Modularized iterative full seismic waveform inversion for 3D-heterogeneous media based on waveform sensitivity kernels

Dissertation
zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften an der Fakultät für Geowissenschaften der Ruhr-Universität Bochum

vorgelegt von

Florian Schumacher

aus Ahlen

Bochum, Dezember 2013

Acknowledgments

Contents

Contents

1 Introduction

Seismic inversion aims at inferring information on physical properties of the earth from seismic observables. All those methods differ mainly in the physical parameterization of the earth model which is inverted for, as well as the kind of seismic data which is explained by that model within their errors. Isotropic or anisotropic elastic constants or wave speeds inside the earth may be of interest on local or global scale, as well as attenuation or petrophysical properties. Some methods try to derive those properties explaining distinct observables derived from measured seismograms, such as arrival times and amplitudes of wave phases, surface wave dispersion, or the splitting of certain wave phases due to anisotropy. Other methods, however, do not reduce the data information to those kinds of distinct observables, but rather aim at explaining each and every wavelet of the measured waveforms by earth structure. Hence, these methods are called *full waveform inversion* methods.

Existing full waveform inversion methods are mainly (linearized) iterative methods, that is they improve the earth model iteratively in the course of the inversion process, starting with some approximation of the true earth. They aim at finding a model explaining the data best within their errors. In general they make use of different kinds of scalar-valued measures of misfit, which quantify how well the data functionals of the currently inverted earth model coincide with the measured data. Those, in turn, are assumed to be data functionals of the true earth model that is searched for. The data functionals of the currently inverted earth model are, in general, synthetically evaluated by numerical or semi-analytical wave propagation methods. In the course of the iterative inversion process, an earth model which explains the data better than the previous model (i.e. for which the misfit functional becomes smaller) is often found using the Fréchet derivative of the misfit functional w.r.t. the model parameters. This derivative is sometimes called *misfit kernel* and quantifies how sensitive the misfit is to changes of certain model parameters. This motivates the general terminology of *sensitivity kernel*.

Since a specific scalar-valued misfit functional combines all waveform data to one scalar quantity in a certain way, the respective misfit kernel only quantifies sensitivity of this specific combination of waveform data. Hence, a different choice of measure of misfit changes this sensitivity quantification and, hence, influences the convergence behaviour of the inversion procedure. This motivates the usage of sensitivity information separately given for each data functional of full seismic waveforms. Such sensitivities quantify the change of a specific data functional w.r.t. change of a specific model parameter. They are given in form of the Fréchet derivatives of the data functionals w.r.t. the model parameters, called *waveform sensitivity kernels*. They constitute the most general form of sensitivity quantification of seismic waveform

data: All sensitivity kernels of misfit functionals or seismic observables based on some combination of full waveform data may be derived from the individual sensitivities of the data functionals. Using waveform sensitivity kernels in a full waveform inversion procedure, hence, raises hope to a better inversion performance and more flexible usage of data information in the course of the inversion process.

   This thesis develops, validates and discusses a full waveform inversion procedure for arbitrary heterogeneous earth models based on waveform sensitivity kernels.
Chapter 2 gives a general introduction to geophysical inverse problems and general solution strategies. Most common methods of seismic (full waveform) inversion are presented and a motivation of the full waveform inversion concept presented in this thesis is given.
Chapter 3 then explicitly derives the waveform sensitivity kernels on which the inversion procedure is based and points to general issues in the course of practically computing them. Some examples of sensitivity kernels and their application in sensitivity analysis are presented, as well.
Chapter 4 derives in detail the full waveform inversion procedure, discussing any issues in its practical realization.
Chapter 5 finally introduces the software package ASKI, in which the inversion procedure was implemented. A validation of the software and the underlying full waveform inversion method is presented in form of a synthetic full waveform inversion conducted by ASKI.
Chapters 6 and 7, thereafter, give a summary and an outlook.

2 General aspects of inverse problems: conventional solution approaches and new concepts

Seismic data generally is acquired as time-dependent values related to ground motion, which is excited by a seismic source and recorded at some receiver position for some component, i.e. direction, of space. In addition to the specific radiation characteristics of the seismic source and the transfer function of the recording instrument, the measured seismograms contain the signature of earth structure through which the seismic waves have travelled from source to receiver. In terms of *cause and action*, the specific structure of material properties inside the earth, as well as the source characteristics, cause the seismic data to have the measured values, neglecting any instrument influences which may be sufficiently known. The *forward problem* consists of predicting the effect from the cause, i.e. predicting the seismogram from known source characteristics and elastic material properties. The *inverse problem*, on the other hand, is to find the cause of a known effect. That is, the geophysical inverse problem, in this case, consists of inferring earth structure and source characteristics from measured seismic data.

Geophysical inverse problems are in general *ill-posed*, i.e. they violate one of the following conditions after Hadamard (Engl et al., 2000, p. 31):

1. For all admissible data, a solution exists.

2. For all admissible data, the solution is unique.

3. The solution depends continuously on the data.

Usually, conditions 2. and 3. are violated, which means in general solutions are non-unique and that errors in the data from recording instruments or ambient noise sources produce instable solutions. These issues have to be handled by, so called, *regularizing* the inverse problem.

2.1 The geophysical inverse problem and general solution strategy

Following Parker (1994) and its notation, the collection of a finite set of geophysical data d_i, $i = 1, \ldots, N$, can be interpreted as the solution of the forward problem defined by N functionals F_i mapping the true earth model $\hat{m} \in V$ from some linear vector space V to $d_i \in \mathbb{R}$, i.e.

$$d_i = F_i\left[\hat{m}\right], \quad i = 1, \ldots, N.$$

Thereby, the model space V is assumed to be equipped with an inner product and normally is of infinite dimension. Its elements are vector-valued functions $\mathbf{f}(\mathbf{s})$ of position $\mathbf{s} \in D \subset \mathbb{R}^3$ and a standard choice of an inner product on V is defined by

$$(\mathbf{f}, \mathbf{g}) = \int_D \mathbf{f}(\mathbf{s}) \cdot \mathbf{g}(\mathbf{s}) \, \mathrm{d}^3 \mathbf{s}$$

(Parker, 1994, eq. 2.02(8)). In case of all functionals F_i being *linear*, the inverse problem, i.e. determining $\hat{m}$ given F_i and d_i, is called a *linear inverse problem*, otherwise it is called a *non-linear inverse problem*.

Let the data predictions for some given earth model m be denoted by $s_i = F_i[m]$. If the measured data values d_i cannot be taken as exact but obey certain uncertainties $\sigma_i > 0$, a measure of fitness is required quantifying whether a model m satisfactorily predicts the data within their errors. Regarding predictions s_i and data d_i as vectors $\mathbf{s}, \mathbf{d} \in \mathrm{E}^N$ from some N-dimensional normed space $\left(\mathrm{E}^N, \|\cdot\|\right)$, such a measure of fit may be defined by the misfit functional

$$E[m] = \left\| \Sigma^{-1} \left(\mathbf{d} - \mathbf{s}\right) \right\|^2, \tag{2.1}$$

where $\Sigma = \mathrm{Diag}\left(\sigma_1, \ldots, \sigma_N\right)$. In order to find a model explaining the data well, however, it is not desirable to simply minimize $E[m]$. This might for example explain uncertainties in the data due to measurement errors by earth structure. Instead, a minimization of the squared model norm $\|m\|^2 = (m, m)$ is sought subject to the condition that misfit $E[m]$ reaches an acceptable tolerance. Assuming the common 2-norm on E^N, this problem may be formulated as minimizing the Lagrange functional

$$E_L[m] = (m, m) + \nu \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(d_i - s_i\right)^2 \tag{2.2}$$

using an additional Lagrange multiplyer $\nu > 0$ (Parker, 1994, eq. 3.02(26)). ν may be interpreted as a *trade-off* parameter, balancing out the two undesirable properties of a model m, namely a large norm and a bad fit of data.

2.2 Seismic full waveform inversion

In seismic full waveform inversion, the solutions $s_i = F_i[m]$ of the forward problem constitute complete seismic waveform information. That is, the s_i sample the dis-

placement field $\mathbf{u}(\mathbf{r}, t)$ at some receiver position $\mathbf{r}$, where $\mathbf{u}$ solves the elastic wave equation

$$\rho \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} - \frac{\partial}{\partial x_j} \left(c_{ijkl} \frac{\partial u_k(\mathbf{x}, t)}{\partial x_l} \right) = f_i(\mathbf{x}, t)$$

for given earth model $m = (\rho, c_{ijkl})$ (subject to certain boundary conditions). Here, $\mathbf{f}(\mathbf{x}, t)$ is the exciting force field and Einstein summation rules apply. The evaluation of any such $F_i[m]$ is non-linear, hence is the inverse problem of determining a model $\hat{m}$ which sufficiently predicts the full waveform data $d_i = F_i[\hat{m}]$. This non-linearity is usually dealt with by linearization techniques or Monte Carlo approaches. The most common linearized and non-linear methods for minimizing either eq. (2.1) or eq. (2.2) (or something motivated by those) are shortly presented in the following, leading to the linearized full waveform inversion procedure based on explicit waveform sensitivity kernels which is developed in this thesis.

2.2.1 Probabilistic non-linear methods

Monte Carlo methods of seismic inversion, or sometimes referred to as "probabilistic tomography" (Resovsky and Trampert, 2003) in general follow the approach of statistically sampling the model space V in order to find sets of models $\hat{m}$ which sufficiently explain the data d_i. General introductions of such methods can be found in Tarantola (1987); Sambridge and Mosegaard (2002). For any chosen model $m_k \in$ V the forward problem is solved, i.e. the data predictions $s_{ik} = F_i[m_k]$, $i = 1, \ldots, N$, are determined with which some misfit functional is evaluated for model m_k, given the measured data values d_i. On the basis of the values of that misfit functional for the samples $m_k \in$ V, certain statistical strategies may be followed to seek a model that best explains the data within their errors. Such a recent probabilistic technique for highly non-linear waveform inversion is called Neighbourhood Algorithm (Sambridge, 1999a,b).

The most severe disadvantage of Monte Carlo methods, however, is to cope with V in general being of infinite dimension. Obviously, searching a model space of high (finite) dimension already demands solving a tremendous amount of forward problems. Dependent on the costs of evaluating functionals F_i, Monte Carlo methods may not yet be feasible in practice, especially for complex earth models (compare Curtis and Lomax, 2001).

2.2.2 Deterministic iterative optimization technique: the adjoint method

A different approach to invert model $\hat{m}$ which sufficiently explains the data is a deterministic search. Starting with some approximation m_0 to the true model $\hat{m}$, a sequence of models $m_1, m_2, \ldots, m_k, \ldots$ is derived which improve the goodness of fit of the data. That is, the models gradually minimize the objective function, e.g. a Lagrange functional, or even solely the misfit.

Some methods utilize the gradient of the objective function to determine a direction

in the model space in which a model that improves the data fit can be found. Early steepest-descent methods (Bamberger et al., 1982; Gauthier et al., 1986) straightly following the gradient direction in general exhibit slow convergence properties. That is why more elaborate pre-conditioned conjugate-gradient schemes (Mora, 1987; Tarantola, 1987; Tromp et al., 2005) are preferred in three-dimensional elastic full waveform inversion (Fichtner et al., 2009; Butzer et al., 2013). These methods iteratively improve the model in direction of the conjugate-gradient for which additionally the gradient of the last iteration is accounted for.

The gradient of the objective function can be explicitly derived by adjoint methods: First, the forward problem $s_i = F_i[m_{k-1}]$ is solved. In addition to the computation of synthetic seismograms, however, the wavefield excited by the seismic source is evaluated throughout the inversion domain. Some methods store the required wavefield information throughout the medium (Butzer et al., 2013). Other time-stepping methods memorize the wavefield at the final time as well as any wavefield history on absorbing boundaries, in order to reconstruct the required wavefield information in the second step of the procedure (Liu and Tromp, 2006). In this second step, the determined data residuals $d_i - s_i$ are backpropagated in time into the inversion domain, simultaneously from all receivers. The resulting wavefield is called *adjoint wavefield*. The gradient of the objective function, then, is derived from combination of that adjoint wavefield with the incident wavefield excited by the seismic source, which was either stored in the previous simulation or is reconstructed on-the-fly in the adjoint simulation. The new model m_k which improves the fit of the data is generally found conducting a line-search in the model space along the conjugate-gradient direction. That is, the distance in the model space along the direction of the conjugate-gradient is determined in which a model can be found that reduces the objective function best. This distance is found solving $s_i = F_i[m]$ for a few models m in direction of the conjugate-gradient and evaluating the objective function accordingly for each model m. Based on a simple interpolation of the objective function, e.g. by a parabola, that model is determined for which the interpolating function has a sensible minimum. For stability reasons, the conjugate-gradient must be pre-conditioned, i.e. artificially suppressed, especially in the vicinity of sources, receivers and free surfaces. This is because the gradient usually exhibits very large values there. Since the backpropagation of the data residuals are done simultaneously from all receivers, their number is of minor significance to the computational effort of adjoint methods. The number of simulations scales solely with a multiple of the number of involved sources.

Better convergence in the minimization of a misfit or Lagrange functional is assured using Gauss-Newton methods (e.g. Akcelik et al., 2002; Epanomeritakis et al., 2008) or Newton-like methods involving full Hessians or approximations of it (Liu and Nocedal, 1989; Pratt et al., 1998; Fichtner and Trampert, 2011), though such methods are computationally more demanding. Full Newton methods (Santosa and Symes, 1988) are even more costly which may be the reason why such methods are not yet widespread in applications to large problems.

2.3 Iterative linearized waveform inversion based on explicit waveform sensitivity kernels

Yet another approach of solving the inverse problem as presented in section 2.1 is to use the Fréchet derivatives of the data functionals F_i. In case of linear inverse problems, i.e. for linear functionals F_i, these can be explicitly computed, as presented in the following. In order to solve the non-linear seismic inverse problem, such Fréchet derivatives are usually applied in linearized iterative procedures.

By the Riesz representation theorem, the linear functionals F_i can be expressed in terms of building the inner product with a representing element $g_i \in V$

$$F_i[m] = (g_i, m)$$

(Parker, 1994, eq. 2.03(1)). The g_i represent the data d_i in the model space and are called *kernels*. Note that kernel g_i is identical with the Fréchet derivative of linear data functional F_i:

$$F_i[m + h] = (g_i, m + h) = F_i[m] + (g_i, h) \quad, m, h \in V.$$

Therefore the g_i are also called *sensitivity kernels*, quantifying the change of prediction $s_i = F_i[m]$ w.r.t. change in the model space. Inversely speaking, one may ask how the model must be changed in order to achieve a certain change in the data. This relation is taken advantage of when using sensitivity kernels to solve geophysical inverse problems. For seismic inversion, the sensitivity kernels g_i are in practice derived from first order Born scattering, i.e. single scattering theory. They depend on the data functionals F_i, which define the type of seismic observable (e.g. travel time delays, amplitude anomalies, ground motion, etc.). Since for the seismic inverse problem, however, the functionals F_i are not linear, kernels g_i^0 are computed in a *linearized* fashion and are only valid in the neighbourhood $U(m^0)$ of a chosen background model m^0, i.e.

$$F_i[m^0 + h] \simeq F_i[m^0] + \left(g_i^0, h\right) \quad, h \in U(m^0).$$

Hence, for particular types of data and background models m^0, specific kinds of sensitivity kernels can be derived: For laterally homogeneous background earth models, for example, Marquering et al. (1998); Friederich (1999) compute sensitivity kernels of shear and surface wave phases, whereas Nissen-Meyer et al. (2007); Zhao and Chevrot (2011) derive sensitivity kernels for full waveforms. Zhao et al. (2005) apply sensitivity kernels for travel time and amplitude anomalies relative to 3D reference models on regional scale. Chen et al. (2007b) use kernels of phase-delay anomalies in full 3D tomography on the Los Angeles Region.

In this thesis, sensitivity kernels of full waveform data functionals are derived from Born scattering theory. Based on those *waveform sensitivity kernels* which can be computed for arbitrary 3D-heterogeneous background models, an iterative full waveform inversion procedure is developed, validated and discussed. Note that waveform sensitivity kernels constitute the most general sensitivity information available for any

seismic observable related to waveforms. That is, the above mentioned sensitivity kernels of travel time and amplitude anomalies etc. can be derived from the waveform sensitivity kernels (e.g. Zhao and Chevrot, 2011).

2.3.1 Iterative inversion procedure

This section 2.3.1 introduces the inversion procedure which is derived throughout the next chapters. The overall non-linear problem of seismic full waveform inversion which is solved by this procedure, states:

> Given a finite set of measured data d_i, $i = 1, \ldots, N$, which represent complete seismic waveforms, find an earth model $\hat{m} \in V$ which adequately predicts the data d_i. $\hspace{2em}$ (A)

Here, the data samples d_i as well as their corresponding data functionals F_i are assumed to represent *spectral* waveform data. Reasons as to why the frequency domain is chosen for inversion are given below in section 3.3.1. A solution to this overall non-linear inverse problem (A) is found by an iterative procedure of solving linearized inverse problems, which is presented schematically in the following in the style of Parker (1994).

Similar to the adjoint methods in section 2.2.2, the iterative procedure derives a sequence of models $m_1, m_2, \ldots, m_k, \ldots$ approximating the desired model $\hat{m}$ based on a given starting model m_0. The choice of the starting model can significantly influence the convergence of the process. However, any criteria on adequately choosing m_0 are not discussed in this work.

In each iteration k, only a data subset, i.e. d_i, $i \in D_k \subset \{1, \ldots, N\}$, is inverted associated with a specific frequency window. For stability of the inversion procedure, the first iteration steps should invert the low frequency content of the data and higher frequencies should be added gradually in later iterations.

Then, for all respective data functionals F_i, $i \in D_k$, sensitivity kernels g_i^{k-1}, $i \in D_k$, are computed explicitly in a linearized sense, valid within a neighbourhood $U(m_{k-1})$ of previously derived model m_{k-1}, i.e.

$$F_i[m_{k-1} + h] \simeq F_i[m_{k-1}] + (g_i^{k-1}, h) \ , \ h \in U(m_{k-1}). \hspace{2em} (2.3)$$

The inversion step of iteration k is motivated by solving the overall inverse problem (A), but for the *subset* of data D_k only. That is, the new model m_k should minimize Langrange functional (2.2) summing only over $i \in D_k$. Writing $m_k = m_{k-1} + m_k^\Delta$ and assuming the linearization (2.3) (i.e. $m_k^\Delta \in U(m_{k-1})$) leads to

$$\begin{aligned} E_L[m_k] &= (m_k, m_k) + \nu \sum_{i \in D_k} \frac{1}{\sigma_i^2} \left(d_i - F_i[m_k]\right)^2 \\ &= (m_k, m_k) + \nu \sum_{i \in D_k} \frac{1}{\sigma_i^2} \left(d_i - F_i[m_{k-1} + m_k^\Delta]\right)^2 \end{aligned}$$

$$= (m_k, m_k) + \nu \sum_{i \in D_k} \frac{1}{\sigma_i^2} \left(d_i - F_i[m_{k-1}] - (g_i^{k-1}, m_k^\Delta) \right)^2$$

$$= (m_k, m_k) + \nu \sum_{i \in D_k} \frac{1}{\sigma_i^2} \left(r_i - (g_i^{k-1}, m_k^\Delta) \right)^2 \quad ,$$

where $r_i = d_i - F_i[m_{k-1}]$ denotes the data residuals in iteration k. (m_k, m_k) is solely dependent on m_k^Δ knowing m_{k-1} and, thus, can be expressed in terms of some regularization operator $(m_k, m_k) = \mathcal{R}(m_k^\Delta)$. Hence, minimizing $E_L[m_k]$ for data subset D_k is equivalent to the minimization

$$\min_{\mathbf{m}_k^\Delta} \left(\mathcal{R}(m_k^\Delta) + \nu \sum_{i \in D_k} \frac{1}{\sigma_i^2} \left(r_i - (g_i^{k-1}, m_k^\Delta) \right)^2 \right) \quad , \tag{2.4}$$

which is actually solved in each iteration k of this inversion procedure. Here a regularization operator $\mathcal{R}$ is used which smoothes the model update m_k^Δ (cf. section 4.2.3). In addition to the minimization of the model norm, namely, Parker (1994, beginning of section 3.05) suggests different kinds of regularization, e.g. minimizing the first or second derivative of the model values in order to smooth the model.

The kernels g_i^{k-1} are solely dependent on model m_{k-1} and predictions $F_i[m_{k-1}]$. For their computation, forward wavefields and strains for every involved seismic source must be propagated into and stored throughout the medium of interest. Similarly, backpropagations from the receiver into the medium are required which by contrast to adjoint methods, however, are realized independently of any data values d_i in form of Green funtions, i.e. impulse responses of the medium. Also, no form of pre-conditioning of the kernels is required here.

2.3.2 Splitting up the inversion process

The independence of the backpropagations from any data values d_i allows to reuse the Green function associated with a receiver position $\mathbf{r}$ (as a form of universal back-propagation) in combination with every forward wavefield from a seismic source that was recorded at $\mathbf{r}$. Thus independent of each other, one simulation per receiver component and one for each source is required. Depending on the number of involved sources and receiver components, this can lead to a lower overall number of required wave propagation simulations compared with adjoint methods (cf. Chen et al., 2007a). The computational efforts of using waveform sensitivity kernels compared with adjoint methods are discussed in section 4.3.4. The advantage of reusing the wavefields, however, requires to store them throughout the inversion domain. Hence, this motivates to separate the required operations in the inversion procedure using waveform sensitivity kernels, which is shortly discussed here.

The solution of the forward problem in a certain iteration of the inversion procedure, i.e. the simulation of wave propagation, is kept separate from the other operations of computing kernels and updating the model. This allows to use different forward

methods, satisfying the requirements of the particular inverse problem to solve. It is even possible to change the kind of forward method from one iteration to the next, e.g. starting with a 1D method if the starting model m_0 is laterally homogeneous, continuing with full 3D methods, etc. Any modularized implementation of the inversion procedure can also be flexibly extended to other forward methods, provided well defined interfaces between the modules are respected.

As well as the wavefields as solutions of the forward problem must be stored on hard disc, it is feasible to store the sensitivity kernels to disc, too. This allows any sensitivity and resolution analysis along with (or separate from) conducting the inversion step in an iteration. Even a full singular value decomposition on the sensitivity matrix may be examined, allowing valuable insight to the resolving power of the chosen data subset. As explained in the following chapters, the amount of storage required for the kernels is relatively small compared with the wavefields due to the combination of wavefield and strain components to sensitivity values for relatively few model parameters and pre-integration onto the inversion domain.

Finally, keeping the inversion step separate in the iterative inversion process allows to repeat solving the minimization (2.4) discarding certain data from D_k or applying different regularization operators $\mathcal{R}$. Section 4.3.1 points out that this is relatively cheap compared with solving the forward problem.

Furthermore, the concept of separation improves the overall computational efficiency of the solution procedure: The forward modelling of the data may be optimally adapted to its frequency content used in a specific iteration. Modelling data of high frequency content, namely, is in general more expensive compared with modelling low frequency data. Also the resolution of the inverted model can be iteratively refined in the course of the inversion process, along with the increase in frequency content. As a result, the number of unknowns can be kept small in the early iterations, hence the computational costs for solving the inverse problem. This way, the whole process of full waveform inversion becomes very flexible and allows for complete control by the scientist in every iteration of the procedure.

3 Waveform kernels and sensitivity analysis

This chapter derives the waveform sensitivity kernels used in the full waveform inversion procedure presented in this thesis. Starting off with Born scattering theory in frequency domain, basic sensitivity relations are derived assuming arbitrary elastic symmetries. The application of those general relations to the specific case of isotropic elasticity yields explicit formulas of waveform sensitivity kernels for such parameterizations. Issues arising in the computation of the kernels in the frequency domain are addressed. Exemplary kernels and potential applications of sensitivity analysis using the kernels are presented thereafter.

Einstein summation convention is assumed in sections 3.1 and 3.2.

3.1 Born scattering theory

The relations in this section 3.1 were derived following unpublished lecture notes by W. Friederich (2006). Subscript indices refer to orthogonal directions in Cartesian space.

3.1.1 The equation of dynamic elasticity

In the frequency domain the elastodynamic equation of motion reads

$$-\rho\omega^2 u_i - \partial_j \left(c_{ijkl} \partial_l u_k \right) = f_i \quad ,$$

where $\mathbf{u}$ is the displacement vector field which is excited by a body force field $\mathbf{f}$ at angular frequency ω in a medium having density ρ and tensor of linear elasticity c. ∂_j denotes the spatial partial derivative in direction j (refer to Aki and Richards, 1980, eq. (2.17) for a derivation in the time domain). $\mathbf{u}$ additionally satisfies the following boundary conditions on the union of all inner fluid-solid boundaries Σ_{FS} and solid-solid boundaries Σ_{SS}

$$[\mathbf{n} \cdot \mathbf{u}]^+_- = 0 \text{ on } \Sigma_{FS},$$
$$[\mathbf{u}]^+_- = 0 \text{ on } \Sigma_{SS}, \tag{3.1}$$
$$[\mathbf{n} \cdot \mathbf{T}(\mathbf{u})]^+_- = 0 \text{ on } \Sigma_{FS} \text{ and } \Sigma_{SS},$$

where $[]_-^+$ denotes the jump condition on the respective surface which has normal vector field $\mathbf{n}$ and $\mathbf{T}(\mathbf{u}) = c_{ijkl}\partial_l u_k$ is the stress tensor associated with displacement field $\mathbf{u}$. Free surface conditions are respected, assuming zero traction above free surfaces.

3.1.2 Betti's theorem

Betti's theorem is one of the reciprocity theorems of the elastodynamic equation and is used to derive an expression for the first order scattered wavefield. Let the elastodynamic differential operator be denoted by $\mathbf{L}$, with $\mathbf{L}(\mathbf{u}) = \mathbf{f}$, i.e.

$$L_i(\mathbf{u}) = -\rho\omega^2 u_i - \partial_j\left(c_{ijkl}\partial_l u_k\right) \quad .$$

Let, furthermore, $\mathbf{u}$ and $\mathbf{v}$ be arbitrary displacement fields given in a possibly disconnected volume Ω which has boundaries Σ with normal vector field $\mathbf{n}$. Then, by partial integration (or rather Gauss's theorem) and product rule of derivatives,

$$\int_\Omega \mathbf{v}\cdot\mathbf{L}(\mathbf{u})\,\mathrm{d}^3\mathrm{V} = \int_\Omega v_i L_i(\mathbf{u})\,\mathrm{d}^3\mathrm{V}$$

$$= \int_\Omega \left[-\rho\omega^2 v_i u_i - v_i\partial_j\left(c_{ijkl}\partial_l u_k\right)\right]\,\mathrm{d}^3\mathrm{V}$$

$$= \int_\Omega \left[-\rho\omega^2 v_i u_i + \left(\partial_j v_i\right)c_{ijkl}\partial_l u_k\right]\,\mathrm{d}^3\mathrm{V} - \int_\Sigma \left(v_i c_{ijkl}\partial_l u_k\right)n_j\,\mathrm{d}^2\mathrm{S}$$

$$= \int_\Omega \left[-\rho\omega^2 v_i u_i + \partial_l\left(u_k c_{ijkl}\partial_j v_i\right) - u_k\partial_l\left(c_{ijkl}\partial_j v_i\right)\right]\,\mathrm{d}^3\mathrm{V} - \int_\Sigma \left(v_i c_{ijkl}\partial_l u_k\right)n_j\,\mathrm{d}^2\mathrm{S}$$

$$= \int_\Omega \left[-\rho\omega^2 v_i u_i - u_k\partial_l\left(c_{ijkl}\partial_j v_i\right)\right]\,\mathrm{d}^3\mathrm{V} + \int_\Sigma \left[\left(u_k c_{ijkl}\partial_j v_i\right)n_l - \left(v_i c_{ijkl}\partial_l u_k\right)n_j\right]\,\mathrm{d}^2\mathrm{S} \quad .$$

Using the symmetry relations in eq. (3.12) of the elasticity tensor, index pairs ij and kl in the volume and first surface integral can be interchanged, yielding

$$\int_\Omega \mathbf{v}\cdot\mathbf{L}(\mathbf{u})\,\mathrm{d}^3\mathrm{V}$$

$$= \int_\Omega \left[-\rho\omega^2 v_i u_i - u_k\partial_l\left(c_{klij}\partial_j v_i\right)\right]\,\mathrm{d}^3\mathrm{V} + \int_\Sigma \left[\left(u_k c_{klij}\partial_j v_i\right)n_l - \left(v_i c_{ijkl}\partial_l u_k\right)n_j\right]\,\mathrm{d}^2\mathrm{S} \quad .$$

Finally, renaming in the same two expressions the index pair (k,i) as (i,k) and (l,j) as (j,l) leads to

$$\int_\Omega \mathbf{v}\cdot\mathbf{L}(\mathbf{u})\,\mathrm{d}^3\mathrm{V}$$

$$= \int_\Omega \left[-\rho\omega^2 u_i v_i - u_i\partial_j\left(c_{ijkl}\partial_l v_k\right)\right]\,\mathrm{d}^3\mathrm{V} + \int_\Sigma \left[\left(u_i c_{ijkl}\partial_l v_k\right)n_j - \left(v_i c_{ijkl}\partial_l u_k\right)n_j\right]\,\mathrm{d}^2\mathrm{S}$$

$$= \int_\Omega \mathbf{u}\cdot\mathbf{L}(\mathbf{v})\,\mathrm{d}^3\mathrm{V} + \int_\Sigma \left[\mathbf{u}\cdot\mathbf{T}(\mathbf{v}) - \mathbf{v}\cdot\mathbf{T}(\mathbf{u})\right]\cdot\mathbf{n}\,\mathrm{d}^2\mathrm{S} \tag{3.2}$$

which is called Betti's theorem. Note the particular symmetry property in terms of the elastodynamic differential operator.

3.1.3 Perturbed equation of motion and perturbed boundary conditions

In Born scattering formulations, scattered waves are produced by slight perturbations from a reference earth model, i.e. changes of density and elastic constants in volume Ω as well as perturbations of the boundaries Σ. The properties of the reference model will always be referred to by a superscript 0, e.g. ρ^0, c_{ijkl}^0, $\mathbf{u}^0$, etc. and perturbations by a superscript Δ, i.e. ρ^Δ, c_{ijkl}^Δ, $\mathbf{u}^\Delta$, etc. This means that the perturbed properties (indicated by superscript 1) may be written as

$$\rho^1 = \rho^0 + \rho^\Delta$$
$$c_{ijkl}^1 = c_{ijkl}^0 + c_{ijkl}^\Delta$$
$$\mathbf{u}^1 = \mathbf{u}^0 + \mathbf{u}^\Delta$$

and the perturbed boundaries are denoted by Σ_{FS}^1 and Σ_{SS}^1. Applying these notations of ρ and c_{ijkl} to the differential operator $\mathbf{L}$ should give intuitive meaning to $\mathbf{L}^0$, $\mathbf{L}^\Delta$ and $\mathbf{L}^1$. Thus, for any body force $\mathbf{f}$ exciting displacement field $\mathbf{u}^0$ in the reference medium the perturbed displacement field $\mathbf{u}^1$ satisfies

$$L_i^1(\mathbf{u}^1) = -\rho^1 \omega^2 u_i^1 - \partial_j \left(c_{ijkl}^1 \partial_l u_k^1 \right) = f_i$$

and in addition fulfils boundary conditions similar to eq. (3.1) on the perturbed boundaries Σ_{FS}^1 and Σ_{SS}^1. The new stress tensors w.r.t. the reference and the perturbed model associated with some displacement field $\mathbf{v}$ now read $\mathbf{T}^0(\mathbf{v}) = c_{ijkl}^0 \partial_l v_k$ and $\mathbf{T}^1(\mathbf{v}) = c_{ijkl}^1 \partial_l v_k$, respectively.

Born scattering means single scattering, i.e. scattered waves are neglected to be scattered again by model perturbations. Neglecting any terms of higher than first order, such as $\mathbf{L}^\Delta \left(\mathbf{u}^\Delta \right)$, and observing that by definition $\mathbf{L}^0 \left(\mathbf{u}^0 \right) = \mathbf{L}^1 \left(\mathbf{u}^1 \right) = \mathbf{f}$, it follows

$$
\begin{aligned}
\mathbf{L}^0 \left(\mathbf{u}^\Delta \right) &= \mathbf{L}^0 \left(\mathbf{u}^1 \right) - \mathbf{L}^0 \left(\mathbf{u}^0 \right) \\
&= \mathbf{L}^0 \left(\mathbf{u}^1 \right) - \mathbf{f} \\
&= \mathbf{L}^0 \left(\mathbf{u}^1 \right) - \mathbf{L}^1 \left(\mathbf{u}^1 \right) \\
&= -\mathbf{L}^\Delta \left(\mathbf{u}^1 \right) \\
&= -\mathbf{L}^\Delta \left(\mathbf{u}^0 \right) - \mathbf{L}^\Delta \left(\mathbf{u}^\Delta \right) \\
&\simeq -\mathbf{L}^\Delta \left(\mathbf{u}^0 \right) \quad .
\end{aligned}
\tag{3.3}
$$

Furthermore, the stress tensor changes by

$$
\begin{aligned}
\mathbf{T}^\Delta &= \mathbf{T}^1 \left(\mathbf{u}^1 \right) - \mathbf{T}^0 \left(\mathbf{u}^0 \right) \\
&= c_{ijkl}^1 \partial_l u_k^1 - c_{ijkl}^0 \partial_l u_k^0 \\
&= \left(c_{ijkl}^0 + c_{ijkl}^\Delta \right) \partial_l \left(u_k^0 + u_k^\Delta \right) - c_{ijkl}^0 \partial_l u_k^0 \\
&= c_{ijkl}^0 \partial_l u_k^\Delta + c_{ijkl}^\Delta \partial_l u_k^0 + c_{ijkl}^\Delta \partial_l u_k^\Delta \\
&\simeq c_{ijkl}^0 \partial_l u_k^\Delta + c_{ijkl}^\Delta \partial_l u_k^0 \quad .
\end{aligned}
\tag{3.4}
$$

3.1.4 The scattered displacement field

The aim of this section is to derive an expression for the scattering of a seismic displacement field $\mathbf{u}$ excited by some force, caused by model perturbations. This can be achieved by applying Betti's theorem in the reference model to a special choice of displacement fields: $\mathbf{u}$ is chosen as $\mathbf{u}^{\Delta}$ and $\mathbf{v}$ as $\mathbf{g}_n^0(\mathbf{r})$, which is a short vector notation of the Green function in the reference model for a single force in direction n at $\mathbf{r}$. The Green function $G_{in}^0(\mathbf{x}, \mathbf{r}) = \left(\mathbf{g}_n^0(\mathbf{r})\right)_i$ satisfies

$$- \rho^0 \omega^2 G_{in}^0(\mathbf{x}, \mathbf{r}) - \partial_j c_{ijkl}^0 \partial_l G_{kn}^0(\mathbf{x}, \mathbf{r}) = \delta_{in} \delta(\mathbf{x} - \mathbf{r}) \quad , \tag{3.5}$$

subject to boundary conditions of eq. (3.1) on the unperturbed boundaries. Note that the exciting force $\delta_{in}\delta(\mathbf{x} - \mathbf{r})$ is located at a single point in space, namely $\mathbf{r}$, since δ denotes the Dirac delta distribution. δ_{in} is the Kronecker delta. The force is also located at the single point $t = 0$ in time, which however is not obvious in the frequency domain formulation of eq. (3.5). Force term $\delta_{in}\delta(\mathbf{x} - \mathbf{r})$, namely, is implicitly multiplied by constant $1 \in \mathbb{R}$, which is the spectrum of the source time function $\delta(t)$.

Making this choice, Betti's theorem reads

$$\int_{\Omega} \mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{L}^0\left(\mathbf{u}^{\Delta}\right) d^3\mathbf{x} = \int_{\Omega} \mathbf{u}^{\Delta} \cdot \mathbf{L}^0\left(\mathbf{g}_n^0(\mathbf{r})\right) d^3\mathbf{x}$$
$$+ \int_{\Sigma^0} \left[\mathbf{u}^{\Delta} \cdot \mathbf{T}^0\left(\mathbf{g}_n^0(\mathbf{r})\right) - \mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^0(\mathbf{u}^{\Delta})\right] \cdot \mathbf{n}^0 d^2\mathbf{S} \quad ,$$

where Σ^0 is the union of all unperturbed boundaries with normal vector field $\mathbf{n}^0$. Note that if some boundaries subdivide Ω into disjoint regions, the integration on Σ^0 runs over both sides of those boundaries. Accounting for $\mathbf{L}^0\left(\mathbf{u}^{\Delta}\right) = -\mathbf{L}^{\Delta}\left(\mathbf{u}^0\right)$ (eq. (3.3)) and $L_i^0\left(\mathbf{g}_n^0(\mathbf{r})\right) = \delta_{in}\delta(\mathbf{x} - \mathbf{r})$ (eq. (3.5)) the above equation yields

$$- \int_{\Omega} G_{in}^0(\mathbf{x}, \mathbf{r}) L_i^{\Delta}\left(\mathbf{u}^0\right) d^3\mathbf{x} = \int_{\Omega} u_i^{\Delta}(\mathbf{x}) \delta_{in} \delta(\mathbf{x} - \mathbf{r}) d^3\mathbf{x}$$
$$+ \int_{\Sigma^0} \left[\mathbf{u}^{\Delta} \cdot \mathbf{T}^0\left(\mathbf{g}_n^0(\mathbf{r})\right) - \mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^0(\mathbf{u}^{\Delta})\right] \cdot \mathbf{n}^0 d^2\mathbf{S} \quad .$$

Integration over $\delta(\mathbf{x} - \mathbf{r})$ extracts the integrand at $\mathbf{r}$, resulting in

$$u_n^{\Delta}(\mathbf{r}) = - \int_{\Omega} G_{in}^0(\mathbf{x}, \mathbf{r}) L_i^{\Delta}\left(\mathbf{u}^0\right) d^3\mathbf{x}$$
$$- \int_{\Sigma^0} \left[\mathbf{u}^{\Delta} \cdot \mathbf{T}^0\left(\mathbf{g}_n^0(\mathbf{r})\right) - \mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^0(\mathbf{u}^{\Delta})\right] \cdot \mathbf{n}^0 d^2\mathbf{S} \quad . \tag{3.6}$$

This already is a first expression for the scattered wavefield $\mathbf{u}^{\Delta}$ which in the following is derived further in a more explicit way. The volume integral may be expanded using

Gauss's theorem as

$$
-\int_{\Omega} G_{in}^0(\mathbf{x}, \mathbf{r}) L_i^{\Delta}(\mathbf{u}^0) \, \mathrm{d}^3\mathbf{x}
$$

$$
= \int_{\Omega} G_{in}^0(\mathbf{x}, \mathbf{r}) \left(\rho^{\Delta} \omega^2 u_i^0(\mathbf{x}) + \partial_j \left(c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right) \right) \mathrm{d}^3\mathbf{x}
$$

$$
= \int_{\Omega} \left[G_{in}^0(\mathbf{x}, \mathbf{r}) \rho^{\Delta} \omega^2 u_i^0(\mathbf{x}) - \left(\partial_j G_{in}^0(\mathbf{x}, \mathbf{r}) \right) c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right] \mathrm{d}^3\mathbf{x}
$$

$$
+ \int_{\Omega} \partial_j \left(G_{in}^0(\mathbf{x}, \mathbf{r}) c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right) \mathrm{d}^3\mathbf{x}
$$

$$
= \int_{\Omega} \left[G_{in}^0(\mathbf{x}, \mathbf{r}) \rho^{\Delta} \omega^2 u_i^0(\mathbf{x}) - \left(\partial_j G_{in}^0(\mathbf{x}, \mathbf{r}) \right) c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right] \mathrm{d}^3\mathbf{x}
$$

$$
+ \int_{\Sigma^0} \left[G_{in}^0(\mathbf{x}, \mathbf{r}) c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right] n_j^0 \, \mathrm{d}^2\mathrm{S} \quad . \tag{3.7}
$$

This adds another surface integral to eq. (3.6), which may be combined with the one containing $\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^0(\mathbf{u}^{\Delta})$:

$$
\int_{\Sigma^0} \left[G_{in}^0(\mathbf{x}, \mathbf{r}) c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) \right] n_j^0 \, \mathrm{d}^2\mathrm{S} + \int_{\Sigma^0} \left[\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^0(\mathbf{u}^{\Delta}) \right] \cdot \mathbf{n}^0 \, \mathrm{d}^2\mathrm{S}
$$

$$
= \int_{\Sigma^0} \left[G_{in}^0(\mathbf{x}, \mathbf{r}) \left(c_{ijkl}^{\Delta} \partial_l u_k^0(\mathbf{x}) + c_{ijkl}^0 \partial_l u_k^{\Delta}(\mathbf{x}) \right) \right] n_j^0 \, \mathrm{d}^2\mathrm{S}
$$

$$
= \int_{\Sigma^0} \left[\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^{\Delta} \right] \cdot \mathbf{n}^0 \, \mathrm{d}^2\mathrm{S} \quad . \tag{3.8}
$$

Incorporating eqs. (3.7) and (3.8) into eq. (3.6) finally leads to an explicit expression for the scattered wavefield $\mathbf{u}^{\Delta}(\mathbf{r})$ at location $\mathbf{r}$:

$$
u_n^{\Delta}(\mathbf{r}) = \int_{\Omega} \left[\omega^2 \rho^{\Delta} G_{in}^0(\mathbf{x}, \mathbf{r}) u_i^0(\mathbf{x}) - \left(\partial_j G_{in}^0(\mathbf{x}, \mathbf{r}) \right) c_{ijkl}^{\Delta} \partial_j u_k^0(\mathbf{x}) \right] \mathrm{d}^3\mathbf{x} \tag{3.9}
$$

$$
+ \int_{\Sigma^0} \left[\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^{\Delta} - \mathbf{u}^{\Delta} \cdot \mathbf{T}^0 \left(\mathbf{g}_n^0(\mathbf{r}) \right) \right] \cdot \mathbf{n}^0 \, \mathrm{d}^2\mathrm{S} \quad .
$$

The surface integral in eq. (3.9) provides terms for sensitivity kernels relating changes in the waveform to the boundary perturbations. Throughout this work, however, it is assumed that boundaries are *not* perturbed, i.e. $\Sigma_{FS}^1 = \Sigma_{FS}^0$ and $\Sigma_{SS}^1 = \Sigma_{SS}^0$, which results in the surface integral being zero, as argued in the following. Since the normal vector field always points outwards, the integrand of the surface integral can be rewritten in terms of jump conditions

$$
\left[\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^{\Delta} - \mathbf{u}^{\Delta} \cdot \mathbf{T}^0 \left(\mathbf{g}_n^0(\mathbf{r}) \right) \right] \cdot \mathbf{n}^0
$$

$$
= \left[\mathbf{g}_n^0(\mathbf{r}) \cdot \mathbf{T}^{\Delta} \cdot \mathbf{n}^0 \right]_-^+ - \left[\mathbf{u}^{\Delta} \cdot \mathbf{T}^0 \left(\mathbf{g}_n^0(\mathbf{r}) \right) \cdot \mathbf{n}^0 \right]_-^+
$$

$$
= \left[\mathbf{n}^0 \cdot \mathbf{T}^{\Delta} \cdot \mathbf{g}_n^0(\mathbf{r}) \right]_-^+ - \left[\mathbf{n}^0 \cdot \mathbf{T}^0 \left(\mathbf{g}_n^0(\mathbf{r}) \right) \cdot \mathbf{u}^{\Delta} \right]_-^+ \quad , \tag{3.10}
$$

assuming that in this case integration is done only over one of possible two sides of the boundaries and noticing that $\mathbf{T}^1$ and $\mathbf{T}^{\Delta}$ are symmetric. Furthermore, $\mathbf{u}^1$ satisfies

boundary conditions on the perturbed boundaries

$$\left[\mathbf{n}^1 \cdot \mathbf{u}^1\right]_-^+ = 0 \text{ on } \Sigma_{FS}^1,$$
$$\left[\mathbf{u}^1\right]_-^+ = 0 \text{ on } \Sigma_{SS}^1, \tag{3.11}$$
$$\left[\mathbf{n}^1 \cdot \mathbf{T}^1\left(\mathbf{u}^1\right)\right]_-^+ = 0 \text{ on } \Sigma_{FS}^1 \text{ and } \Sigma_{SS}^1 \quad,$$

by definition. Since, however, the perturbed boundaries are assumed to equal the unperturbed ones, $\mathbf{n}^1$ equals $\mathbf{n}^0$ everywhere and the perturbed properties $\mathbf{u}^1$ and $\mathbf{T}^1$ fulfil the above conditions (3.11) on the unperturbed boundaries. By subtracting boundary conditions for $\mathbf{u}^0$ and $\mathbf{T}^0$, those conditions also hold true for the perturbations $\mathbf{u}^\Delta$ and $\mathbf{T}^\Delta$, assuming the linearization in eq. (3.4) and taking the linearity of the boundary conditions into account. Observing that additionally $\mathbf{g}_n^0(\mathbf{r})$ satisfies eq. (3.1) on the unperturbed boundaries, it follows that the expressions inside the brackets in eq. (3.10) are continuous across the unperturbed boundaries. Hence their jump vanishes, and therefore does the surface integral in eq. (3.9) in case of unperturbed boundaries.

Although not always written, it should be kept in mind that model parameters ρ, c_{ijkl} are of course dependent on spatial variable $\mathbf{x}$ and eq. (3.9) holds true for the spectral wavefields evaluated at angular frequency ω.

Looking at eq. (3.9), note that $u_i^0(\mathbf{x})$ represents the forward propagation from the seismic source to the scattering point $\mathbf{x}$ and $G_{in}^0(\mathbf{x}, \mathbf{r})$ represents the backward propagation from the receiver $\mathbf{r}$ to scattering point $\mathbf{x}$ inside the medium. This kind of backpropagation, however, does not intuitively reflect the act of scattering. But involving the reciprocity property of Green functions $G_{in}^0(\mathbf{x}, \mathbf{r}) = G_{ni}^0(\mathbf{r}, \mathbf{x})$ (which can be derived from Betti's theorem), this backpropagation can be interpreted as scattered waves originating at scattering point $\mathbf{x}$, observed at receiver $\mathbf{r}$. However, the computation of Green functions for sources throughout the medium (at all potential scattering points $\mathbf{x}$) evaluated at just a few receiver positions $\mathbf{r}$ is not feasible from a numerical point of view. That is why the formulation of eq. (3.9) in terms of backpropagations $G_{in}^0(\mathbf{x}, \mathbf{r})$ from receivers into the medium is used in practical considerations.

3.2 Sensitivity kernels for scattered wavefields

In linear elasticity, Hook's law gives the general relationship between the Cauchy stress tensor $\mathbf{T}$ and the infinitesimal strain tensor ϵ through the elasticity tensor c by

$$T_{ij} = c_{ijkl}\epsilon_{kl} \quad,$$

(Landau and Lifschitz, 1966, eq. (10,3)). The symmetries of the stress and strain tensors along with the reversibility property of elastic energy imply certain symmetries of the elasticity tensor,

$$c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij} \tag{3.12}$$

which reduce the number of $3^4 = 81$ independent elastic constants to 21 (cf. Landau and Lifschitz, 1966, eq. (10,2)). The elasticity tensor, hence may by the well-established Voigt notation be rewritten as

$$c_{ijkl} = c_{\alpha\beta}$$

using a mapping for the tensor indices defined by

$$
\begin{array}{ccccccc}
ij & = & 11 & 22 & 33 & 23 & 13 & 12 \\
& & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\alpha & = & 1 & 2 & 3 & 4 & 5 & 6
\end{array}
\qquad (3.13)
$$

Assuming the tensor symmetries of eq. (3.12) and the map in eq. (3.13), as well as the strain notations

$$
\begin{aligned}
e_{ij} = e_{ji} &= \frac{1}{2}\left(\partial_j u_i^0\left(\mathbf{x}\right) + \partial_i u_j^0\left(\mathbf{x}\right)\right) \\
\gamma_{ij} = \gamma_{ji} &= \frac{1}{2}\left(\partial_j G_{in}^0\left(\mathbf{x},\mathbf{r}\right) + \partial_i G_{jn}^0\left(\mathbf{x},\mathbf{r}\right)\right) \quad,
\end{aligned}
\qquad (3.14)
$$

the following simplification of term $\left(\partial_j G_{in}^0\left(\mathbf{x},\mathbf{r}\right)\right) c_{ijkl}^\Delta \partial_j u_k^0\left(\mathbf{x}\right)$ in eq. (3.9) may be derived.

$$\left(\partial_j G_{in}^0\left(\mathbf{x},\mathbf{r}\right)\right) c_{ijkl}^\Delta \partial_j u_k^0\left(\mathbf{x}\right) =$$

$$c_{11}^\Delta\left(\gamma_{11}e_{11}\right) \;+\; c_{12}^\Delta\left(\gamma_{11}e_{22} + \gamma_{22}e_{11}\right) \;+\; c_{13}^\Delta\left(\gamma_{11}e_{33} + \gamma_{33}e_{11}\right) \;+$$
$$c_{14}^\Delta\left(2\gamma_{11}e_{23} + 2\gamma_{23}e_{11}\right) \;+\; c_{15}^\Delta\left(2\gamma_{11}e_{13} + 2\gamma_{13}e_{11}\right) \;+$$
$$c_{16}^\Delta\left(2\gamma_{11}e_{12} + 2\gamma_{12}e_{11}\right) \;+$$

$$c_{22}^\Delta\left(\gamma_{22}e_{22}\right) \;+\; c_{23}^\Delta\left(\gamma_{22}e_{33} + \gamma_{33}e_{22}\right) \;+\; c_{24}^\Delta\left(2\gamma_{22}e_{23} + 2\gamma_{23}e_{22}\right) \;+$$
$$c_{25}^\Delta\left(2\gamma_{22}e_{13} + 2\gamma_{13}e_{22}\right) \;+\; c_{26}^\Delta\left(2\gamma_{22}e_{12} + 2\gamma_{12}e_{22}\right) \;+$$

$$c_{33}^\Delta\left(\gamma_{33}e_{33}\right) \;+\; c_{34}^\Delta\left(2\gamma_{33}e_{23} + 2\gamma_{23}e_{33}\right) \;+\; c_{35}^\Delta\left(2\gamma_{33}e_{13} + 2\gamma_{13}e_{33}\right) \;+ \qquad (3.15)$$
$$c_{36}^\Delta\left(2\gamma_{33}e_{12} + 2\gamma_{12}e_{33}\right) \;+$$

$$c_{44}^\Delta\left(4\gamma_{23}e_{23}\right) \;+\; c_{45}^\Delta\left(4\gamma_{23}e_{13} + 4\gamma_{13}e_{23}\right) \;+\; c_{46}^\Delta\left(4\gamma_{23}e_{12} + 4\gamma_{12}e_{23}\right) \;+$$

$$c_{55}^\Delta\left(4\gamma_{13}e_{13}\right) \;+\; c_{56}^\Delta\left(4\gamma_{13}e_{12} + 4\gamma_{12}e_{13}\right) \;+$$

$$c_{66}^\Delta\left(4\gamma_{12}e_{12}\right)$$

The explicit formula eq. (3.15), relating perturbations in all 21 independent elastic constants to changes in the data, along with term $\omega^2 \rho^\Delta G_{in}^0\left(\mathbf{x},\mathbf{r}\right) u_i^0\left(\mathbf{x}\right)$ in eq. (3.9), permits the computation of waveform sensitivity kernels for arbitrary elastic model parameterizations. For parameterizations $P = \{p_1, \ldots, p_N\}$ with $p_N = \rho$ in which

parameters p_i, $1 \leq i \leq N - 1$, linearly depend on the elastic constants $c_{\alpha\beta}$, expressions K^p, $p \in P$, of waveform sensitivity kernels can be derived which fulfil

$$u_n^\Delta(\mathbf{r}) = \int_\Omega \left[\sum_{p \in P} p^\Delta(\mathbf{x}) K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r}) \right] \mathrm{d}^3\mathbf{x} \quad , \tag{3.16}$$

where kernel value $K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r})$ quantifies how sensitive component n of the displacement spectrum $\mathbf{u}$ excited by seismic source $\mathbf{s}$ recorded at receiver position $\mathbf{r}$ at angular frequency ω is with respect to changes of model parameter $p \in P$ at scattering point $\mathbf{x}$.

In the following, kernels for exemplary isotropic parameterizations are derived explicitly.

3.2.1 Kernels for isotropic parameterizations

For isotropic elastic parameterizations, the elasticity tensor c has only two independent parameters and may, e.g. with help of the Lamé moduli λ and μ, be written as

$$c_{ijkl} = \lambda \delta_{ij}\delta_{kl} + \mu \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} \right) \quad ,$$

(compare Landau and Lifschitz, 1966, eqs. (4,4), (4,6), (10,3)). This leads to

$$c_{11} = c_{22} = c_{33} = \lambda + 2\mu$$
$$c_{44} = c_{55} = c_{66} = \mu$$
$$c_{23} = c_{13} = c_{12} = \lambda$$

and $c_{\alpha\beta} = 0$ otherwise. In terms of Lamé moduli, eq. (3.15) hence becomes

$$\partial_j G_{in}^0(\mathbf{x}, \mathbf{r}) \, c_{ijkl}^\Delta \partial_j u_k^0(\mathbf{x}) =$$

$$\lambda^\Delta \left[(\gamma_{11} + \gamma_{22} + \gamma_{33})(e_{11} + e_{22} + e_{33}) \right] + \tag{3.17}$$

$$\mu^\Delta \left[2(\gamma_{11}e_{11} + \gamma_{22}e_{22} + \gamma_{33}e_{33}) + 4(\gamma_{23}e_{23} + \gamma_{13}e_{13} + \gamma_{12}e_{12}) \right] \quad ,$$

such that formulas for waveform sensitivity kernels of Lamé moduli and density in case of isotropic parameterization $P_{\text{Lamé}} = \{\lambda, \mu, \rho\}$ may be defined as

$$K_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) = -(\gamma_{11} + \gamma_{22} + \gamma_{33})(e_{11} + e_{22} + e_{33})$$

$$K_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) = -2(\gamma_{11}e_{11} + \gamma_{22}e_{22} + \gamma_{33}e_{33}) - 4(\gamma_{23}e_{23} + \gamma_{13}e_{13} + \gamma_{12}e_{12}) \tag{3.18}$$

$$K_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r}) = \omega^2 G_{in}^0(\mathbf{x}, \mathbf{r}) \, u_i^0(\mathbf{x}) \quad ,$$

with which eq. (3.16) becomes

$$u_n^\Delta(\mathbf{r}) = \int_\Omega \left[\lambda^\Delta K_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) + \mu^\Delta K_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) + \rho^\Delta K_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r}) \right] \mathrm{d}^3\mathbf{x} \quad . \tag{3.19}$$

In an analogous way, kernels for any isotropic parameterization derived from the 21 components of the elasticity tensor c may be derived. In seismic applications, the seismic velocities v_p and v_s are of central interest as physical material properties. It is well known that they are related to the Lamé moduli and density by

$$v_p = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad ,$$

$$v_s = \sqrt{\frac{\mu}{\rho}} \quad ,$$

implying

$$\lambda = \rho\, v_p^2 - 2\rho\, v_s^2 \quad ,$$

$$\mu = \rho\, v_s^2 \quad .$$

Linearizing these relations allows to derive kernels for the isotropic parameterization $P_{\mathrm{Veloc}} = \{v_p, v_s, \rho\}$ with which eq. (3.19) may be rewritten in an analogous linear fashion: Linearization by the total derivatives of λ and μ as functions of v_p, v_s and ρ leads to

$$\mathrm{d}\lambda = \frac{\partial\lambda}{\partial v_p}\,\mathrm{d}v_p + \frac{\partial\lambda}{\partial v_s}\,\mathrm{d}v_s + \frac{\partial\lambda}{\partial\rho}\,\mathrm{d}\rho$$

$$= 2\rho v_p\,\mathrm{d}v_p - 4\rho v_s\,\mathrm{d}v_s + \left((v_p)^2 - 2\,(v_s)^2\right)\mathrm{d}\rho \quad ,$$

$$\mathrm{d}\mu = \frac{\partial\mu}{\partial v_p}\,\mathrm{d}v_p + \frac{\partial\mu}{\partial v_s}\,\mathrm{d}v_s + \frac{\partial\mu}{\partial\rho}\,\mathrm{d}\rho$$

$$= 2\rho v_s\,\mathrm{d}v_s + (v_s)^2\,\mathrm{d}\rho \quad ,$$

which in superscript notation of this chapter translates to

$$\lambda^\Delta = 2\rho^0 v_p^0\, v_p^\Delta - 4\rho^0 v_s^0\, v_s^\Delta + \left((v_p^0)^2 - 2\,(v_s^0)^2\right)\rho^\Delta \quad ,$$

$$\mu^\Delta = 2\rho^0 v_s^0\, v_s^\Delta + (v_s^0)^2\,\rho^\Delta \quad .$$
(3.20)

By inserting eq. (3.20) into eq. (3.19), finally expressions K^{v_p}, K^{v_s}, K^ρ for sensitivity kernels of P_{Veloc}-parameters v_p, v_s, ρ may be derived from the $P_{\mathrm{Lamé}}$ kernels $\tilde{K}^\lambda$, $\tilde{K}^\mu$, $\tilde{K}^\rho$. Equation (3.19), hence, becomes

$$u_n^\Delta(\mathbf{r}) = \int_\Omega \Big[\left(2\rho^0 v_p^0\, v_p^\Delta - 4\rho^0 v_s^0\, v_s^\Delta + \left((v_p^0)^2 - 2\,(v_s^0)^2\right)\rho^\Delta\right)\tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r})$$

$$+ \left(2\rho^0 v_s^0\, v_s^\Delta + (v_s^0)^2\,\rho^\Delta\right)\tilde{K}_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r})$$

$$+ \rho^\Delta\,\tilde{K}_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r})\Big]\,\mathrm{d}^3\mathbf{x}$$

$$= \int_\Omega \Big[v_p^\Delta\left(2\rho^0 v_p^0\,\tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r})\right)$$

$$+ v_s^\Delta\left(2\rho^0 v_s^0\,\tilde{K}_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) - 4\rho^0 v_s^0\,\tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r})\right)$$

$$+ \rho^\Delta\left(\left((v_p^0)^2 - 2\,(v_s^0)^2\right)\tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) + (v_s^0)^2\,\tilde{K}_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) + \tilde{K}_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r})\right)\Big]\,\mathrm{d}^3\mathbf{x}$$

$$= \int_\Omega \left[v_p^\Delta K_{n\mathbf{s}}^{v_p}(\mathbf{x}, \mathbf{r}) + v_s^\Delta K_{n\mathbf{s}}^{v_s}(\mathbf{x}, \mathbf{r}) + \rho^\Delta K_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r}) \right] \mathrm{d}^3\mathbf{x} \quad , \tag{3.21}$$

introducing

$$K_{n\mathbf{s}}^{v_p}(\mathbf{x}, \mathbf{r}) = 2\rho^0 v_p^0 \, \tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) \quad ,$$

$$K_{n\mathbf{s}}^{v_s}(\mathbf{x}, \mathbf{r}) = 2\rho^0 v_s^0 \, \tilde{K}_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) - 4\rho^0 v_s^0 \, \tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) \quad , \tag{3.22}$$

$$K_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r}) = \tilde{K}_{n\mathbf{s}}^\rho(\mathbf{x}, \mathbf{r}) + \left(\left(v_p^0\right)^2 - 2\left(v_s^0\right)^2 \right) \tilde{K}_{n\mathbf{s}}^\lambda(\mathbf{x}, \mathbf{r}) + \left(v_s^0\right)^2 \tilde{K}_{n\mathbf{s}}^\mu(\mathbf{x}, \mathbf{r}) \quad .$$

Note that although density ρ is a parameter in both parameterizations $P_{\text{Lamé}} = \{\lambda, \mu, \rho\}$ and $P_{\text{Veloc}} = \{v_p, v_s, \rho\}$, the respective sensitivity kernels $\tilde{K}^\rho$ and K^ρ are different, since the seismic velocities are dependent on density. Hence, when deriving sensitivity kernels for some specific parameterization, always the complete set of model parameters must be taken into account.

3.3 Practical considerations and implementation

Applying the formulas derived in the previous sections in a practical scheme and implementing them in software means to cope with practical issues and questions of numerical performance. In this section, some key aspects in the computation of the waveform sensitivity kernels derived above are addressed. Issues on their practical use in the full waveform inversion process are laid out in more detail in chapter 4.

3.3.1 Choosing the frequency domain

Just as the Born scattering theory was presented in the frequency domain in section 3.1, for several reasons it also makes sense in practice to compute the waveform sensitivity kernels in the frequency domain.

First of all, a linearized iterative full waveform inversion procedure normally processes data for a certain frequency band (in each iteration), which can be discretized using generally few frequencies, only. Starting with low frequencies and gradually adding high frequency information contained in the data stabilizes the waveform inversion. That is why the application of the waveform kernels in the full waveform inversion concept as presented in chapter 4 iteratively requires waveform sensitivity values for certain frequency bands. In this sense, a frequency-domain quantification of sensitivity constitutes a natural concentration on the essence of information needed. This is very important, since it significantly reduces the usually very large amount of storage capacities required to hold the wavefields in memory and to write them to files.

Furthermore, the frequency domain computation of sensitivity kernels from forward wavefields and Green tensors results in multiplication operations of complex numbers, which is rather cheap compared with the analogous convolutions in the time domain. There is some computational cost, however, for time stepping forward codes to transform the time series into the frequency domain in the first place. But since wavefields and Green tensors are reused for the computation of sensitivity kernels for different paths of the dataset (explained in section 4.2.2), it pays off to transform the wavefields to the frequency domain once and then do cheap multiplications.

Besides, for time-stepping methods of seismic wave propagation it would, in most cases, be nearly impossible to produce the required time-domain wavefields throughout the inversion domain. Since often a much finer time sampling is used than imposed by the Nyquist theorem, either heavily oversampled time series would have to be held in memory and written to file, or the time series would have to be resampled. Resampling, however, would require anti-aliasing filtering on-the-fly when computing Green functions for single force sources, since impulse-like source time functions in general produce very high frequencies in the wavefield. For such time-stepping methods, a Fourier transform applied on-the-fly solves those problems and permits the processing of realistic inversion domains in terms of memory, as is done in case of using SPECFEM3D for ASKI (cf. section 5.1).

3.3.2 Choosing the frequency discretization properly

From a theoretical point of view, it appears irrelevant whether the sensitivity relations derived above are applied in the time domain or the frequency domain, since both domains are connected by the invertible Fourier transform as an operator on integrable functions. However, dealing with finite and discretized signals imposes certain stability rules, as presented in this section.

In practice, for a recorded seismic signal of finite length T the sensitivity kernel should relate the arrival of scattered waves within time interval $[0, T]$ to potential scatterers inside the medium from which the scattering could have originated. For the computation of the kernel, the required forward wavefield propagating from seismic source into the medium and Green function propagating from the receiver position into the medium *each* must represent a propagation time of T. This is because it is far from easy for complex media to even approximately predict at a specific point in the medium the maximum time span for which some forward or backward propagation must be registered (not knowing about the propagation of the other one): Scattering may happen arbitrarily close to source and receiver (i.e. arbitrarily close to maximum record length T) and the waves travelling between source, scatterer and receiver through a complex medium are in general of various wave types travelling at different wave speeds. Especially since the Green functions as a sort of universal backpropagation should be recombined with every forward wavefield (cf. section 4.2.2), the Green functions must account for all possible wave propagation of all forward fields to potential scatterers, which only can be assured a priori by choosing a propagation time of T for both, Green function and forward wavefield.
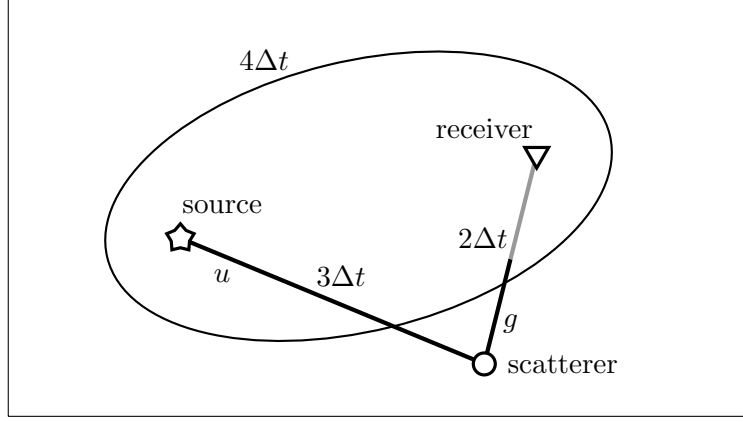
Figure 3.1: Simple sketch showing the example setting of source, scatterer and receiver, where forward wavefield u (source $\rightarrow$ scatterer) has a travel time of $3\Delta t$ and Green function g (receiver $\rightarrow$ scatterer) has a travel time of $2\Delta t$ (compare figs. 3.2a and 3.2c). Since the sum of the travel times $3\Delta t + 2\Delta t = 5\Delta t$ is larger than record length $T = 4\Delta t$, the scatterer is outside of the ellipse representing the boundary inside which scattering arrives at times earlier than T.

The finiteness of the signals immediately implies that they can be represented in the frequency domain by a discrete Fourier series at frequency step $\Delta f = \frac{1}{T}$, assuming all time-domain signals to be periodic with period T. Additionally, the measured data is in practice discretely given at some time stepping Δt, such that the periodic discrete time series may be uniquely represented in the frequency domain by their *discrete* Fourier transforms, which themselves are periodic with frequency period $\frac{1}{\Delta t}$. The computation of the sensitivity kernels in the frequency domain by the *product* of the frequency spectra of forward wavefield and Green function (as well as their strains, cf. eqs. (3.9), (3.15) and (3.22)) harbors a problem when using the frequency discretization of $\Delta f = \frac{1}{T}$, as demonstrated by the following simplified example.

In the simple setup as depicted in fig. 3.1, it is assumed that seismic waves were recorded at the receiver for total record length $T = 4\Delta t$, assuming a time stepping of some Δt. Hence, scattered waves should be modelled within the recorded time interval $[0, T]$ in order to explain possible data residuals by scatterers in the model. Propagating the forward wavefield u and the Green function g into the medium for propagation time T, the travel times of $3\Delta t$ and $2\Delta t$, respectively, are observed at the specified scattering point in the medium as displayed in figs. 3.2a and 3.2c. The product $K = U \cdot G$ of the spectra U, G of waveforms u, g represents the sensitivity kernel in the frequency domain (see fig. 3.3), which in time domain equivalently computes as the circular convolution of u and g and produces sensitivity at time $t = 1\Delta t$ (see fig. 3.2e). This, however, does not reflect the scattering correctly and could cause the kernel to *wrongly* explain possible data residuals at time $1\Delta t$ by model anomalies at the remote scattering point from which the scattering actually arrives *later* at time $5\Delta t > T$. Obviously, the discrete computation of the kernel from forward wavefield and Green function *both* exhibiting a propagation time T produces sensi-

(a) u recorded at scatterer

(b) u' recorded at scatterer, padded with zeros

(c) g recorded at scatterer

(d) g' recorded at scatterer, padded with zeros

(e) k computed from circular convolution of u and g

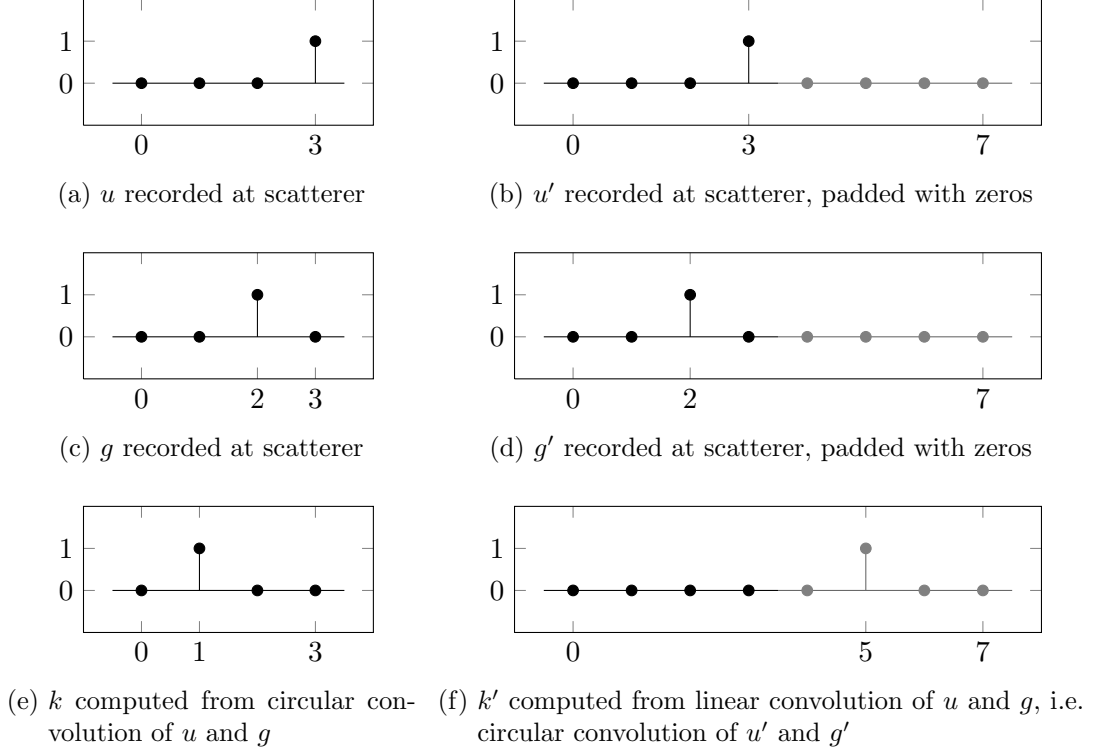(f) k' computed from linear convolution of u and g, i.e. circular convolution of u' and g'

Figure 3.2: Normalized waveforms and their convolutions (representing kernels) of the simple example setting (compare fig. 3.1). Forward wavefield u and Green function g recorded at the scattering point are assumed to be periodic signals of period $T = 4\Delta t$. Obviously, the circular convolution k of u and g does not reflect the scattering correctly since the arrival of the scattered wave is wrapped to the beginning of the kernel at time $t = 1\Delta t \triangleq 5\Delta t$ (modulo 4) due to the periodicity of the signals with period $T = 4\Delta t$. Padding u and g with zeros after record length T and, hence, extending the period of u' and g' to $2T$, yields a correct convolution k' representing the scattered wave arriving at the receiver at correct time $t = 5\Delta t$. Note that k and k' actually need to be scaled by a factor Δt in order to account for proper integration.

tivity values *outside* the sketched ellipse in fig. 3.1. As a consequence, the sensitivity kernel as a function of space contains signatures (at the remote points in the medium) of scattered waves arriving at the receiver at times *up to* $2T$. Hence, the extended time interval $[0, 2T]$ must be considered in order to correctly represent the scattering. This can be achieved by padding the measured data as well as u and g with zeros in interval $[T, 2T]$ assuming now a time period of $2T$ (see figs. 3.2b and 3.2d). This is equivalent to choosing a frequency discretization given by frequency step $\frac{\Delta f}{2}$ for the representation of the padded waveforms u' and g' in the frequency domain (compare figs. 3.3a and 3.3b). The product $K' = U' \cdot G'$ of the spectra U', G' of the padded waveforms u', g' now correctly represents all relevant scattering as seen in fig. 3.2f and is equivalent to a *linear* convolution of the short waveforms u, g of period T.

f	$0\Delta f$		$1\Delta f$		$2\Delta f$		$3\Delta f$	
$U(f)$	1		i		-1		$-i$	

f	$0\frac{\Delta f}{2}$	$1\frac{\Delta f}{2}$	$2\frac{\Delta f}{2}$	$3\frac{\Delta f}{2}$	$4\frac{\Delta f}{2}$	$5\frac{\Delta f}{2}$	$6\frac{\Delta f}{2}$	$7\frac{\Delta f}{2}$
$U'(f)$	1	$\frac{-1-i}{\sqrt{2}}$	i	$\frac{1-i}{\sqrt{2}}$	-1	$\frac{1+i}{\sqrt{2}}$	$-i$	$\frac{-1+i}{\sqrt{2}}$

(a) spectral values of U and U' at frequencies f

f	$0\Delta f$		$1\Delta f$		$2\Delta f$		$3\Delta f$	
$G(f)$	1		-1		1		-1	

f	$0\frac{\Delta f}{2}$	$1\frac{\Delta f}{2}$	$2\frac{\Delta f}{2}$	$3\frac{\Delta f}{2}$	$4\frac{\Delta f}{2}$	$5\frac{\Delta f}{2}$	$6\frac{\Delta f}{2}$	$7\frac{\Delta f}{2}$
$G'(f)$	1	$-i$	-1	i	1	$-i$	-1	i

(b) spectral values of G and G' at frequencies f

f	$0\Delta f$		$1\Delta f$		$2\Delta f$		$3\Delta f$	
$K(f)$	1		$-i$		-1		i	

f	$0\frac{\Delta f}{2}$	$1\frac{\Delta f}{2}$	$2\frac{\Delta f}{2}$	$3\frac{\Delta f}{2}$	$4\frac{\Delta f}{2}$	$5\frac{\Delta f}{2}$	$6\frac{\Delta f}{2}$	$7\frac{\Delta f}{2}$
$K'(f)$	1	$\frac{-1+i}{\sqrt{2}}$	$-i$	$\frac{1+i}{\sqrt{2}}$	-1	$\frac{1-i}{\sqrt{2}}$	i	$\frac{-1-i}{\sqrt{2}}$

(c) spectral values of K and K' at frequencies f, note that $K = U \cdot G$, $K' = U' \cdot G'$

Figure 3.3: The tables contain the discrete (periodic) spectra of the respective discrete (periodic) time series shown in fig. 3.2 (denoted by respective capital letters). The spectral values can be computed by the discrete Fourier transform for which $\Delta f = \frac{1}{T} = \frac{1}{4\Delta t}$ by convention. Applying the inverse discrete Fourier transform to these spectra will result in the respective time series from fig. 3.2. This figure demonstrates that padding with zeros in the time domain is equivalent to an interpolation in the frequency domain. Hence, only the *complete* spectral information of K' at all frequencies $n\frac{\Delta f}{2}$, $n = 0, \ldots, 7$ represents the scattering correctly, since K and K' are indistinguishable at frequencies $n\Delta f$, $n = 0, \ldots, 3$. Note that the spectra U, U', G, G' actually need to be scaled by a factor Δt in order to account for proper integration (hence, K, K' are missing a factor Δt^2).

Of course one could use a frequency step of $\Delta f = \frac{1}{T}$ (assuming period T) when at the same time *discarding* scattering points outside the $4\Delta t$ ellipse, in this example. In media different from a homogeneous halfspace, however, this boundary is not of ellipsoidal shape anymore and may even "overlap" with itself: In case there is some

reflecting boundary close enough to source and receiver, there might be points in the medium from which scattering might arrive within *both* time intervals, the valid one $[0, T]$ and the invalid one $[T, 2T]$. This is the case if the forward wavefield and/or Green function may travel to that point by different paths which take different times (one significantly shorter than the other). Potential scattering at such points can *only* correctly be taken into account by discriminating times smaller than T from later ones, that is incorporating spectral information at the fine sampling of $\frac{\Delta f}{2}$. This is particularly well demonstrated below in fig. 3.10f, where presumably some reflection phase contaminates the complete relevant scattering domain between source and receiver.

In conclusion, it is obviously relevant when transforming spectral kernels back to the time domain to use such a fine frequency discretization of the spectra (as will be shown in fig. 3.10). However, in case of deriving and applying those kernels in the frequency domain only (e.g. in the full waveform inversion method presented in chapter 4) the *same* fine sampling is required just as well: Although this is not as obvious as in the time domain, any spectral sampling lower than $\frac{1}{2T}$ ignores important information (equivalently assumes periodicity of the corresponding time series of a period smaller than $2T$). Hence, *also* in frequency-domain application of the sensitivity kernels an improper spectral sampling runs the risk of incorrectly interpreting the scattering, e.g. producing model artefacts in the waveform inversion. The choice of a frequency step even smaller than $\frac{1}{2T}$, i.e. interpolating the spectra beyond this limit, should still picture the scattering correctly since equivalently more zero waveform is added to data, u and g.

Note that these rules on the proper sampling of the frequency spectra apply to all methods of seismic wave propagation in both, time and frequency domain. The rules only arise from the frequency domain computation of the kernels by multiplication of spectra (equivalently *circular* convolutions) and the requirement of *discretizing* the spectral domain in practice, as well as the property of reusing the Green functions for reasons of efficiency and the structure of the medium being in general not homogeneous. In case of such kernels being computed directly in the time domain from wavefields by a linear convolution, the topic of this section obviously does not need to be discussed at all.

3.4 Examples

As already explained above, the waveform sensitivity kernels quantify for some given background model how much a waveform (seen as data functionals of the same background model) changes if the background model parameters are perturbed. In the frequency domain, however, waveform data functionals are in general not very intuitive to understand for human beings. So, a quantification of how sensitive a complex-valued spectrum of a seismogram is at a specific frequency with respect to changes in specific model parameters may not be very informative to look at, as demonstrated in section 3.4.2.

In geophysics, however, time-domain data, such as seismograms, give a much more intuitive understanding of the underlying wave propagation inside the medium compared with frequency-domain data, due to the experience of collecting such data in the time domain. This consideration motivates looking at waveform sensitivity kernels in the time domain relating residual waveforms at specific points in time to changes in the structural earth model. This way, potential locations of origin of any scattered waves occurring in measured seismograms may be detected on the basis of the intuitive understanding of how seismic waves propagate in time through the given background model. A synthetic example of such a scenario is presented in section 3.4.3.

At first, however, the next section shortly presents the general setup of acquisition geometry and earth model for which the example kernels were computed.

3.4.1 A simple synthetic example: the setup

The sensitivity kernel in frequency and time domain as shown below was computed by ASKI (section 5.1) for the following simple geometric setup and background earth model. The spectra of the wavefields necessary to derive the kernels were computed using SPECFEM3D Cartesian 2.1 (Tromp et al., 2008) for ASKI. The time-domain kernel then was derived by inverse Fourier transform of the frequency-domain kernel.

Into a simple layered background model (fig. 3.4a), a model anomaly was placed (fig. 3.4b) which scatters the seismic waves excited by a vertical point force acting downwards at the source position (yellow dot) on a free surface. Those scattered waves are detected at the receiver position (red dot) in the seismic displacement field by comparing the waveforms computed for the background model with those computed for the perturbed model. The latter are interpreted here as the seismic measurements. All waveforms in the synthetic examples of this section, however, are computed by SPECFEM3D using the source-time-function as shown in fig. 3.5.

3.4.2 Kernels in the frequency domain

Figures 3.6 and 3.7 show spectral sensitivity kernels (eq. (3.22)) for different data samples and different model parameters of the chosen isotropic model parameterization $P_{\text{Veloc}} = \{v_p, v_s, \rho\}$. Each image displays the sensitivity of a specific spectral datum (e.g. real part of upwards displacement spectrum at frequency 5 Hz) with respect to changes of a specific elastic constant (i.e. v_p or v_s) at all points inside the medium. Hence, large absolute values of such a sensitivity kernel around a certain point in the medium result in significant change of the respective spectral datum if the respective elastic constant is perturbed around that point. Thereby, the sign of the change in the datum is the product of signs of sensitivity and model anomaly, as can be seen in eq. (3.21) from the multiplicative terms in the integrand. Keep in mind, however, that the Born theory on which those kernels are based assumes single scattering. That is, scattered waves generated at model perturbations are assumed to
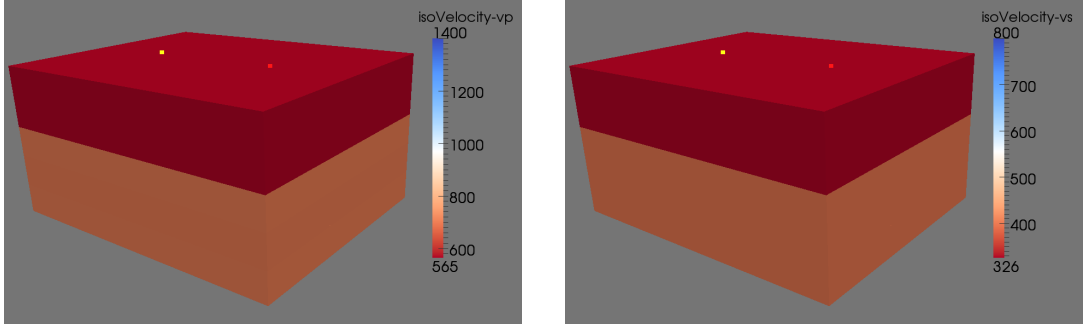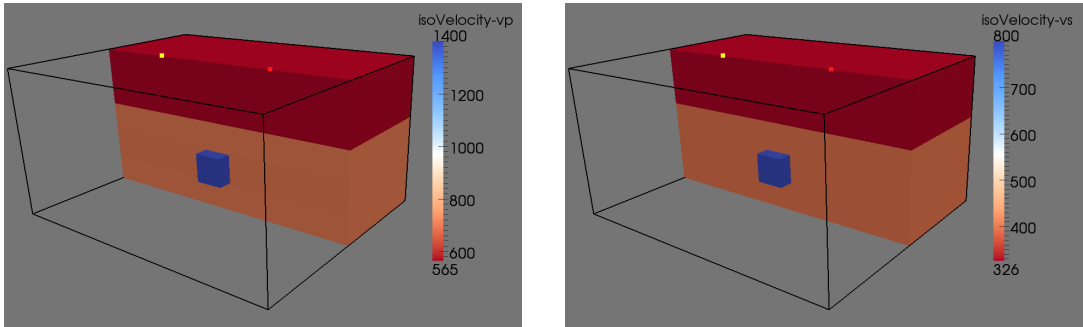
32

(a) v_p and v_s background earth model, slow layer over a faster halfspace.

(b) Fast anomaly in v_p and v_s placed into the halfspace.

Figure 3.4: Isotropic background earth model and model perturbation. Source (yellow) and receiver (red) are placed on a free surface. The seismic velocities are given in $\frac{\text{m}}{\text{s}}$. Density is not of importance and was always chosen as $1.4 \cdot v_s$ (in $\frac{\text{kg}}{\text{m}^3}$).

travel to the receiver through the background medium: it is not supported that such scattered waves are again perturbed by yet another model perturbation.

It can be clearly seen in the figures that sensitivity mainly concentrates at source, receiver and in-between, as well as on the free surface. Furthermore, the general structure of the kernels correlates with wavelength, i.e. the smaller the wavelength the finer the kernel is structured. This becomes obvious by comparing sensitivity kernels of v_p and v_s at the same frequency or comparing kernels of the same parameter at different frequencies, as well as observing that the wave speeds in the halfspace are larger than in the upper layer. The antisymmetry of the kernels for the transverse data component (fig. 3.7) may be explained by the different polarization of scattered waves arriving at the receiver from different sides of the source-receiver line. In general, sensitivity kernels for S-velocity are experienced to have higher absolute values than kernels for P-velocity. This can also be seen in the example kernels of this section, where here additionally the strong influence of surface waves needs to be accounted for.

Despite the above explanations, it remains not very intuitive to understand the propagation of potential scattered waves occurring in complex-valued spectral data

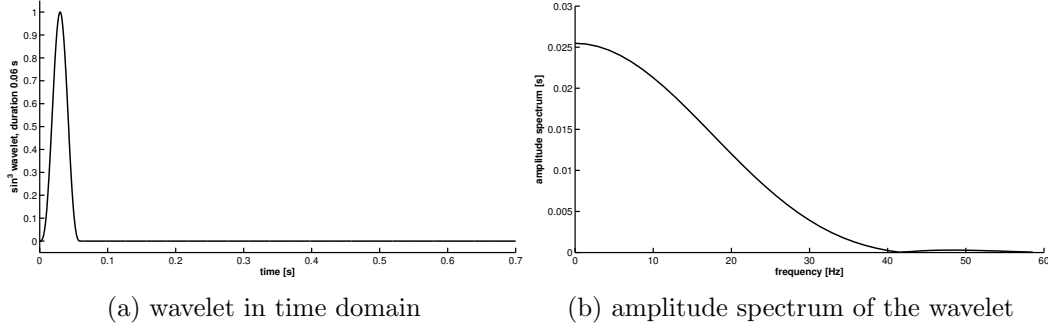(a) wavelet in time domain   (b) amplitude spectrum of the wavelet

Figure 3.5: The source time function used for the SPECFEM3D simulations: $\sin^3\left(t\,\frac{\pi}{t_d}\right)$, $t < t_d = 0.06$s, zero otherwise.

from the spectral sensitivity kernels by only looking at them. That is why in the following the scattering is examined in the time domain.

3.4.3 Where do the scattered waves come from?

In this section, the potential of time-domain sensitivity kernels in the prediction of possible locations of origin of scattered waves observed in measured seismic data is discussed by means of the synthetic example setting described in section 3.4.1. This instrument of looking at time-distributed sensitivity is, however, not meant to achieve a complete image of the ground model as done by the full waveform inversion presented in chapter 4. It is rather intended to provide a first intuitive understanding of the underground and the resolving power of the measured data.

Observing a residual waveform comparing measured and synthetic data (here shown in fig. 3.8), the time-domain waveform sensitivity kernels provide for each time step those potential locations in the model domain at which perturbations may have excited scattered waves explaining the data residuals in terms of linearized single scattering. Figure 3.9 presents some interesting snapshots of those time-dependent sensitivity distributions which were computed by an inverse Fourier transform from the spectral kernels presented above. It can be seen that indeed v_s-sensitivity of correct sign propagates through the upper part of the model anomaly at times of non-zero data residuals. However, without the knowledge of the true model anomaly (as provided in this example) it is not certain from where the observed scattered waves originate after all, since there is simultaneous high v_s-sensitivity at other points of the domain. Also, a combination of changes in all model parameters may have caused the scattered waves instead of just v_s perturbations, not to mention that any elastic model parameterization can only approximate the true earth. Instead, this kind of sensitivity analysis can provide a first intuitive understanding of the resolving power of the data, assuming a sufficiently good knowledge of the earth model: The kernels highlight those subdomains of the model to which the specific data are very sensitive and indicate other places at which probably no scattering is produced by the incident seismic waves, dependent on the propagation time of those scattering.
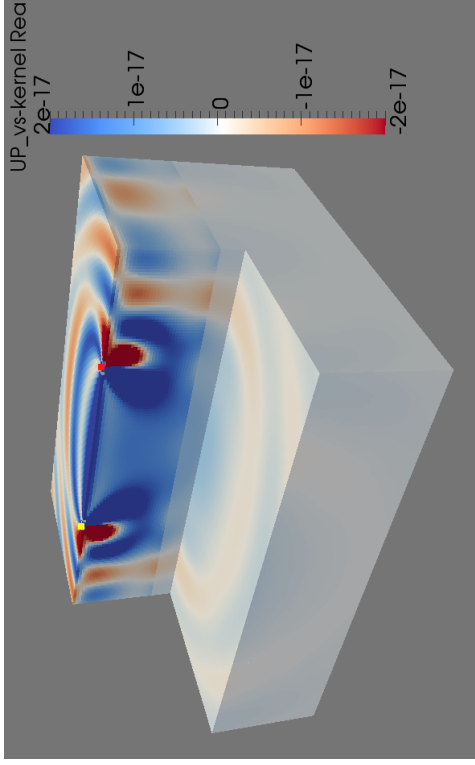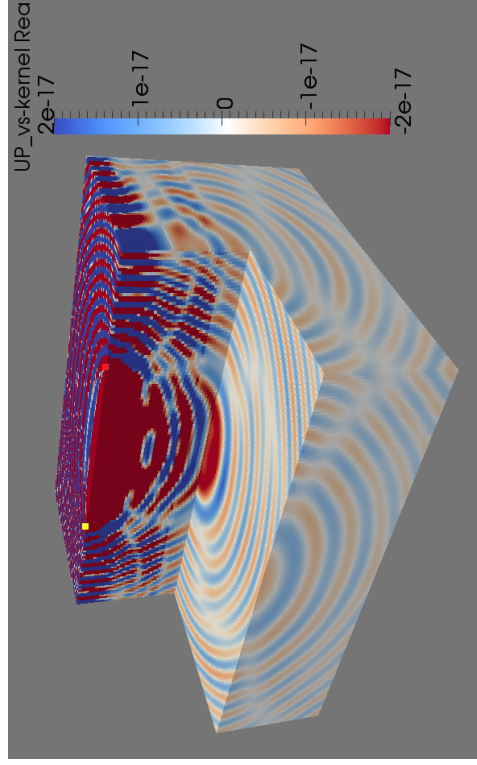
(a) real part of the v_p kernel at 5Hz

(b) real part of the v_s kernel at 5Hz

(c) real part of the v_p kernel at 25Hz

(d) real part of the v_s kernel at 25Hz

Figure 3.6: Spectral sensitivity kernel of the real part of the vertical upwards displacement component at different frequencies and for different model parameters. It can be clearly seen that sensitivity concentrates at source, receiver and in-between, as well as the free surface. The general structure of the kernels is consistent with wavelength, i.e. the smaller the wavelength (dependent on wave speed and frequency) the finer the structure becomes.

(a) real part of the v_p kernel at 5Hz
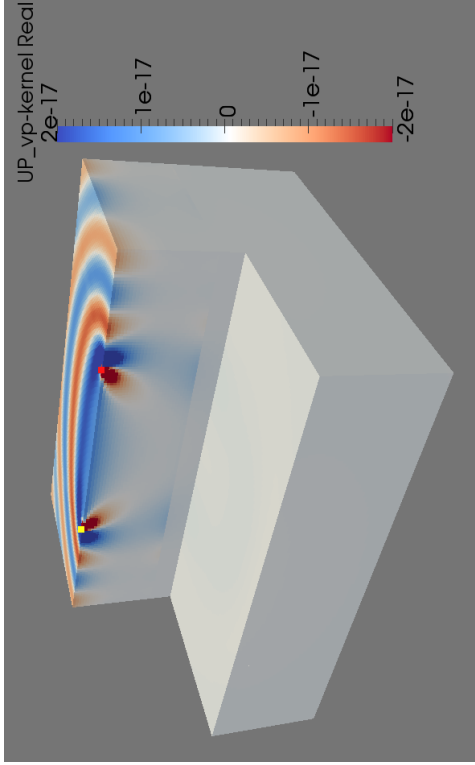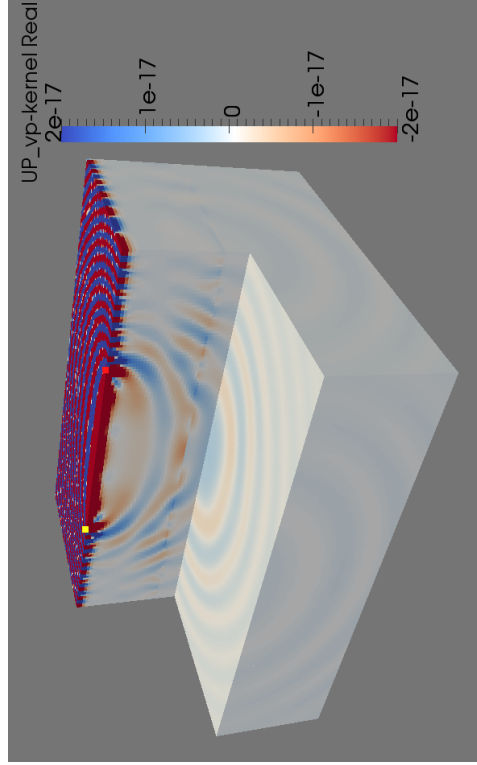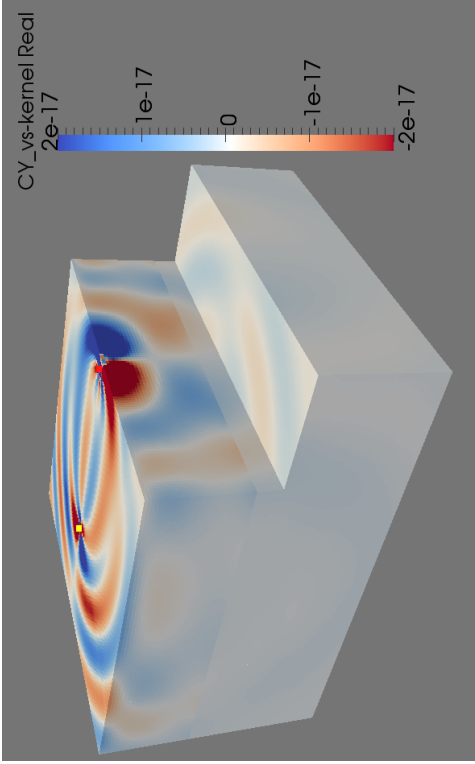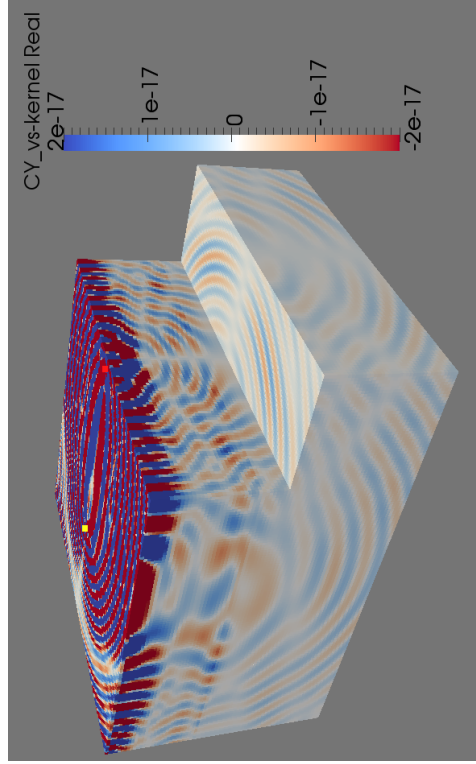
(b) real part of the v_s kernel at 5Hz
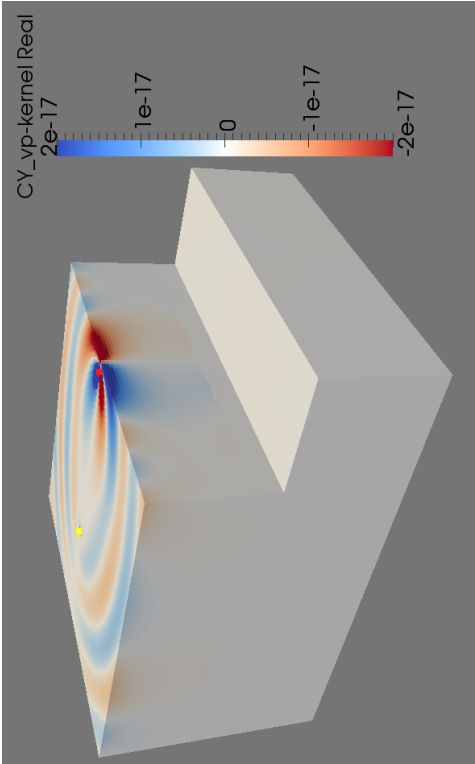
(c) real part of the v_p kernel at 25Hz
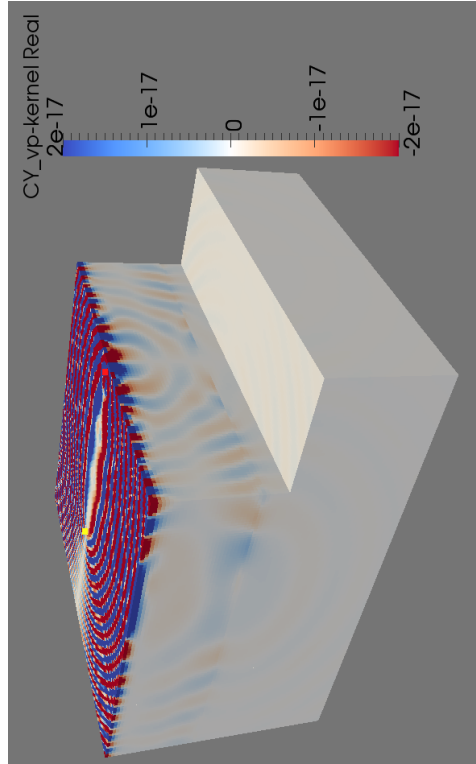
(d) real part of the v_s kernel at 25Hz

Figure 3.7: Spectral sensitivity kernel of the real part of the transverse displacement component at different frequencies and for different model parameters. It can be clearly seen that sensitivity concentrates at source, receiver and in-between, as well as the free surface. The general structure of the kernels is consistent with wavelength, i.e. the smaller the wavelength (dependent on wave speed and frequency) the finer the structure becomes. Note the antisymmetry in terms of a change in sign when crossing the source-receiver line.

Regarding the efficiency of computing the waveform kernels in the frequency domain (section 3.3.1), here *complete* waveform information is required instead of wavefields in just a small frequency band. Hence, the computation of a time-domain sensitivity kernel is, in terms of memory and storage, even more demanding than the computation of a kernel for the last iteration steps of the full waveform inversion procedure presented in chapter 4. In case of looking at time-domain kernels for just a few source-receiver combinations, however, this does not constitute an issue.

In addition, the time-domain waveform sensitivity kernels as derived here by an inverse Fourier transform demonstrate the issue of insufficient sampling of spectral kernels as was introduced in section 3.3.2. Figure 3.10 compares time-domain kernels derived from spectral kernels at the two frequency discretizations $\Delta f = \frac{1}{T}$ and $\Delta f = \frac{1}{2T}$. In case of the coarse frequency discretization, i.e. assuming periodicity of period T, the sensitivity at times later than T is clearly wrapped to early times in the kernel (compare figs. 3.10d and 3.9f). When using $\Delta f = \frac{1}{2T}$, hence assuming period $2T$, those artefacts vanish. In the frequency domain, on the other hand, this issue does not become evident in terms of artefacts in the kernel of a certain datum but in terms of the absence of certain data samples. That is why the insufficient sampling of the spectra is not obvious in the frequency domain. However, this problem would not occur in case of computing the time-domain sensitivity kernels directly in the time domain by a *linear* convolution of wavefields, as mentioned above.

As another possible application of such time-domain sensitivity kernels, one could think of a comparison of the kernels with active seismic measurements directly in the field. This assumes, however, that an approximate background model is sufficiently known and that kernels and synthetic data were computed for known acquisition geometry and source mechanisms prior to the field experiment. Also the measured data needs to be pre-processed in the field in order to be sufficiently comparable with the synthetics. Provided all this, a first intuitive understanding of the underground could be gained already in the field by observing the sensitivity at the times of measured data residuals. In case of more than one potential acquisition geometries in question, the comparison of data residuals and kernels for a few test shots with just few receivers may support the decision for that acquisition being most promising in terms of illuminating the desired model subdomain.
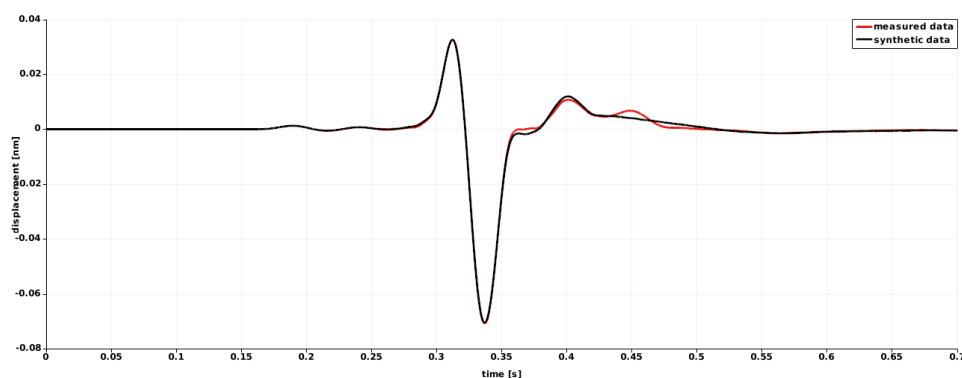
Figure 3.8: Seismic measurement (red, synthetically computed w.r.t. the perturbed model) and synthetics (computed w.r.t. the unperturbed layered background model): Vertical upwards ground displacement at the receiver position. Models and acquisition as described in section 3.4.1.

(a) v_s kernel at time 0.0 s

(b) v_s kernel at time 0.18 s

(c) v_s kernel at time 0.33 s

(d) v_s kernel at time 0.45 s

(e) v_s kernel at time 0.47 s

(f) v_s kernel at time 0.66 s
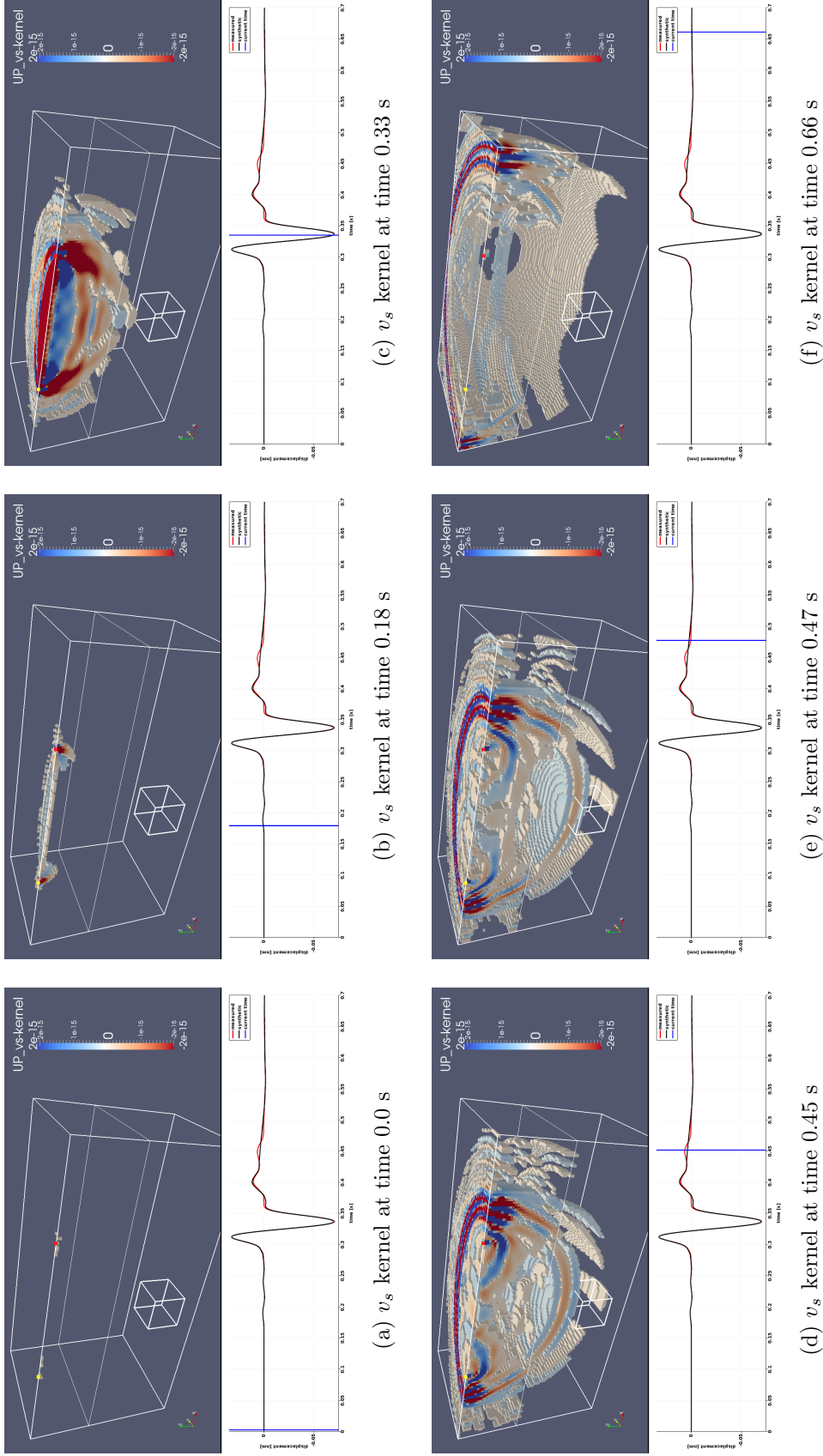
Figure 3.9: Time-domain waveform kernel: v_s-sensitivity of the upward displacement seismogram at different times. Below the kernels, these measured (red) and synthetic (black) seismograms are plotted indicating the scattered waves as the residual waveform (cf. fig. 3.8). A cut through the model domain is shown similar to fig. 3.4b and sensitivity values close to zero are not plotted. Note the sensitivity propagating through the anomaly at arrival times of the scattered waves and that positive (negative) sensitivity indeed causes positive (negative) change in the data, given the positive model anomaly.

(a) using small $\Delta f = \frac{1}{2T}$ at time 0.0 s

(b) using small $\Delta f = \frac{1}{2T}$ at time 0.18 s

(c) using small $\Delta f = \frac{1}{2T}$ at time 0.33 s

(d) using large $\Delta f = \frac{1}{T}$ at time 0.0 s

(e) using large $\Delta f = \frac{1}{T}$ at time 0.18 s

(f) using large $\Delta f = \frac{1}{T}$ at time 0.33 s
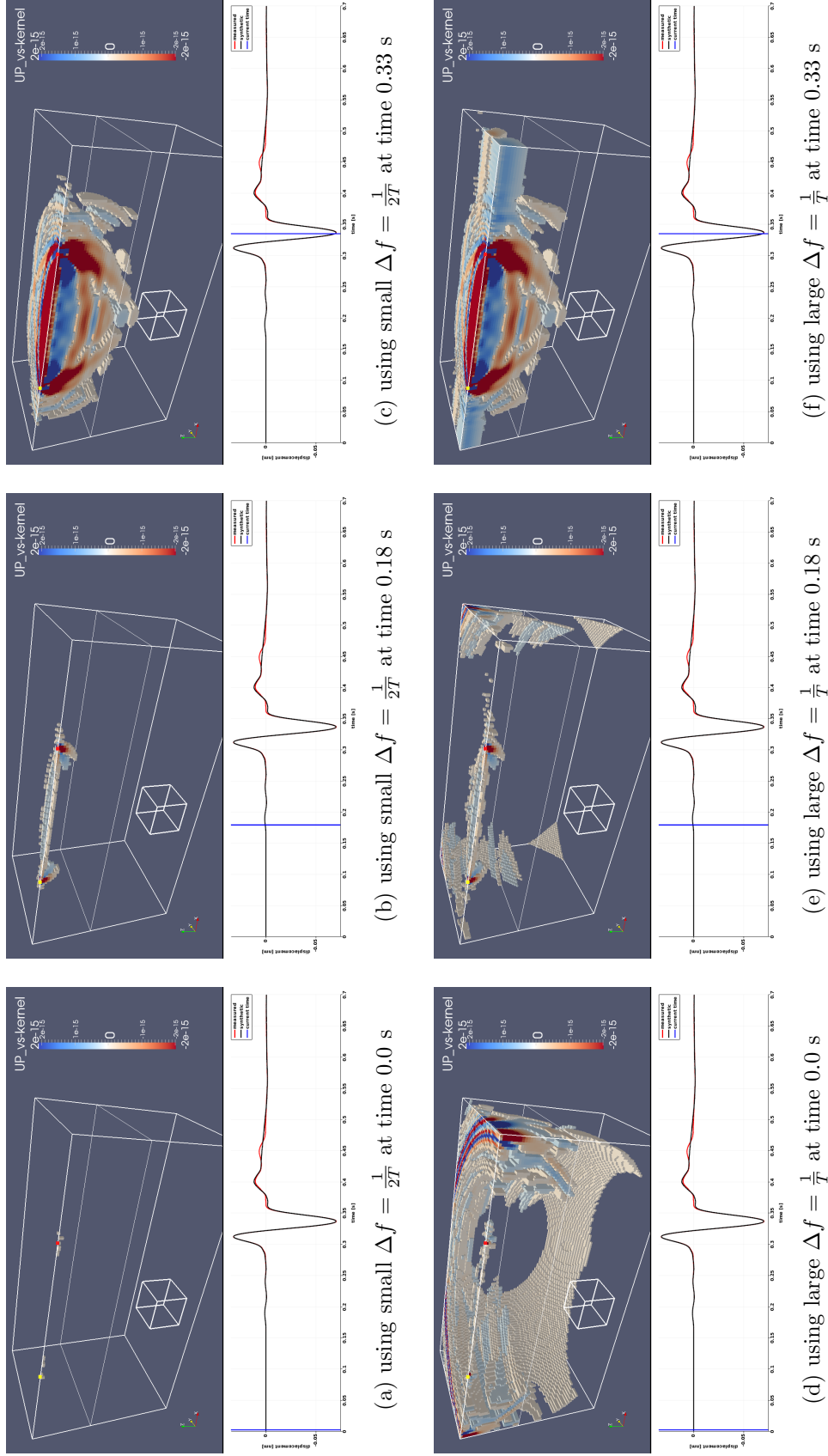
Figure 3.10: Time-domain waveform kernel at different times, derived by inverse Fourier transform from spectral kernels of two different frequency discretizations, demonstrating the necessity of choosing a proper frequency step (compare section 3.3.2) : 3.10a, 3.10b and 3.10c are based on a stable frequency discretization of $\Delta f = \frac{1}{2T}$, while 3.10d, 3.10e and 3.10f are based on insufficiently sampled spectral kernels using $\Delta f = \frac{1}{T}$. Note that the time kernel based on the insufficiently sampled frequency kernel is indeed periodic with period T, by comparing figs. 3.10d and 3.9f. The artefacts in 3.10f are probably caused by late arriving boundary reflections from the SPECFEM3D simulation domain.

4 Iterative full waveform inversion procedure

Based on the sensitivity kernels as derived in the previous chapter 3, this chapter 4 presents in detail the modularized iterative full seismic waveform inversion procedure motivated in section 2.3.

Section 4.1 defines the form of data used in the inversion process and introduces certain notations of data related quantities. The following section 4.2 presents the inversion method in general, introducing individual steps and referring to distinct characteristics. Section 4.3 thereafter discusses certain issues and opportunities arising in practical use and implementation of the method.

4.1 Data representation

The application of a full waveform inversion procedure requires a set of waveform data which can be inverted. The full waveform inversion aims at explaining the complete information content of the data by earth structure (and source characteristics). This section will characterize the form of waveform data used in this waveform inversion process.

Central in the linearized iterative inversion concept is the linear scattering eq. (3.9), which relates changes in the model parameters p^Δ to changes in the displacement spectra $\mathbf{u}^\Delta(\mathbf{r})$ as a data related quantity. Below in section 4.2, $u_n^\Delta(\mathbf{r}) = u_n^1(\mathbf{r}) - u_n^0(\mathbf{r})$ will be interpreted as the difference between recorded data $u_n^1(\mathbf{r})$ and data functional $u_n^0(\mathbf{r})$ which is synthetically computed w.r.t. some iteratively updated reference earth model. Remember, as motivated in section 3.3.1, that the computation of sensitivity kernels and the underlying scattering theory were derived in the frequency domain. The displacement field $\mathbf{u}^0(\mathbf{r})$ in eq. (3.9) is excited by some seismic source and recorded at receiver position $\mathbf{r}$ for receiver component n, which does not need to be parallel to a coordinate axis, but can be interpreted as any direction of space (e.g. local north or vertical down). Equation (3.9), furthermore, holds true for an arbitrary angular frequency ω, which results in a scattering formulation for the complete seismic signal, hence a "full waveform" relation in the frequency domain. As data, therefore, complete seismic signals represented in the frequency domain are used in this inversion concept being discretized as described in the following.

A data sample d has a specific value of spectral ground displacement and is associ-

ated to a seismic source which has excited the displacement field, a receiver position and component at which the displacement was recorded and a frequency $f = \frac{\omega}{2\pi}$ at which the displacement spectrum is evaluated. Since the complex-valued relation in eq. (3.9) finally will be split into real and imaginary part, as presented below in section 4.2.3, the data samples are chosen to be real-valued and are additionally characterized by their complex component. The process of data collection usually involves receivers which record seismic waves excited by many sources. This may result in seismograms of different length that are to be used as seismic data, e.g. dependent on source-receiver distances in case of global seismology. Considering section 3.3.2, for time series of a length of T the corresponding frequency spectrum should be discretized by $\Delta f \leq \frac{1}{2T}$. However, for reasons of reusing Green functions for many sources in the course of modelling the data (referred to later on in sections 4.2.1 and 4.2.2), it is necessary to define *one global* frequency step $\Delta \hat{f}$ with which *all* frequencies f of *all* data samples contained in the data set can be represented (i.e. f is some multiple of $\Delta \hat{f}$). The largest Δf that fulfils the stability criterion of section 3.3.2 for *all* seismograms of the data set obviously is $\Delta \hat{f} = \frac{1}{2T_{\max}}$, where $T_{\max}$ is the maximum record length of seismograms contained in the data set. Finally, for each recorded seismogram a separate frequency window (i.e. maximum and minimum frequency) discretized by $\Delta \hat{f}$ is to define the actual data to be inverted. This frequency window is dependent on the meaningful frequency content of the seismogram, i.e. on any characteristics of seismic source, source-receiver distance, instrument and noise level.

Based on these general properties of the data, the notation describing the data set as used in the following can be defined, assuming numerous seismic data traces have been recorded involving certain sets of seismic sources, receivers, possible directions of receiver components and assuming, furthermore, that all involved frequencies are represented by a certain finite set of discrete frequencies. Let the total number of involved seismic sources be N_S and their positions be denoted by $\mathbf{s}_l$, $1 \leq l \leq N_S$. Let the total number of involved receivers be N_R and their positions be denoted by $\mathbf{r}_m$, $1 \leq m \leq N_R$. Let the total number of involved receiver components be N_C and the components be denoted by n, $1 \leq n \leq N_C$. Let the total number of involved frequencies be N_F and the frequencies be denoted by $f_p = \frac{\omega_p}{2\pi}$, $1 \leq p \leq N_F$. Let the 2 possible complex components be denoted by $\mathfrak{c}_q$, $1 \leq q \leq 2$. Then, a specific data sample of the data set is characterized by a tuple $(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q)$, the collection of which may have some fixed mapping $\mathrm{idx}^{\mathrm{d}}$, assigning each tuple a unique index $i = \mathrm{idx}^{\mathrm{d}}(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q)$ in range $1 \leq i \leq N_D$, where

$$N_D = N_S \cdot N_R \cdot N_C \cdot N_F \cdot 2$$

is the total maximum number of possible data samples. Hence, the ith data sample

$$d_i \quad , \quad i = \mathrm{idx}^{\mathrm{d}}(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q) \ ,$$

denotes the real or imaginary part (dependent on $\mathfrak{c}_q$) of the displacement spectrum $\mathbf{u}$ at frequency f_p excited by seismic source $\mathbf{s}_l$ recorded at receiver $\mathbf{r}_m$ in direction of component n. Obviously, not necessarily all N_D data samples may be used for inversion since they were not recorded, of bad quality or left out due to frequency windowing. That is why data samples d_i for certain subsets $i \in D \subseteq \{1, \ldots, N_D\}$ are considered in the following.

4.2 Full waveform inversion procedure

This section 4.2 presents the practical realization of the full waveform inversion concept using waveform sensitivity kernels which was motivated in section 2.3. As presented in section 2.3.1, a strategy of iteratively improving some initial earth model by solving linearized inverse problems is followed in order to realize the non-linear full waveform inversion process.

First of all, an elastic parameterization of the earth, i.e. a specific symmetry of the tensor c of linear elasticity, must be chosen along with a suitable method for seismic wave propagation which can handle this symmetry. Furthermore, a starting earth model $\mathbf{m}_0$ comprising parameters of this parameterization must be chosen initiating the iterative inversion procedure. Forward methods supporting only laterally homogeneous earth models may only be suitable for conducting the first iteration step, provided that such a 1D model is used as the starting model $\mathbf{m}_0$. This is because in each iteration $k = 1, 2, \ldots$ an in general full 3D-heterogeneous model $\mathbf{m}_k$ will be produced which improves the earth model $\mathbf{m}_{k-1}$ and on which wave propagation simulations will be based during iteration $k+1$. Anelastic properties of the Earth may be inverted for using respective sensitivity kernels for those kinds of model parameters, which are not derived in this thesis. However, even in case of inverting for elastic constants and density only, anelasticity may be accounted for passively by the wave propagation method using reasonable constant anelastic parameters which are not inverted for.

Before iteration k, some data subset defined by some $D_k \subseteq \{1, \ldots, N_D\}$ is chosen which will be inverted in the iteration. For reasons of regularization, those subsets of data should contain low frequency data in early iterations, gradually increasing the frequency content in the course of the iterative procedure. Thereby, the number of data samples may iteratively increase preserving the data of the previous iterations, i.e. $D_1 \subset D_2 \subset D_3 \subset \ldots$. This might produce inversion results which are more stable than in case of the subsets being pairwise disjoint or only sparsely overlapping, which on the other hand could enhance the computational performance. The model update $\mathbf{m}_k^\Delta = \mathbf{m}_k - \mathbf{m}_{k-1}$ is, then, derived from kernel eq. (3.16) with help of data samples $\{d_i | i \in D_k\}$ (interpreted as perturbed displacement $\mathbf{u}^1$) and corresponding waveform sensitivity kernels related to those data. In terms of scattering, the currently inverted earth model $\mathbf{m}_{k-1}$ is taken as the background model on which the kernels are based and for which synthetic data functionals $s_i = F_i[\mathbf{m}_{k-1}]$, $i \in D_k$, (interpreted as unperturbed displacement $\mathbf{u}^0$) corresponding to d_i are computed. By means of the sensitivity kernels, then, a change in the model $\mathbf{m}_k^\Delta = \mathbf{m}_k - \mathbf{m}_{k-1}$ is derived that best explains the data residuals $r_i = d_i - s_i$, $i \in D_k$, (interpreted as the scattered waves $\mathbf{u}^\Delta$) subject to regularizing smoothing constraints.

In the following, the individual steps necessary to conduct an arbitrary k^{th} iteration of the iterative full waveform inversion procedure are presented and discussed in more detail, also referring to practical aspects of implementation. This includes solving the forward problem, computing the sensitivity kernels, solving an inverse problem

subject to regularization constraints and connecting the iterations properly.

4.2.1 Solving the forward problem

Given some data subset D_k and reference earth model $\mathbf{m}_{k-1}$ which was inverted in the previous iteration (or the starting model $\mathbf{m}_0$, if $k = 1$), this section describes the solution of the forward problem in the course of iteration k of the full waveform inversion. Solving the forward problem, in terms of the general opening remarks of chapter 2, means to model the data, i.e. evaluate the propagation of seismic waves in the medium given a known earth model, which in this case is the currently inverted reference earth model $\mathbf{m}_{k-1}$.

Aiming at computing waveform sensitivity kernels w.r.t. reference model $\mathbf{m}_{k-1}$, section 3.2 concludes that for arbitrary symmetry of tensor c, the following quantities need to be known throughout the domain of interest: For an individual data sample d_i, $i = \mathrm{idx}^{\mathrm{d}}\left(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q\right) \in D_k$, the sensitivity kernel values which quantify the change in d_i due to changes in the model parameters can be computed from the real or imaginary part (dependent on $\mathfrak{c}_q$) of

- the displacement spectrum originating from seismic source $\mathbf{s}_l$ and its strains

- the Green function spectrum for a single force source in direction n of receiver $\mathbf{r}_m$ and its strains

both given at frequency f_p. Therefore, the computation of these quantities for all data samples d_i, $i \in D_k$ constitutes the solution of the forward problem in iteration k of the inversion procedure. Additionally, corresponding synthetic data samples s_i should be computed as data functionals w.r.t. model $\mathbf{m}_{k-1}$, which are used for the computation of the iteration specific data residuals $r_i = d_i - s_i$, as required in section 4.2.3. Those can be consistently derived at the same time when simulating the propagation of the forward wavefield from the seismic sources into the medium. It is important to notice for the computation of sensitivity kernels, that the Green function spectrum representing the backward propagation from receiver component $(\mathbf{r}_m, n)$ to scattering points inside the medium is *not* dependent on the particular data values d_i. It only depends on the background model and the meta information of receiver $\mathbf{r}_m$ and component n. For this reason, the displacement spectra for all sources as well as the Green functions for all receiver components which are involved in the current data subset D_k can be computed *independently* of each other at the required frequencies. This is particularly different from the adjoint method of full waveform inversion (remember section 2.2.2) and can imply a smaller number of required wave propagation simulations (cf. section 4.3.4).

In case of solely inverting given data associated with selected receiver components, only the respective Green functions are required. These can be computed by applying single force sources in the direction of the recorded data at the receiver positions (e.g. local north on the surface of the earth). However, for a general sensitivity analysis of

the earth model involving data components for which there was no data recorded (e.g. in order to evaluate whether it would be desirable to have data for those components), it may make sense to compute a complete Green tensor for three linearly independent components for the involved receivers $\mathbf{r}_m$. In this case, any data component can be modeled by linearly combining the Green functions of the independent components adequately.

4.2.2 Connecting forward and inverse problem: sensitivity kernels

Using the wavefields produced by the forward wave propagation method, as discussed in the previous section 4.2.1, waveform sensitivity kernels are now computed as derived in section 3.2.

For individual data sample d_i, $i = \mathrm{idx}^{\mathrm{d}}(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q) \in D_k$ inverted in iteration k, the respective waveform sensitivity kernel values $K_{n\mathbf{s}_l}^p(\mathbf{x}, \mathbf{r}_m)$ are computed at frequency f_p for each parameter $p \in P$ of the chosen model parameterization P. As seen in eqs. (3.9) and (3.15) and demonstrated in section 3.2.1, the formulas for the kernel computation contain multiplications of the displacement spectrum $u_i(\mathbf{x})$ originating at source $\mathbf{s}_l$ with the Green function spectrum $G_{in}(\mathbf{x}, \mathbf{r}_m)$ originating in direction n at receiver $\mathbf{r}_m$ as well as multiplications of their strains.

Due to the Green function spectra being independent of any actual data values d, they may be efficiently reused in the computation of the kernels for *all* data that were recorded at the respective receiver component. That is, Green function spectrum $\mathbf{g}_n(\mathbf{r})$ of particular receiver component $(\mathbf{r}, n)$ may be recombined by *all* spectra $\mathbf{u}$ for sources $\mathbf{s}$ for which there are data samples d_i, $i = \mathrm{idx}^{\mathrm{d}}(\mathbf{s}, \mathbf{r}, n, f_p, \mathfrak{c}_q) \in D_k$ in iteration k. As already indicated above, this motivates the choice of a global frequency discretization: $\mathbf{g}_n(\mathbf{r})$, namely, must in this case represent a time window of size $T_{\max}^{n\mathbf{r}}$, where $T_{\max}^{n\mathbf{r}}$ is the maximum length of seismograms recorded at receiver component $(\mathbf{r}, n)$. Alternatively speaking, the spectrum $\mathbf{g}_n(\mathbf{r})$ must be discretized by a frequency step $\Delta f^{n\mathbf{r}} \leq \frac{1}{2T_{\max}^{n\mathbf{r}}}$ in order to additionally meet the stability conditions of section 3.3.2. However in order to avoid artefacts in the kernels and be able to perform the multiplications in the discretized frequency domain, all spectra $\mathbf{u}$ for events $\mathbf{s}$ recorded at $(\mathbf{r}, n)$ must be discretized by the same frequency step $\Delta f^{n\mathbf{r}}$. A particular record of some event $\mathbf{s}$ being significantly shorter than $T_{\max}^{n\mathbf{r}}$ implies oversampling of the respective spectrum $\mathbf{u}$. The correct interpretation of the arrival of scattered waves, though, can only be assured in that case using the fine sampling, i.e. assuming periodicity of the respective time series of period $2T_{\max}^{n\mathbf{r}}$, padded by zeros if necessary (compare section 3.3.2). Since all events $\mathbf{s}$ recorded at $(\mathbf{r}, n)$ potentially were also recorded at receivers other than $\mathbf{r}$, this ultimately results in the requirement of using *one global* frequency step for discretization of all involved spectra, namely $\Delta \hat{f} = \frac{1}{2T_{\max}}$ as defined in section 4.1. Hence, oversampling of spectra can only be avoided if all inverted seismograms have the same record length.

Although the kernels are computed within (but not necessarily throughout) the forward modelling domain from the solution of the forward problem, they are in-

tended to relate the data to model perturbations in the inversion domain. Hence those two domains need to be connected in a sensible way. The requirements on the characterizations of the forward modelling and the inversion domain regarding their interconnection are discussed now in the following, resulting in a generalized approach to relate forward and inversion domains.

Independent grids for forward and inverse problem

The general principle of separation in the waveform inversion procedure (section 2.3.2) encourages to keep the spatial description of the inversion domain completely separate from any spatial description of the solution of the forward problem. The resulting powerful advantages can only be realized by taking special care in certain respects, which here will be concentrated on.

Keeping the solution of forward and inverse problem separate, arbitrary methods for modelling wave propagation may solve the forward problem in this waveform kernel inversion. Methods which discretize the spatial domain in form of any sort of point or volume grids generally establish the resolution of the grid dependent on the following quantities. The degree of approximation of the wavefield in space and resolution in time (or frequency) must be taken into account, as well as the earth model used for modelling the wave propagation and any stability criteria dependent on the methods themselves. That is, the methods primarily must respect the "resolution of the data" (in terms of wavelength) *globally* throughout the medium *without knowing* (at the time of solving the forward problem) which power the data actually carry to resolve structure in certain parts of the inversion domain. It, hence, basically does not make any sense to simply use the global resolution of the forward wavefield as the model resolution for inversion without applying any regularizing control mechanisms on the model resolution. Thus, complete control over the model resolution at any point in the model domain which is independent of solving the forward problem is a desirable powerful mechanism. Furthermore, by separating the spatial discretization of the inversion domain from the solution of the forward problem, this solution can also be computed by non-grid wave propagation methods which approximate the wavefield by modal approaches using basis functions.

In special cases of forward methods which use some modal approximation of the forward wavefields by expanding them in terms of certain basis functions, the waveform kernels as well could be expressed in terms of these functions. In those cases, any sets of localizing wavelets for which the integration with the modal basis functions of the forward problem is known analytically (or can be easily computed), could be used for spatial description of the inverted model. This kind of model description would allow for varying resolution inside the model domain but would also be specifically dependent on the modal description of the forward solution, hence, the forward method. But since in the modularized concept followed in this work *no* assumptions should be made on the way of solving the forward problem, a general restriction on *nodal*, i.e. point-wise descriptions of the forward wavefields is inevitable. Since modal basis functions can always be evaluated on arbitrary point sets, it does not mean any

harm to introduce a point grid on which the wavefields and strains must be computed. However, *no* assumptions are made on the location of the grid points. They are defined by every forward method itself.

The arbitrary definition of a point forward grid consequently discards an analytical relation between forward and inversion grid and implies a numerical one. The integral relations derived in section 3.2 contain the model (update) parameters p^Δ as functions of space. This motivates a model description in terms of subsets of the model domain on which the model values are expressed by certain base functions. This kind of model expansion may range from global ansatz functions defined on the whole inversion domain Ω to localizing wavelets defined on (small) subsets of Ω. It only needs to be assured that their product with the kernels can be numerically integrated over the model (sub)domain, given the kernel values on the forward grid points. In the following, a strongly localizing model description is examined, assuming a disjoint subdivision $\dot{\cup}_e \Omega_e^k = \Omega$ of the inversion domain Ω into a set of volumetric cell elements Ω_e^k onto which parameter perturbations $p^\Delta \equiv p_{ek}^\Delta$ are constant. This is equivalent to expanding the model in terms of discontinuous "boxcar" wavelets w_e^k defined by

$$w_e^k(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \Omega_e^k \\ 0 & \text{if } \mathbf{x} \notin \Omega_e^k \end{cases} \quad .$$

This localizing definition of disjoint volumetric inversion grid cells comes closest to a nodal model description, since the particular model parameter values p_{ek}^Δ can be associated with some center of Ω_e^k. As an important feature of the separation of forward and inverse problem, the model description can be chosen *anew* in each iteration k (as indicated by the notation Ω_e^k). Thereby, the resolution of the model, i.e. the spatial resolution of the model expansion functions (here the general size of cells Ω_e^k), can be iteratively refined dependent on the frequency content of the data contained in D_k. With this definition of an inversion grid, eq. (3.16) becomes

$$u_n^\Delta(\mathbf{r}) = \int_\Omega \left[\sum_{p \in P} p^\Delta K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r}) \right] \mathrm{d}^3\mathbf{x}$$

$$= \sum_e \sum_{p \in P} \int_{\Omega_e^k} p^\Delta K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r}) \, \mathrm{d}^3\mathbf{x}$$

$$= \sum_e \sum_{p \in P} p_{ek}^\Delta \int_{\Omega_e^k} K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r}) \, \mathrm{d}^3\mathbf{x} \quad . \tag{4.1}$$

This approach allows to pre-integrate the sensitivity kernels onto the inversion grid before applying the above scattering relation containing the kernel related coefficients $\int_{\Omega_e^k} K_{n\mathbf{s}}^p(\mathbf{x}, \mathbf{r}) \, \mathrm{d}^3\mathbf{x}$ in any sensitivity analysis or inversion.

For the numerical integration of the sensitivity kernels given on the forward grid onto the inversion grid cells, a very general definition of integration rules is required, since no assumptions should be made on the shape of the inversion grid cells and their relation to the forward point grid. Such rules form the key in connecting forward and inversion grid and, hence, constitute a central component in the modularized concept of full waveform inversion presented in this thesis. For any other choice of model expansion functions, analogous rules for numerical integration must be provided.

General integration weights - the connection of forward and inversion grid

A description of integration weights as presented in this subsection is also found in Schumacher (2013). Integration weights meeting the general requirements posed above can be found in Levin (1999). There, rules for integration over arbitrary volumes of functions given on arbitrary point sets are computed following a "composite rule strategy": For subsets of the integration domain, integration formulas are derived which are as local and as stable as possible and are exact for polynomials of a certain fixed degree. The integrals of these polynomials over the subsets should be easily computable. Following Levin (1999) and, as far as sensible, their notation, linear integration weights A_i are now derived which numerically approximate the integrals $\int_{\Omega_e^k} K_{n\mathbf{s}}^p (\mathbf{x}, \mathbf{r}) \, \mathrm{d}^3 \mathbf{x}$ in eq. (4.1) by

$$\int_{\Omega_e^k} K_{n\mathbf{s}}^p (\mathbf{x}, \mathbf{r}) \, \mathrm{d}^3 \mathbf{x} \simeq \sum_{i=1}^{N} A_i K_{n\mathbf{s}}^p (\mathbf{x}_i, \mathbf{r}) \quad , \tag{4.2}$$

assuming the sensitivity kernel values are given on some set of points $\{\mathbf{x}_i\}_{i=1}^{N} \subset \Omega_e^k$ inside a cell.

The weights A_i are constructed by subdividing the cell $\Omega_e^k = \bigcup_{k=1}^{K} E_k$ into K disjoint subsets E_k and summing

$$A_i = \sum_{k=1}^{K} A_i^{(k)} \ , \ 1 \leq i \leq N \, ,$$

where separately for each E_k the N local weights $A_i^{(k)}$, $1 \leq i \leq N$, are calculated as follows: Fixing some polynomial degree m, a basis $\{q_j\}_{j=1}^{J}$ of the space Π_m of all polynomials in $\mathbb{R}^3$ with maximum total degree m is chosen, where $J = \binom{3+m}{m}$ is the dimension of space Π_m. $A_i^{(k)}$ are then defined as the components $a_i = A_i^{(k)}$ of vector $\mathbf{a} = D^{-1} E \left(E^t D^{-1} E \right)^{-1} \mathbf{c}$, where

$$D = 2\mathrm{Diag}\left\{ \eta \left(\|\mathbf{x}^* - \mathbf{x}_1\| \right), \dots, \eta \left(\|\mathbf{x}^* - \mathbf{x}_N\| \right) \right\} \quad ,$$
$$E_{i,j} = q_j (\mathbf{x}_i) \, , \quad 1 \leq i \leq N, \, 1 \leq j \leq J$$

and vector $\mathbf{c}$ contains the integrals of polynomials q_j over subset E_k, i.e. $c_j = \int_{E_k} q_j$. Furthermore, $\eta(r) = \exp \left(r^2 / h^2 \right)$ is a fast increasing weight function which gives the localizing properties of the weights, h is approximately the diameter of subset E_k and $\mathbf{x}^*$ is some center of it. This composite local approach of calculating global integration weights involves K solutions of a full linear system of order J.

One possible practical realization of the computation of the integration weights is to assume a few standardized geometries of inversion grid cells, e.g. some standard cube, tetrahedron etc., which every volumetric cell of the inversion grid can be transformed to. For any Ω_e^k containing wavefield sampling points $\{\mathbf{x}_i\}_{i=1}^{N}$ it, then, suffices to know the transformations $T(\mathbf{x}_i) = \boldsymbol{\xi}_i \in \Omega_{\mathrm{std}}$ of the $\mathbf{x}_i \in \Omega_e^k$ into a respective standard cell Ω_{std}, as well as Jacobian values $\mathcal{J}_i = \mathcal{J}_{T^{-1}} (\boldsymbol{\xi}_i)$ of the inverse transformation

T^{-1} evaluated at transformed points $\boldsymbol{\xi}_i$. Using these ingredients, the approximate integration in eq. (4.2) can be expressed in terms of integration weights A_i^{std} computed by the above method for points $\boldsymbol{\xi}_i$ in standard cell Ω_{std}:

$$
\begin{aligned}
\int_{\Omega_e^k} K_{n\mathbf{s}}^p(\mathbf{x},\mathbf{r})\,\mathrm{d}^3\mathbf{x} &= \int_{\Omega_{\text{std}}} K_{n\mathbf{s}}^p\left(T^{-1}(\boldsymbol{\xi}),\mathbf{r}\right) \mathcal{J}_{T^{-1}}(\boldsymbol{\xi})\,\mathrm{d}^3\boldsymbol{\xi} \\
&\simeq \sum_{i=1}^N A_i^{\text{std}}\, K_{n\mathbf{s}}^p\left(T^{-1}(\boldsymbol{\xi}_i),\mathbf{r}\right) \mathcal{J}_{T^{-1}}(\boldsymbol{\xi}_i) \\
&= \sum_{i=1}^N A_i^{\text{std}}\, K_{n\mathbf{s}}^p(\mathbf{x}_i,\mathbf{r})\, \mathcal{J}_i \quad .
\end{aligned}
\tag{4.3}
$$

This way, weights A_i for cell Ω_e^k which may have arbitrary geometric shape, can in practice be realized by weights $A_i^{\text{std}}\mathcal{J}_i$ calculated for cell Ω_{std} of known standardized geometric shape, given the knowledge about how to transform Ω_{std} into Ω_e^k.

The general separation of forward and inversion grid without any further assumptions only becomes possible due to this kind of general integration rules which, theoretically, do not assume anything about the point and volume geometry, except a sufficient number N of independently distributed points $\mathbf{x}_i$ inside each inversion grid cell in order to assure solvability of the systems of equations of order J. This way, the pre-integration onto the inversion grid cells of the sensitivity kernels computed on the forward grid indeed connects the spatial descriptions of forward and inverse problem.

4.2.3 Solving a structural inverse problem: updating the model

Solving a structural inverse problem according to the general opening remarks of chapter 2, means to deduce the structural earth model which caused the data functionals to have the respective measured values. In the kth iteration of this iterative full waveform inversion procedure, the step of solving the inverse problem finds an improvement $\mathbf{m}_k^\Delta = \mathbf{m}_k - \mathbf{m}_{k-1}$ of the model from data subset $\{d_i | i \in D_k\}$ as well as the solution of the forward problem. This model improvement is derived by exploiting the discretized scattering relation in eq. (4.1) for each datum d_i, $i \in D_k$, interpreting data residuals $r_i = d_i - s_i$ as the scattered wavefield $\mathbf{u}^\Delta$.

The above introduction of inversion grid $\{\Omega_e^k\}_e$ along with model parameterization P that is inverted for in the waveform inversion define a discrete description of the structural earth model for iteration k. Hence, model update $\mathbf{m}_k^\Delta$ can be discretely defined by the unknown values p_{ek}^Δ which may be sorted by some iteration-dependent fixed index mapping idx_k^{m} similar to the indexing of the data. For $j = \text{idx}_k^{\text{m}}(p,e)$ this introduces the notation

$$
m_{jk}^\Delta = p_{ek}^\Delta \quad .
$$

The discretization of the model domain results in an equation of type (4.1) for each data residual r_i, $i \in D_k$, which are linear in $p_{ek}^\Delta = m_{jk}^\Delta$ allowing to set up a linear

system of equations

$$r_i = K_{ij} m_{jk}^\Delta \quad , i \in D_k \, . \tag{4.4}$$

For each $i = \mathrm{idx}^\mathrm{d}\left(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q\right)$ and $j = \mathrm{idx}_k^\mathrm{m}(p, e)$, the entry K_{ij} of system matrix $\mathbf{K}$ is by eq. (4.1) defined by

$$K_{ij} = \begin{cases} \int_{\Omega_e^k} \Re\left(K_{n\mathbf{s}_l}^p\left(\mathbf{x}, \mathbf{r}_m\right)\right) \mathrm{d}^3\mathbf{x} & \text{if } \mathfrak{c}_q \text{ indicates real part,} \\ \int_{\Omega_e^k} \Im\left(K_{n\mathbf{s}_l}^p\left(\mathbf{x}, \mathbf{r}_m\right)\right) \mathrm{d}^3\mathbf{x} & \text{if } \mathfrak{c}_q \text{ indicates imaginary part,} \end{cases} \tag{4.5}$$

where $\Re$ and $\Im$ denote real and imaginary part, respectively, of the complex-valued kernel which is evaluated at frequency f_p. Solving linear system (4.4) in some way yields a model update vector $\mathbf{m}_k^\Delta$. It becomes clear now why the data samples as well as the respective sensitivity kernel values were defined to be real-valued: Only a real-valued linear system of equations assures a real-valued solution $\mathbf{m}_k^\Delta$, which is required in case of parameterizing the earth model by real-valued elastic constants and density. Splitting up the complex numbers by real and imaginary part is straightforward and was chosen because usually complex numbers are dealt with in this representation. In fact, any other real-valued representation of complex numbers could be chosen as well.

Regularizing the inverse problem by additional constraints

However, the optimal mathematical solution of this (possibly overdetermined / underdetermined) linear system of equations is usually not physically sensible. Therefore, additional constraints should be added in form of extra equations as suggested in the following. The application of additional smoothing constraints to linear system (4.4) has several advantages. By such a reduction of degrees of freedom, the inverse problem is regularized. Although this could also be achieved by choosing a coarser inversion grid $\{\Omega_e^k\}_e$ (i.e. larger, hence, fewer inversion grid cells), smoothing avoids solutions with large localized spikes which are usually produced in a mathematically optimal solution in the absence of any physics. Such strong and spiky model perturbations would, furthermore, run the risk of violating the linearization assumptions.

Any form of linear smoothing of a specific model update value $m_{\hat{j}k}^\Delta$, $\hat{j} = \mathrm{idx}_k^\mathrm{m}(\hat{p}, \hat{e})$, is to force it to equal some average of values m_{jk}^Δ, $j = \mathrm{idx}_k^\mathrm{m}(\hat{p}, e)$, of same parameter $\hat{p}$ on cells e neighbouring cell $\hat{e}$. That is,

$$m_{\hat{j}k}^\Delta = \sum_{j \in N_{\hat{e}}} w_j m_{jk}^\Delta \quad , \tag{4.6}$$

where set $N_{\hat{e}} = \left\{\,\mathrm{idx}_k^\mathrm{m}(\hat{p}, e)\,|\,e \text{ is neighbour of } \hat{e}\right\}$ and w_j are cell specific averaging weights depending on cells $\hat{e}$ and e, with $\sum_{j \in N_{\hat{e}}} w_j = 1$. The introduction of such a smoothing equation for each model value eventually means a coupling of all model values of the certain parameter $\hat{p}$. In order to even increase the influence of the other model values on the smoothed value, the "neighbourhood" over which a particular model value is smoothed could be extended to any larger zone around cell $\hat{e}$.

In case of cubic inversion grid cells aligned along orthogonal directions in Cartesian space, the form of smoothing eqs. (4.6) may be motivated by the following. Choosing the six direct face neighbours for definition of set $N_{\hat{e}}$ and constant weights $w_j = \frac{1}{6}$, (4.6) is equivalent to the vanishing of the Laplace operator Δ at cell $\hat{e}$ in finite difference approximation applied to the model values as a function of space:

$$\Delta m^{\Delta}(\hat{e}) = \frac{\partial^2 m^{\Delta}}{\partial x^2}(\hat{e}) + \frac{\partial^2 m^{\Delta}}{\partial y^2}(\hat{e}) + \frac{\partial^2 m^{\Delta}}{\partial z^2}(\hat{e}) = 0 \quad,$$

assuming a notation $m^{\Delta}(\hat{e})$ of the model update value of a fixed parameter $\hat{p}$ as a function on the inversion grid cells, evaluated at the center of cell $\hat{e}$. The equivalence can be seen from separately approximating the partial second derivatives by finite differences. E.g. assuming a distance h of center cell $\hat{e}$ to both its x-direction neighbour cells e_{x+} and e_{x-},

$$\frac{\partial^2 m^{\Delta}}{\partial x^2}(\hat{e}) = \frac{\frac{m^{\Delta}(e_{x+}) - m^{\Delta}(\hat{e})}{h} - \frac{m^{\Delta}(\hat{e}) - m^{\Delta}(e_{x-})}{h}}{2h}$$

$$= -\frac{m^{\Delta}(\hat{e})}{h^2} + \frac{m^{\Delta}(e_{x+}) + m^{\Delta}(e_{x-})}{2h^2} \quad. \tag{4.7}$$

Summing eqs. (4.7) for all directions of space (always assuming the same distance h) and setting to zero, indeed yields

$$0 = -3\frac{m^{\Delta}(\hat{e})}{h^2} + \frac{1}{2h^2}\sum_{\substack{e \text{ is } x\text{-nb.} \\ \text{or } y\text{-nb.} \\ \text{or } z\text{-nb.}}} m^{\Delta}(e)$$

$$\Leftrightarrow \qquad m^{\Delta}(\hat{e}) = \frac{1}{6}\sum_{\substack{e \text{ is } x\text{-nb.} \\ \text{or } y\text{-nb.} \\ \text{or } z\text{-nb.}}} m^{\Delta}(e) \quad,$$

which is eq. (4.6) for the chosen direct face neighbours and constant weights $w_j = \frac{1}{6}$.

On external boundaries of the inversion domain where there are actually neighbours missing, as well as possible internal boundaries over which the model values should not be smoothed, different choices of the weights w_j may define different kinds of "boundary conditions". Demonstrated by use of the regular Cartesian inversion grid cells as in the above paragraph, two kinds of conditions are discussed in the following, here called *continuity boundary conditions* and *zero boundary conditions*. First of all assume that neighbour e_{x-} of a center cell $\hat{e}$ is missing, e.g. at either an outer or inner boundary of the model domain. Choosing the regular average, i.e. constant average weights for all present neighbours of cell $\hat{e}$, eq. (4.6) gives

$$m^{\Delta}(\hat{e}) = \frac{1}{5}m^{\Delta}(e_{x+}) + \frac{1}{5}\sum_{\substack{e \text{ is } y\text{-nb.} \\ \text{or } z\text{-nb.}}} m^{\Delta}(e)$$

$$\Leftrightarrow \qquad \frac{6}{5}m^{\Delta}(\hat{e}) - \frac{1}{5}m^{\Delta}(\hat{e}) = \frac{1}{5}m^{\Delta}(e_{x+}) + \frac{1}{5}\sum_{\substack{e \text{ is } y\text{-nb.} \\ \text{or } z\text{-nb.}}} m^{\Delta}(e)$$

$$\Leftrightarrow \qquad \frac{6}{5}m^\Delta\left(\hat{e}\right) = \frac{1}{5}m^\Delta\left(e_{x+}\right) + \frac{1}{5}m^\Delta\left(\hat{e}\right) + \frac{1}{5}\sum_{\substack{e \text{ is } y\text{-nb.} \\ \text{or } z\text{-nb.}}}m^\Delta(e)$$

$$\Leftrightarrow \qquad m^\Delta\left(\hat{e}\right) = \frac{1}{6}m^\Delta\left(e_{x+}\right) + \frac{1}{6}m^\Delta\left(\hat{e}\right) + \frac{1}{6}\sum_{\substack{e \text{ is } y\text{-nb.} \\ \text{or } z\text{-nb.}}}m^\Delta(e) \quad ,$$

implicitly assuming the presence of the missing neighbour e_{x-} having model value $m^\Delta\left(e_{x-}\right) = m^\Delta\left(\hat{e}\right)$. It can be analogously seen for an arbitrary number of missing neighbours that, hence, the choice of the regular average with constant weights over all present neighbours assume some kind of *continuity boundary conditions* with which central value $m^\Delta\left(\hat{e}\right)$ is continued across the boundary. Note, however, that in case of internal boundaries this does *not* mean to smooth over the boundary, since the implicitly assumed continuity value is allowed to be different at either side of the boundary. Hence, those *continuity boundary conditions* are suitable for inner boundaries across which smoothing is not desired. On outer boundaries of the model domain, on the other hand, different kind of conditions may be sensible to apply, especially in case of inverting for changes of the model values. Outside the model domain, namely, it should make sense to assume zero changes since a reasonable model domain must completely contain the domain which influences the data. Assuming the presence of missing outer neighbours e having model values $m^\Delta(e) = 0$, leads in the simple Cartesian geometry of the cells to constant weights $w_j = \frac{1}{6}$ *regardless* of the number of actually present neighbours. This kind of *zero boundary conditions* may be sensibly applied on outer boundaries of the model domain, except for free surfaces, smoothing the model values on that boundaries additionally down to zero.

In case of more elaborate inversion grid cell geometry, Wardetzky et al. (2007) give an overview over different approaches to define discrete Laplace operators in the broader field of computer graphics. Such approaches may, mostly derived for two-dimensional settings, account for direction of and distance to the particular neighbours.

Writing eqs. (4.6) as a matrix multiplication $\mathbf{R}\,\mathbf{m}_k^\Delta = 0$ introduces the smoothing regularization matrix $\mathbf{R}$: For arbitrary parameter p and $i = \mathrm{idx}_k^\mathrm{m}(p,e)$

$$R_{ij} = \begin{cases} -1 & \text{if } j = i, \\ w_j & \text{if } j \in N_e, \\ 0 & \text{otherwise.} \end{cases}$$

Adding equations $\mathbf{R}\,\mathbf{m}_k^\Delta = 0$ to the system (4.4), yields and overdetermined linear system of equations

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{K} \\ \mathbf{R} \end{pmatrix} \mathbf{m}_k^\Delta \quad .$$

Here, $\mathbf{r}$ must not be confused with the notation of receivers but denotes the vector $\mathbf{r} = (r_i)_{i \in D_k}$ of data residuals consistent with the rows of kernel matrix $\mathbf{K}$. $\mathbf{0}$ denotes

a zero vector of sufficient dimensionality. A solution $\hat{\mathbf{m}}_k^\Delta$ of such a system is generally found in a least squares sense by minimizing

$$\left\| \begin{pmatrix} \mathbf{r} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{K} \\ \mathbf{R} \end{pmatrix} \hat{\mathbf{m}}_k^\Delta \right\|^2 = \min_{\mathbf{m}_k^\Delta} \left\| \begin{pmatrix} \mathbf{r} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{K} \\ \mathbf{R} \end{pmatrix} \mathbf{m}_k^\Delta \right\|^2$$

$$= \min_{\mathbf{m}_k^\Delta} \left(\left\| \mathbf{r} - \mathbf{K}\,\mathbf{m}_k^\Delta \right\|^2 + \left\| \mathbf{R}\,\mathbf{m}_k^\Delta \right\|^2 \right) \quad .$$

It becomes clear now that the additional regularization equations cause a penalty term in the objective function that is to be minimized. The influence of this penalty term $\| \mathbf{R}\,\mathbf{m}_k^\Delta \|^2$ in the course of interactively conducting the iterative inversion process can be controlled by introducing a global scaling factor γ on scaling eqs. (4.6). In terms of the minimization problem, this finally means to solve

$$\min_{\mathbf{m}_k^\Delta} \left(\left\| \mathbf{r} - \mathbf{K}\,\mathbf{m}_k^\Delta \right\|^2 + \gamma^2 \left\| \mathbf{R}\,\mathbf{m}_k^\Delta \right\|^2 \right) \quad , \tag{4.8}$$

or equivalently solve

$$\begin{pmatrix} \mathbf{r} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{K} \\ \gamma\mathbf{R} \end{pmatrix} \mathbf{m}_k^\Delta \tag{4.9}$$

in a least-squares sense.

The final minimization problem was already motivated in section 2.3.1. Data uncertainties as in eq. (2.4) are discussed later in section 4.3.3 (cf. eq. (4.11)). Note that the smoothing intensity factor γ corresponds to $\frac{1}{\sqrt{\nu}}$, where ν is the Lagrange multiplier introduced in eq. (2.2). In order to additionally force the model update values $\mathbf{m}_k^\Delta$ not only to be smooth but to be small as well, the diagonal terms of regularization matrix $\mathbf{R}$ could be increased which, however, is not done here.

Updating the model

Having deduced a vector of model update values $\mathbf{m}_k^\Delta$ as presented above, eventually the currently inverted earth model $\mathbf{m}_{k-1}$ must be updated. Since the spatial descriptions of the model values given by choice of inversion grid $\{\Omega_e^k\}_e$ may be defined anew in each iteration step, $\mathbf{m}_{k-1}$ and $\mathbf{m}_k^\Delta$ must be properly connected in order to obtain the new updated model $\mathbf{m}_k$.

Assume $\mathbf{m}_{k-1}$ was derived on inversion grid $\{\Omega_e^{k-1}\}_e$ of the previous iteration $k-1$, or is given on some starting grid in case of starting model $\mathbf{m}_0$. Then, already for the solution of the forward problem in iteration k and possibly for the kernel computation (e.g. in case of isotropic parameterization $P_{\text{Veloc}} = \{v_p, v_s, \rho\}$), model $\mathbf{m}_{k-1}$ must be known on the forward grid on which the kernels are computed. The next section 4.2.4 presents strategies how this can be accomplished. Every inversion grid cell Ω_e^k, however, is connected by the general integration rules to those points $\{\mathbf{x}_i\}_{i=1}^N \subset \Omega_e^k$ of the forward grid contained in it (derived above, eq. (4.2)). Thus, by

help of integration weights A_i the model values $\mathbf{m}_{k-1}(\mathbf{x}_i)$ of reference model $\mathbf{m}_{k-1}$ given on forward grid points contained in Ω_e^k may be interpolated onto cell Ω_e^k of the new inversion grid by

$$m_{j,k-1} = \sum_{i=1}^{N} \frac{A_i}{\sum_{n=1}^{N} A_n} \, \mathbf{m}_{k-1}(\mathbf{x}_i) \quad ,$$

where $j = \mathrm{idx}_k^{\mathrm{m}}(p,e)$ are the respective indices of model update values m_{jk}^{Δ} defined on cell e for each parameter p. Finally, the inverted model update values m_{jk}^{Δ} can be imposed onto the interpolated reference model values $\mathbf{m}_{j,k-1}$ producing the new inverted earth model $\mathbf{m}_k$ given on the inversion grid $\{\Omega_e^k\}_e$.

In general, the model resolution should become finer in the course of the iterative inversion process adding higher frequencies to the data, i.e. cells Ω_e^k are smaller than cells Ω_e^{k-1}. Hence, the above interpolation of the reference model values $m_{k-1}(\mathbf{x}_i)$ given on the forward grid points onto inversion grid cells Ω_e^k should *not* remove any relevant model structure by smoothing it out. The general smoothing property of the interpolations between the different forward and inversion grids in the course of the iterations, however, is an important stability feature. This is explained and demonstrated later in the example section 5.2.

4.2.4 Connecting the iterations: flexible 3D interpolation from inversion grid to new forward grid

The separation of forward and inverse problem in the iterative inversion procedure demands to transfer the inverted model $\mathbf{m}_k$ given on the inversion grid $\{\Omega_e^k\}_e$ to the model description as required by the forward method in the next iteration $k+1$ to solve the forward problem.

For grid-based nodal methods, an interpolation of inverted model values m_{jk} given on the inversion grid cells Ω_e^k onto the the new forward grid suits to connect the iterations. The volumetric inversion grid is, in this case, regarded as a nodal point grid, assigning the model values to some centers of the cell elements. A very general way of interpolation from one point grid to another is given by Shepard (1968) where an interpolation method is presented using weighted averages based on inverse-distance weighting which accounts for issues of nearby points, direction and slope. Many interpolation methods based on Shepard (1968) or modifications of it may be found in the literature (e.g. Franke and Nielson, 1980; Łukaszyk, 2004; Masjukov and Masjukov, 2005).

For forward methods using a modal description of the model, the required coefficients of the respective model basis functions could be computed by evaluating those functions at the centers of the inversion grid cells to which the inverted model values are associated. If additionally the inverted model was represented by more complex ansatz functions than the here chosen "boxcar" wavelets, a translation of the inverted model description to the one required by the forward method might become more

complicated and is not discussed here.

4.3 Discussion and issues in practice and implementation

In the practical realization of the iterative full waveform inversion procedure presented in the previous section 4.2 various aspects need to be considered regarding implementation, the connection between time and frequency domain and data processing. Those characteristics are discussed now in general, before in section 5.1 the explicit software implementation ASKI is presented with which the synthetic waveform inversion of section 5.2 is conducted.

4.3.1 Modularization and object-oriented programming

The general concept of separation, as motivated in section 2.3.2 and presented in section 4.2 encourages modularization and object-oriented strategies of programming when realizing the inversion procedure in form of software implementations. As a side-effect, the produced software is sustainable and easy to maintain, adapt, extend and develop. The full waveform inversion procedure based on waveform sensitivity kernels as presented in this thesis was implemented in the software package ASKI (section 5.1) following a modularized concept by object-oriented programming in which a general abstraction of units of functionality separates independent operations, as presented in the following.

The forward problem is solved by external forward modelling codes simulating seismic wave propagation by arbitrary methods and writing the required wavefield spectra (cf. section 4.2.1) as files to hard disc. The access to those files is defined in generic modules which have distinct software interfaces branching to method dependent routines which specifically implement how to deal with any method related output and structure. This way, the characteristics of each forward method is optimally taken into account, as no general assumptions on possible files and structures are made, e.g. on the kind of output – parallel/serial – or the way to communicate the wavefields. Some methods may derive quantities like regular point grids using just a few numbers, others may need to read in complete unstructured grids and configurations. Installing these modularized interfaces, the software package can be extended by arbitrary forward codes which themselves may optimally define concepts of communicating information as long as they meet the requirements of the software interfaces to the rest of the software package.

Since the integration rules centrally connect the wavefields given on the forward grid and the integrated kernels on the inversion grid, any modularized realization in terms of software should account for a close interconnection of forward grid, integration weights and inversion grid. In ASKI, the inversion grid module is central: it is generically designed, branching to submodules for specific types of inversion grids, dependent on the general geometrical shape of the cells and the characteristics of

the inversion domain. Spherical and Cartesian inversion domains may be supported using any kinds of coordinates like polar, cylindrical, or Cartesian ones. All specific inversion grid modules honor generally defined interfaces which include locating the forward grid points inside the inversion grid cells, providing their transformed points and Jacobian values w.r.t. the standard cell of respective type, as well as transforming point and volume coordinates to a geometric standard reference frame for reasons of graphical plotting. This indeed demonstrates the independence of forward and inverse problem, since any type of inversion grid may be used in connection with any forward code, provided the well defined interfaces to the inversion grid submodules such as coordinate formats of forward grid points etc. are honored.

In addition to the property of generally combining arbitrary kinds of forward methods with arbitrary shapes of inversion grids, the object-oriented approach at the same time allows for special handling of special forward methods. Some methods may provide both, a point grid on which the forward problem is solved as well as integration operations onto some volume element structure, like finite element methods, for example. For those kinds of forward methods, a method specific inversion grid submodule is able to provide a distinct structure of inversion grid cells and knowledge about the contained forward grid points and can serve to take advantage of optimal integration accuracy providing own integration weights. Furthermore, the general performance is enhanced if no computations need to be conducted to localize forward grid points inside inversion grid cells and to construct integration rules. This, however, is of lesser significance to the whole process.

In the software package ASKI, the kernel computation is done by a separate binary program reading in the wavefield spectra from the files written by the forward code. The kernels are pre-integrated onto the inversion grid cells, yielding the entries of the kernel matrix (as referred to in sections 4.2.2 and 4.2.3). As pointed out in section 2.3.2, it is feasible to store the entries of the kernel matrix to disc just as well, since this opens very flexible possibilities of sensitivity and resolution analysis as well as variable hands-on regularization of the inverse problem. The use of different elastic (and possibly anelastic) model parameterizations for which kernels are computed, was realized by an independent software module characterizing such model parameterizations. Additionally, those parameterizations must be supported by the modules computing the waveform kernels and modules providing the unperturbed reference model.

Finally, the regularized inverse problem is solved in each iteration by yet another separate ASKI module which reads in the kernel matrix, defines smoothing constraints and is quite cheap compared with solving the forward problem (compare table 5.1 in section 5.2). By the general separation, hence, the inversion step of each iteration can be repeated at relatively low costs applying different intensities of smoothing or detecting and removing data samples which cannot be fitted well (cf. section 4.3.3). This allows the user to produce sensible models in each iteration step by full manual control over inverted data samples and any auxiliary conditions.

4.3.2 Solving the forward problem in practice: issues of numerical time-stepping methods

In general there are (semi-)analytical and numerical methods in time or frequency domain used to solve the forward problem, i.e. to simulate seismic wave propagation within some earth model. For complex 3D-heterogeneous media, however, only numerical methods are suited as there are no general analytical solutions of the wave equation in those cases. In order to compute Green function spectra as required for the sensitivity kernels, wavefields excited by single force sources that are impulsive in time must be simulated (cf. eq. (3.5)). Analytical and frequency-domain methods can represent such forces by analytical incorporation of δ-distributions in their equations. This section 4.3.2 discusses certain issues of numerical time-stepping methods related to approximating such force distributions and transforming the wavefields to the frequency domain. The approximation of point forces in space (i.e. $\delta(\mathbf{x} - \mathbf{r})$) in numerical grid-based methods, however, is a different issue (though widely established in practice) and is not discussed here.

Numerical time-stepping methods solve the wave equation iteratively stepping forward in time, using an explicit discretized source time function with which the force field is excited. Naturally, an infinite impulse cannot be represented sufficiently by discretized samples in a computer. For stability reasons, most time-stepping methods must approximate such an impulse by some narrow continuous source time function (e.g. a thin Gaussian). This function should exhibit the property that its integral is 1, just like the integral over a δ-impulse. Numerical filtering test with the time-stepping spectral-element method SPECFEM3D (Tromp et al., 2008), however, showed that using narrow continuous approximations to a δ-distribution still can produce significant phase shifts. A solution to this issue is given by a deconvolution technique. Before this technique is described below, however, the computation of the required wavefields and strains in the frequency domain by numerical time-stepping forward methods is discussed next.

The time-domain wavefields simulated by numerical time-stepping forward methods must be transformed to the frequency domain by a Fourier transform. Since often a much finer time sampling is used than imposed by the Nyquist theorem, optimized algorithms like the Fast Fourier Transform may, however, not be feasible to use here: For any such transformation requiring the complete time series it may not be sensible in practice to keep heavily oversampled time series in memory, since the memory requirements soon become very large for larger problems. Resampling the time series, on the other hand, would require anti-aliasing filtering on-the-fly when simulating Green functions: The narrow source time functions applied for this purpose, namely, produce in general very high frequencies in the wavefield. Hence, a Fourier transform applied on-the-fly can be used for such methods, as done by SPECFEM3D for ASKI. Additionally, data for only a small frequency window may be inverted in each iteration (cf. introduction to section 4.2) which requires kernels, thus waveforms, at few specific frequencies only. Therefore, a direct evaluation of the Fourier integral by numerical integration could even be more efficient than transformation algorithms optimized to

produce complete spectra:

$$\int_{-\infty}^{\infty} s(t)\, e^{-i\omega t}\, \mathrm{d}t \;\simeq\; \Delta t \sum_{j=1}^{N_T} s(t_j)\, e^{-i\omega t_j} \quad , \tag{4.10}$$

where the time domain (hence time series $s(t)$) is discretized by N_T samples t_j of equal spacing Δt. Note that implicitly this approximation assumes $s(t) = 0$ outside the time interval $[t_1, t_{N_T}]$ and the summation can be conducted on-the-fly during the simulation of wave propagation.

In order to finally compute Green function spectra correctly, numerical time-stepping methods should simulate the wave propagation for some stable approximation of a δ-impulse source time function. Afterwards the resulting transformed wavefield spectra must be corrected for this approximation, as already motivated above. This can be achieved by deconvolving the explicitly chosen source time function from the wavefield spectrum. This deconvolution is accomplished easily in the frequency domain dividing the complex-valued wavefield spectrum by the spectrum of the source time function at each frequency. For numerical reasons, the complex numbers by which is divided should be close to $1 \in \mathbb{R}$, which is one reason why the methods should choose a stable approximation to an infinite impulse which is already close to that property. There are alternatives to approximate impulsive source time functions. For example an approximate Heaviside can be chosen as a source time function (realized by a steep error function), simultaneously deriving the wavefield w.r.t. time, as is done by SPECFEM3D for ASKI.

For time-domain methods, some explicit simulation times must be chosen to produce the forward wavefields originating from seismic sources as well as the Green functions originating from the receiver components. Although the produced spectra must be discretized by some global frequency step $\Delta \hat{f}$ (cf. sections 3.3.2 and 4.2.2), these simulation times do not necessarily need to equal $\frac{1}{2\Delta \hat{f}}$, but can be chosen individually smaller under following conditions. Assuming the notation of section 4.2.2, the simulation time for the Green function of receiver component $(n, \mathbf{r})$ can be chosen as $T_{\max}^{n\mathbf{r}}$. For the forward wavefield from source $\mathbf{s}$ it can be chosen as $T_{\max}^{\mathbf{s}}$, which consistently defines the maximum length of seismograms for event $\mathbf{s}$ recorded at any receiver component. Although both, $T_{\max}^{n\mathbf{r}}$ and $T_{\max}^{\mathbf{s}}$, may be smaller than $T_{\max} = \frac{1}{2\Delta \hat{f}}$, a direct Fourier transform like eq. (4.10) for fine frequency sampling by $\Delta \hat{f}$ automatically accounts for padding zero waveform at (all) times larger than the simulation time. This way, the simulation times are efficiently kept as small as possible. The resulting kernels correctly represent the potential scattering, since the forward and Green function spectra throughout the medium reflect the correct (maximum) propagation time of the waves. Hence, the fine spectral discretization by the global $\Delta \hat{f}$ imposed by the kernel stability criterion of section 3.3.2 does *not* effect simulation times, which is a significant factor of the overall computational efficiency. It only requires evaluation and storage of the spectra for each of the frequencies.

4.3.3 Data related aspects

This section 4.3.3 collects some aspects of the practical realization of the waveform inversion related to data such as source and receiver influences, data uncertainties and impacts of the separation principle on data selection.

Modelling transfer functions of sources and receivers

In practice, the data may not be measurements of ground displacement as assumed formally here in the derivation of the waveform inversion process. In addition to the perturbations of the structural earth model that are to be detected by the inversion, the measured data values are influenced by any characteristics of seismic sources and recording systems. The time evolution of the source, i.e. the source time function, and the instrument responses of the receiver components can be modelled by appropriate filters applied to the synthetic wavefields in order to model their influences on the measured data. In the frequency domain this filtering is easily accomplished by the multiplication of the forward wavefield and Green function spectra by the respective complex-valued filter constants of source and receiver transfer functions. For this kind of frequency-domain inversion, however, the recorded data measured in the time domain must always be transferred to the frequency domain.

If such source and receiver responses are not known and cannot be approximated, they may be inverted for jointly in the course of the iterative inversion process. Forbriger (2003) e.g. iteratively inverts for the source wavelet by determining appropriate field-device responses in each forward calculation of synthetic predictions. When waveform inversion is done using sensitivity kernels as in this work, it is advantageous to compute impulse responses of the medium not only for the backpropagations from the receivers but also for the forward propagations from the sources, i.e. forward wavefield spectra and synthetic data. Thus, source and receiver transfer functions can be directly multiplied with the pre-integrated frequency-domain sensitivity kernels. Also the synthetics $(s_i)_{i \in D_k}$ can be thus adapted to the data before computing data residuals $r_i = d_i - s_i$. The complete separation in the inversion process of forward modelling and updating the model, hence, allows to iteratively invert for source or receiver transfer functions along with the structural inversion without much effort: After computing universal impulse response solutions of the forward problem in a certain iteration and storing those spectra to disc, any inversion of source wavelet or receiver responses may be done separately by external tools as an intermediate step, possibly updating the responses found in the previous iteration. Thereafter, the model update can be derived applying the newly updated filters to the sensitivity kernels and synthetic predictions .

Alternatively, any direct hands-on "inversion" of source and receiver responses can always be achieved by repeatedly solving the linear system (4.9) applying slightly modified filter values at a time and manually decide from looking at the result what filters to use. Below it is pointed out that solving this linear system is relatively cheap

compared with the solution of the forward problem.

Data uncertainties

Since measured data are contaminated by noise, it is desirable to downweight the contribution of the particular data samples to the misfit according to the noise. For this purpose, the uncertainties $\sigma_i > 0$ can be introduced for each datum d_i to quantify its quality. They are easily accounted for in the waveform inversion by scaling equations (4.4) by factor $\frac{1}{\sigma_i}$. The minimization problem (4.8), then, changes to

$$\min_{\mathbf{m}_k^\Delta} \left(\left\| \Sigma^{-1} \left(\mathbf{r} - \mathbf{K}\,\mathbf{m}_k^\Delta \right) \right\|^2 + \gamma^2 \left\| \mathbf{R}\,\mathbf{m}_k^\Delta \right\|^2 \right) \quad , \tag{4.11}$$

where $\Sigma = \mathrm{Diag}(\sigma_i)$ is a diagonal matrix constructed from the uncertainties σ_i, $i \in D_k$, corresponding to the order of data samples inverted in iteration k (i.e. the rows of $\mathbf{K}$ and $\mathbf{r}$). Note that the choice of the frequency domain for inversion is in particular suitable for such uncertainty quantification, since the uncertainties themselves usually vary with frequency.

Selection of data

As already indicated above at the end of section 4.3.1, the complete separation in the inversion process of updating the model from solving the forward problem and computing the kernels allows to test for different data subsets in a certain iteration by repetitive conduction of the inversion step: Just a new linear system for the changed data subset needs to be set up by reading (different) sensitivity kernel values from file and the regularized system has to be solved, which in general is very cheap compared with the solution of the forward problem. This offers the chance of discarding certain data values, e.g. which are of worst quality, and comparing the resulting model to the one yielded by incorporating those data.

By the general modularized approach, it is even thinkable in each iteration step to do a full singular value decomposition of the kernel matrix $\mathbf{K}$ in order to detect those data which are most sensitive to the derived model perturbations and only use them for inversion. After all, the data alone decides on which features of the model can be resolved.

"Part-waveform" inversion

As a side note, inverting data in the frequency domain potentially allows for time-windowing, too. This permits to conduct some kind of "part-waveform" inversion. If only certain time windows in a certain measured seismogram should be explained by earth structure, a specific windowing could be applied to the measured seismograms

as well as respective time-domain synthetics *prior to* transforming them to the frequency domain, producing equivalent data residuals in the frequency domain which contain only wavelet information of that specific time window. In order to appropriately deduce model updates from those time-windowed data residuals, the sensitivity kernels must be computed in the frequency domain accounting for the respective time windows: Windowing in the time domain, i.e. multiplication of the time series by some certain windowing function is equivalent to a convolution operation in the frequency domain, by the convolution theorem. Therefore, the spectrum of the time windowing function must be computed by Fourier transform. This spectrum then is convolved in the frequency domain with the wavefield spectra of forward and backward propagation prior to multiplying them to compute kernels. At this stage, however, source and receiver responses must already be accounted for, which makes it complicated to combine this approach of time windowing with the flexible modelling of source and receiver transfer functions as presented above. Furthermore, complete spectral information of the waveforms is required in this case and not only a small frequency window, which ruins the computational advantages presented in section 3.3.1.

This kind of inversion of certain windows of waveforms has not yet been implemented or tested and is not discussed any further here.

4.3.4 Expenses and efficiency: storage is vital

So far in this thesis, many advantages of the inversion process separating the steps of solving forward and inverse problem have been presented. Those benefits, however, must be paid by the demands presented now in the following.

As already referred to in section 2.3, Chen et al. (2007a) compare the "adjoint-wavefield" method with the "scattering-integral" method which is comparable to the method of full waveform inversion presented here. They showed that the adjoint-wavefield method in conjugate-gradient implementation after Tromp et al. (2005) needs overall 6–7 iterations to realize one iteration of the scattering-integral method. Furthermore it requires a number of wave propagation simulations solely dependent on the number of sources. The scattering-integral method, on the other hand, separately requires one simulation for each source and one per receiver component. Hence, only when inverting datasets which have a significantly higher number of involved receiver components than sources, the adjoint-wavefield method needs less simulations than the scattering-integral method. Investigating Chen et al. (2007a) yields that therefore the number of involved receiver components must be least 35–42 times as high as the number of involved sources. In all other cases of less involved receiver components, the scattering-integral method is more efficient in terms of numbers of simulations. In order to take advantage of that, however, all wavefields needs to be store throughout the medium. This requires a tremendous amount of disc space and many input / output operations compared with the adjoint-wavefield method.

In the same way here, the access to permanent storage devices is vital for the application of the full waveform inversion procedure presented in this thesis. Especially

the forward wavefield and Green function spectra require much storage in the course of the iterations (compare table 5.1 below in section 5.2). By allowing to redefine the resolution of the forward grid in every iteration step according to frequency content of the currently inverted data, the storage requirements can be reduced to absolute necessities during the inversion process.

In comparison to the wavefields, the kernels require much smaller amounts of storage. This is due to the combination of wavefield and strain components to sensitivity values for relatively few model parameters, even though there is a separate kernel for each source-receiver pair. Also the pre-integration of the kernels onto the inversion grid cells yields a reduction of storage, since the forward grid on which the kernels are computed is in general much finer sampled than the inversion grid.

In order to take advantage of the separation in form of flexible sensitivity analysis or repeating the inversion step using different regularization or data subsets, large amounts of data must be read from disc. This implies that large storage capacities need to be accessible in a convenient way. This can be realized, for instance, by installing large scratch discs on the parallel high performance computing systems on which such a waveform inversion is conducted.

As mentioned above, Chen et al. (2007a) observe that the scattering-integral method requires 6–7 times fewer iterations compared with the adjoint-wavefield method in conjugate-gradient implementation. Due to the relatively small number of iterations to conduct, it is feasible to keep the number of automated decisions as small as possible and let the user *manually* derive the model updates in the course of the iterative process, having all the presented control mechanisms at hand.

5 Implementation and validation

In this chapter, finally an application of the full waveform inversion procedure is presented in form of a synthetic example which demonstrates the previously discussed merits and issues. The exemplary waveform inversion was conducted by the modularized software package ASKI in which the inversion procedure was implemented and which is shortly introduced in the beginning.

5.1 The software package ASKI

The modularized computation of waveform sensitivity kernels along with sensitivity analysis tools and the full waveform inversion procedure, as presented in the previous chapters, was implemented in the software package ASKI – Analysis of Sensitivity and Kernel Inversion. It consists of numerous binary programs and scripts which communicate via file interfaces and follow the modularized concept of the inversion procedure. It is implemented in the fashion of a software library containing about 20000 lines of code and about 5500 lines of documenting commentary written in modern FORTRAN and PYTHON. Those programming languages provide object-oriented features allowing for the required modularization. It, hence, is sustainably programmed and may be easily maintained and extended to further functionality. ASKI is well documented and freely available under the terms of the GNU general public license (Schumacher, 2013).

ASKI provides very generalized file interfaces to external forward modelling codes. So far, the 3D spectral-element code SPECFEM3D (Tromp et al., 2008) and the 1D semi-analytical code GEMINI (Friederich and Dalkolmo, 1995) in both, Cartesian and spherical framework are supported. The ASKI code library contains several generic modules for handling the forward grid, the physical properties of the medium on the forward grid, as well as the displacement fields, Green tensors and strains provided on the forward grid. A set of forward code specific submodules then branch from each of the generic modules to routines which specifically define how to deal with any method related output and structure. ASKI can be easily extended to other forward methods due to its modularized object-oriented structure.

For the computation of the frequency-domain waveform sensitivity kernels, ASKI currently supports the two isotropic elastic parameterizations $P_{\mathrm{Veloc}} = \{v_p, v_s, \rho\}$ and $P_{\mathrm{Lamé}} = \{\lambda, \mu, \rho\}$. It applies the respective formulas derived in section 3.2.1. The general integration weights after Levin (1999) presented in section 4.2.2 are implemented for polynomial degrees $m \in \{1, 2, 3\}$ on both, a cubic and a tetrahedral

standard inversion grid cell. This allows for pre-integration of the kernels on hexahedral and tetrahedral (or mixed-type) unstructured inversion grids. At the moment, several kinds of inversion grids are supported, all assuming a disjoint set $\{\Omega_e^k\}_e$ of volumetric inversion grid cells as motivated in section 4.2.2: A module for simple Cartesian inversion grids consisting of layers of cuboidal-shaped cells with layer dependent refinement is provided by ASKI that can be used in combination with any supported forward method (i.e. any forward grid). This module automatically locates the forward grid inside the inversion grid cells and provides the required Jacobians for the computation of integration rules. The same is done by an analogous, though more elaborate module, supporting (mixed) hexahedral / tetrahedral inversion grids constructed by external meshing tools. This allows to define inversion domains of complex geometries. Additionally, ASKI provides an inversion grid module for the specific forward method SPECFEM3D, using its finite elements as the inversion grid cells and allowing to use its internal integration rules. The kernels are written in a database-like binary file format which can be read by certain software modules of the ASKI software library. Those modules are used in several binary programs working on the entries of the sensitivity matrix to implement certain operations of sensitivity analysis, an inverse Fourier transform of the kernels to the time domain, as well as the iterative full waveform inversion procedure.

The linear system of equations (4.9) is set up using a separate module to define the regularization constraints. It is solved by LAPACK libraries (Anderson et al., 1999) in a serial way, i.e. this operation, which can become intense, is not yet parallelized. ASKI provides modularized mechanisms to flexibly define data subsets D_k for which the kernel sensitivity matrix is set up. Source and receiver filters as well as the frequency domain measured data are defined by external text files which can be created or modified by certain provided ASKI modules as well as arbitrary customized external tools. For plotting, ASKI provides output of kernels and models in VTK file formats (Schroeder et al., 2003). This common file format can be visualized for example using the software PARAVIEW (Squillacote, 2007), as done throughout this work.

5.2 Synthetic full waveform inversion

This section finally presents a synthetic example of an iterative full waveform inversion based on waveform sensitivity kernels. It is conducted by the modularized program package ASKI using the spectral-element code SPECFEM3D Cartesian 2.1 for solving the forward problem. It is published as an example of ASKI providing some intermediate wavefield and kernel files and including a detailed documentation on how to reproduce the results presented in the following (cf. Schumacher, 2013).

Although not yet provided by the current release version 0.3 of ASKI, such a waveform inversion may be done on spherical scale just as well and kernel computation was implemented and tested also in spherical framework. However, a small scale Cartesian example in crosshole geometry is preferred here for simplicity.

Remember that the waveform sensitivity kernels are not intuitive to look at for human beings (section 3.4.2). Despite the general issues of a frequency-domain inversion that must be taken into account, as pointed out above, it is not an issue for a computer to handle complex values, since simply equations are set up and solved.

5.2.1 Example data set and acquisition

The inverted dataset in this example is synthetically computed by SPECFEM3D. The true elastic earth model is a homogeneous isotropic half space with a free surface into which a block of mixed slow and fast velocity anomalies is placed, as presented in fig. 5.1. This figure also shows the acquisition geometry: Two lines of vertical boreholes containing sources and receivers are placed 50m apart. Three boreholes (SA, SB, SC) each contain four single force sources pointing into the direction of the line of the four receiver boreholes (RA, RB, RC, RD), each of which contains three 3-component receivers. The vertical 2D arrays of sources and receivers have a lateral distance of 10m to the anomaly cube and are otherwise aligned with the cube face in terms of depth and second lateral direction.

Figure 5.2a displays the source time function which is used in the SPECFEM3D simulations to produce the data seismograms for a record length of 0.25s applying a force of 1kN. Those seismograms are afterwards transformed to the frequency domain by a discrete Fourier transform. Data for all twelve sources recorded at all twelve receivers in all 3 directions of space are contained in the data set for the frequency window as indicated in fig. 5.2b. Since the recording time is 0.25s, the frequency window of 36Hz – 76Hz is discretized by frequency step $\Delta \hat{f} = \frac{1}{2 \cdot 0.25 \text{s}} = 2\text{Hz}$. In notation of section 4.1, hence, $N_S = N_R = 12$, $N_C = 3$, $N_F = 21$. Since, of course, the complete information (i.e. both complex components) are used, the total number of data samples here is $N_D = N_S \cdot N_R \cdot N_C \cdot N_F \cdot 2 = 18144$. No additional noise is applied to the data, thus choosing uncertainties $\sigma_i = 1$ for all $i \in D$ implicity accounting for no uncertainties at all (referring to section 4.3.3).

5.2.2 Procedure and particular iteration steps in detail

Starting off with the homogeneous background, twelve iteration steps are conducted using ASKI, manually deriving new models by the procedure explained in section 4.2. The model domain is indicated by the surrounding box in fig. 5.1 and measures laterally 100m × 100m and 75m in depth, bounded on top by a free surface. As an isotropic elastic model parameterization, seismic velocities v_p, v_s and density ρ are chosen. The latter, however, is kept unperturbed in the true model and is not inverted for.

ASKI's internal simple Cartesian grid is used as an inversion grid. In each iteration, the highest possible order of integration is assured depending on the distribution of forward grid points inside the cells (usually order $m = 3$, rarely order $m = 2$ for certain

(a) v_p anomaly

(b) v_s anomaly
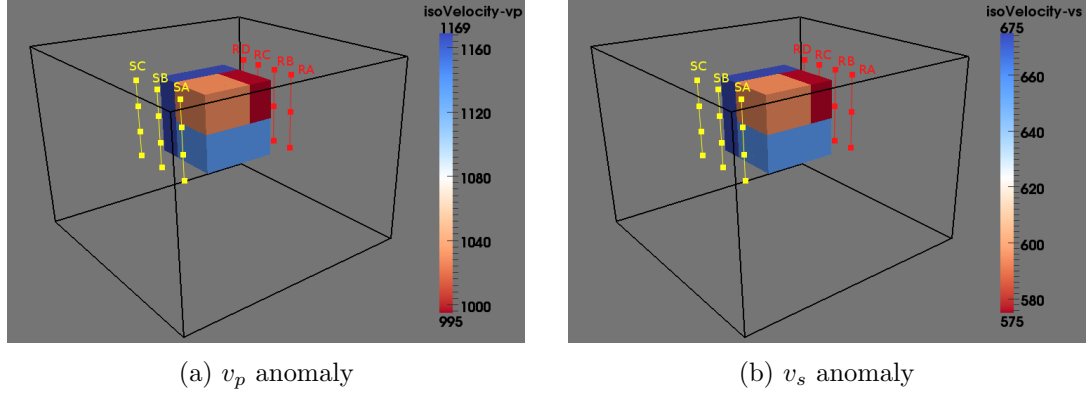
Figure 5.1: The images show P-velocity and S-Velocity anomalies (in $\frac{\mathrm{m}}{\mathrm{s}}$) that are placed into a homogeneous isotropic elastic background model of constant parameters $\rho = 875\,\frac{\mathrm{kg}}{\mathrm{m}^3}$, $v_p = 1082\,\frac{\mathrm{m}}{\mathrm{s}}$ and $v_s = 625\,\frac{\mathrm{m}}{\mathrm{s}}$. The anomaly consists of $\pm 5\%$ and $\pm 8\%$ velocity perturbations. The cube has an edge length of 30m and is placed 10m below the free surface. Density is not perturbed and will not be inverted for.
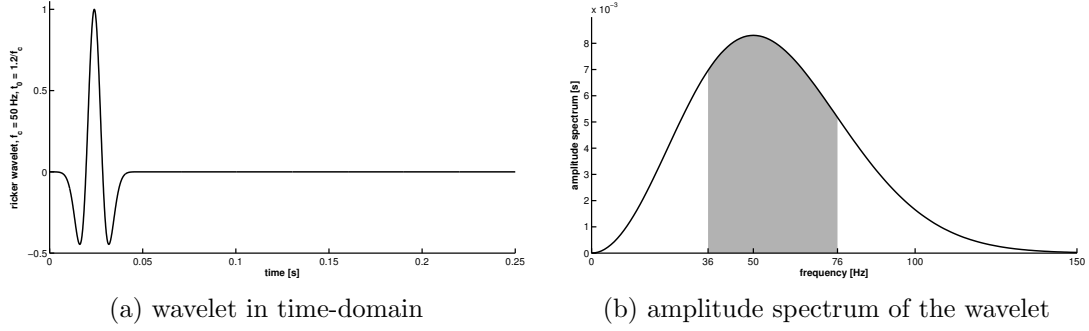

(a) wavelet in time-domain

(b) amplitude spectrum of the wavelet

Figure 5.2: The source time function used for the inversion: Ricker wavelet with center frequency of $f_c = 50$Hz (time shift $\frac{1.2}{f_c}$) and its amplitude spectrum indicating the frequency window of 36Hz – 76Hz for which data is inverted.

cells containing less points). As already said above, the forward problem is solved using SPECFEM3D. The simulation domain in use, however, is chosen larger than the inversion domain, namely $200\mathrm{m} \times 200\mathrm{m} \times 150\mathrm{m}$ (bounded on top by the free surface), in order to account for any reflections at the non-perfect absorbing boundaries (Stacey, 1988). The wavefields and Green tensors are stored throughout the inversion domain only on all inner points of the spectral elements (GLL polynomial degree 4 in all dimensions, i.e. $3^3 = 81$ inner points per element). For reasons of redundancy, the face and corner points of the spectral elements are excluded since they collapse with points of neighbouring elements. Those inner points constitute the "forward grid" on which the wavefields for kernel computation are computed and which the number $N_{\mathrm{forw}}^{\mathrm{grid}}$ refers to in the following. In this case, however, it is not equal to the total grid on which seismic wave propagation is simulated. All SPECFEM3D simulations are conducted in MPI parallelization on 48 cores on 4 physical machines each having two 6-core CPUs equipped with AMD Opteron™ 2435 Processors (maximum frequency

2.60GHz per core). The machines are connected via standard gigabit Ethernet and have enough RAM memory available (64565 MiB per machine, i.e. about 5380 MiB per core) so that no memory swapping is required. The large amount of produced output for wavefields and kernels is written to (and afterwards read from) the machines' scratch discs.

As suggested in section 4.3.3, impulse responses are computed for *both* forward wavefields for sources and backpropagations from receivers by simulating a very narrow Gaussian source time function. This is realized by SPECFEM3D for ASKI in form of using a steep error function (i.e. an approximation to a Heaviside step) as the source time function while at the same time differentiating the wavefield w.r.t. time. This approximation of an impulse is deconvolved before writing out the wavefield spectra as suggested in section 4.3.2. In order to correctly model the source response of the data, the complex-valued spectrum of the Ricker wavelet is assumed to be known and applied as a filter to synthetic data and kernels before solving the inverse problem.

Smoothing conditions

As said above, the inversion grid consists of the regularly aligned rectangular cuboids. In terms of eq. (4.6), all direct face neighbours of a cell $\Omega_{\hat{e}}^k$ are used to define conditions on the model values $m_{\hat{j}k}^\Delta$, $\hat{j} = \mathrm{idx}_k^{\mathrm{m}}(p, \hat{e})$, separately for each parameter p. On the boundaries of the inversion domain, *zero boundary conditions* are applied by choosing constant $w_j = \frac{1}{6}$ in eq. (4.6) independent of the actual number of direct face neighbours of the cell. This is done even at the free surface since the current ASKI version does not support the realization of the different kinds of conditions at different boundaries. These zero conditions were experienced to be more stable in this example compared with the continuity boundary conditions, where sometimes model contributions were found to be produced at the outer boundaries of the inversion domain. This may be due to the inversion domain being relatively small in this example.

Dependent on the particular iteration k, a smoothing intensity factor γ is chosen as introduced in section 4.2.3 (subsection about regularization). The values of γ chosen in each iteration of this example are given below in table 5.1. A standard ASKI feature, however, is to derive in each iteration an additional scaling factor γ_p dependent on model parameter p defined by

$$\gamma_p = \max_{\substack{i \in D_k \\ j \in M_p}} |K_{ij}| \quad ,$$

where M_p is the set of all indices $j = \mathrm{idx}_k^{\mathrm{m}}(p, e)$ of model values corresponding to model parameter p and arbitrary cell of inversion grid $\{\Omega_e^k\}_e$. This is done in order to account for different possible orders of magnitude of sensitivity values K_{ij} due to the different physical dimensions of the respective parameter p. Hence, each smoothing equation for a particular model value $m_{\hat{j}k}^\Delta$, $\hat{j} = \mathrm{idx}_k^{\mathrm{m}}(\hat{p}, \hat{e})$, is finally scaled by the product $\gamma \gamma_{\hat{p}}$ of global smoothing intensity factor γ (only dependent on the iteration step) and parameter dependent constant $\gamma_{\hat{p}}$.

Iteration statistics

The subsets of data which are inverted in the iterations k vary by frequency content only, i.e. always all data paths for all sources and receiver components (and both complex components, of course) are used. The particular frequency windows are given by the discrete frequencies in table 5.1. For reasons of computational costs, the frequency windows do not contain all previous frequencies, but a moving window of 5 discrete frequencies is chosen in an overlapping fashion.

Figure 5.3 displays 2D slices through the inversion domain, plotting the forward point grid on which the waveform spectra are produced inside the volumetric inversion grid cells. Note the iterative refinement of both, the forward and inversion grid in the course of the iterative inversion. This supports a stable inversion along with the choice of inverted data subsets, i.e. starting with low frequencies and continuing to higher ones.

The refinement of forward and inversion grid also becomes evident by the increase of numbers $N_{\mathrm{forw}}^{\mathrm{grid}}$ and $N_{\mathrm{inv}}^{\mathrm{grid}}$ in table 5.1, which gives a good overview of the development of various quantities throughout the iterations of the inversion. It is clearly seen that due to this iterative refinement, and the gradual increase of frequency content of the inverted data subset, the computational costs for all operations, i.e. solving the forward problem as well as computing the kernels and deriving a model update, are low in the beginning and iteratively increase. Among these three steps, however, in each iteration the solution of the forward problem is most expensive in terms of both, computation time and storage requirements. Compared with the wavefields, the pre-integrated kernels roughly require only 3.5%–2.5% of disc space and their computation time is neglectable. The relatively short computation times for solving the linear system of equations, finally demonstrate the feasibility of repeating this step testing different smoothing intensities or data subsets. Solving linear systems which are significantly larger than the one in the last iteration, however, justifies to use parallelized linear system solvers to increase performance, as suggested in chapter 7.

For the choice of inversion grid resolution, the involved wavelengths are taken into account along with a heuristic testing of cell sizes. Furthermore, the frequency content and smoothing intensities in each iteration step are chosen intuitively by hand in order to yield sensible model improvements.

5.2.3 Discussion of the inversion results

In the following now, the results are evaluated and discussed with regard to the concept of full waveform inversion as presented in this work.

		inverted data subset		forward problem			kernel computation				inverse problem			
k	N_f	f_p [Hz]	$|D_k|$	N_{forw}^{grid}	S_{forw}	T_{forw}^{48c}	N_{inv}^{grid}	N_{mval}	S_{kern}	$\sim T_{kern}^{1c}$	N_{row}	N_{col}	γ	$\sim T_{inv}^{1c}$
1	1	36	864	59168	260.01	41.6	384	768	2.53	few sec	1632	768	7.5	few sec
2	3	36, 38, 40	2592	77315	849.41	59.2	567	1134	11.21	⋮	3726	1134	11.0	⋮
3	3	36, 38, 40	2592	77315	849.41	61.6	700	1400	13.84	⋮	3992	1400	11.0	some sec
4	5	36, 38, 40, 42, 44	4320	91238	1603.80	72.0	968	1936	31.90	30 sec	6256	1936	11.0	⋮
5	5	40, 42, 44, 46, 48	4320	91238	1603.80	78.4	1296	2592	42.71	30 sec	6912	2592	9.0	⋮
6	5	44, 46, 48, 50, 52	4320	162000	2847.66	124.8	1960	3920	64.59	55 sec	8240	3920	13.0	2 min
7	5	48, 50, 52, 54, 56	4320	162000	2847.66	131.2	2475	4950	81.57	55 sec	9270	4950	9.0	⋮
8	5	52, 54, 56, 58, 60	4320	217800	3828.52	174.4	3072	6144	101.25	75 sec	10464	6144	5.0	⋮
9	5	56, 58, 60, 62, 64	4320	217800	3828.52	185.3	3757	7514	123.82	75 sec	11834	7514	5.0	6 min
10	5	60, 62, 64, 66, 68	4320	329232	5787.29	283.2	4536	9072	149.50	2 min	13392	9072	9.0	8 min
11	5	64, 66, 68, 70, 72	4320	329232	5787.29	286.4	5054	10108	166.57	2 min	14428	10108	5.0	11 min
12	5	68, 70, 72, 74, 76	4320	329232	5787.29	354.4	6000	12000	197.75	2 min	16320	12000	5.0	14 min

Table 5.1: Some statistics about the iteration steps: k = iteration step; N_f = number of frequencies; f = frequencies [Hz] occurring in data subset; $|D_k|$ = number of data samples in data subset; N_{forw}^{grid} = number of points in forward grid; S_{forw} = total amount of storage [MiB] of forward wavefields; T_{forw}^{48c} = total runtime [min] to solve forward problem (parallel on 48 cores); N_{inv}^{grid} = number of inversion grid cells; N_{mval} = number of free model values; S_{kern} = total amount of storage [MiB] of pre-integrated kernel values; T_{kern}^{1c} = approximate runtime of kernel computation (serial on 1 core); N_{row} = number of rows of LSE (including smoothing conditions); N_{col} = number of columns of LSE; γ = smoothing intensity factor; T_{inv}^{1c} = approximate runtime of LSE solver (serial on 1 core)

(a) iteration step 1 (b) iteration step 2 (c) iteration step 3

(d) iteration step 4 (e) iteration step 5 (f) iteration step 6

(g) iteration step 7 (h) iteration step 8 (i) iteration step 9

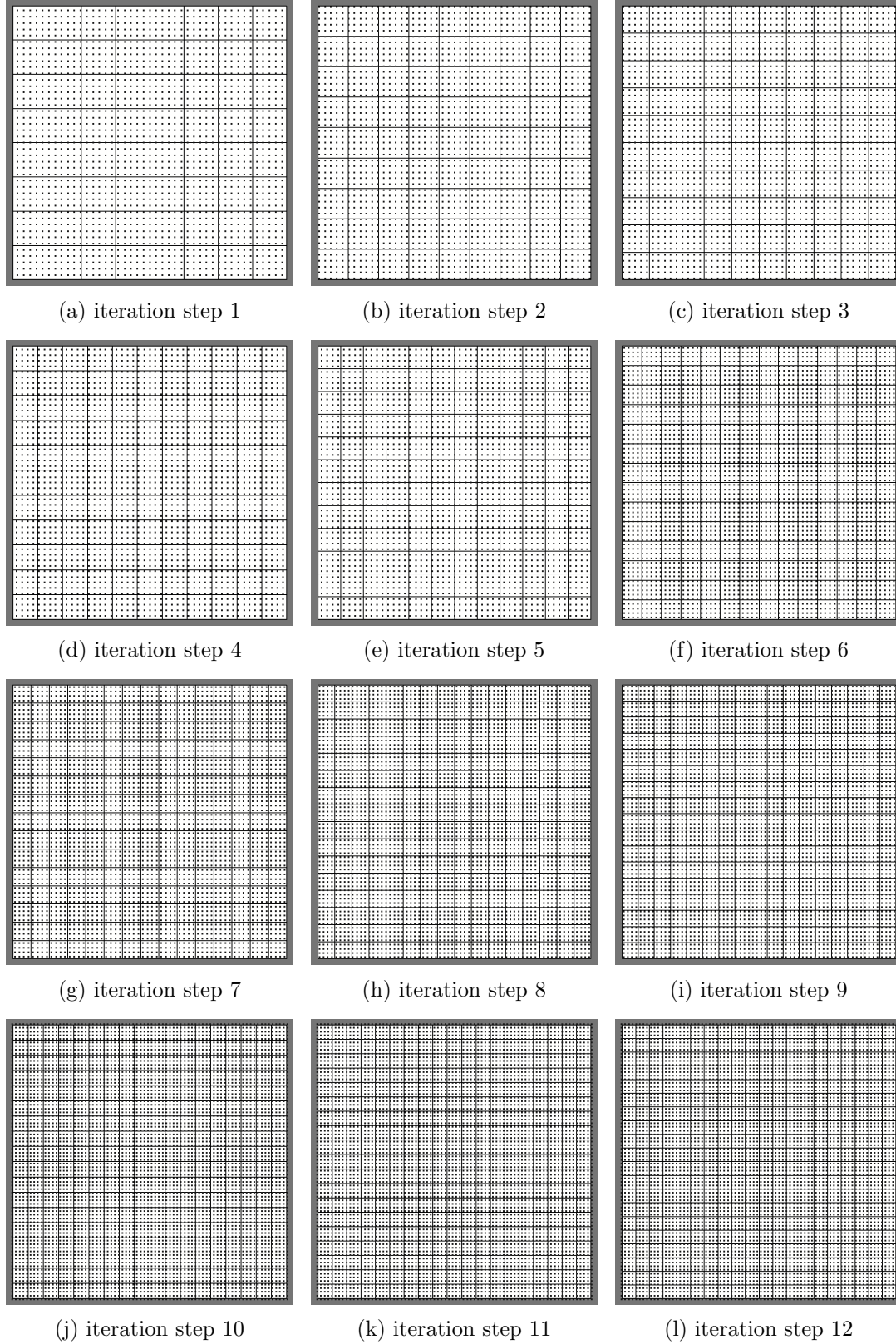(j) iteration step 10 (k) iteration step 11 (l) iteration step 12

Figure 5.3: Slices through the inversion grid (view from above) showing distribution of forward point grid inside in the inversion grid cells for each iteration step. Note that the resolution of the forward grid, as well as the inversion grid, is adaptively refined throughout the inversion.

Inverted models

Figures 5.6 and 5.7 show the resulting P-velocity models in the course of the iterative inversion by different slices through the inversion domain and figs. 5.8 and 5.9 show the respective S-velocity models. It can be clearly seen that the velocity anomalies are resolved very well by the full waveform inversion in spite of some of the anomalous structure being smaller than the wavelengths at the dominating frequency of 50Hz, e.g. the -8% P-velocity perturbation of size $20\text{m} \times 10\text{m} \times 15\text{m}$ (dominating wavelength 21.64m). However, the P-velocity model produces some artefacts that do not occur in the S-velocity model (e.g. compare roughness of models in figs. 5.6d and 5.8d). This may be due to the larger wavelengths of the P-waves in combination with the fact that the current ASKI version uses the same inversion grid resolution and smoothing intensity factor γ for all parameters p, i.e. here for both v_p and v_s.

Note the smoothing effect of the interpolations from one inversion grid to the next via interpolation onto the forward grid. This smoothing effect is actually essential for the convergence of the process, which is in particular seen by model values on inversion grid cells which are close to the boundary of the anomaly cube (e.g. fig. 5.9d, left panel of iterations 11 and 12): for cells intersected by the anomaly boundaries, the model values are smoothed out a bit to the background model (or across internal boundaries in the anomaly), whereas for iterations in which the anomaly is geometrically represented in shape by the volumes of the inversion grid cells, more or less sharp boundaries are seen (assuming constant model values on the cells). This way, indeed the model stably converges into the right place in the course of the iterative refinement of the inversion grid.

Regularization

Table 5.1 lists the smoothing intensity factors γ used in the iterations. The different quality of the derived models shows, that the data is not capable of resolving the P- and the S-velocity models equally well and that the description of the P-velocity model contains too many degrees of freedom, overall. It, thus, may be helpful to allow for parameter specific resolution of the inversion grid, as well as independent smoothing for each parameter. In this case, the smoothing of parameter v_p had to be stronger than for v_s. Now it becomes obvious that it may not be sensible in this case to apply the the additional scaling factor γ_p as defined above in section 5.2.2: First of all, the physical dimensions of parameters v_p and v_s are the same, hence the physical dimensions of the respective kernels, which does not justify an additional scaling due to physical dimensions. Secondly, in general the absolute sensitivity values for parameter v_p are smaller than for v_s due to smaller amplitudes of the P-waves compared with S-waves. Thus, in general $\gamma_{v_p} < \gamma_{v_s}$ which has a counterproductive effect, since a stronger smoothing of parameter v_p is required. Therefore, in the future regularization criteria related to wavelength should be found. At best, an inversion grid $\{\Omega_{ep}^k\}_{e,p}$ is used allowing independent inversion grid cells (i.e. model resolution) for each parameter p.

However, by use of such more general inversion grids $\{\Omega_{ep}^k\}_{e,p}$ any additional regularization conditions such as smoothing would *not* become obsolete, as demonstrated in fig. 5.10. It shows the inverted P-velocity and S-velocity model after iteration step 6 on slices as in figs. 5.6 and 5.8 for different choice of smoothing intensity factors γ as well as for the case of not applying any additional smoothing to the kernel system. It becomes obvious that mathematically optimal solutions of the overdetermined / underdetermined kernel linear system may in general not be physically sensible and require additional physically motivated constraints.

Fit of the data

Just like the velocity models are reproduced quite well, the time-domain seismic data is fitted equally well. Exemplary for one source, namely the lowermost source in the central borehole SB, the development of the iterative fit of waveforms is presented in fig. 5.4 showing seismograms for three different receivers: In despite of a significant misfit of the seismograms for starting and true model, iteratively a continuous improvement can be observed throughout the complete time interval. Signal is produced correctly at times where there was no signal before (fig. 5.4c $0.05\text{s} \le t \le 0.1\text{s}$). Phases are fittet (figs. 5.4a and 5.4b $t = 0.11\text{s}$), as well as amplitudes (fig. 5.4a $t = 0.14\text{s}$; fig. 5.4b $t = 0.16\text{s}$; fig. 5.4c $t = 0.11\text{s}$). Even completely distorted wavelets, seeming to be out of phase are reconstructed in principle (figs. 5.4a and 5.4a $0.1\text{s} \le t \le 0.15\text{s}$). This indeed demonstrates that not just certain attributes of the seismograms like certain arrivals etc. but the complete waveforms are fittet by the full waveform inversion procedure, as hoped for.

However, the quantity that is actually fitted by the inversion method in some iteration k is not time-domain data but the frequency-domain data residuals $r_i = d_i - s_i$, $i \in D_k$. It is, therefore, interesting to observe the normalized misfit $\bar{\chi}^2$ of the data in frequency domain defined by

$$\bar{\chi}^2 = \frac{\sum_{i \in S} |r_i|^2}{\sum_{i \in S} |d_i|^2} \quad , \tag{5.1}$$

where $S \subseteq \{1, \dots, N_D\}$ represents some data subset. Figure 5.5 shows this normalized misfit computed after each iteration for certain fixed frequency contents of the data. It demonstrates that the normalized misfit decreases in the course of the inversion and the fit stabilizes very fast at low frequencies. Furthermore, the misfit reductions at higher frequencies not yet used in the early iterations of the inversion demonstrate, that changes in the structural model affect all frequencies contained in the data.

Note that in iterations 2 and 3 data for the same frequency content (36Hz, 38Hz, 40Hz) is inverted and that the inversion grid is almost the same, as well. Observing that neither the inverted models nor the normalized misfit are improved in iteration 3, it is reasonable to assume that the complete information content of the data at those frequencies 36Hz–40Hz is already accounted for in iteration 2. This raises the hope that by the particular data subsets as chosen here in the iterations, indeed the

complete information content of the data in the total frequency window of 36Hz – 76Hz is accounted for properly.

Also it can be clearly seen in fig. 5.5 that the fit of low frequencies does not get worse in later iterations when those frequencies are no longer part of the inverted data subsets. This legitimates to use a moving frequency window and not preserve all previously used frequencies. As already referred to in section 5.2.2, this is done for computational reasons, since choosing $D_1 \subset D_2 \subset \ldots$ would eventually blow up the number of data $|D_k|$, i.e. the number N_{row} of rows of the regularized linear system.

Kernel based inversion in discussion

Referring to table 5.1, the advantages of the approach of separation in the inversion procedure become obvious. The iterative refinement of the forward and the inversion grid does not only stabilize the inversion process but allows to optimally exploit in each iteration step the computational resources in terms of runtime and storage, dependent on the complexity of the currently inverted model. This way, the first few iterations may run much faster than the later ones. As already discussed above in section 4.3.4, this must on the other hand be paid by the cost of storing the waveforms on hard disc. Although the total amount of required storage is already optimally reduced by choosing the frequency domain, some operations need to write large amounts of output (e.g. producing the wavefields and strains) and others need to read in large files such as for kernel computation or setting up the sensitivity matrix. Memory, however, is not an issue in general and is not referred to here.

Moreover, it is evident that the implemented full waveform inversion procedure based on waveform sensitivity kernels does not produce any artefacts at sources, receivers and free surfaces, as gradients of misfit functionals do. Thus, no kind of pre-conditioning is required using ASKI.
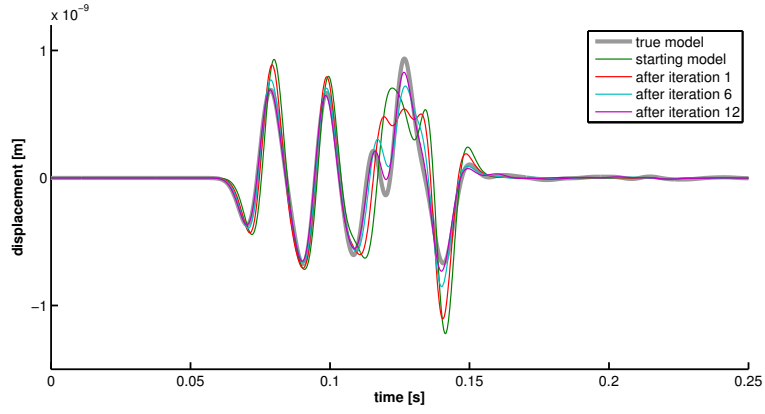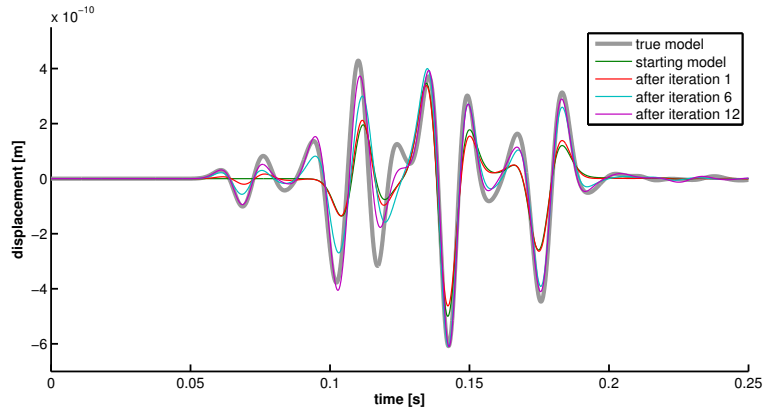
(a) Inline component of topmost receiver in borehole RD
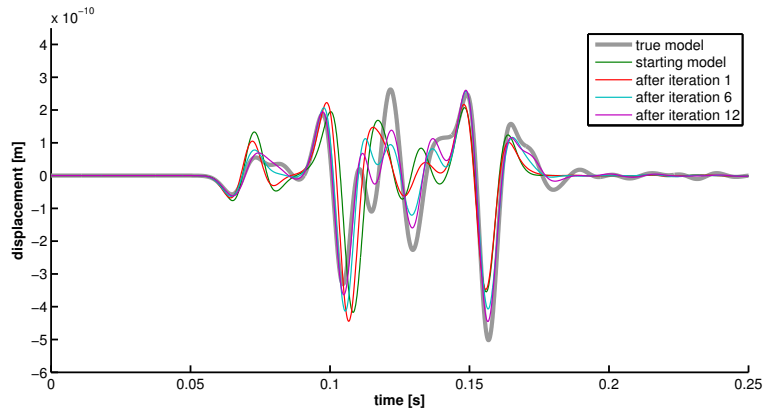
(b) Transverse component of central receiver in borehole RC

(c) Vertical upwards component of lowermost receiver in borehole RB

Figure 5.4: Seismograms for lowermost source in borehole SB at different receiver components. Refer to fig. 5.1 for nomenclature. The figures show the (synthetically computed) measured data, i.e. seismograms w.r.t. the true earth model, in comparison to synthetic seismograms w.r.t. the starting model and some intermediate inverted models.
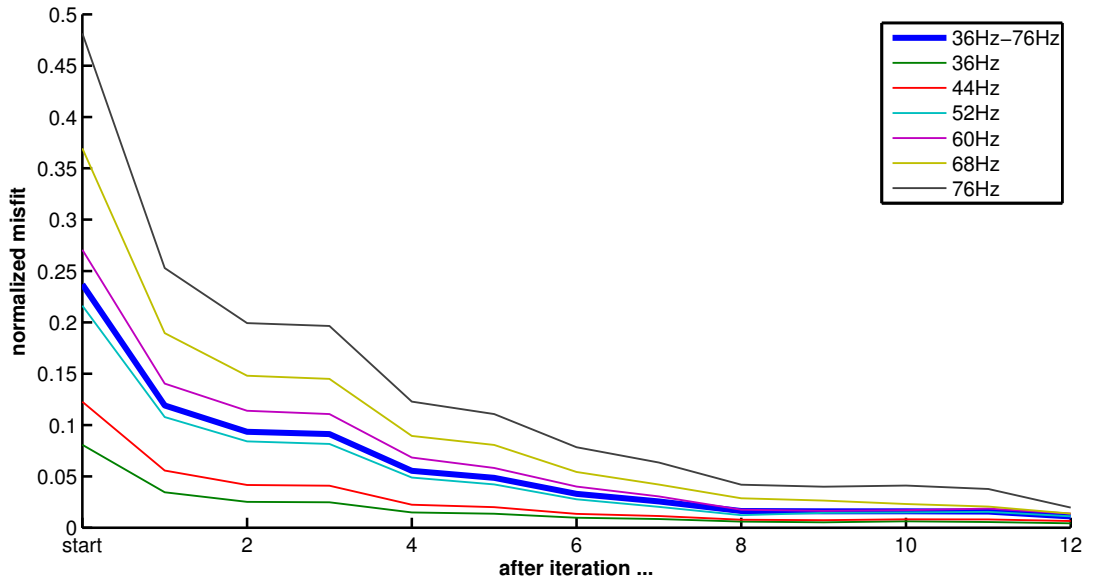
Figure 5.5: The development of the normalized misfit $\bar{\chi}^2$ (as defined in eq. (5.1)) in the course of the iterative inversion for different subsets S of data, containing single frequencies only (thin lines). The thick blue line represents the normalized misfit computed for the complete data set, i.e. the complete frequency content. Note, that the fit stabilizes very fast at low frequencies. Furthermore, it can be seen that changes in the structural model effect all frequencies of the data, since at early iteration steps also the fit for the higher frequencies decrease even if not used yet.
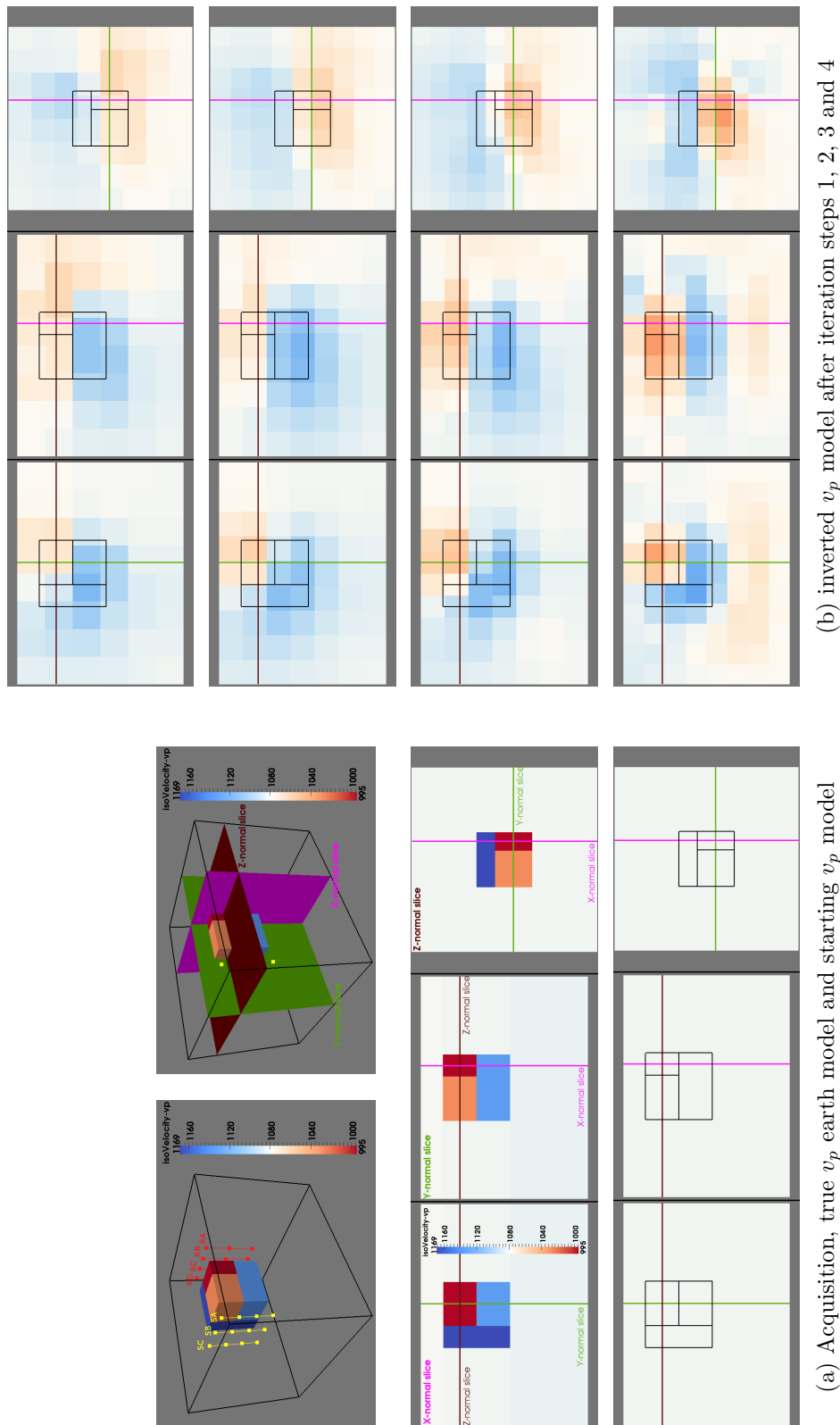
(a) Acquisition, true v_p earth model and starting v_p model

(b) inverted v_p model after iteration steps 1, 2, 3 and 4

Figure 5.6: Inverted P-velocity model results after the first iteration steps shown on slices centered in the slow anomaly block.
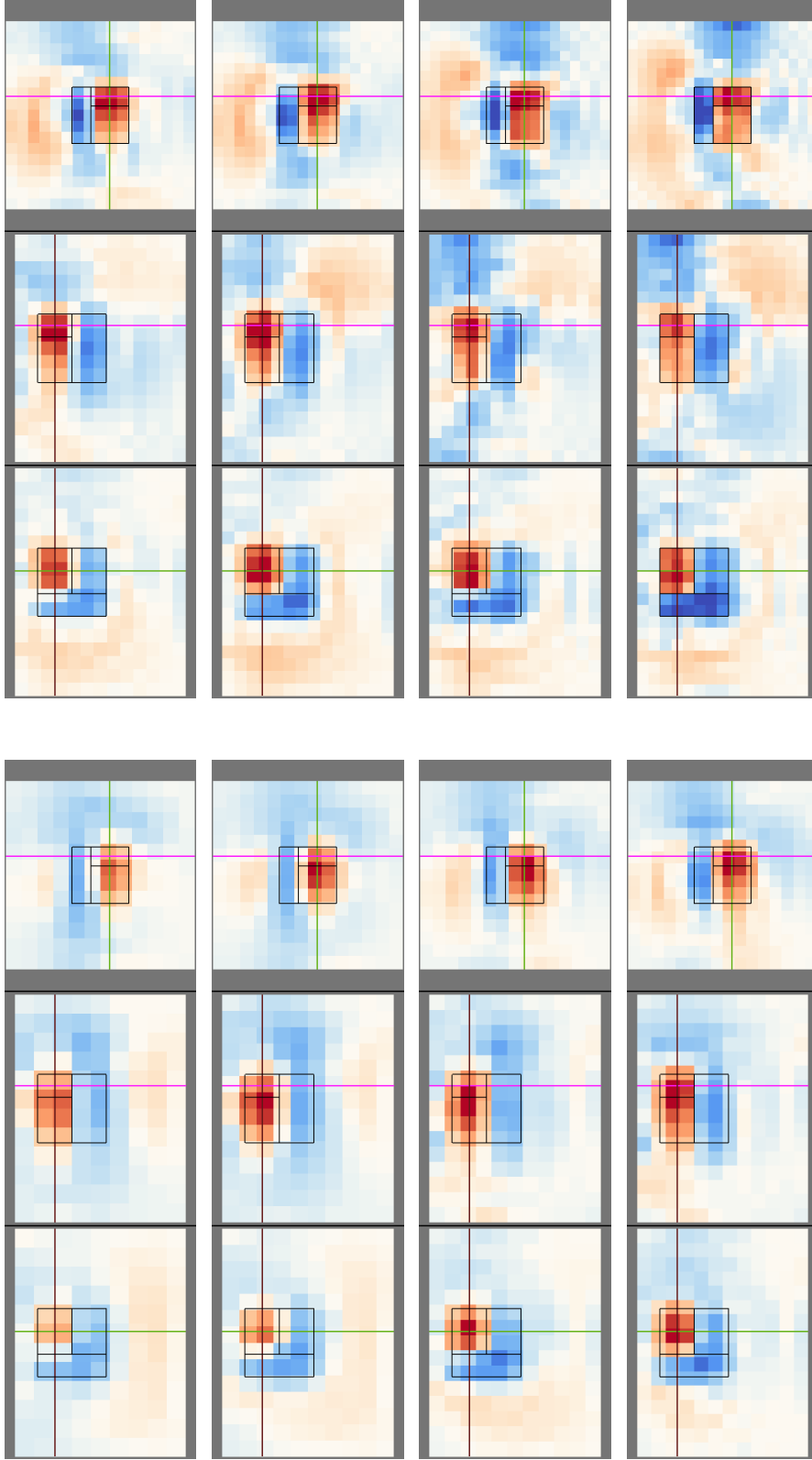
(d) inverted v_p model after iteration steps 9, 10, 11 and 12

(c) inverted v_p model after iteration steps 5, 6, 7 and 8

Figure 5.6: Continued: Inverted P-velocity model results after the rest of iteration steps on slices through slow anomaly block.

(a) Acquisition, true v_p earth model and starting v_p model
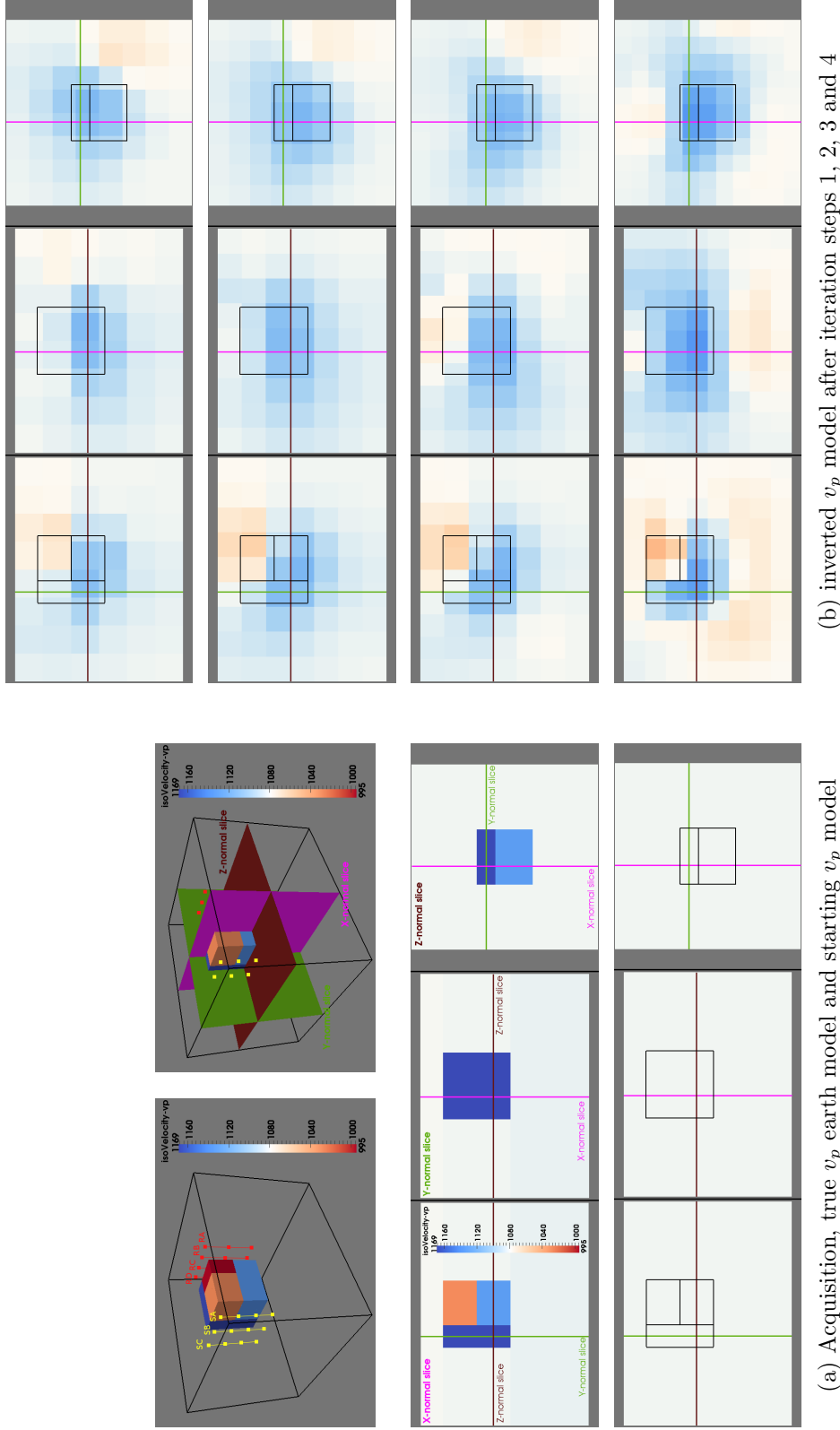

(b) inverted v_p model after iteration steps 1, 2, 3 and 4

Figure 5.7: Inverted P-velocity model results after the first iteration steps shown on slices centered in the fast part of the anomaly.

(d) inverted v_p model after iteration steps 9, 10, 11 and 12

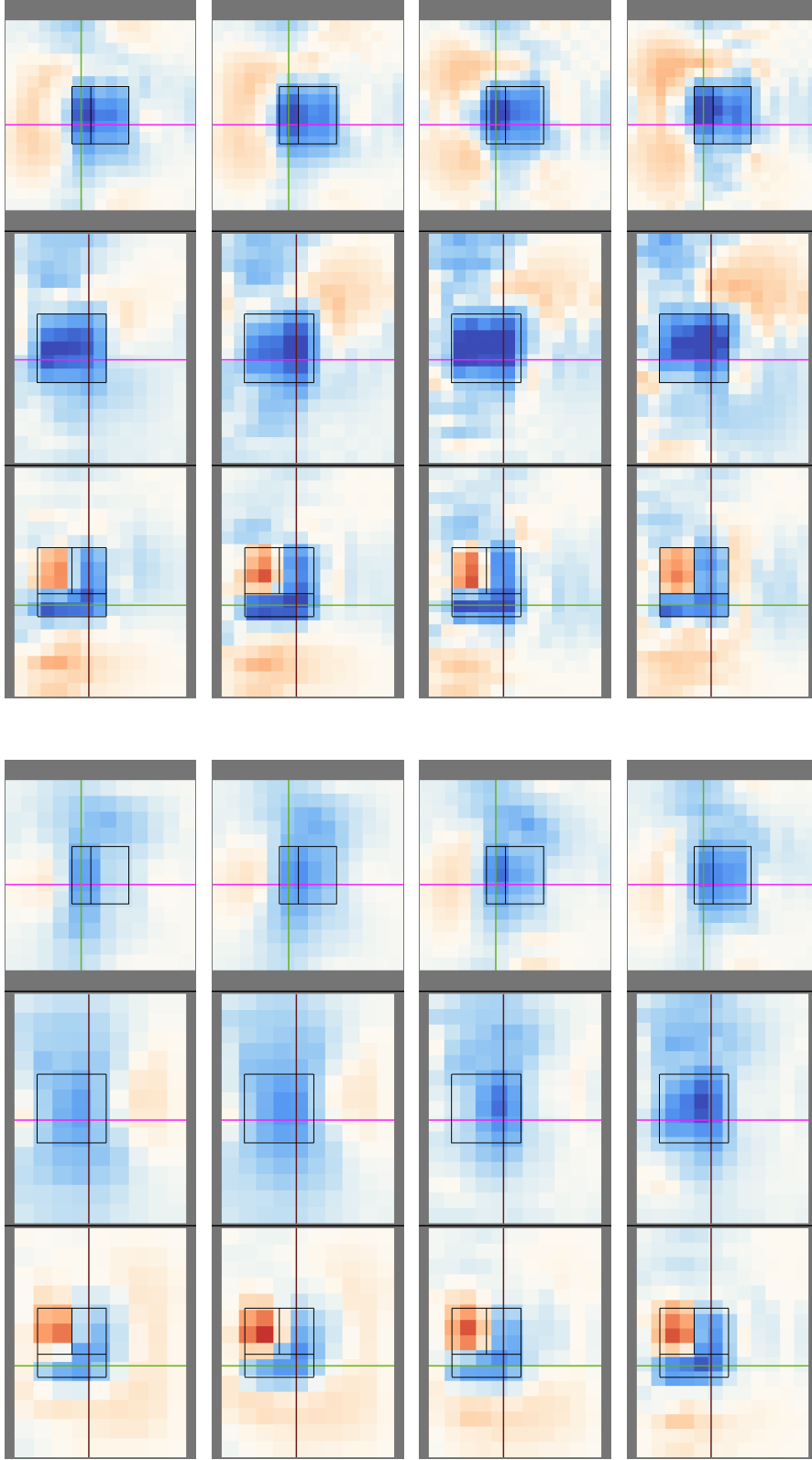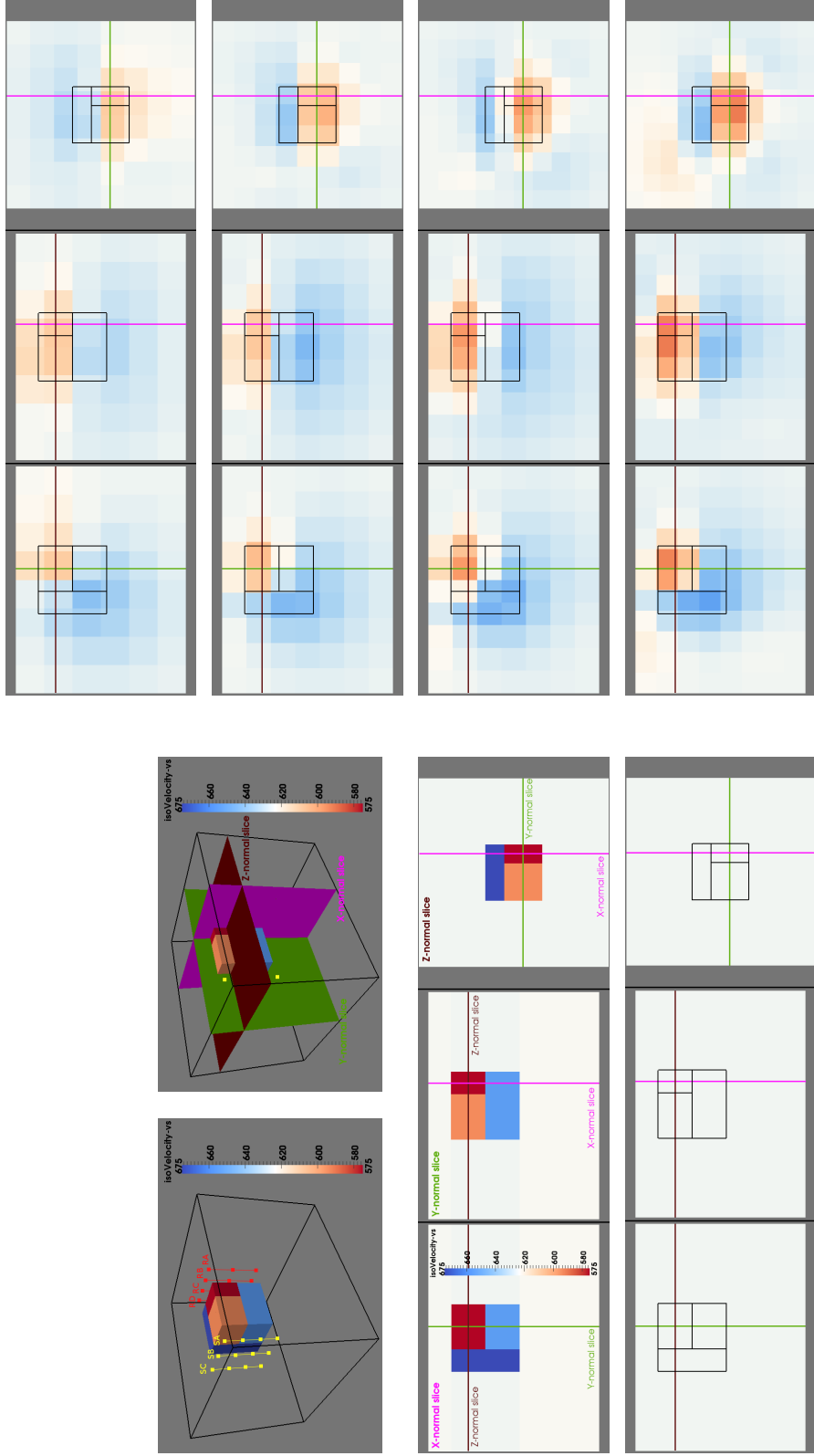(c) inverted v_p model after iteration steps 5, 6, 7 and 8

Figure 5.7: Continued: Inverted P-velocity model results after the rest of iteration steps on slices through fast part of anomaly.

(a) Acquisition, true v_s earth model and starting v_s model

(b) inverted v_s model after iteration steps 1, 2, 3 and 4

Figure 5.8: Inverted S-velocity model results after the first iteration steps shown on slices centered in the slow anomaly block.

(c) inverted v_s model after iteration steps 5, 6, 7 and 8

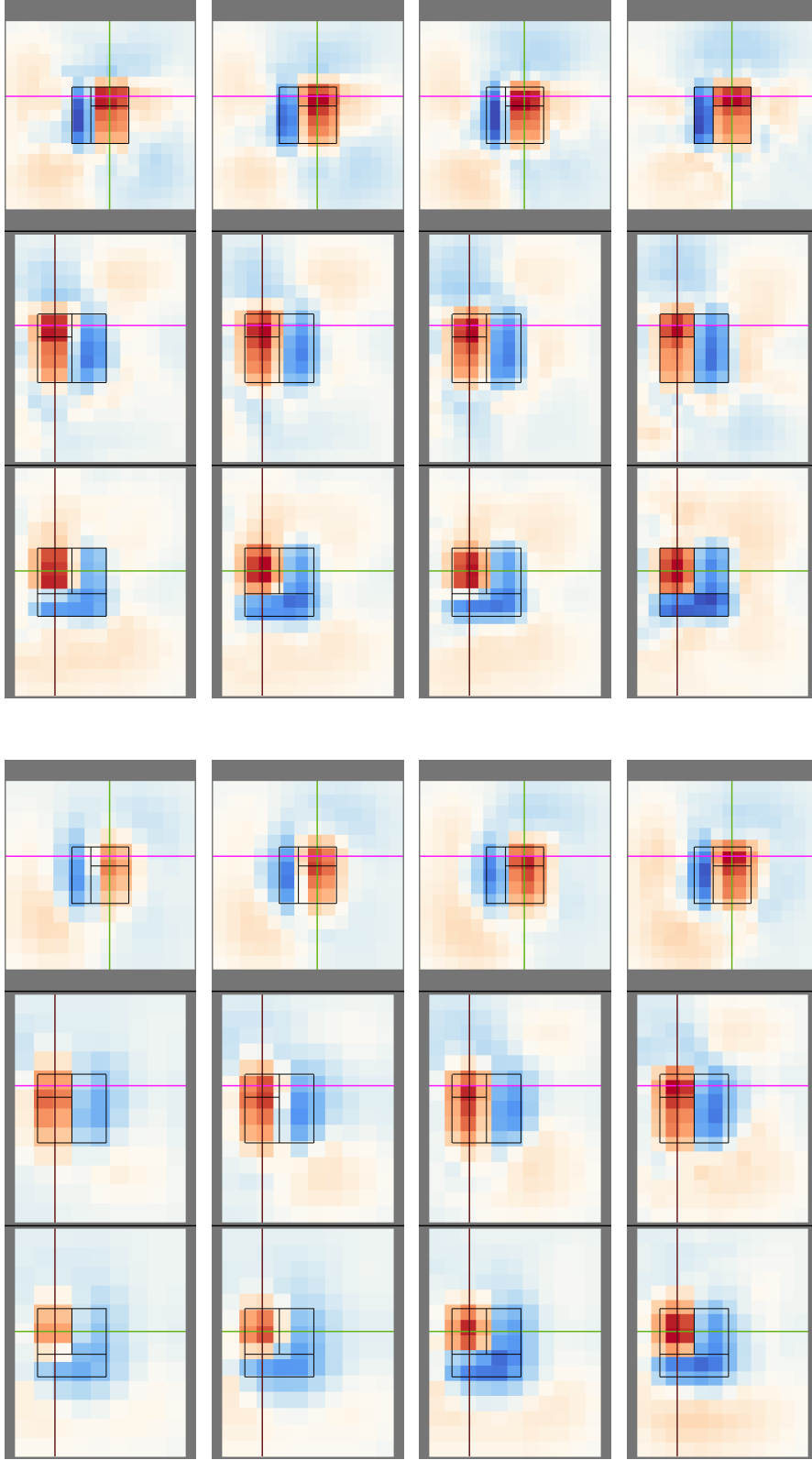(d) inverted v_s model after iteration steps 9, 10, 11 and 12
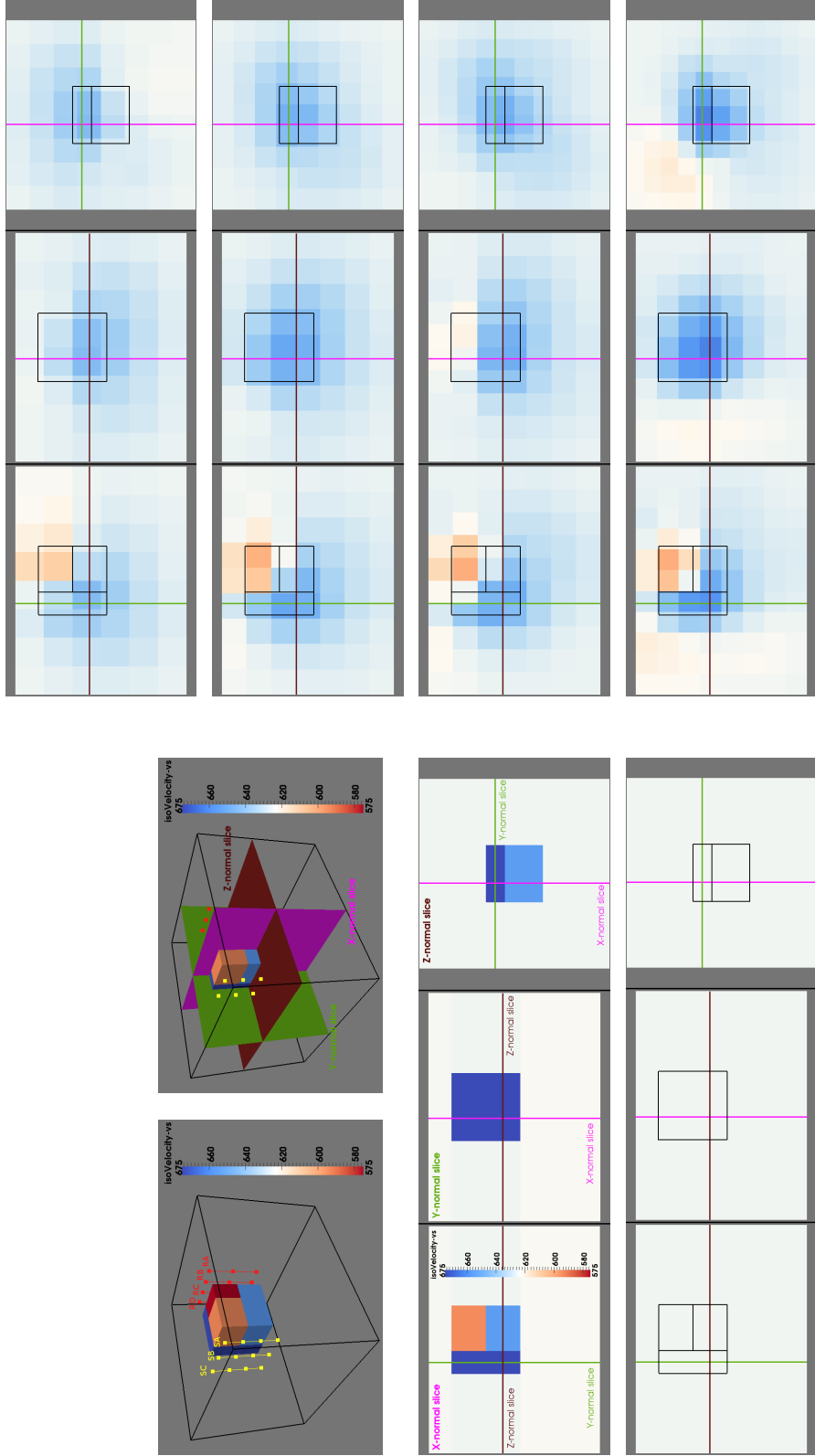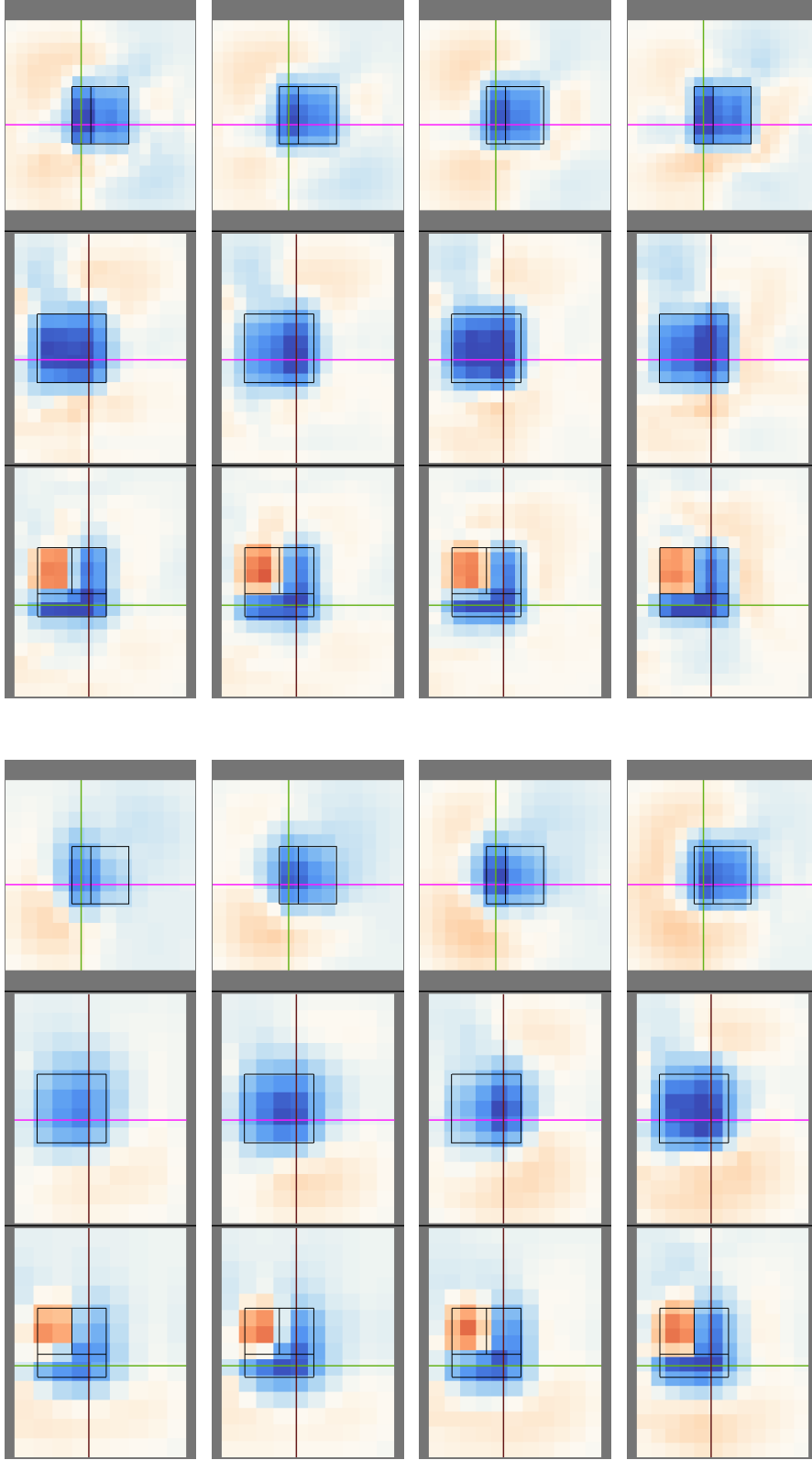
Figure 5.8: Continued: Inverted S-velocity model results after the rest of iteration steps on slices through slow anomaly block.

(a) Acquisition, true v_s earth model and starting v_s model

(b) inverted v_s model after iteration steps 1, 2, 3 and 4

Figure 5.9: Inverted S-velocity model results after the first iteration steps shown on slices centered in the fast part of the anomaly.

(d) inverted v_s model after iteration steps 9, 10, 11 and 12

(c) inverted v_s model after iteration steps 5, 6, 7 and 8

Figure 5.9: Continued: Inverted S-velocity model results after the rest of iteration steps on slices through fast part of anomaly.

(a) inverted v_p model: no smoothing (top) and $\gamma = 1.0$, 13.0, 100.0

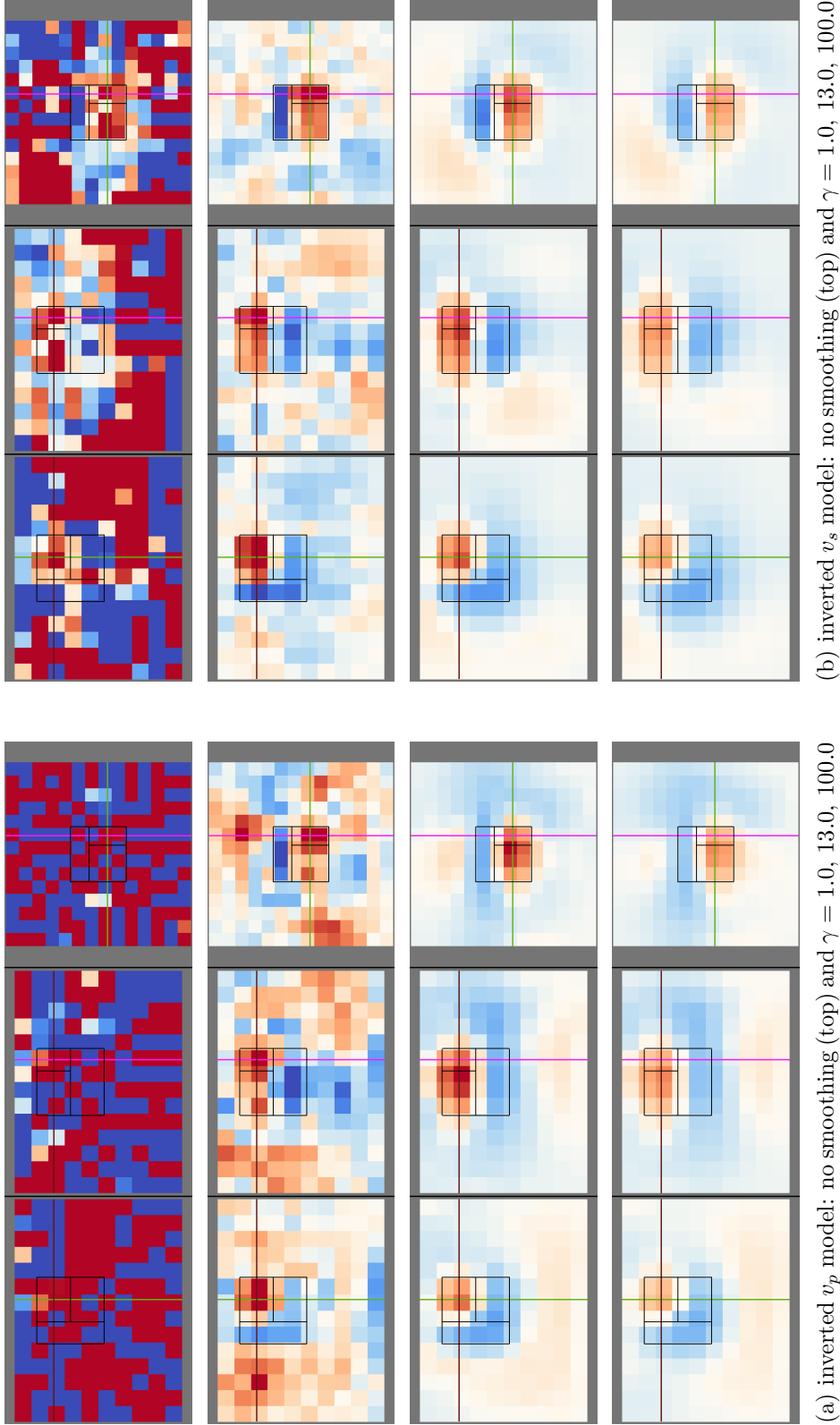(b) inverted v_s model: no smoothing (top) and $\gamma = 1.0$, 13.0, 100.0

Figure 5.10: Inverted P-velocity and S-velocity model after iteration step 6 on slices as in figs. 5.6 and 5.8 for different choice of smoothing intensity factors γ. The top image in the subfigures, respectively, show the inverted model derived with *no* smoothing constraints applied. This clearly demonstrates that some mathematically optimal solution of the linear system is in general not physically sensible and that some additional constraints must be applied to the kernel linear system. The intensity of smoothing clearly increases with increasing γ. Also, for same γ the P-velocity model is rougher than the S-velocity model.

6 Summary

This thesis develops, validates and discusses a full waveform inversion procedure for arbitrary heterogeneous earth models based on waveform sensitivity kernels.

The sensitivity kernels quantify in a linearized sense the influence of elastic earth model parameters and density on frequency-domain full waveform data functionals. The waveform data constitute spectra of the displacement field originating from a seismic source and recorded at a receiver position in some direction of space in a certain frequency band. Based on Born scattering (i.e. single scattering) theory, general expressions are derived of how sensitive a certain spectral sample is w.r.t. changes of density or the 21 independent elastic constants at a certain point in the medium. Explicit formulas of kernels for two simple isotropic earth model parameterizations are given, namely density and Lamé parameters as well as density and seismic velocities. The kernels are expressed in terms of multiplications of wavefield and strain spectra. Required are the displacement field and strain spectrum originating from the respective seismic source evaluated at the scattering point, as well as the Green function spectrum and its strains originating from the respective receiver component. In order to properly represent potential scattering arriving at the receiver in a time window $[0, T]$, it is emphasized that forward wavefield and Green function spectra *each* must represent a propagation time of T throughout the medium. For the calculation of kernels by the multiplication of discretized frequency spectra (equivalently circular convolutions), then, a frequency sampling of $\Delta f \leq \frac{1}{2T}$ is required in order to correctly represent potential scattering at any point in the medium. This is demonstrated by an example.

Based on the waveform kernels as the Fréchet derivatives of linearized data functionals, a linearized iterative full seismic waveform inversion procedure is developed, in which the solution of the forward problem, the computation of sensitivity kernels and the derivation of a model update are held completely separate. A sufficiently good starting model is required to initiate the process. For stability reasons, the first iterations should involve data subsets of low frequencies, gradually adding data at higher frequencies throughout the iterations. The separation in particular keeps the model description for the forward problem and the description of the inverted model update independent in each iteration. Hence, the resolution of the inverted model as well as the complexity of solving the forward problem can be iteratively increased, respecting the frequency content of the inverted data subset in the particular iteration. This optimizes the computational effort of both, solving the forward and solving the inverse problem in each iteration. However, arbitrary unstructured volume and point grids must be connected properly. This is achieved in form of generalized integration rules and 3D-unstructured interpolation methods, which are presented in this work.

The model update is inferred solving a minimization problem in a least-squares sense, thus resulting in Gauss-Newton convergence of the overall iterative process.

The principle of separation naturally encourages modularization and object-oriented programming for the practical realization of the iterative inversion procedure. It was implemented in the modularized software package ASKI – Analysis of Sensitivity and Kernel Inversion. As a form of toolbox, ASKI does not implement an internal wave propagation code for solving the forward problem, but instead provides a generalized interface to arbitrary external forward modelling codes. So far, the 3D spectral-element code SPECFEM3D and the 1D semi-analytical code GEMINI in both, Cartesian and spherical framework are supported. ASKI is freely available under the terms of the GNU general public license. It is implemented in the fashion of a software library, defining encapsulated units of functionality which interact by well defined interfaces. This allows for flexible implementation of different kinds of tools for data processing, sensitivity analysis, and the full waveform inversion procedure developed in this thesis. Therefore, any software based on the toolbox ASKI is sustainable and easy to maintain, adapt, extend and develop. For visualization ASKI provides output in the VTK file format, which can be displayed using common external software. The independent tools of ASKI, each tackling a distinct task, however, must communicate via file output / input. Especially the spectral wavefields required for kernel computation and produced by the forward modelling codes, usually require large amounts of disc space, which again needs to be accessed e.g. by the kernel computation program. Hence, large storage capacities need to be accessible in a convenient and fast way. Storing the complete sensitivity matrix to hard disc, however, allows arbitrary sensitivity and resolution analysis tools to work on that data, which permits the scientist full manual control over each step in a customized sensitivity analysis and full waveform inversion procedure.

In conclusion, the presented full seismic waveform inversion method constitutes a very flexible alternative to conventional full waveform inversion methods. It provides full control over the inversion process which exhibits fast convergence properties. The implementation of the method in form of the modularized toolbox ASKI offers the potential for sensitivity and resolution analysis as well as full waveform inversion in applications ranging from continental to local scale and from seismological to engineering interests.

7 Outlook

7.1 Resolution analysis and sensitivity focussing

As stated in the introduction, seismic inversion aims at inferring information on physical properties of the earth from a usually finite set of seismic observables. By only finite measurements, however, it is not possible to uniquely constrain a model from the model space which usually has infinite dimension. Depending on the distribution of seismic sources and receivers, some parts of the earth (in between sources and receivers) might be constrained well, whereas physical properties of material far away from sources and receivers cannot be inferred satisfactorily. The waveform sensitivity kernels derived in this thesis connect the data functionals and the model, determining for a certain datum those model values which the datum is very sensitive to. It, hence, is desirable to apply the sensitivity kernels for the purpose of resolution analysis.

Backus and Gilbert (1968) suggest to quantify the power of a finite set of data functionals to resolve the earth model at a certain point in space on the basis of their representers in the model space, i.e. their sensitivity kernels. Applying this method to full waveform data functionals, the complete model information contained in the data can be made visible. Since the kernel relations are applied linearly everywhere, however, this resolving power of the non-linear seismic data functionals can only be inferred in a linearized sense. For future work, the pre-integrated waveform sensitivity kernels as derived in this thesis may be used in the fashion of Backus and Gilbert (1968) to estimate the model resolution of the data as suggested in the following.

Using pre-integrated kernels for resolution analysis after Backus and Gilbert

Generalizing to three dimensions, Backus and Gilbert (1968) consider linear averages of the earth model at a certain point $\mathbf{x}_0$ in the model domain Ω

$$\langle m \rangle_{\mathbf{x}_0} = \int_\Omega A(\mathbf{x}_0, \mathbf{x})\, m(\mathbf{x})\, \mathrm{d}^3\mathbf{x} \quad ,$$

where A is an averaging kernel satisfying

$$\int_\Omega A(\mathbf{x}_0, \mathbf{x})\, \mathrm{d}^3\mathbf{x} = 1 \quad .$$

Assuming furthermore the given data functionals F_i, $i = 1, \ldots, N$, to be linear and having representers g_i, i.e. $F_i[m] = (g_i, m)$, they prove that there exist coefficients

$a_i(\mathbf{x}_0)$ such that

$$A(\mathbf{x}_0, \mathbf{x}) = \sum_{i=1}^{N} a_i(\mathbf{x}_0) g_i(\mathbf{x}) \quad .$$

Here, the notation of Parker (1994) as used in chapter 2 is preferred, for consistency. The measured data $d_i = F_i[\hat{m}]$, hence, perfectly constrain the true earth model $\hat{m}$ at point $\mathbf{x}_0 \in \Omega$ if $a_i(\mathbf{x}_0)$ can be found such that the averaging kernel $A(\mathbf{x}_0, \mathbf{x})$ would be the Dirac delta distribution $\delta(\mathbf{x} - \mathbf{x}_0)$. With only a finite set of data, however, this cannot be achieved in practice, such that one can only hope $A(\mathbf{x}_0, \mathbf{x})$ to have a high, narrow peak at $\mathbf{x}_0$ and have small values elsewhere. The width of the peak then measures the resolving power of the data d_i in the vicinity of $\mathbf{x}_0$.

Backus and Gilbert (1968) suggest to derive the averaging kernel A by choosing a measure of "δ-ness" for A and solving an optimization problem which penalizes solutions poorly satisfying that criterion. They suggest to choose a function $J(\mathbf{x}_0, \mathbf{x})$ which vanishes at $\mathbf{x}_0$ and increases monotonically with distance to $\mathbf{x}_0$. A, then, can be found by minimizing

$$\int_{\Omega} J(\mathbf{x}_0, \mathbf{x}) \, |A(\mathbf{x}_0, \mathbf{x})|^2 \, \mathrm{d}^3\mathbf{x} \quad \text{subject to} \quad \int_{\Omega} A(\mathbf{x}_0, \mathbf{x}) \, \mathrm{d}^3\mathbf{x} = 1 \quad . \tag{7.1}$$

For reasons of efficiency, it is desirable to utilize the pre-integrated kernels as derived for the inversion procedure to solve the above minimization problem. Following the notation of the previous chapters, it may be of interest to know in a certain iteration k how well the data subset D_k is able to resolve a specific model update value $m_{\hat{j}k}^{\Delta}$, $\hat{j} = \mathrm{idx}_k^{\mathrm{m}}(\hat{p}, \hat{e})$. The first intuitive approach is to pathologically choose for all $\mathbf{x}_0 \in \Omega_{\hat{e}}^k$

$$J(\mathbf{x}_0, \mathbf{x}) = \begin{cases} 0 & , \text{if } \mathbf{x} \in \Omega_{\hat{e}}^k \\ 1 & , \text{if } \mathbf{x} \notin \Omega_{\hat{e}}^k \end{cases} \quad . \tag{7.2}$$

Then, a linear combination

$$\tilde{A}(\hat{j}, j) = \sum_{i \in D_k} a_i(\hat{j}) K_{ij} \tag{7.3}$$

of the pre-integrated entries of the kernel matrix $\mathbf{K}$ (defined in eq. (4.5)) is sought which closely focusses at $j = \hat{j}$. This already accounts implicitly for the separation of different model parameters p: The pre-integrated averaging kernel $\tilde{A}(\hat{j}, j)$, $\hat{j} = \mathrm{idx}_k^{\mathrm{m}}(\hat{p}, \hat{e})$, namely, should not only concentrate in space at inversion grid cell $\hat{e}$, but should as well not be contaminated by contributions from parameters p other than $\hat{p}$ (cf. Backus and Gilbert, 1968, eq. 4.10). Consistently reformulating eq. (7.2) yields the Kronecker delta $\tilde{J}(\hat{j}, j) = \delta_{\hat{j}j}$. In this pre-integrated discretized formulation, eq. (7.1) now becomes the problem to minimize

$$\sum_{j \neq \hat{j}} \left| \sum_{i \in D_k} a_i(\hat{j}) K_{ij} \right|^2 \quad \text{subject to} \quad \sum_{j} \sum_{i \in D_k} a_i(\hat{j}) K_{ij} = 1 \quad . \tag{7.4}$$

Let $\mathbf{K}'_{\hat{j}}$ denote the modified kernel matrix which misses column $\hat{j}$ and $\mathbf{k} = (\sum_j K_{ij})_{i \in D_k}$ denote the sum of the columns of $\mathbf{K}$. Then, additionally incorporating the data uncertainty matrix Σ as in section 4.3.3, eq. (7.4) finally becomes to solve the linear system

$$\begin{pmatrix} \left(\Sigma^{-1}\mathbf{K}'_{\hat{j}}\right)^T \\ \nu \left(\Sigma^{-1}\mathbf{k}\right)^T \end{pmatrix} \mathbf{a}(\hat{j}) = \begin{pmatrix} \mathbf{0} \\ \nu \end{pmatrix} \quad , \tag{7.5}$$

in a least-squares sense, where ν is a Lagrange multiplier, and $\mathbf{a}(\hat{j})$ and $\mathbf{0}$ are vector notations for $(a_i(\hat{j}))_{i \in D_k}$ and a suitable number of zeros, respectively. ν, again, can be understood as an intensity factor. Note that linear system (7.5) is in general overdetermined if the number of unknown model values exceeds the number of data, which should usually be the case. The average kernel (7.3) corresponding to the determined solution $\mathbf{a}(\hat{j})$ now quantifies how well the data residuals $r_i = d_i - F_i[m_{k-1}]$, $i \in D_k$, resolve the model update value $m_{\hat{j}k}^{\Delta}$.

Focussing on a model subdomain

For some reasons suggested below, it can be of interest to determine a linear combination of the data which is insensitive to model perturbations outside some larger focussing subdomain $\Omega_F \subset \Omega$, instead of trying to resolve the model at a singular point $\mathbf{x}_0 \in \Omega$ as discussed above. Let the focussing subdomain be the union of certain inversion grid cells $e \in E_F$, i.e.

$$\Omega_F = \bigcup_{e \in E_F} \Omega_e \quad .$$

Depending on the selection of model parameters $p \in P_F \subseteq P$ for which the sensitivity should be focussed, the index subset

$$J_F = \left\{ \mathrm{idx}^{\mathrm{m}}(p, e) \mid p \in P_F \text{ and } e \in E_F \right\}$$

is defined. Motivated by the derivation of eq. (7.5), let $\mathbf{K}'_{J_F}$ denote the modified kernel matrix which misses columns $j \in J_F$. For a given data subset $S \subset D$, then, the following linear system can be set up and solved in a least-square sense:

$$\begin{pmatrix} \left(\Sigma^{-1}\mathbf{K}'_{J_F}\right)^T \\ \nu \left(\Sigma^{-1}\mathbf{k}\right)^T \end{pmatrix} \mathbf{a}(J_F) = \begin{pmatrix} \mathbf{0} \\ \nu \end{pmatrix} \quad . \tag{7.6}$$

This yields focussing coefficients $a_i(J_F)$, $i \in S$, for which ideally

$$\sum_{i \in S} a_i(J_F) \, K_{ij}$$

vanishes for $j \notin J_F$. Note that for the computation of the kernel matrix, some background model m_0 must be assumed and all sensitivity relations only hold true

in a linearized sense in a neighbourhood of model m_0. Hence, any observed focussed misfit

$$E_F[m_0] = \sum_{i \in S} a_i(J_F) \frac{1}{\sigma_i} (d_i - F_i[m_0])^2$$

caused by perturbing m_0 should solely be due to changes in parameters $p \in P_F$ inside subdomain Ω_F.

One possible application of the focussed misfit functional $E_F[m_0]$ could be seismic reconnaissance in mechanized tunnelling. In order to properly adjust the tunnel boring machine, it is vital to detect perturbations in the structural model directly in front of the tunnel headwall which significantly differ from the presumed geological background model. It is, however, not trivial by the limited possibilities of deploying sources and receivers to conduct a satisfactory seismic inversion process inverting for a detailed distribution of specific physical properties. Focussing the sensitivity of the collected data d_i in the volume Ω_F in front of the tunnel headwall, any unforeseen perturbations of high material contrasts, such as rock inside soft soil or inclusions of water or air would show in the focussed misfit E_F. Knowing the presumed background model and data acquisition before boring the tunnel, allows to pre-compute the coefficients $a_i(J_F)$ and to compute the misfit E_F at low computational costs on site incorporating the collected (pre-processed) data d_i. In fact it is vital to have the misfit functional E_F fastly available on site, since the perturbations must be detected before the tunnel boring machine enters the focussing domain Ω_F. Model perturbations outside Ω_F should not significantly influence the misfit E_F. It is important to note that for this kind of data focussing the influence of model domain outside Ω_F is not just ignored but fully accounted for by simulating seismic wave propagation in the *complete* surrounding of the tunnel. This includes any strong surface influences which in general disturb scattering effects from deep reflectors.

Moreover, the coefficients $a_i(J_F)$ might help to determine an optimal acquisition geometry which in active seismic experiments focusses the sensitivity of the measured data set in some desired model subdomain Ω_F. Any small coefficient $a_i(J_F)$, namely, corresponds to a data sample d_i, $i = \text{idx}^{\text{d}}(\mathbf{s}_l, \mathbf{r}_m, n, f_p, \mathfrak{c}_q)$, which is not very sensitive w.r.t. changes of model parameters $p \in P_F$ inside volume Ω_F. Hence, the collection of data from source $\mathbf{s}_l$ at receiver component $(\mathbf{r}_m, n)$ at frequency f_p does not contribute to the desired illumination of volume Ω_F. Remembering that the solution of eq. (7.6) is rather cheap compared with computing the entries of the kernel matrix, different data subsets S can be tested at relatively low cost, e.g. varying the positions of sources and receivers. Note, however, that a relatively good background model must be known, since all sensitivity relations again hold in a linearized sense only.

7.2 Extending the functionality of software package ASKI

Since the software package ASKI is constructed in a modularized fashion by object-oriented programming features, it can be easily extended to other kinds of functionality. For example, any sensitivity analysis and kernel focussing tools as suggested

above in section 7.1 can be implemented in form of new modules and binary programs using the basic operations on kernels and kernel matrices already provided by the ASKI software library. In order to improve the performance of operations on very large kernel matrices, the parallelized LAPACK libraries SCALAPACK (Blackford et al., 1997) will be fully incorporated in the software package.

It is desirable to always account for the full resolving power of data subset D_k in each iteration k of the full waveform inversion procedure. That is, the size of inversion grid cells Ω_{ep}^k, possibly depending on parameter p, as well as smoothing γ should be chosen optimally. For this purpose, the "δ-ness" of every model value could be examined, as described in section 7.1. At remote points in the model domain, the averaging kernels should smear out significantly and they should overlap for neighbouring model values. This behaviour reflects the expected linear dependency of the unknowns. It is, however, not advisable to visually select by the smearing which model values to discard and which to keep. This can be done in a more strategic approach by conducting a singular value decomposition (SVD) of the regularized kernel matrix. Since SCALAPACK provides a parallelized SVD operation, it seems feasible to consider doing an SVD even for large sensitivity matrices.

As already mentioned in section 5.1, ASKI can be easily extended to other forward modelling codes, providing specific ASKI submodules that define the interface to the new code. It may happen that in specific applications of full waveform inversion or sensitivity analysis a specialized forward solver is required, e.g. due to specific geometries or physical properties of the medium. This application can still benefit from any sensitivity analysis tools and the full waveform inversion method provided by ASKI after incorporating the forward solver into the ASKI software package. Since the internal interfaces of the ASKI software modules are very general, it might even be possible to solve other geophysical inverse problems using ASKI, like a full waveform inversion of ground penetrating radar. Only a new model parameterization must be defined along with particular routines calculating the respective kernels.

But also for seismic full waveform inversion, the support of other kinds of anisotropic elastic or even anelastic model parameterizations can be added to ASKI. In particular, it should be straightforward in the frequency domain to implement sensitivity kernels for attenuation of the elastic constants. Time-dependent relaxation of stress and strain results in a convolution operation for their relation. In frequency domain, hence, it involves multiplication operations of complex-valued elastic constants $c_{ijkl}(\omega)$ which become frequency dependent. By linearly recombining the complex-valued scattering expressions derived in this thesis, one should be able to state explicit (linearized) formulas of sensitivity kernels for the attenuation of particular elastic constants at specific frequencies.

8 Bibliography

Akcelik, V., Biros, G., and Ghattas, O. (2002). Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation. In *Supercomputing, ACM/IEEE 2002 Conference*, pages 41–41. IEEE.

Aki, K. and Richards, P. G. (1980). *Quantitative Seismology, Theory and Methods.* W. H. Freeman and Company.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., and Sorensen, D. (1999). *LAPACK Users' Guide.* Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition.

Backus, G. and Gilbert, F. (1968). The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, 16(2):169–205.

Bamberger, A., Chavent, G., Hemon, C., and Lailly, P. (1982). Inversion of normal incidence seismograms. *Geophysics*, 47(5):757–770.

Blackford, L. S., Choi, J., Cleary, A., D'Azevedo, E., Demmel, J., Dhillon, I., Dongarra, J., Hammarling, S., Henry, G., Petitet, A., Stanley, K., Walker, D., and Whaley, R. C. (1997). *ScaLAPACK Users' Guide.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

Butzer, S., Kurzmann, A., and Bohlen, T. (2013). 3D elastic full-waveform inversion of small-scale heterogeneities in transmission geometry. *Geophysical Prospecting*, 61(6):1238–1251.

Chen, P., Jordan, T. H., and Zhao, L. (2007a). Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1):175–181.

Chen, P., Zhao, L., and Jordan, T. H. (2007b). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, 97(4):1094–1120.

Curtis, A. and Lomax, A. (2001). Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2):372–378.

Engl, H. W., Hanke, M., and Neubauer, A. (2000). *Regularization of inverse problems.* Kluwer Acad. Publ., Dordrecht.

8 Bibliography

Epanomeritakis, I., Akelik, V., Ghattas, O., and Bielak, J. (2008). A newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015.

Fichtner, A., Kennett, B. L., Igel, H., and Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.

Fichtner, A. and Trampert, J. (2011). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798.

Forbriger, T. (2003). Inversion of shallow-seismic wavefields: II. Inferring subsurface properties from wavefield transforms. *Geophysical Journal International*, 153(3):735–752.

Franke, R. and Nielson, G. (1980). Smooth interpolation of large sets of scattered data. *International Journal for Numerical Methods in Engineering*, 15(11):1691–1704.

Friederich, W. (1999). Propagation of seismic shear and surface waves in a laterally heterogeneous mantle by multiple forward scattering. *Geophys. J. Int.*, 136:180–204.

Friederich, W. and Dalkolmo, J. (1995). Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of Green's function in the frequency domain. *Geophys. J. Int.*, 122:537–550.

Gauthier, O., Virieux, J., and Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7):1387–1403.

Landau, L. and Lifschitz, E. (1966). *Lehrbuch der Theoretischen Physik*, volume 7. Akad.-Verl., Berlin. 2., berichtigte Aufl.

Levin, D. (1999). Stable integration rules with scattered integration points. *Journal of Computational and Applied Mathematics*, 112:181–187.

Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Liu, Q. and Tromp, J. (2006). Finite-frequency kernels based on adjoint methods. *Bull. Seismol. Soc. Am.*, 96(6):2383–2397.

Łukaszyk, S. (2004). A new concept of probability metric and its applications in approximation of scattered data sets. *Computational Mechanics*, 33(4):299–304.

Marquering, H., Nolet, G., and Dahlen, F. (1998). Three-dimensional waveform sensitivity kernels. *Geophysical Journal International*, 132(3):521–534.

Masjukov, A. V. and Masjukov, V. V. (2005). Multiscale modification of Shepard's method for multivariate interpolation of scattered data. In *Proceedings: 10th International Conference "Mathematical Modelling and Analysis 2005" and 2nd International Conference "Computational Methods in Applied Mathematics"*, pages 467–472. Technika, Vilnius.

Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52(9):1211–1228.

Nissen-Meyer, T., Dahlen, F., and Fournier, A. (2007). Spherical-earth fréchet sensitivity kernels. *Geophysical Journal International*, 168(3):1051–1066.

Parker, R. L. (1994). *Geophysical inverse theory*. Princeton university press.

Pratt, G., Shin, C., and Hicks (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362.

Resovsky, J. and Trampert, J. (2003). Using probabilistic seismic tomography to test mantle velocity–density relationships. *Earth and Planetary Science Letters*, 215(1):121–134.

Sambridge, M. (1999a). Geophysical inversion with a neighbourhood algorithm–I. Searching a parameter space. *Geophysical Journal International*, 138(2):479–494.

Sambridge, M. (1999b). Geophysical inversion with a neighbourhood algorithm–II. Appraising the ensemble. *Geophysical Journal International*, 138(3):727–746.

Sambridge, M. and Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3).

Santosa, F. and Symes, W. W. (1988). Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology. *Inverse problems*, 4(1):211.

Schroeder, W., Martin, K., and Lorensen, B. (2003). The Visualization Toolkit: An Object-Oriented Approach To 3D Graphics, 3rd version.

Schumacher, F. (2013). *ASKI– Analysis of Sensitivity and Kernel Inversion, user manual*. `http://www.rub.de/aski`.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, ACM '68, pages 517–524, New York, NY, USA. ACM.

Squillacote, A. H. (2007). *The ParaView guide: a parallel visualization application*. Kitware.

Stacey, R. (1988). Improved transparent boundary formulations for the elastic-wave equation. *Bulletin of the Seismological Society of America*, 78(6):2089–2097.

Tarantola, A. (1987). *Inverse problem theory: Methods for data fitting and model parameter estimation*. Elsevier.

Tromp, J., Komatitsch, D., and Liu, Q. (2008). Spectral-Element and Adjoint Methods in Seismology. *Communications in Computational Physics*, 3(1):1–32.

Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1):195–216.

Wardetzky, M., Mathur, S., Kälberer, F., and Grinspun, E. (2007). Discrete Laplace Operators: No Free Lunch. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, SGP '07, pages 33–37, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

Zhao, L. and Chevrot, S. (2011). An efficient and flexible approach to the calculation of three-dimensional full-wave fréchet kernels for seismic tomography–I. Theory. *Geophysical Journal International*, 185(2):922–938.

Zhao, L., Jordan, T. H., Olsen, K. B., and Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, 95(6):2066–2080.

Curriculum Vitae

Schullaufbahn

1995–2004	Gymnasium St. Michael, Ahlen
	Erwerb des Abiturs

Hochschulstudium

2004–2009	Studium der Mathematik an der TU Kaiserslautern
09/2007–06/2008	Auslandsstudium in Bath (England) im Zuge des Erasmus Programms
08/2009	Erwerb des Diploms in Mathematik, Schwerpunkt *Geomathematik*
	Diplomarbeit über ein potenzialtheoretisches Randwerteproblem mit Titel *Multiscale Processing of Geodetic Boundary Data*, Betreuung Prof. Willi Freeden
seit 10/2009	Wissenschaftlicher Mitarbeiter am Institut für Geologie, Mineralogie und Geophysik an der Ruhr-Universität Bochum