_j}^{-2} - v_{P0}^{-2} \right)^2, \tag{5.22}$$

where N is the number of cells in the subsurface model considered, v_{P_i} represents the ith element of the v_P model, and v_{P0} represents the background value of the v_P model. By minimizing ψ, I attempt to find the viable model which maximally removes changes in v_P from the background, thus maximally violating my hypothesis. For this minimization, I use three outer-loop iterations of the null-space shuttling approach, each with 20 inner-loop iterations. The calculated shuttle for this choice of penalty term is shown in Figure 5.3. This shuttle mainly consists of a decrease in v_P, and a change in Q_P around the anomaly, suggesting that in this case, the data support some ambiguity between v_P and Q_P variables; Q_P variables are able to compensate for a reduction in

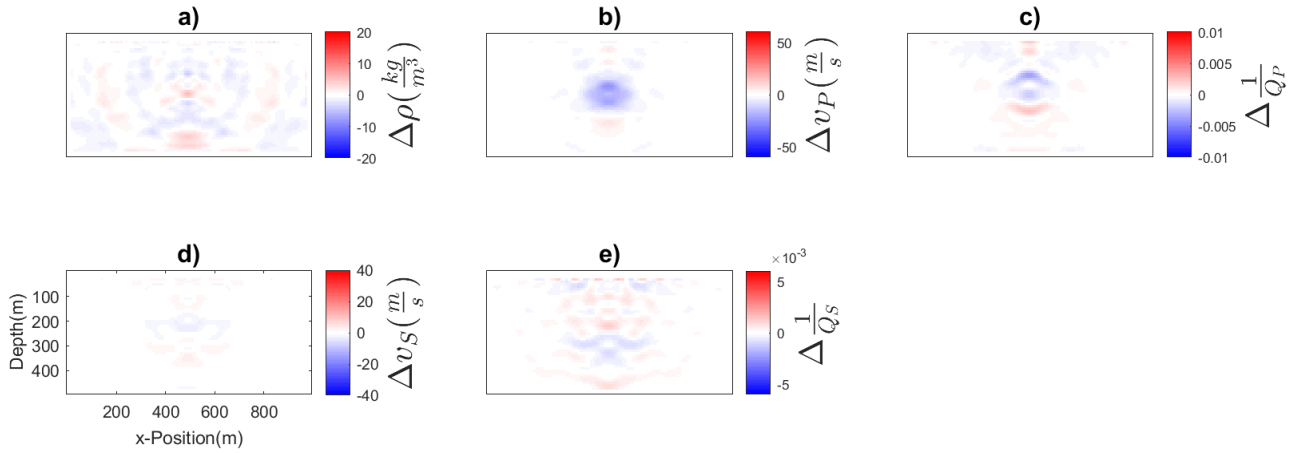

Figure 5.3: Optimal null-space shuttle for eliminating v_P anomaly.

the v_P anomaly. After applying this shuttle to the inversion result, I obtain the model shown in Figure 5.4. The shuttled model slightly reduces the amplitude of the v_P anomaly and slightly increases that of the Q_P anomaly (demonstrated by the small amplitudes in Figure 5.3), while keeping an objective function value within 1% (a chosen line-search tolerance) of the inversion output objective. This model represents a lower bound on the maximum uncertainty in the stated hypothesis; models exist which violate the hypothesis to at least this extent (equivalently, this is a upper bound on the minimum value ψ can take in the acceptable model space). The relatively limited reduction in the v_P anomaly represented by this model suggests that the data insist on the presence of a significant v_P anomaly; no viable models were found without such variations. This conclusion is supported by the hypothesis function, which only drops from 1.00 for the inversion output, to 0.78 for the shuttled model. Our knowledge of the true model in this example (where a v_P anomaly is, in fact, present) is in agreement with the uncertainty analysis in this case.

In a contrasting example, we may obtain the inversion result in Figure 5.2 and hypothesize that it is ρ rather than v_P which must differ from the background to match the measured data. The hypothesis function in this case is that of the previous example with ρ replacing v_P, and is given by:

$$\psi = \sum_{j=1}^{N} \left(\rho_j - \rho_0\right)^2, \tag{5.23}$$

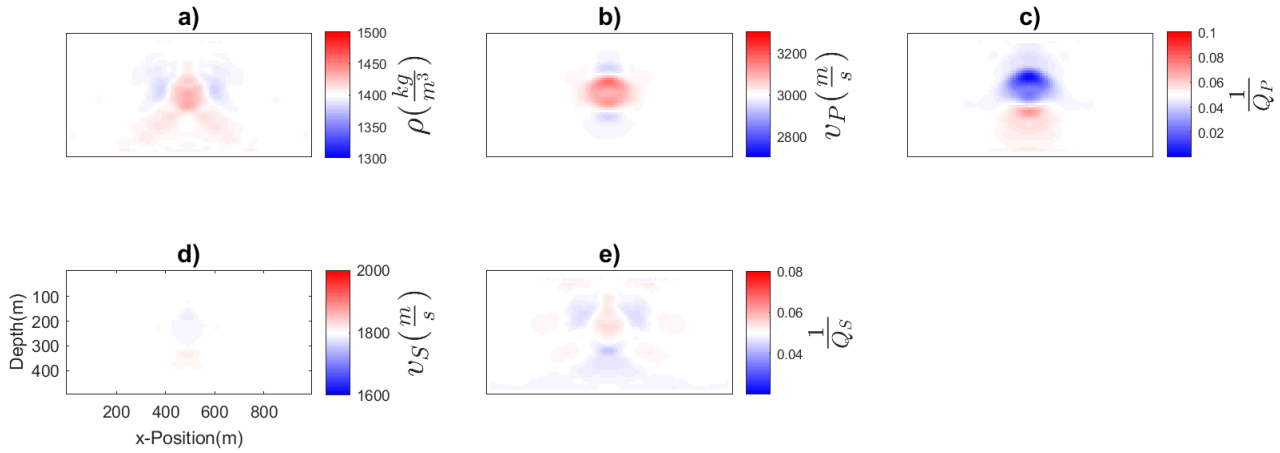

Figure 5.4: Shuttled model for eliminating v_P anomaly.

To calculate the ψ-minimizing shuttle in this case, the same approach was used as in the previous example. The shuttled result using this approach is shown in Figure 5.5. This model differs substantially from the inversion result in its estimate of ρ, recovering little to no anomaly in this parameter. This suggests that reduction or elimination of the variations of ρ from the background can be achieved without changing the objective function in this case. Once again, this is reflected in the values of the hypothesis function ψ: 1.00 for the inversion result, and 0.22 for the shuttled model. This substantial reduction in ψ indicates that viable models were found to violate the stated hypothesis in this case. In turn, this means that there is relatively low confidence in the presence of ρ changes from the background: viable models without this feature exist. The results in this case are again consistent with the true model, which in fact contains no ρ anomaly.

5.4.3 Gas sand example

The model I consider for the second example is shown in Figure 5.6. This model is based on the Marmousi model, and contains a layer structure which defines similar geometry for each physical property considered. The ratios of these properties are variable throughout the model, however, and differ substantially from the background trends in an anomalous region representing a gas-saturated sand at about 200 m depth, and 600 m in x-position. In this example, I focus on characterizing the

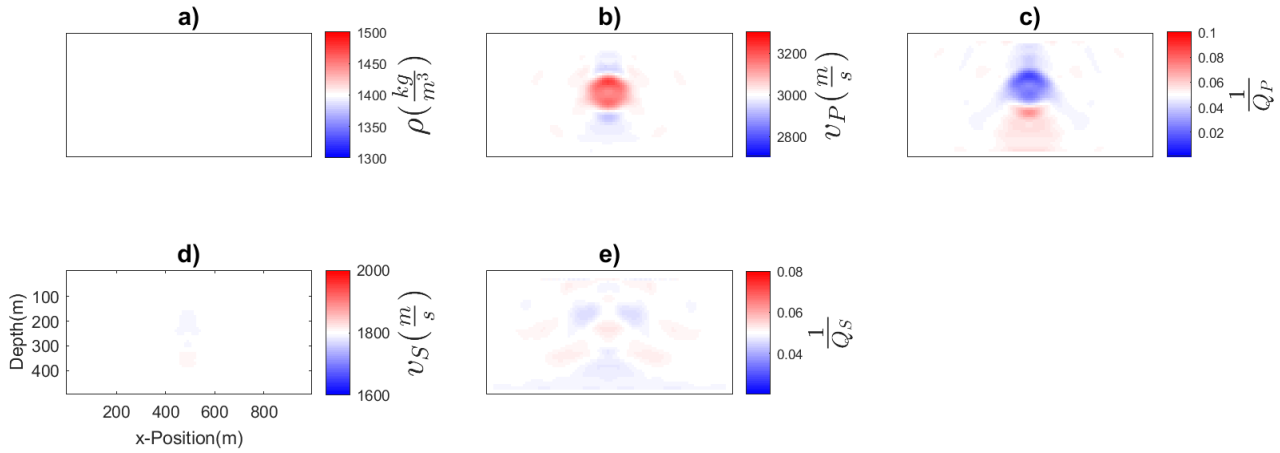

Figure 5.5: Equivalent-objective model for eliminating ρ anomaly.

confidence of the inversion in this anomaly.

I will characterize uncertainty for inversions using two different acquisition geometries with this model to better understand how different types of data can influence the confidence in inversion results. First, I investigate a surface acquisition geometry, where sources and multi-component receivers are present only on the surface, followed by a hybrid VSP geometry, where receivers are present in a vertical column representing a well in addition to the surface sources and receivers. For both acquisition geometries, 74 evenly spaced explosive sources and 148 evenly spaced multi-component receivers are be considered along the top of the model. For the hybrid acquisition, 58 multi-component receivers in a column from the top to the bottom of the model are considered at 200 m from the right edge of the model as well. The initial model for v_P, v_S, and ρ for both inversions is obtained by averaging the true model in the x direction. The low v_P and ρ anomaly is not included in this averaging, and so is not represented in the background model. The initial models for Q_P and Q_S are set as constants, given the difficulty in estimating these properties. The background model is shown in Figure 5.7. For both inversions, I consider data from 1 Hz to 20 Hz, in ten frequency bands, with five frequencies per band. For each band, one iteration of truncated Gauss-Newton optimization was applied, with 20 inner loop iterations per FWI iteration.

The inversion result for the surface-only data case is shown in Figure 5.8. This model provides

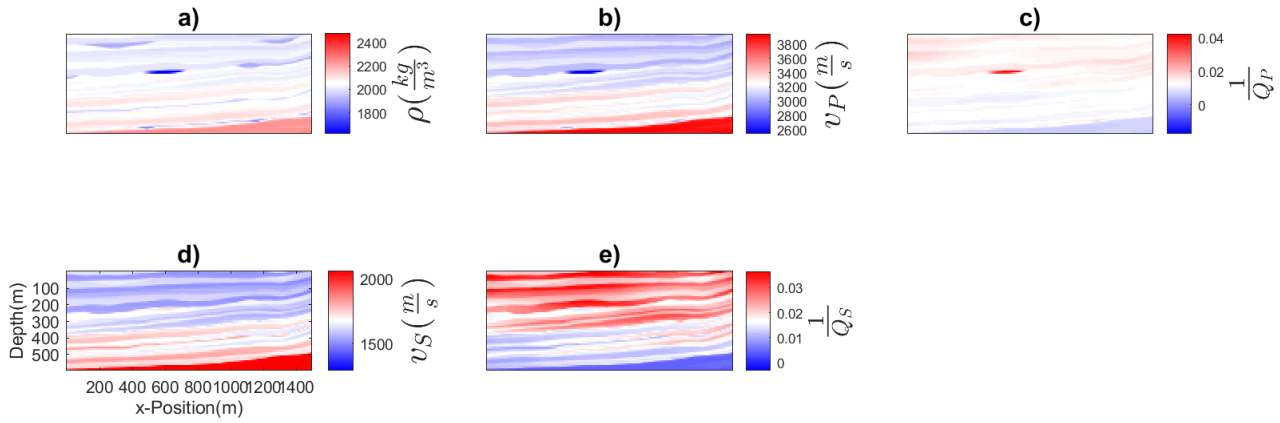

Figure 5.6: True model for synthetic tests, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

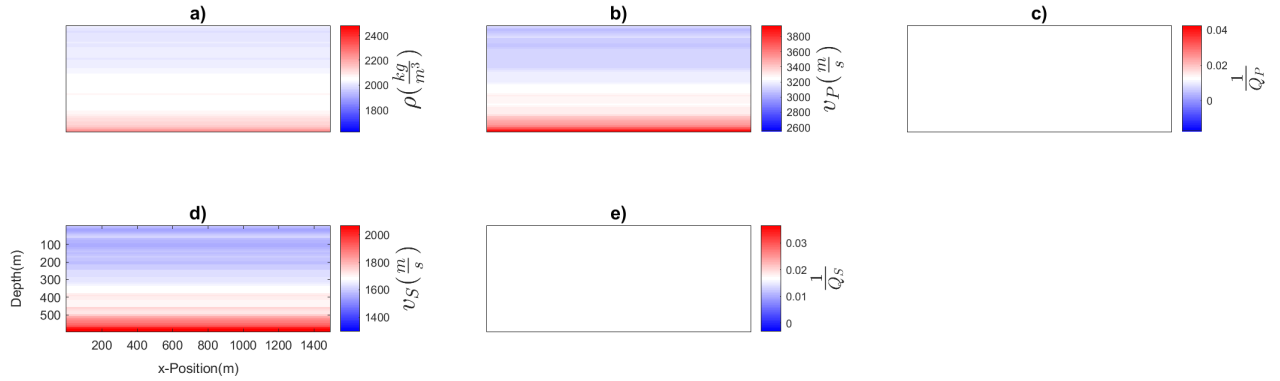

Figure 5.7: Initial model for inversions, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

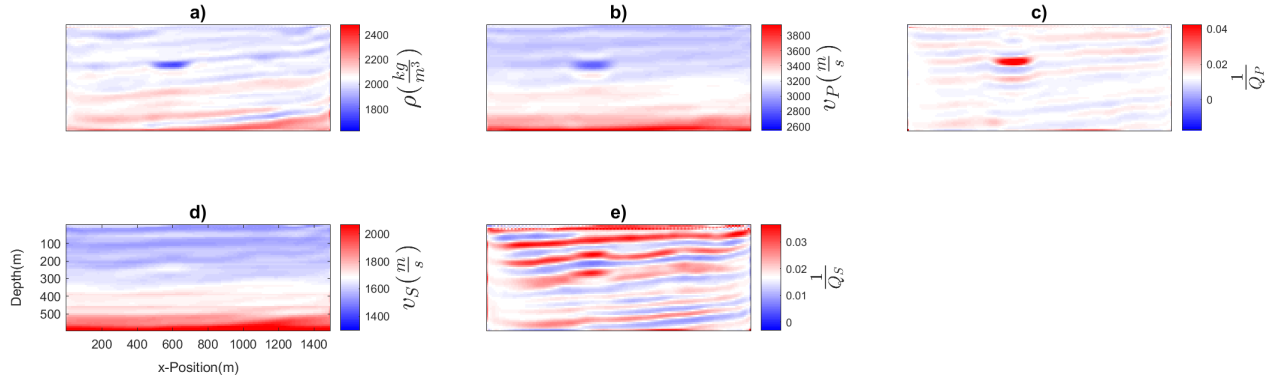

Figure 5.8: Inversion output for surface-type acquisition, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

relatively good recovery of the true density given the available wavelengths in the data. The recovered v_P and Q_P models seem to lack the resolution of the density result, but the gas sand feature is recovered. The estimates of v_S and Q_S are poor here, especially in the deep parts of the model. While the specific strategy used to obtain the result in Figure 5.8 could be altered, our focus here is on quantifying the confidence of the inversion in specific features. With this focus, I note that a prominent feature of both the true model and the inversion result is the gas-sand region of anomalously low ρ, v_P and Q_P. Let us suppose that the presence or absence of this anomaly is crucial to the interpretation of the inversion result. In the context of this procedure, we are relatively uninterested in the uncertainty of the inversion result in an arbitrary direction of model space, but keenly interested in the confidence with which this anomaly has been recovered. Accordingly, I seek to estimate the confidence in the hypothesis that this feature is present in the true model. To assess this confidence, I define the hypothesis function ψ to be large when such a feature is present in the inversion, but negligible otherwise:

$$\begin{aligned}\psi = \sum_{j=1}^{x_i \in x^*} & b_1 \left(\rho(x_j) - \bar{\rho}(x_j)\right)^2 \mathscr{H}\left(\rho(x_j) - \bar{\rho}(x_j)\right) \\ & + b_2 \left(v_P^{-2}(x_j) - \bar{v}_P^{-2}(x_j)\right)^2 \mathscr{H}\left(v_P^{-2}(x_j) - \bar{v}_P^{-2}(x_j)\right) \\ & + b_3 \left(Q_P^{-1}(x_j) - \bar{Q}_P^{-1}(x_j)\right)^2 \mathscr{H}\left(Q_P^{-1}(x_j) - \bar{Q}_P^{-1}(x_j)\right),\end{aligned} \tag{5.24}$$

where $\mathscr{H}$ is the Heaviside step function, b_1, b_2, and b_3 are scaling terms, x^* are the locations where the anomaly is recovered in the inversion, and $\bar{\rho}$, $\bar{v}_P^{-2}$, and $\bar{Q}_P^{-1}$ are the values of these properties in the initial model. Because we are interested in a change in each of these parameters, I choose b_1, b_2, and b_3 to weigh the initial contribution of each parameter equally.

With b_1, b_2, and b_3 chosen to weight the anomalies in each of ρ, v_P and Q_P equally, I can investigate which, if any, of these parameters are constrained by the inversion to be low at the location this anomaly. I estimate the optimal shuttle for this objective by following the procedure in Algorithm 1, using three outer-loop iterations of ten inner-loop iterations each. The resulting shuttle for this choice of penalty term is shown in Figure 5.9. This shuttle represents the step which should maximally reduce ψ without changing the objective function value. For this inversion, the

calculated shuttle represents a substantial violation of the stated hypothesis: while ψ was normalized to 1.00 for the inversion result, the shuttled model (shown in Figure 5.10) has a substantially lower $\psi = 0.395$. This means that the anomaly could be substantially smaller and still satisfy the FWI objective function just as well as the inversion output (the shuttled model has $\phi = 1.007 * \phi_0$). It is important to note the limitations on the information conveyed by this shuttle. Because the optimization of the shuttle is only approximate, this result does not represent the maximum uncertainty in the hypothesis contained in the FWI objective function, but instead a lower limit on this maximum. In other words, this shuttle shows that the inversion result is ambiguous (in terms of the data- and prior-fitting represented in the FWI objective function) with models violating the hypothesis at least as much as $\psi = 0.395$, but does not guarantee that there are no ambiguous models violating the hypothesis even more.

While the ψ value of the shuttled model is the most quantitatively informative result of this test, other details of the shuttle can be interpreted as well. In particular, the relative contributions of different parameters to the shuttle may be suggestive of the uncertainties in these parameters. The inversion output had $\psi = 1$, with contributions of each parameter through equation 5.24 equal to 0.333. After shuttling, the relative contribution of each parameter to ψ was reduced: to 0.141 from ρ, to 0.212 from v_P^{-2}, and to 0.043 from Q_P^{-1}. This means the shuttle approximating the minimizer of ψ demonstrates a large uncertainty in the Q_P^{-1} anomaly and less uncertainty in v_P^{-2} and ρ anomalies. Once again, these are lower limits on the maximum uncertainty represented in the FWI objective function, so while the v_P^{-2} anomaly may be more uncertain than reflected by a 0.256 ψ contribution, it is at least this uncertain.

Because the null-space shuttling approach I propose here quantifies uncertainty in a specific hypothesis, investigating more than one hypothesis requires that more than one instance of shuttling. Suppose as a second example that a key feature of the inversion result is the v_P/v_S ratio low at the location of the anomaly. To investigate this feature, I pose the hypothesis that the v_P/v_S ratio at this location is substantially lower than the average of this ratio in the rest of the study area. This

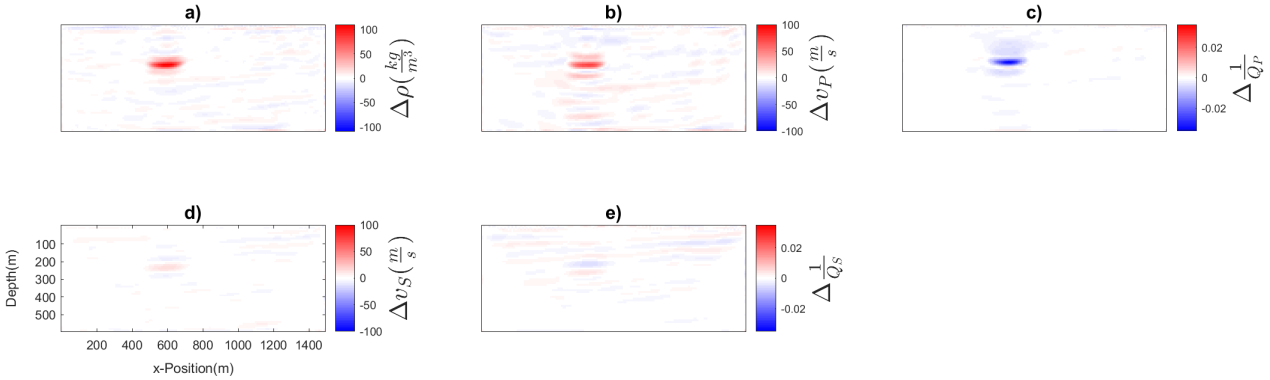

Figure 5.9: Optimal null-space shuttle for minimizing anomaly magnitude in ρ, v_P and Q_P, surface-only acquisition.

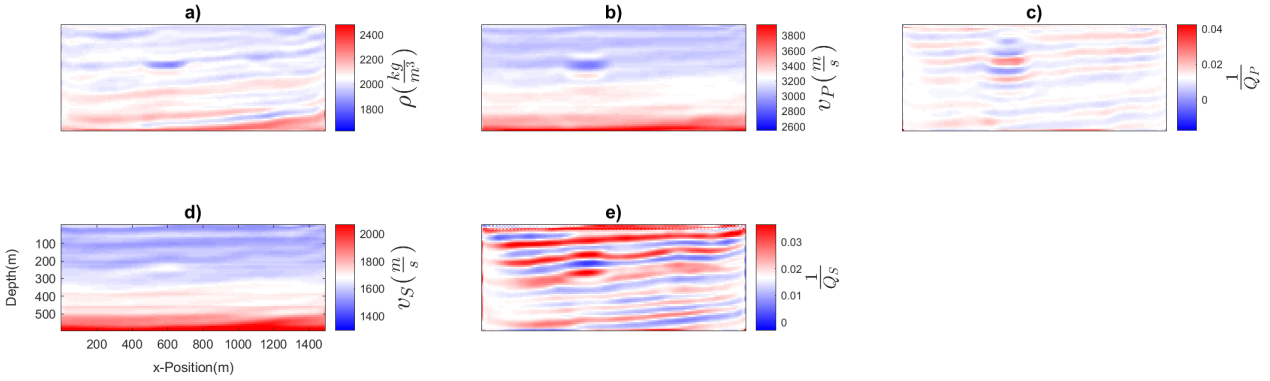

Figure 5.10: Equivalent-objective model minimizing anomaly magnitude in ρ, v_P and Q_P, surface-only acquisition. Compare to Figures 5.8, 5.9.

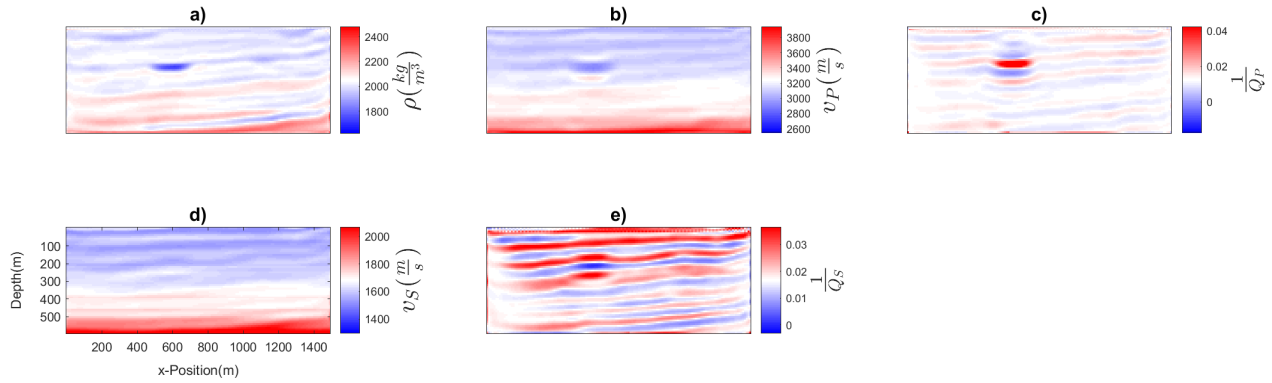

Figure 5.11: Equivalent-objective model minimizing v_P/v_S ratio anomaly, surface-only acquisition. Compare to Figure 5.8.

leads to a hypothesis function I define as

$$\psi = \sum_{j=1}^{x_j \in x^*} - \left(\frac{v_P(x_j)}{v_S(x_j)} - \tilde{\sigma} \right)^2, \tag{5.25}$$

where $\tilde{\sigma}$ is the average v_P to v_S ratio in the rest of the model. This choice of hypothesis function pushes toward models with anomaly v_P/v_S ratios similar to those in the rest of the model. I estimate the shuttle minimizing this function using three outer-loop iterations and ten inner-loop iterations. After applying the calculated shuttle, I obtain the model shown in Figure 5.11. In this case the inversion result has a ψ value of 1.000, while the shuttled model has $\psi = 0.399$. Again, this drop represents a lower bound on the maximum uncertainty, so the v_P/v_S ratio could be significantly higher at the anomaly while still satisfying the data- and prior-fitting requirements of the inversion. This shuttle establishes a greater relative uncertainty for v_P, where the minimum in the anomaly region increased from about 2810 m/s to 2940 m/s than for v_S, where changes were on the order of ± 10 m/s. Notably, both the Q_P and Q_S estimates near the anomaly change considerably for this shuttle, suggesting that an increase in v_P/v_S ratio can be partially compensated by a change in Q.

The data available in a given inversion problem can have a substantial impact on the confidence of the inversion in different model features. To investigate the effect of an acquisition geometry on our confidence in FWI results, I repeat the examples explored above, but now with additional data representing down-hole receivers. The same inversion strategy as above was used in this case, and

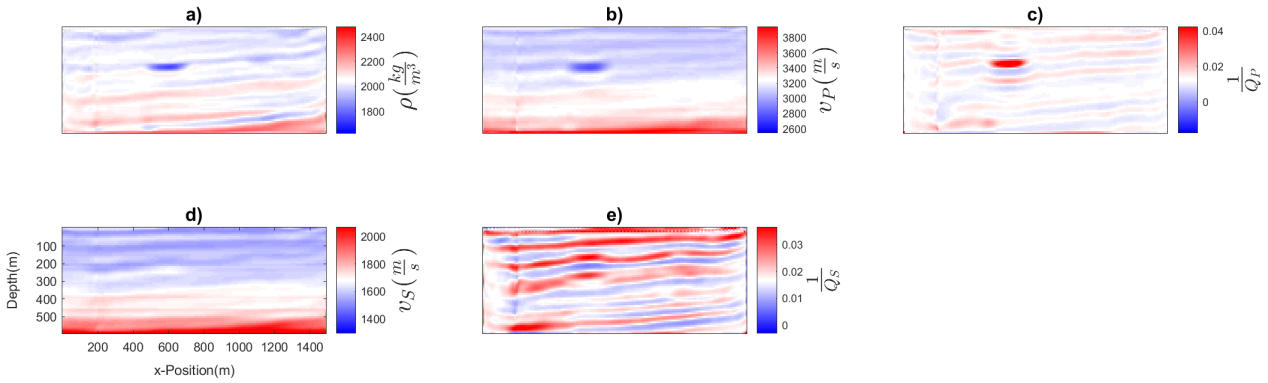

Figure 5.12: Inversion output for surface and VSP-type acquisition, defined in terms of ρ, Q values, and v_P, v_S at reference frequency ω_0.

the inversion result is shown in Figure 5.12. The major changes in the inversion result that these additional data bring are higher amplitude Q estimates in the bottom left of the model, and artifacts near the receiver locations, at about $x = 200$m.

To assess the confidence of the new inversion in the recovered anomaly, I again define ψ by equation 5.24. I approximate the shuttle which minimizes ψ using the same procedure as for the surface-only data case, and applying it, obtain the shuttled model in Figure 5.13. In this case, the additional information from the VSP has done relatively little to constrain ψ, which falls from 1.000 for the inverted model, to 0.438 in the shuttled model, little different from the $\psi = 0.395$ obtained in the surface-only data case. The relative contributions to ψ from the different parameters in this case are informative: the contributions from ρ and Q_P^{-1} are 0.147 and 0.036 as compared to 0.141 and 0.043 in the VSP and surface data only cases respectively, representing relatively small changes, while the contribution from v_P^{-2} increases to 0.256 from 0.212. This suggests that access to the VSP data in the inversion has reduced the ability of null-space shuttles to simultaneously reduce the amplitude of the ρ and Q_P^{-1} anomalies and the v_P^{-2} anomaly.

In a final example, I test the confidence of the hybrid VSP inversion in the low v_P/v_S anomaly it recovers. Calculating the optimal shuttle in the same way as for the surface-only inversion, I obtain the shuttled model in Figure 5.14. In this case, as in the case of surface-only data, the shuttled model shows a large degree of uncertainty in the hypothesis. In this case, the hypothesis

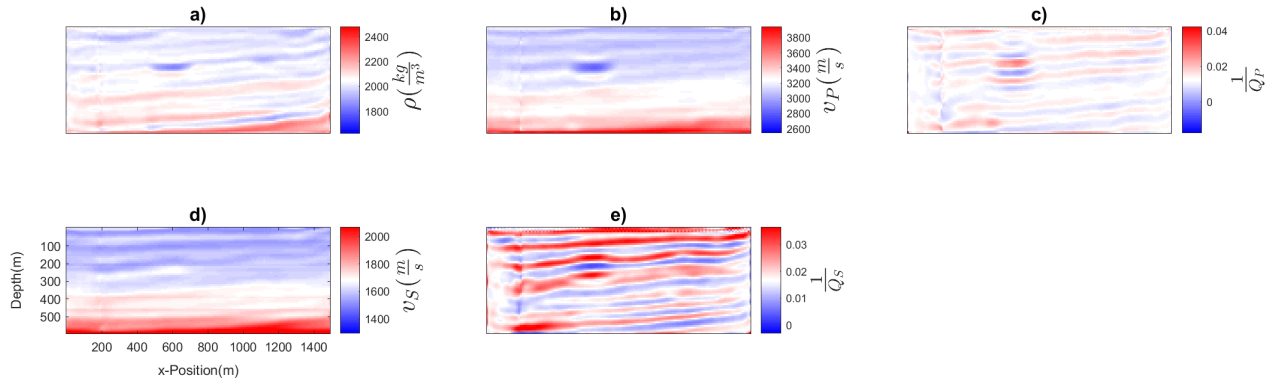

Figure 5.13: Equivalent-objective model minimizing anomaly magnitude in ρ, v_P and Q_P, surface and VSP acquisition. Compare to Figures 5.12, 5.10.

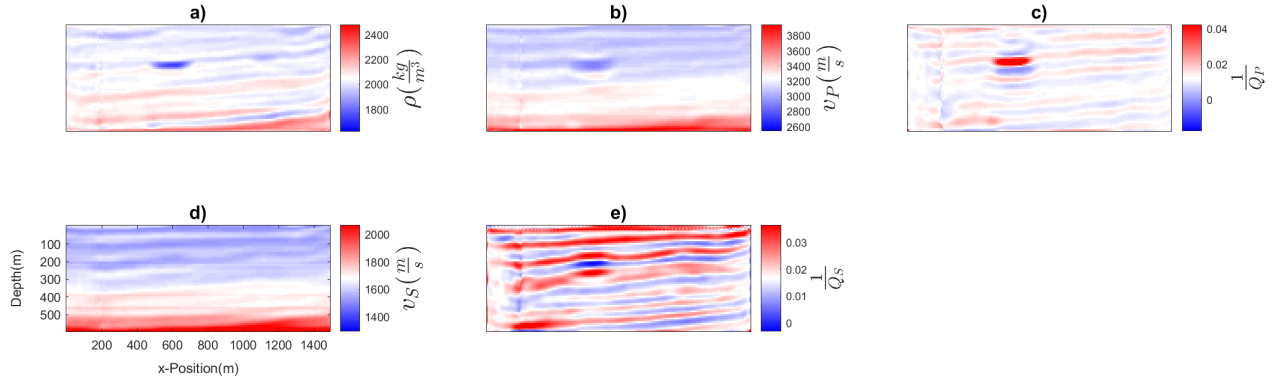

Figure 5.14: Equivalent-objective model minimizing v_P/v_S ratio anomaly, surface-only acquisition. Compare to Figures 5.12, 5.11.

function drops from 1.000 for the inversion result to 0.392 for the shuttled model. This is consistent with the ψ of 0.399 for the shuttled model in the surface-only case, suggesting that the VSP data do little to better constrain the v_P/v_S ratio of the anomaly in this case.

5.5 Discussion

A comparison of the approach I propose here and uncertainty quantification strategies which draw samples from an approximate Bayesian posterior distribution is important to understanding the role of this method. Posterior distribution sampling methods based on the Bayesian approach of Tarantola (2005) require estimates of both the most likely model and the posterior covariance

matrix in order to generate samples from a posterior probability distribution. Typically, these approaches assume that the most likely model is the one recovered in the inversion, and focus primarily on how to accurately estimate the posterior covariance matrix. The approach I discuss here instead focuses on quantifying uncertainty in the selection of the most probable model. This means that this approach concerns a type of uncertainty which is additional to that characterized by the posterior covariance matrix: no matter how accurately this matrix is calculated it does not account for our uncertainty in the location of the center of the posterior probability distribution. I do not attempt a probabilistic characterization of this uncertainty; I instead attempt to find the model that maximally violates a given hypothesis while still remaining an equally valid 'most probable model' as the inversion output.

In an FWI objective function, there are often features other than the quality of data-fit which play a major role in determining how good or bad an inversion result is. These usually take the form of regularization terms, which inform the objective function of a priori knowledge about the subsurface. Just as this a priori information is useful in preventing inversions from obtaining unacceptable results, it prevents null-space shuttles from equating the fitness of models on the basis of data fit alone. This means it is important to include an informative regularization term in the objective function when considering this approach for uncertainty characterization.

It is important for the computational expense of the approach I discuss here that in equation 5.9 I consider an optimization over $\mathbf{M}$' and not $\mathbf{M}$. In other words, I do not consider all viable models in minimizing ψ, but only the marginally viable ones (which are related by null-space shuttles). While it is certain that the largest possible step that can be taken in any given direction while remaining viable will result in a model in $\mathbf{M}$', it is not certain that the model representing the largest possible uncertainty in ψ does. By choosing to only consider null-space shuttles, I make the assumption that the model minimizing ψ lies outside of, or near $\mathbf{M}$'. I judge this assumption to be appropriate for two reasons. Firstly, except in the case of complicated, specific hypothesis functions or simple media, a minimum of ψ can generally be expected to lie outside $\mathbf{M}$', simply

because the complexities actually recovered by the inversion are unlikely to minimize any but a highly specific function. Secondly, the consequences of discovering a minimum of ψ inside $\mathbf{M}$' appear to be largely acceptable, as evidenced by the first ball model example I considered, where exactly this issue occurred, but uncertainty estimates were reasonable regardless.

The computational cost incurred in the uncertainty estimation for each hypothesis in the complex examples I considered here was on the order of 70 wavefield propagation problems per frequency in the band considered. In comparison, the cost of the inversion itself was closer to 200 wavefield propagation problems per frequency in the bands considered, so meaningful uncertainty estimates were obtained at less computational cost than the inversion itself. It also important to note that the cost of assessing the uncertainty is tunable; by changing the number of inner and outer iterations used, the computational cost and accuracy in representing an upper uncertainty bound can be altered. Given the targeted nature of this approach, computational cost might also be reduced by limiting the dimensionality of the model to describe only the features which will play a major role in the hypothesis function.

If there is high confidence in the large-scale behavior of the subsurface in a given area, and the key questions about the subsurface to be answered through use of seismic data are well understood, this approach may also have value in providing an estimate of how well a given acquisition geometry will be able to constrain given features of the model. This could be useful for acquisition design by providing an estimate of whether sufficient confidence is obtained with a potential geometry.

5.6 Conclusions

General uncertainty in FWI is difficult to quantify, and generally requires large amounts of computation and storage for accurate treatment. The targeted approach to uncertainty quantification investigated here explores the uncertainty in a stated hypothesis about the model. I quantify this uncertainty by finding the marginally viable model which maximally violates the stated hypothesis. The problem of finding this model can be treated as an optimization over null-space shuttles

from the inversion output. If appropriately defined, our confidence in the hypothesis considered can inform us about our degree of certainty in the key decision-altering features of an inversion result. The computational cost of this approach is tunable and can be small relative to that of the full inversion. The numerical examples I consider here suggest that while the inversions seem to constrain the recovered P- and S- wave velocities reasonably well, there is substantial uncertainty in density, and especially Q. This approach offers a viable way to assess the impacts of effects like cross-talk on our confidence in inversion results.

Chapter 6

Conclusions

6.1 Summary

Attenuation effects play a major role in seismic wave propagation, and their treatment in full waveform inversion is important if the correct treatment of seismic amplitudes is to be achieved. While the concept of QFWI has been investigated, there are major outstanding issues which complicate the treatment of attenuation. In this thesis, I have sought to propose strategies to help address some of the chief challenges which present themselves in an attenuative full-waveform inversion. Specifically, I have presented methods for mitigating the impacts of uncertainty in attenuation physics, for quantifying expected cross-talk between both co-located and spatially separated viscoelastic variables, and for quantifying an important class of inversion uncertainty, which relates to imperfect optimization features (such as cross-talk) in viscoelastic FWI. In this section I will briefly summarize the major findings of my numerical investigation of these methods.

Because seismic data cannot constrain seismic velocities and Q which vary freely in both space and frequency, it is generally necessary to use a specific attenuation physics model to better constrain the frequency dependence of these variables in inversion. In Chapter 3, my focus was developing a strategy to cope with the problems that arise when the physics model selected is not a good representation of true subsurface attenuation. The approach I suggested was intermediate between an inflexible insistence on the assumed physics model and a freely varying Q and velocity. This approach enforced the assumed physics only within small frequency bands, and recovered one subsurface model for each frequency band considered. In numerical examples, I inverted data generated with the attenuation and dispersion associated with a single standard linear solid, assuming a Kolsky-Futterman model of attenuation in the inversions. In this framework, the flexible approach I proposed was significantly better able to generate a meaningful estimate of the true subsurface

properties. I proposed that the band size in this approach controlled a tradeoff between attenuation model flexibility and inversion cross-talk.

In Chapter 4, I proposed a strategy for estimating the extent of cross-talk expected in viscoelastic FWI, given different acquisition settings and optimization strategies. This work is important in viscoelastic FWI because cross-talk between Q and velocity variables is very pronounced, and because the conventionally used radiation pattern analysis is uninformative about key features of cross-talk. While radiation patterns suggest that frequency and phase information are the keys to distinguishing velocity and Q variables, they provide no obvious information about how this information changes cross-talk as a function of acquisition or optimization strategy. Radiation pattern analysis is alow uninformative about cross-talk between spatially separated variables, which may be important if there is cross-talk with an obscuring Q region. I proposed a strategy of differencing numerical simulations to determine the contribution of different subsets of variables, and especially the cross-talk. Using this approach, I determined that transmission information plays an important role in suppressing cross-talk between Q variables and the corresponding velocities. I demonstrated that cross-talk between elastic properties and spatially separated Q variables is an important, and sometimes dominant mode of cross-talk with Q. I also showed that differences in source-type between explosive and directed point forces can have major effects on cross-talk.

Several cross-talk modes investigated in Chapter 4 were not significantly reduced even with very computationally intensive optimization strategies (especially with surface-only acquisition). Given that compuational costs are often a limitation in FWI, these cross-talk modes are likely to be present in many inversion results. In Chapter 5 I proposed a targeted uncertainty quantification strategy which specifically targeted uncertainty associated with imperfect optimization effects, like cross-talk. By focusing the uncertainty quantification on a specific hypothesis about the inversion result, this approach allowed for computationally affordable uncertainty estimates, based on a null-space shuttling approach. I demonstrated this approach on numerical viscoelastic inversion examples. The results of these tests suggested relatively high confidence in velocity variables, and

much less confidence in the recovered estimates of density and Q. This type of tool could be very useful for assessing confidence in key inversion features that may be subject to cross-talk.

6.2 Future Work

In this thesis, I presented several techniques designed to help solve some of the specific problems faced in QFWI. Several substantial obstacles to the effective implementation of QFWI remain, however. Perhaps the most significant of these obstacles is the problem of cross-talk. In Chapter 3, I showed that large frequency bands are key for reducing cross-talk, in Chapter 4, I identified acquisition geometries and optimization strategies that are effective in reducing Q-related cross-talk, and in Chapter 5, I proposed a tool which can help quantify the uncertainties introduced by cross-talk. The fact remains, however, that with typical acquisition geometries and affordable computation levels, cross-talk is substantial in QFWI, and can adversely affect the quality and reliability of results. Improved optimization strategies, objective function formulations or model parameterizations which allow for cross-talk to be more efficiently reduced are key for making QFWI effective in the absence of significant transmission data.

With respect to the work discussed in Chapter 3, there are several outstanding questions which could be the subjects of additional research. While I proposed that the band size in the flexible approach controlled a tradeoff between mitigating the effects of Q model uncertainty and cross-talk, an effective means of assessing the degree of Q model uncertainty in a given problem would be very helpful for deciding on this tradeoff term. Another interesting question relates to the definition of cross-talk when the attenuation model considered is incorrect. Cross-talk as I define it here relates to the mis-attribution of data residuals to the wrong model residuals. This definition works well when the cross-talk is between parameters considered in the inversion, and is easily extended to the case where the cross talk is from a model variable not represented in the inversion into one that is represented, but it seems very unclear how to define cross-talk where one parameter is represented inadequately. This isn't the same as the case in which a parameter not included in the

inversion contributes to cross-talk, because in some sense we would like the data residuals caused by attenuation effects to be attributed to the model variables controlling attenuation, but there is no well-defined model residual for a parameter which does not accurately represent the effects it is meant to. Defining a metric of cross-talk which would allow for an analysis like the one in Chapter 4 while also allowing for attenuation model errors would help to better analyze this part of the QFWI problem.

While the uncertainty quantification approach I present in Chapter 5 should be useful for quantifying several different types of uncertainty, it is not a method that is capable to quantifying uncertainties arising from possible errors in the attenuation physics. In fact, while strategies exist for coping with wave physics errors in FWI exist, no strategies that I am aware of attempt to quantify the associated inversion result uncertainties. Determining a way to estimate these uncertainties would significantly advance the confidence that could be placed in QFWI uncertainty estimates.

Another project which could be done to expand on the work developed in Chapter 5 would be to determine better ways of framing the hypotheses used in the targeted uncertainty quantification. The hypothesis functions tested in Chapter 5 were relatively crude examples of metrics that could be used to assess key aspects of the inversion. These metrics could have been better optimized to be more informative about whether the hypotheses represented were upheld or violated. Application to field data examples could also offer the opportunity to develop this technique with respect to key uncertainties in a real setting.

Bibliography

Aki, K. and P. G. Richards, 2002, Quantitative seismology: University Science Books, 2nd edition.

Alkhalifah, T. and R. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, no. 3, R91–R101.

Anagaw, A. Y., 2014, Full waveform inversion using simultaneous encoded sources based on first- and second-order optimization methods.

Annan, A., 2005, Ground-penetrating radar, *in* Near-surface geophysics, 357–438. Society of Exploration Geophysicists.

Barnes, C. and M. Charara, 2009, The domain of applicability of acoustic full-waveform inversion for marine seismic data: Geophysics, **74**, no. 6, WCC91–WCC103.

Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of cross-well data: Geophysical Prospecting, **56**, 897–906.

Baysal, E., D. D. Kosloff, and J. W. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

Berenger, J., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of computational physics, **114**, 185–200.

Best, A., C. McCann, and J. Sothcott, 1994, The relationships between the velocities, attenuations and petrophysical properties of reservoir sedimentary rocks 1: Geophysical Prospecting, **42**, 151–178.

Biot, M. A., 1956, Theory of propagation of elastic waves in a fluid-saturated porous solid: J. Acous. Soc. Am., **28**, 168–191.

Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, 2007, Structure of the California coast ranges and San Andreas fault at SAFOD from seismic waveform inversion and reflection imaging: Journal of Geophysical Research: Solid Earth, **112**.

Borcherdt, R. D., 1973, Energy and plane waves in linear viscoelastic media: Journal of Geophys-

ical Research, **78**, 2442–2453.

Bording, R. P., A. Gersztenkorn, L. R. Lines, J. A. Scales, and S. Treitel, 1987, Applications of seismic travel-time tomography: Geophysical Journal International, **90**, 285–303.

Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Bregman, N., R. Bailey, and C. Chapman, 1989, Crosshole seismic tomography: Geophysics, **54**, 200–215.

Broyden, C. G., 1970, The convergence of a class of double-rank minimization algorithms 1. general considerations: IMA Journal of Applied Mathematics, **6**, 76–90.

Bui-Thanh, T., O. Ghattas, J. Martin, and G. Stadler, 2013, A computational framework for infinite-dimensional Bayesian inverse problems part i: The linearized case, with application to global seismic inversion: SIAM Journal on Scientific Computing, **35**, A2494–A2523.

Bunks, C., F. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Carcione, J., 2007, Wavefields in real media: Elsevier Publishing Co., New York, NY.

Casula, G. and J. Carcione, 1992, Generalized mechanical model analogies of viscoelastic behaviour: Bolletina di Geofisica Teorica ed Applicata, **34**, 235–256.

Chen, J., C. R. Wilson, and B. D. Tapley, 2006, Satellite gravity measurements confirm accelerated melting of Greenland ice sheet: Science, **313**, 1958–1960.

Choi, Y., D. Min, and C. Shin, 2008, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media: Geophysical Prospecting, **56**, 863–881.

da Silva, N. V. and G. Yao, 2017, Wavefield reconstruction inversion with a multiplicative cost function: Inverse Problems, **34**, 015004.

Daily, W., A. Ramirez, D. LaBrecque, and J. Nitao, 1992, Electrical resistivity tomography of vadose water movement: Water Resources Research, **28**, 1429–1442.

Deal, M. M. and G. Nolet, 1996, Nullspace shuttles: Geophysical Journal International, **124**, 372–380.

Debens, H. A., M. Warner, A. Umpleby, N. V. da Silva, et al., 2015, Global anisotropic 3D FWI: Presented at the Annual International Meeting, Society of Exploration Geophysicists.

Downton, J. E. and C. Ursenbach, 2006, Linearized amplitude variation with offset (AVO) inversion with supercritical angles: Geophysics, **71**, no. 5, E49–E55.

Eliasson, P. and A. Romdhane, 2017, Uncertainty quantification in waveform-based imaging methods - a Sleipner CO2 monitoring study: Energy procedia, **114**, 3905–3915.

Engquist, B. and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, **12**, 979–988.

Esser, E., L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2018, Total variation regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences, **11**, 376–406.

Fabien-Ouellet, G., E. Gloaguen, and B. Giroux, 2017, Time domain viscoelastic full waveform inversion: Geophysical Journal International, **209**, 1718–1734.

Fang, Z., C. Da Silva, R. Kuske, and F. J. Herrmann, 2018, Uncertainty quantification for inverse problems with weak partial-differential-equation constraints: Geophysics, **83**, no. 6, R629–R647.

Fatti, J. L., G. C. Smith, P. J. Vail, P. J. Strauss, and P. R. Levitt, 1994, Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the geostack technique: Geophysics, **59**, 1362–1376.

Fichtner, A. and T. v. Leeuwen, 2015, Resolution analysis by random probing: Journal of Geophysical Research: Solid Earth, **120**, 5549–5573.

Fichtner, A. and J. Trampert, 2011, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1624.

Fichtner, A. and A. Zunino, 2019, Hamiltonian nullspace shuttles: Geophysical research letters,

46, 644–651.

Fletcher, R., 1970, A new approach to variable metric algorithms: The computer journal, **13**, 317–322.

Fu, L., B. Guo, and G. T. Schuster, 2018, Multiscale phase inversion of seismic data: Geophysics, **83**, no. 2, R159–R171.

Futterman, W., 1962, Dispersive body waves: Journal of Geophysical Research, **67**, 5279–5291.

Gholami, Y., R. Brossier, S. Operto, A. Ribodetti, and J. Virieux, 2013, Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? part 1: Sensitivity and trade-off analysis: Geophysics, **78**, no. 2, R81–R105.

Goldfarb, D., 1970, A family of variable-metric methods derived by variational means: Mathematics of computation, **24**, 23–26.

Goodway, B., T. Chen, and J. Downton, 1997, Improved avo fluid detection and lithology discrimination using lamé petrophysical parameters;"$\lambda\rho$","$\mu\rho$", & "λ/μ fluid stack", from p and s inversions, *in* SEG Technical Program Expanded Abstracts 1997, 183–186. Society of Exploration Geophysicists.

Gray, S. H. and W. P. May, 1994, Kirchhoff migration using eikonal equation traveltimes: Geophysics, **59**, 810–817.

Hak, B. and W. Mulder, 2010, Migration for velocity and attenuation perturbations: Geophysical Prospecting, **58**, 939–952.

——— 2011, Seismic attenuation imaging with causality: Geophysical Journal International, **184**, 439–451.

Hicks, G. and R. Pratt, 2001, Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.

Huang, G. and W. W. Symes, 2015, Full waveform inversion via matched source extension, *in* SEG Technical Program Expanded Abstracts 2015, 1320–1325. Society of Exploration Geophysicists.

Innanen, K., 2014, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, 717–734.

Innanen, K. A. and A. B. Weglein, 2007, On the construction of an absorptive-dispersive medium model via direct linear inversion of reflected seismic primaries: Inverse Problems, **23**, 2289–2310.

Jiracek, G. R., 1990, Near-surface and topographic distortions in electromagnetic induction: Surveys in Geophysics, **11**, 163–203.

Johnston, D. H., M. Toksöz, and A. Timur, 1979, Attenuation of seismic waves in dry and saturated rocks: II. Mechanisms: Geophysics, **44**, 691–711.

Kalita, M., V. Kazei, Y. Choi, and T. Alkhalifah, 2019, Regularized full-waveform inversion with automated salt flooding: Geophysics, **84**, no. 4, R569–R582.

Kamath, N. and I. Tsvankin, 2016, Elastic full-waveform inversion for vti media: Methodology and sensitivity analysis: Geophysics, **81**, no. 2, C53–C68.

Kamath, N., I. Tsvankin, and E. Díaz, 2017, Elastic full-waveform inversion for vti media: A synthetic parameterization study: Geophysics, **82**, no. 5, C163–C174.

Kamei, R. and R. Pratt, 2013, Inversion strategies for visco-acoustic waveform inversion: Geophysical Journal International, **194**, 859–884.

Köhn, D., D. D. Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, 2012, On the influence of model parameterization in elastic full waveform tomography: Geophysical Journal International, **191**, 325–345.

Kolsky, H., 1956, The propagation of stress pulses in viscoelastic solids: Philosophical Magazine, **1**, 693–710.

Kuang, W. and J. Bloxham, 1997, An earth-like numerical dynamo model: Nature, **389**, 371–374.

Kühl, H. and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: Geophysics, **68**, 262–273.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference

on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Liu, D. C. and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization: Mathematical programming, **45**, 503–528.

Liu, H., D. Anderson, and H. Kanamori, 1976, Velocity dispersion due to anelasticity; implications for seismology and mantle composition: Geophys. J. Roy. Astr. Soc, **47**, 41–58.

Liu, Q. and D. Peter, 2019, Square-root variable metric based elastic full-waveform inversion—part 2: uncertainty estimation: Geophysical Journal International, **218**, 1100–1120.

——— 2020, Square-root variable metric-based nullspace shuttle: A characterization of the nonuniqueness in elastic full-waveform inversion: Journal of Geophysical Research: Solid Earth, **125**, e2019JB018687.

Luo, S. and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion, *in* SEG Technical Program Expanded Abstracts 2011, 2788–2792. Society of Exploration Geophysicists.

Luo, Y., Y. Ma, Y. Wu, H. Liu, and L. Cao, 2016, Full-traveltime inversion: Geophysics, **81**, no. 5, R261–R274.

Ma, Y., D. Hale, B. Gong, and Z. Meng, 2012, Image-guided sparse-model full waveform inversion: Geophysics, **77**, no. 4, R189–R198.

Malinowski, M., S. Operto, and A. Ribodetti, 2011, High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion: Geophysical Journal International, **186**, 1179–1204.

Mavko, G. and A. Nur, 1975, Melt squirt in the astenosphere: J. Geophys. Res., **80**, 1444–1448.

Maxwell, S. C., J. Rutledge, R. Jones, and M. Fehler, 2010, Petroleum reservoir characterization using downhole microseismic monitoring: Geophysics, **75**, no. 5, 75A129–75A137.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion:

Geophysical Journal International, **205**, 345–377.

Métivier, L., R. Brossier, S. Operto, and J. Virieux, 2015, Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach: SEG Expanded Abstracts, 1198–1203.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, no. 2, B401–B437.

Moradi, S. and K. Innanen, 2016, Scattering of homogeneous and inhomogeneous seismic waves in low-loss viscoelastic media: Geophysical Journal International, **202**, 1722–1732.

Mulder, W. and B. Hak, 2009, An ambiguity in attenuation scattering imaging: Geophysical Journal International, **178**, 1614–1624.

Nash, S. G., 1985, Preconditioning of truncated-Newton methods: SIAM Journal on Scientific and Statistical Computing, **6**, 599–616.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Nocedal, J. and P. S. Wright, 2006, Numerical optimization: Springer, 2nd edition.

Nolet, G., 1987, Seismic tomography: With applications in global seismology and exploration geophysics, volume **5**: Springer Science & Business Media.

Odebeatu, E., J. Zhang, M. Chapman, E. Liu, and X. Y. Li, 2006, Application of spectral decomposition to detection of dispersion anomalies associated with gas saturation: The Leading Edge, **2**, 206–210.

O'Doherty, R. F. and N. A. Anstey, 1971, Reflections on amplitudes: Geophysical Prospecting, **19**, 430–458.

Oh, J. and T. Alkhalifah, 2016, The scattering potential of partial derivative wavefields in 3-D elastic orthorhombic media: an inversion prospective: Geophysical Journal International, **206**, 1740–1760.

Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, 2013, A

guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, 1040–1054.

Owusu, J. C., O. Podgornova, M. Charara, S. Leaney, A. Campbell, S. Ali, I. Borodin, L. Nutt, and H. Menkiti, 2015, Anisotropic elastic full-waveform inversion of walkaway vertical seismic profiling data from the Arabian Gulf: Geophysical Prospecting, **64**, 38–53.

Pan, W., Y. Geng, and K. A. Innanen, 2018, Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and Hussar land data set application: Geophysical Journal International, **213**, 1305–1333.

Pan, W., K. A. Innanen, G. F. Margrave, M. C. Fehler, X. Fang, and J. Li, 2016, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, **81**, no. 5, R275–R291.

Parker, R. L., 2003, Ideal bodies for mars magnetics: Journal of Geophysical Research: Planets, **108**.

Plessix, R., A. Stopin, H. Kuehl, V. Goh, and K. Overgaag, 2016, Visco-acoustic full waveform inversion: Annual International Meeting, European Association of Geoscientists & Engineers, 78th EAGE Conference and Exhibition 2016, cp–490.

Plessix, R.-E., P. Milcik, H. Rynia, A. Stopin, K. Matson, and S. Abri, 2013, Multiparameter full-waveform inversion: marine and land examples: The Leading Edge, **Sep**, 1030–1038.

Pratt, R. G., 1990, Frequency-domain elastic wave modeling by finite differences; a tool for cross-hole seismic imaging: Geophysics, **55**, 626–632.

Pratt, R. G., C. Shin, and G. Hick, 1998, Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion: Geophysical journal international, **133**, 341–362.

Pride, S. R., J. G. Berryman, and J. M. Harris, 2004, Seismic attenuation due to wave-induced flow: Journal of Geophysical Research: Solid Earth, **109**.

Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal, 2011, On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study:

Geophysical Journal International, **187**, 1495–1515.

Quan, Y. and J. Harris, 1997, Seismic attenuation tomography using the frequency shift method: Geophysics, **62**, 895–905.

Ren, H., G. Goloshubin, and F. J. Hilterman, 2009, Poroelastic analysis of amplitude-versus-frequency variations: Geophysics, **74**, no. 6, N41–N48.

Sabaka, T. J., N. Olsen, and R. A. Langel, 2002, A comprehensive model of the quiet-time, near-earth magnetic field: phase 3: Geophysical Journal International, **151**, 32–68.

Sams, M., J. Neep, M. Worthington, and M. King, 1997, The measurement of velocity dispersion and frequency-dependent intrinsic attenuation in sedimentary rocks: Geophysics, **62**, 1456–1464.

Sears, T. J., S. Singh, and P. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.

Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: Mathematics of computation, **24**, 647–656.

Shin, C. and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: Geophysics, **71**, no. 3, R31–R42.

Smith, G. and P. Gidlow, 1987, Weighted stacking for rock property estimation and detection of gas: Geophysical prospecting, **35**, 993–1014.

Song, Z.-M. and P. R. Williamson, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data; part 1, 2.5-D modeling method: Geophysics, **60**, 784–795.

Stolt, R. H., 1978, Migration by fourier transform: Geophysics, **43**, 23–48.

Stolt, R. H. and A. Weglein, 2012, Seismic imaging and inversion: Volume 1: Application of linear inverse theory: Cambridge University Press.

Sudha, K., M. Israil, S. Mittal, and J. Rai, 2009, Soil characterization using electrical resistivity tomography and geotechnical investigations: Journal of Applied Geophysics, **67**, 74–79.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2009, Adjoint tomography of the southern california

crust: Science, **325**, 988–992.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

——— 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

——— 2005, Inverse problem theory and methods for model parameter estimation, volume **89**: siam.

Thurin, J., R. Brossier, and L. Métivier, 2019, Ensemble-based uncertainty estimation in full waveform inversion: Geophysical Journal International, **219**, 1613–1635.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

Ursin, B. and T. Toverud, 2002, Comparison of seismic dispersion and attenuation models: Studia Geophysica et Geodaetica, **46**, 293–320.

Van Camp, M., O. de Viron, A. Watlet, B. Meurers, O. Francis, and C. Caudron, 2017, Geophysics from terrestrial time-variable gravity measurements: Reviews of Geophysics, **55**, 938–992.

Van Gestel, J.-P., J. H. Kommedal, O. I. Barkved, I. Mundal, R. Bakke, and K. D. Best, 2008, Continuous seismic surveillance of Valhall field: The Leading Edge, **27**, 1616–1621.

Van Leeuwen, T. and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, no. 2, R63–R77.

Vigh, D., W. Starr, and K. K. Dingwall, 2009, 3D prestack time domain full waveform inversion: Annual International Meeting, European Association of Geoscientists & Engineers, 71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009, cp–127.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

Wang, Y., L. Dong, Y. Liu, and J. Yang, 2016, 2D frequency-domain elastic full-waveform inversion using the block-diagonal pseudo-hessian approximation: Geophysics, **81**, no. 5, R247–R259.

Ward, S. H. and G. W. Hohmann, 1988, Electromagnetic theory for geophysical applications, *in* Electromagnetic Methods in Applied Geophysics: Voume 1, Theory, 130–311. Society of Exploration Geophysicists.

Warner, M. and L. Guasch, 2014, Adaptive waveform inversion: theory: SEG Expanded Abstracts.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, C. Win, et al., 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59–R80.

Wessel, P. and A. B. Watts, 1988, On the accuracy of marine gravity measurements: Journal of Geophysical Research: Solid Earth, **93**, 393–413.

White, J. E., 1975, Computed seismic speeds and attenuation in rocks with partial gas saturation: Geophysics, **40**, 224–232.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016, A review on the systematic formulation of 3-D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, **207**, 129–149.

Yang, Q., A. Malcolm, H. Rusmanugroho, and W. Mao, 2019, Analysis of radiation patterns for optimized full waveform inversion in fluid-saturated porous media: Geophysical Journal International, **216**, 1919–1937.

Zhou, W., B. F. Beck, and A. L. Adams, 2002, Effective electrode array in mapping karst hazards in electrical resistivity tomography: Environmental geology, **42**, 922–928.

Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas, 2016, A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration: Geophysics, **81**, no. 5, R307–R323.