DIPLOMARBEIT

Rotational ground motion measurements in seismology and their application to the structural inverse problem

Moritz Bernauer

Mathematisches Institut
Ludwig-Maximilians-Universität München

1. Betreuer: Prof. Dr. Heiner Igel
2. Betreuer: Prof. Dr. Albert Sachs

DANKSAGUNG

Zuerst möchte ich denjenigen Menschen danken, die dazu beigetragen haben, dass gerade diese Diplomarbeit zustandegekommen ist.
Das Gemeinschaftsprojekt zwischen Mathematischem Institut und dem Department of Earth and Environmental Siences der Ludwig-Maximilians Universität München währe ohne dem Einverständnis von Herrn Prof. Dr. Hans-Jürgen Schneider in seiner Funktion als Vorsitzender des Prüfungsausschusses nicht möglich gewesen.
Die Entscheidung meine Diplomarbeit im Bereich der mathematischen Seismologie zu schreiben haben zum einen fachliches Interesse, aber auch ein unkomplizierter und angenehm direkter Umgang mit meinem ersten Betreuer Prof. Dr. Heiner Igel leicht gemacht. Eine sympathische Atmosphäre in einer internationalen Arbeitsgruppe mit globalen Kontakten haben mich nie daran zweifeln lassen, die richtige Wahl getroffen zu haben.
In besonderer Weise danken möchte ich Andreas Fichtner. Seine fachliche Kompetenz und vor allem die Selbstverständlichkeit Wissen zu teilen waren sehr wertvoll für diese Arbeit.
Mein Dank gilt auch Herrn Prof. Dr. Albert Sachs, der sich freundlicherweise als zweiter Betreuer zur Verfügung gestellt hat.

Abstract

The objective of this work is to develop a new method of seismic tomography that combines standard translational and new rotational ground motion measurements - embedded in a mathematically consistent framework.

The theory is based on the definition of an apparent shear wave speed, β_a, which is the ratio of the rms displacement velocity and the rms displacement rotation. β_a is only for a plane S-wave in unbounded, homogeneous media equal to the true shear wave speed, β. For the calculation of the sensitivity of β_a with respect to β (sensitivity kernel) the adjoint method is used (see chapter 3). The adjoint method provides the exact first derivative of a physical observable, in this case of the apparent shear wave speed, β_a.

Moreover, the computation of sensitivity kernels requires a solution of the elastic wave equation. This was already done by Fichtner and Igel (2008) using a spectral element method, which is quite accurate but suffering from numerically expensive algorithms. The new approach in this work is the computation of sensitivity kernels using the classical ray method (see chapter 2 and 4). The method includes the concept of characteristic curves for non-linear partial differential equations. The benefits are fast and flexible algorithms.

It is shown that the sensitivity of β_a with respect to β is confined to a volume sorrounding the receiver and that sensitivity kernels can be regarded as functions of frequency (see chapter 4). Being independent of the magnitude and timing of an event makes β_a-measurements an attractive observable for seismic tomography. Further sensitivity kernels are calculated in elementary models as well as in complex media with a 1-D velocity gradient leading to the result that the complexity of the model is not meaningful for the sensitivity kernel. It is the angle of incident of the incoming wave field at the receiver that affects sensitivity kernels significantly.

Finally, the sensitivity kernels are used to solve several synthetic inverse problems (see chapter 5). Due to the singularity of the sensitivity kernel at

the receiver it is problematic to use sensitivity kernels directly for inversions. Therefore a pre-conditioner for sensitivity kernels is presented that allows successful inversions. Based on an iterative minimisation of the least squares misfit between *observed* and *synthetic* data it can be demonstrated that the inversion method converges to acceptable results. In the context of seismic tomography with rotational ground motion measurements a central statement of the work is that successful inversions depend mainly on the choice of optimal frequency bands together with efficient pre-conditioners.

Contents

List of Figures

Chapter 1

General Introduction

Already more than two thousand years ago the greek scientist Eratosthenes of Cyrene determined the Earth's radius. But the question of what happens in the interior of the Earth was for a long time governed by mythology. This is not surprising because explicit tomographic Earth models are only available since information provided by seismic waves can be interpreted by computational seismology techniques.

The objective of this work is to contribute to the development of a new method of seismic tomography that combines standard translational and new rotational ground motion measurements - embedded in a mathematically consistent framework.

Chapter 1 is dedicated to a review of some basic knowledge of classical continuum mechanics in seismology. This is necessary to introduce fundamental expressions, to establish terminological precision and to define some notation conventions. Furthermore it should be emphasised that recent developments in modern ring-laser technology gave the impulse for the following considerations. So a brief survey of modern ring-laser technology in seismology completes the chapter.

1.1 Elasticity theory in seismology

The basic question that underlies this investigation is: What do seismograms tell us about Earth's structure? In principle a seismogram is a kinetic energy impulse recorded by a receiver (seismometer). Before this kinetic energy impulse arrives at the receiver it was initialised by a source (earthquake, explosion etc.) somewhere in the Earth or on its surface and then transported through the medium between source and receiver. The energy transport is achieved by continuous local oscillations of particles between source and receiver. That is the definition of a mechanical wave. In seismology mechanical waves are called seismic waves.
To describe the motion of particles in a continuum each particle is specified by its original position $\mathbf{x}$ at some reference time t_0 (Lagrangian description). We use the cartesian coordinate system, $\mathbf{x} = (x_1, x_2, x_3)$. The vector distance of a particle at time t from the position $\mathbf{x}$ that it occupies at some reference time t_0 is called displacement $\mathbf{u}$. The displacement is a function of space and time, in symbols:

$$\mathbf{u} = \mathbf{u}(\mathbf{x}, t).$$

The displacement $\mathbf{u}$ is a vector field. To calculate the gradient (grad), divergence (div) and curl of $\mathbf{u}$ we adopt common notation in vector calculus:

$$grad(\mathbf{u}) = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{pmatrix} = \partial_j u_i = \nabla \mathbf{u}$$

$$div(\mathbf{u}) = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = \partial_i u_i = \nabla \cdot \mathbf{u}$$

$$curl(\mathbf{u}) = \begin{vmatrix} i & j & k \\ \partial_1 & \partial_2 & \partial_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = \hat{\mathbf{e}}_i \epsilon_{ijk} \partial_j u_k = \nabla \times \mathbf{u}$$

$$div(grad(\mathbf{u})) = \partial_j \partial_j u_i = \nabla \cdot \nabla \mathbf{u} = \nabla^2 \mathbf{u}.$$

The distortion of a medium is analysed by the strain tensor e_{ij}. In the case of an infinitesimal displacement gradient, $|\partial_j u_i| \ll 1$, the strain tensor can be written as

$$e_{ij} \equiv \tfrac{1}{2}(\partial_i u_j + \partial_j u_i).$$

The stress tensor τ_{ij} describes internal forces acting between adjacent points within a continuum. Under the assumption of infinitesimally small particle motion the Earth can be regarded as an elastic medium. An elastic medium possesses a natural state with zero strain and stress to which it will return when applied external forces are removed. Thermodynamic arguments then lead to a linear stress-strain relation, which is just a generalisation of Hooke's law

$$\tau_{ij} = c_{ijkl} e_{kl}.$$

The tensor c_{ijkl} is called the elasticity tensor. Finally elasticity theory results in the general wave equation, valid in classical continuum mechanics for infinitesimal motions in an elastic medium with linear stress-strain relation and density distribution ρ

$$\rho \partial_t^2 u_i - \partial_j c_{ijkl} \partial_k u_l = 0.$$

Until now external body forces (seismic sources) were neglected. The description of a seismic source $\mathbf{f}$ is based on the unidirectional unit impulse applied to one particular point at $\mathbf{x} = \boldsymbol{\xi}$ and time $t = \tau$ in the direction of the x_n -axis:

$$f_i(\mathbf{x}, t) = F \delta(\mathbf{x} - \boldsymbol{\xi}) \delta(t - \tau) \delta_{in}. \tag{1.1}$$

F is the strength of the impulse, δ the Dirac delta function and δ_{in} the Kronecker delta.

Let now $\mathbf{L}$ be a vector differential operator defined on the components of $\mathbf{u}$ by

$$(\mathbf{L}(\mathbf{u}))_i \equiv \rho \partial_t^2 u_i - \partial_j c_{ijkl} \partial_k u_l. \tag{1.2}$$

Hence the equation for elastic displacement is

$$\mathbf{L}(\mathbf{u}) = \mathbf{f}. \tag{1.3}$$

1.2 Green's function and Representation theorem

This section shows how one can represent the displacement at a general point in space and time in terms of the quantities that originated the motion. These techniques are needed mostly in chapter 4.
The description of the displacement $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ throughout the volume V with surface S for a unidirectional unit impulse is called the elastodynamic Green's function. If the unit impulse is applied at $\mathbf{x} = \boldsymbol{\xi}$ and time $t = \tau$ in the direction of the x_n -axis, then we denote the ith component of displacement at general (**x**,t) by $G_{in}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$. G_{in} satisfies the equation

$$\rho\partial_t^2 G_{in} = \delta_{in}\delta(\mathbf{x} - \boldsymbol{\xi})\delta(t - \tau) + \partial_j c_{ijkl}\partial_l G_{kn} \tag{1.4}$$

As initial conditions we always use that $\mathbf{G}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$ and $\partial_t\mathbf{G}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$ are zero for $t \leq \tau$ and $\mathbf{x} \neq \boldsymbol{\xi}$. $\mathbf{G}$ is uniquely specified by the boundary conditions on S. If the boundary conditions are independent of time, then the following reciprocal relation for source and receiver times is valid

$$\mathbf{G}(\mathbf{x}, t; \boldsymbol{\xi}, \tau) = \mathbf{G}(\mathbf{x}, t - \tau; \boldsymbol{\xi}, 0) = \mathbf{G}(\mathbf{x}, -\tau; \boldsymbol{\xi}, -t) \tag{1.5}$$

In order to express the displacement $\mathbf{u}$ due to body forces $\mathbf{f}$ throughout V and to boundary conditions S there is a fundamental theorem in seismology called Representation theorem. It states a way in which the displacement $\mathbf{u}$ at a certain point is made up from contributions due to the force $\mathbf{f}$ throughout V, plus contributions due to the traction $\mathbf{T}(\mathbf{u}, \mathbf{n})$ and to the displacement $\mathbf{u}$ itself on S. In symbols:

$$\begin{aligned} u_n(\mathbf{x}, t) &= \int_{-\infty}^{\infty} d\tau \int\int\int_V f_i(\boldsymbol{\xi}, \tau) G_{in}(\boldsymbol{\xi}, t - \tau; \mathbf{x}, 0) dV(\boldsymbol{\xi}) \\ &+ \int_{-\infty}^{\infty} d\tau \int\int_S \Big(G_{in}(\boldsymbol{\xi}, t - \tau; \mathbf{x}, 0) T_i(\mathbf{u}(\boldsymbol{\xi}, \tau), \mathbf{n}) \\ &- u_i(\boldsymbol{\xi}, \tau) c_{ijkl} n_j G_{kn,l}(\boldsymbol{\xi}, t - \tau; \mathbf{x}, 0) \Big) dS(\boldsymbol{\xi}) \end{aligned} \tag{1.6}$$

It is not the aim of that work to proof this theorem nor to derive all later

results step by step beginning at that level but the Representation theorem is fundamental especially for considerations in chapter 4. Therefore it should be mentioned at this point. For a detailed proof and further derivations concerning the Representation theorem the reader is refered to e.g. Aki/Richards (2002).

Before we actually start with the theory some words shall be said to modern ring-laser instruments that initiated the approach of doing inversion with rotations.

1.3 The ring-laser

Figure 1.1: *Ring-laser (taken from: Schreiber, U., Ring Laser Principles and Design Considerations)*

Seismological observations are still primarily based on three-component translational ground motion measurements (displacement, velocity or acceleration recorded in three orthogonal directions, usually N-S, E-W, up-down). But for a complete description of the motion of a deformable body at a given point in the context of infinitesimal deformation, one needs three components of translation, six components of strain and three components of rotation. The measurement of the deformation of the Earth (strain) is also quite routinely used.
But practically rotational motions induced by seismic waves have been ignored for a long time, first because rotational effects where thought to be small, and second because sensitive measuring devices where not available. This has been changed since modern ring-laser technology has provided the means to develop instruments that allow in principle the observation of rotational motions in a wide frequency band and epicentral distance range. Yet

the fact that inertial seismometers commonly used in seismology are sensitive to and thus polluted by rotational motions should motivate further investigation of this new observable. As an example ring-laser recordings of the Sumatra earthquake (26.12.2004) are shown in figure 1.2.

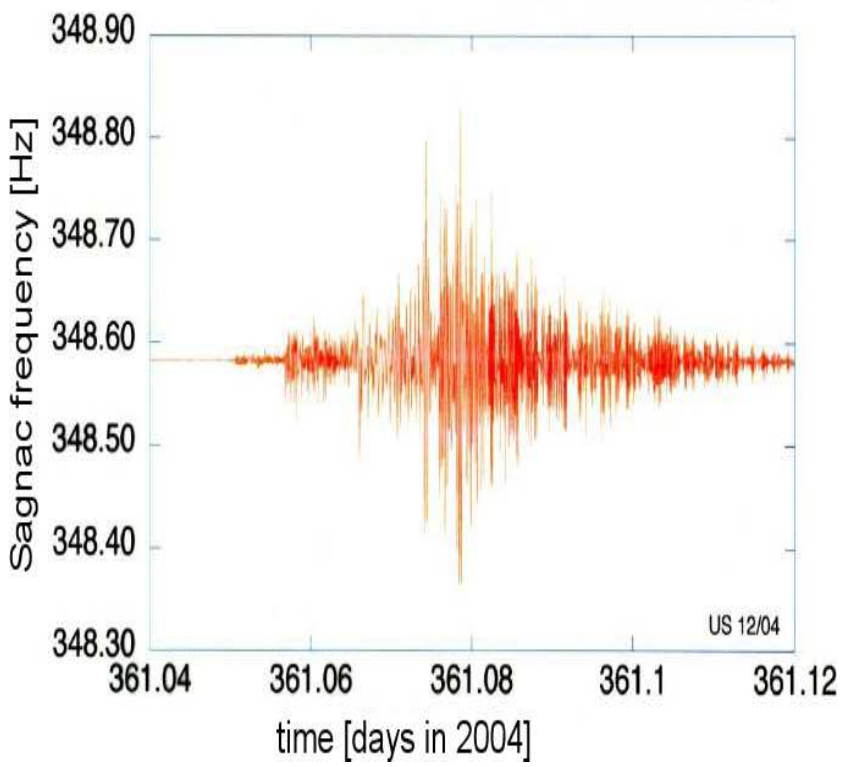


Figure 1.2: *Ring-laser data from the Sumatra earthquake (vertical component, taken from Zürn, W., Das verheerende Sumatra-Beben am 26.12.2004 aus seismologischer Sicht)*

Ring-laser gyroscopes were developed primarily to observe variations in Earth's absolute rotation rate. But they turened out to be sufficiently accurate to record seismic rotations, too.
The measurement principle of a ring-laser is sketched in figure 1.3. The instrument is mounted horizontally and rigidly attached to the ground. Two counter-rotating single-mode laser beams interfere to generate a beating in case the system rotates with respect to the surface normal.

The Sagnac beat frequency δf is directly proportional to the rotation rate $\mathbf{\Omega}$ around the surface normal $\mathbf{n}$ of the ring laser system, expressed by the Sagnac equation

$$\delta f = \frac{4A}{\lambda P}\mathbf{n} \cdot \mathbf{\Omega} \tag{1.7}$$

where P is the perimeter of the instrument, A the area and λ the laser wavelength.

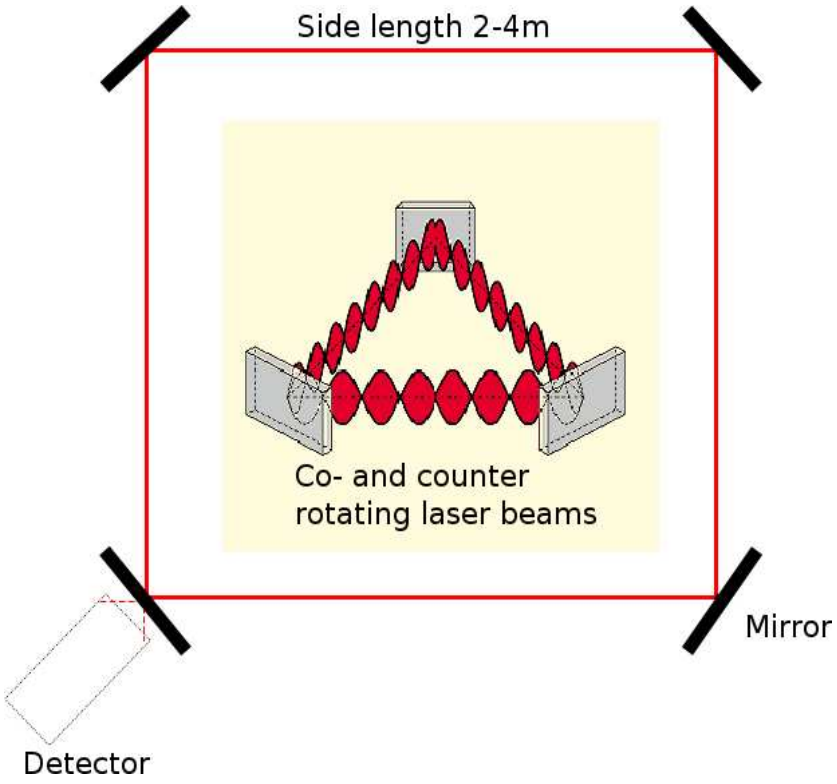


Figure 1.3: *Ring-laser concept (taken from Schreiber, U., Ring Laser Principles and Design Considerations)*

Simpler low-cost ring-laser instruments with equivalent sensitivity have been developed in order to capture signals not only from teleseismic but also from local events.

Chapter 2

Wave equation and ray theory

Anisotropy plays an important role in realistic Earth models. Nevertheless from now on we restrict ourselves to isotropic models. This simplification does not affect the results of this work. As this is a completely new approach in seismic tomography the focus is still on the establishment of a consistent theory as well as on a physical interpretation of its results. For this purpose simple models are the most practical ones. Therefore in this chapter the elastic wave equation will be adapted to an isotropic Earth model. Then the wave equation is solved by the ray method, because the ray method is an efficient tool that makes it possible to run various fast but meaningful computational experiments.

2.1 The isotropic elastic wave equation

In the absence of external forces the general elastic wave equation is

$$\rho\partial_t^2 u_i - \partial_j c_{ijkl}\partial_k u_l = 0.$$

From the symmetry of the stress and strain tensors and a thermodynamic argument it follows that the maximum number of independent elements of the elasticity tensor c_{ijkl} is 21. In an isotropic body this reduces further to

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\delta_{ik}\delta_{jl} + \mu\delta_{il}\delta_{jk}. \tag{2.1}$$

The material parameters λ and μ are in general functions of position. According to the wave equation we get

$$\partial_j c_{ijkl}\partial_k u_l = \partial_i(\lambda\partial_k u_k) + \partial_j(\mu\partial_i u_j + \mu\partial_j u_i).$$

We assume that spatial variations of the material parameters are much smaller than variations of the strain e_{ij} ($\equiv \frac{1}{2}(\partial_i u_j + \partial_j u_i)$). Hence

$$\partial_j c_{ijkl} \partial_k u_l = \lambda \partial_i \partial_k u_k + \mu \partial_j \partial_i u_j + \mu \partial_j \partial_j u_i = (\lambda + 2\mu) \nabla \nabla \cdot \mathbf{u} - \mu \nabla \times \nabla \times \mathbf{u},$$

where we used the identity $\nabla^2 \mathbf{u} = \nabla \nabla \cdot \mathbf{u} - \nabla \times \nabla \times \mathbf{u}$.
This yields the isotropic elastic wave equation for smoothly varying material parameters λ and μ

$$\rho \partial_t^2 \mathbf{u} - (\lambda + 2\mu) \nabla \nabla \cdot \mathbf{u} + \mu \nabla \times \nabla \times \mathbf{u} = 0. \tag{2.2}$$

This equation can be further simplified using the following theorem:

Theorem 2.1.1 (Helmholtz Theorem)**.**
Let $\boldsymbol{u} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3, \boldsymbol{x} \longmapsto \boldsymbol{u}(\boldsymbol{x})$ *be a differentiable vector field with* $\lim_{|\boldsymbol{x}| \to \infty} \boldsymbol{u}(\boldsymbol{x}) = 0$. *Then there exists a unique decomposition of* $\boldsymbol{u}$ *by a scalar field* φ *and a vector field* $\boldsymbol{\phi}$ *with* $\nabla \cdot \boldsymbol{\phi} = 0$*:*

$$\boldsymbol{u} = -\nabla \varphi + \nabla \times \boldsymbol{\phi}.$$

φ *and* $\boldsymbol{\phi}$ *can be written as*

$$\begin{aligned} \varphi(\boldsymbol{x}) &= \frac{1}{4\pi} \int \int \int_{\mathbb{R}^3} \frac{\nabla \cdot \boldsymbol{u}(\boldsymbol{y})}{|\boldsymbol{x} - \boldsymbol{y}|} dV(\boldsymbol{y}) \\ \boldsymbol{\phi}(\boldsymbol{x}) &= \frac{1}{4\pi} \int \int \int_{\mathbb{R}^3} \frac{\nabla \times \boldsymbol{u}(\boldsymbol{y})}{|\boldsymbol{x} - \boldsymbol{y}|} dV(\boldsymbol{y}) \end{aligned}$$

So the displacement field $\mathbf{u}$ can be written as the sum of a curl-free scalar field φ and a divergence-free vector field $\boldsymbol{\phi}$. Now apply the $\nabla\times$ operator respectively the $\nabla\cdot$ operator to $\mathbf{u}$ for separating the displacement field into a divergence-free respectively a curl-free part.
Under the assumption that variation of density ρ is also small we get:

$$divergence - free: \quad \partial_t^2 \boldsymbol{\phi} - \beta^2 \nabla^2 \boldsymbol{\phi} = 0 \tag{2.3}$$

$$curl - free: \quad \partial_t^2 \varphi - \alpha^2 \nabla^2 \varphi = 0 \tag{2.4}$$

with wave velocities: $\beta = \sqrt{\frac{\mu}{\rho}}$ and $\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}}$.

Obviously is $\alpha > \beta$. That is why in seismology the body wave with no rotation is called primary wave (P-wave) and the body wave with no change in volume is called secondary wave (S-wave or shear wave).
The object of this work is to find a way how to include rotational ground motion measurements in seismic tomography. So further investigation concernes only S-waves. From now on with $\mathbf{u}$ we denote the displacement field generated only by S-waves. And we modify the vector differential operator $\mathbf{L}$ defined on the components of $\mathbf{u}$ to:

$$\mathbf{L}(\mathbf{u}) \equiv \rho\partial_t^2\mathbf{u} - \mu\nabla^2\mathbf{u}. \tag{2.5}$$

2.2 Ray theory

The topic of this section is the solution of the elastic wave equation based on high-frequency asymptotic ray theory. The derivation of the equations is kept in a compact form and can be found in detail in Červený, V. (2001). The ray method yields only an approximation of the actual wave field, because travel times and amplitudes are calculated only along certain ray paths. Ray properties (reflections, refractions, etc.) are based on analogy with geometrical optics. Figure 2.1 shows rays reflected from two selected interfaces, recorded at receivers distributed evenly along a line profile.

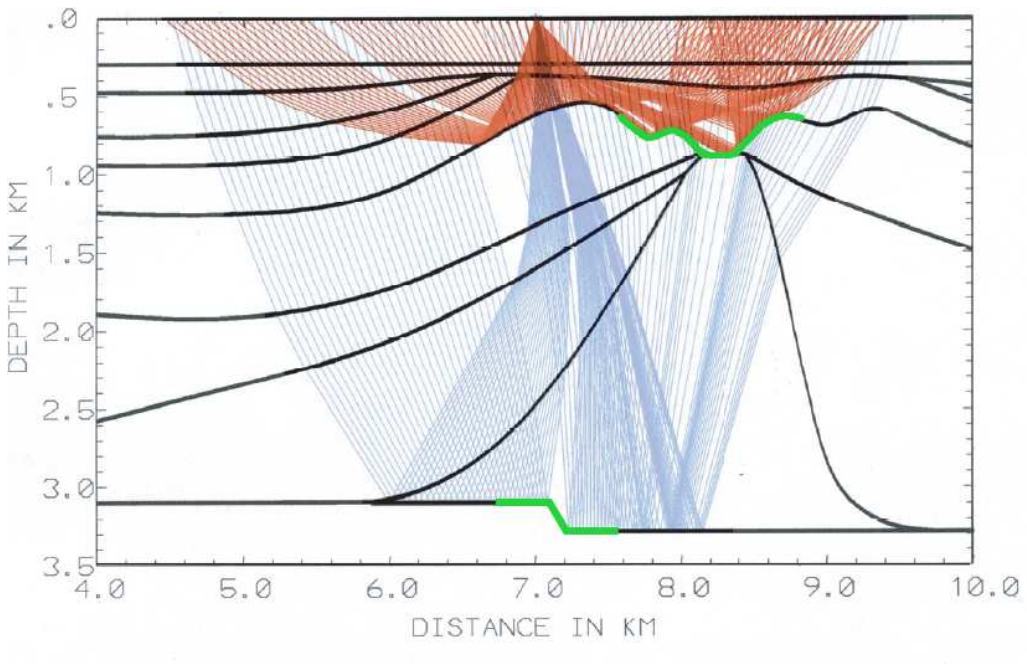


Figure 2.1: *Rays of two reflected waves generated by one shot point; taken from Johanna Brokešová, Asymptotic Ray Method in Seismology - A Tutorial, 2006*

In this model some parameter variations are rather abrupt (green marks), leading to so called *shadow zones* without rays and *caustic points*, where infinitely many rays simultaneously converge at one single point. In both cases application of ray theory becomes problematic. To avoid this situations the model has to be smooth by means of small parameter variations with respect to the wavelength.
In this work we do not have to care about caustics and shadow zones, be-

cause we calculate sensitivities only in homogeneous models or models with 1-D velocity gradient.

In ray theory the solution of the elastic wave equation at point $\mathbf{x}$ and time t is assumed in the form of an asymptotic *ray series* expansion:

$$\mathbf{u}(\mathbf{x},t) = \sum_{k=0}^{\infty} \mathbf{U}_k(\mathbf{x}) F_k(t - \tau(\mathbf{x})). \tag{2.6}$$

The $\mathbf{U}_k$'s are the vectorial amplitudes, τ is the so-called eikonal, which can be interpreted as travel time, and $F(t')$ is an analytic signal, i.e. a complex-valued function whose real and imaginary parts form a Hilbert pair

$$F(t') = f(t') + i\mathcal{H}[f(t')] \tag{2.7}$$

For example the exponential function $exp(it') = cos(t') + isin(t')$ with $\mathcal{H}[cos(t')] = sin(t')$ is an analytic signal.
In the frequency-domain, the ray series corresponding to (2.6) has the form

$$\mathbf{u}(\mathbf{x},\omega) = 2h(\omega)f(\omega)exp[-i\omega(\tau(\mathbf{x}))] \sum_{k=0}^{\infty} \mathbf{U}_k(\mathbf{x})(-i\omega)^{-k} \tag{2.8}$$

with the Fourier transform $f(\omega) = \mathcal{F}[f(t)]$ and the Heaviside function $h(\omega)$. The frequency-domain series is assumed to be an asymptotic power series for high ω. Typically the error decreases with growing ω for any fixed number of terms in the series. On the other hand, when ω is fixed, the error may grow for N greater than a certain value. Thus we are concerned with what happens as ω tends to ∞ rather than what happens as the number k of terms grows.
In practice it is common to consider only the leading term in the series. In the following we restrict ourselves only to the *zero-order ray solution*

$$\mathbf{u}(\mathbf{x},t) = \mathbf{U}(\mathbf{x}) F(t - \tau(\mathbf{x})) \tag{2.9}$$

with ray amplitude $\mathbf{U}$ and travel time τ. Surfaces $\tau(\mathbf{x}) = const.$ represent wavefronts. An important quantity is the gradient of τ, called the slowness vector $\mathbf{p}$. It is a vector perpendicular to the wavefront and given by

$$p_i(\mathbf{x}) = \frac{\partial \tau(\mathbf{x})}{\partial x_i} = \frac{n_i^{\tau}(\mathbf{x})}{c(\mathbf{x})} \tag{2.10}$$

where $\mathbf{n}^{\tau}$ is the unit vector perpendicular to the wavefront and c the phase velocity.

Specifying in 2.9 for example $F(t) = exp(-i\omega t)$, for some fixed frequency, we obtain the so-called time-harmonic solution

$$\mathbf{u}(\mathbf{x}, t, \omega) = \mathbf{U}(\mathbf{x}) exp[-i\omega \cdot (t - \tau(\mathbf{x}))] \tag{2.11}$$

In order to determine the functions $\mathbf{U}(\mathbf{x})$ and $\tau(\mathbf{x})$, we require the solution to fit asymptotically the corresponding equation of motion. Substituting the zero-order ray solution (2.9) into the isotropic S-wave equation (2.5) without body forces and collecting the terms with the same time derivative of F yields

$$-\ddot{F} N_i(\mathbf{U}, \nabla\tau) + \dot{F} M_i(\mathbf{U}, \nabla\tau) - F L_i(\mathbf{U}) = 0 \tag{2.12}$$

and in the frequency-domain

$$-(i\omega)^2 N_i(\mathbf{U}, \nabla\tau) + (i\omega)^1 M_i(\mathbf{U}, \nabla\tau) - (i\omega)^0 L_i(\mathbf{U}) = 0 \tag{2.13}$$

in both cases with

$$\begin{aligned} N_i(\mathbf{U}, \nabla\tau) &= \mu U_i \partial_j \tau \partial_j \tau - \rho U_i \\ M_i(\mathbf{U}, \nabla\tau) &= \mu[2\partial_j U_i \partial_j \tau + U_i \partial_j \partial_j \tau] \\ L_i(\mathbf{U}) &= \mu \partial_j \partial_j U_i \end{aligned}$$

Equations (2.12) and (2.13) represent now the equation of motion. Because of different orders of derivatives of F the solution has to satisfy the conditions $\mathbf{N} = 0$, $\mathbf{M} = 0$ and $\mathbf{L} = 0$. In general it is not possible to satisfy all of the conditions, since they represent nine equations while there are only four parameters (three components of $\mathbf{U}$ and τ). However, in high-frequency approximation only the first two conditions are used to determinate the eikonal and the ray amplitude. This is directly justified by equation (2.13) where the terms with ω^2 and ω^1 dominate over the one with ω^0 for high frequencies.
The condition $\mathbf{N} = 0$ yields the so-called *eikonal equation*. Besides τ it constrains also the direction (polarisation) of $\mathbf{U}$. The magnitude of $\mathbf{U}$ is determined from the *transport equation*, given by $\mathbf{M} = 0$.

The eikonal equation

If we define a matrix $\mathbf{\Gamma}$ by $\Gamma_{ik} := \frac{\mu}{\rho}\delta_{ik}p_l p_l$ we can write

$$N_i(\mathbf{U}, \nabla\tau) = \mu U_i p_j p_j - \rho U_i = (\Gamma_{ji} - \delta_{ji})U_i = 0 \tag{2.14}$$

Regarding (2.14) as an eigenvalue problem of the matrix $\mathbf{\Gamma}$ leads to the useful eikonal equation

$$|\mathbf{p}|^2 = \frac{1}{\beta^2} \tag{2.15}$$

To solve the eikonal equation we use the method of characteristic curves. In the concept of characteristics the static Hamilton-Jacobi equation

$$H(x_i, \tilde{p}_i) = 0 \quad , \quad \tilde{p}_i = \partial_i \boldsymbol{\psi} \tag{2.16}$$

plays an important role. $\boldsymbol{\psi}$ is the function to be determined, so in our case $\boldsymbol{\psi}$ is $\boldsymbol{\tau}$ because $p_i = \partial_i \boldsymbol{\tau}$. In the formalism $\tilde{p}_i' s$ and $x_i' s$ are treated as independent variables. The equation (2.16) is solved by the use of characteristic curves, which satisfy the so-called canonical equations

$$\frac{dx_i}{du} = \frac{\partial H}{\partial \tilde{p}_i} \quad , \quad \frac{d\tilde{p}_i}{du} = -\frac{\partial H}{\partial x_i} \quad , \quad \frac{d\tau}{du} = \tilde{p}_k \frac{\partial H}{\partial \tilde{p}_k} \quad , \quad i = 1, 2, 3 \tag{2.17}$$

with a flow parameter u, i.e. $x_i = x_i(u)$ and $p_i = p_i(u)$, along the curve. If we choose

$$H(x_i, p_i) = \frac{1}{\eta}[(p_k p_k)^{\frac{\eta}{2}} - \beta^{-\eta}] \quad , \quad \eta \in \mathbb{Z} \backslash \{0\} \tag{2.18}$$

the eikonal equation is satisfied along the characteristic curve of H. In this way the eikonal equation (non-linear partial differential equation) is replaced by a system of seven ordinary differential equations (2.17). The characteristic curve (6-dimensional) is obtained from the first six coupled equations of the system (2.17). The projection of this 6D curve into the 3D space, i.e. the geometrical trajectory of the characteristic curve, is called the *seismic ray*. So the first six equations of the system (2.17), we call them *ray tracing system*, provide ray paths, travel times and slowness vectors.
If we insert (2.18) into the canonical equations (2.17), we obtain the ray tracing sytem

$$\frac{dx_i}{du} = \beta^{2-\eta} p_i \quad , \quad \frac{dp_i}{du} = \frac{1}{\eta}\frac{\partial}{\partial x_i}\frac{1}{\beta^\eta} \tag{2.19}$$

and

$$\frac{d\tau}{du} = (p_k p_k)^{\frac{\eta}{2}} = \frac{1}{\beta^\eta} \tag{2.20}$$

The parameter $\eta \in \mathbb{Z}$ controls the physical meaning of the flow parameter u. For seismological applications reasonable $\eta's$ are for example

- $\boldsymbol{\eta = 0}$
  In this case (2.18) is not defined and must be replaced by the limit for $\eta \to 0$. Using the l'Hospital rule we obtain for H in the limit

  $$H(x_i, p_i) = \frac{1}{2} ln(p_k p_k) + ln(\beta) = \frac{1}{2} ln(\beta^2 p_k p_k) \tag{2.21}$$

  Then it follows directly from $\frac{d\tau}{du} = p_k \frac{\partial H}{\partial p_k}$ that u is equal to the travel time τ. And the ray tracing system has the form

  $$\frac{dx_i}{d\tau} = \beta^2 p_i \quad , \quad \frac{dp_i}{d\tau} = -\frac{1}{\beta} \frac{\partial}{\partial x_i} \beta \tag{2.22}$$

- $\boldsymbol{\eta = 1}$
  Here u has the meaning of the arclength s along the ray and we have the equations

  $$\frac{dx_i}{ds} = \beta p_i \quad , \quad \frac{dp_i}{ds} = -\frac{1}{\beta^2} \frac{\partial}{\partial x_i} \beta \quad , \quad \frac{d\tau}{ds} = \frac{1}{\beta}. \tag{2.23}$$

Note that to calculate rays and travel times in isotropic medium, it is sufficient to know the distribution of the velocity β.
The calculation of sensitivity kernels in chapter 4 is restricted to homogeneous models and models with 1-D velocity gradient in x_3-direction. In both cases we set as initial slowness condition $p_{20} = 0$, so that rays are situated in the $x_1 - x_3$ plane. The ray calculation is described completely by the four ray tracing equations for x_1, x_3 and p_1, p_3 with corresponding initial conditions $x_{10}, x_{30}, p_{10}, p_{30}$.
In the homogeneous case calculation of travel times is easy because rays are straight lines with constant slowness vector $\mathbf{p}_0$. The wavefronts are spheres.

In a 1-D model with linearly increasing velocity in the x_3-direction, i.e. $\beta(x_3) = \beta_0 + bx_3$, it can be shown that rays are segments of circles. The ray path is given by

$$[x_1 - x_{10} - \frac{p_{30}\beta_0}{bp}]^2 + [x_3 - x_{30} + \frac{\beta_0}{b}]^2 = \frac{1}{p^2 b^2} \tag{2.24}$$

which is the equation of a circle with the center at $[x_{10} + (pb)^{-1}p_{30}\beta_0; x_{30} - \beta_0(b^{-1})]$ and radius $(pb)^{-1}$. Here x_{10} and x_{30} are the coordinates of the initial point , it is $\beta_0 = \beta(x_{30}),\quad p_{30} = (\frac{1}{\beta_0^2 - p^2})^{1/2}$ and $p_1 = p_{10} = const.$ along a certain ray. The wavefronts are again spheres expressed by

$$[x_1 - x_{10}]^2 + [x_3 - x_{30} + \frac{\beta_0}{b}(1 - cosh(b(\tau - \tau_0)))]^2 = \frac{\beta_{0^2}}{b^2} sinh^2(b(\tau - \tau_0)). \quad (2.25)$$

From the equation for the wavefronts we can derive the travel time $\tau(S, R)$ from any point S to any other point R in the model with constant gradient of velocity b. Denote by β_S the velocity at S, by β_R the velocity at R, and by r the distance between S and R. Then

$$\tau(S, R) = \frac{1}{b} arsinh[\frac{br}{(\beta_R \beta_S)^{1/2}}(1 + \frac{b^2 r^2}{4\beta_S \beta_R})^{1/2}] \quad (2.26)$$

Now we can calculate ray paths and travel times. The magnitude of amplitudes is derived by the transport equation as follows.

The transport equation

The transport equation is given by the condition $M_i(\mathbf{U}, \nabla\tau) = \mu[2\partial_j U_i \partial_j \tau + U_i \partial_j \partial_j \tau] = 0$. Hence

$$2(\nabla \mathbf{U}) \cdot (\nabla\tau) + \mathbf{U}\ (\nabla^2\tau) = 0 \tag{2.27}$$

As we are only interested in the magnitude A of amplitudes it is sufficient to consider

$$2(\nabla A) \cdot (\nabla\tau) + A\ (\nabla^2\tau) = 0. \tag{2.28}$$

We solve the transport equation in the context of *ray tubes*. A ray tube is a familiy of rays with ray parameters $\mathbf{p}_I$ being in the interval $(\mathbf{p}_1, \mathbf{p}_1 + d\mathbf{p}_1) \times (\mathbf{p}_2, \mathbf{p}_2 + d\mathbf{p}_2)$. A part of such a tube, bounded by two wavefronts at times t_0 and t is shown in figure 2.2

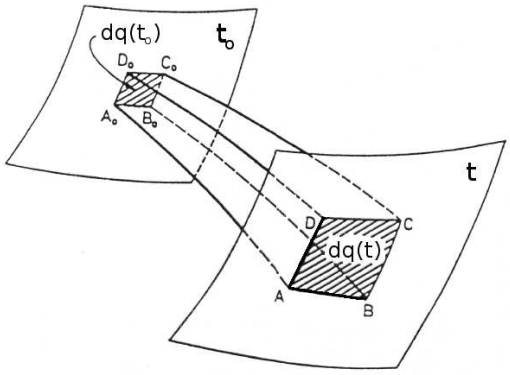


Figure 2.2: *Section of a ray tube (taken from Červený, V. (2001)*

Integration of the transport equation over the volume of an infinitesimally fine ray tube together with the Gauss theorem yields the continuation formula

$$\frac{A_0^2 dq_0}{\beta_0} = \frac{A_1^2 dq_1}{\beta_1} = const. \tag{2.29}$$

This means for example that the magnitude of the amplitude increases for a focusing ray.
Again it is not difficult to derive properties for the homogeneous case, where $\beta_0 = \beta_1$, hence $A_0^2 dq_0 = A_1^2 dq_1$.

Applying the theorem on intersecting lines

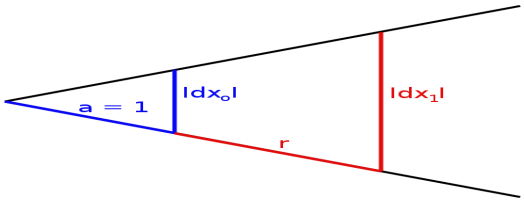


Figure 2.3: *With the theorem on intersecting lines it can be shown in an elementary way that in homogeneous media along a certain ray the magnitude of the amplitude is proportional to* $\frac{1}{r}$.

yields

$$A_1 = \sqrt{\frac{dq_0}{dq_1}} A_0 = \frac{|dx_0|}{|dx_1|} A_0 = \frac{1}{r} A_0. \tag{2.30}$$

So along a certain ray the magnitude of the amplitude is proportional to $\frac{1}{r}$. In practical applications we will transfer this property to the 1-D velocity gradient model. Deviation of the exact solution is supposed to be negligible compared e.g. to the accuracy of approximating the wavefield by the zero-order ray solution.

At the end of chapter 2 we are able to solve the isotropic wave equation by determing travel times and amplitudes using ray theory. In the next chapter we introduce the mathematical background for the computation of the exact first derivative of a physical observable. This is the adjoint method, which will be a fundamental tool for the calculation of sensitivity kernels.

Chapter 3

The adjoint method

Our issue is to reconstruct the Earth's structure combining rotational and translational ground motion measurements associated with S-waves. The elastic wave equation for S-waves $\rho\partial_t^2\mathbf{u} - \rho\beta^2\nabla^2\mathbf{u} = \mathbf{f}$ shows that material properties are reflected by the S-wave velocity $\beta = \sqrt{\frac{\mu}{\rho}}$. So detection of local variations of β in our Earth model is the key for doing structural inversion.

The following simple example confirms that the ratio of the displacement velocity amplitude $|\dot{\mathbf{u}}|$ and the rotation amplitude $|\nabla \times \mathbf{u}|$ is an attractive quantity for inversion with rotations:

Example: Plane S-wave
In a homogeneous isotropic medium a plane S-wave is represented via:

$$\mathbf{u} = \mathbf{A}cos(\mathbf{kx} - wt) = \begin{pmatrix} 0 \\ A_2 cos(k_1 x_1 - wt) \\ 0 \end{pmatrix} \tag{3.1}$$

with propagation in x_1-direction, polarisation in x_2-direction and phase velocity $\beta = \frac{\omega}{|\mathbf{k}|}$.

Hence

$$\begin{aligned} |\dot{\mathbf{u}}| &= \omega A_2 sin(k_1 x_1 - wt) \\ |\nabla \times \mathbf{u}| &= k_1 A_2 sin(k_1 x_1 - wt) \end{aligned}$$

and

$$\frac{|\dot{\mathbf{u}}|}{|\nabla \times \mathbf{u}|} = \frac{\omega}{k_1} = \frac{\frac{2\pi}{T}}{\frac{2\pi}{\lambda}} = \beta \tag{3.2}$$

The plane wave in the example exists only in homogeneous unbounded media. As measurements are normally done at the surface of the Earth and realistic models are not homogeneous we cannot correlate $|\dot{\mathbf{u}}|/|\nabla \times \mathbf{u}|$ directly with the pointwise S-wave speed β. But in general the ratio of the amplitudes of displacement velocity and rotation always keeps the unit of a velocity. This fact and the simple example from above justify the assumption that the ratio under consideration is affected by the S-wave velocity structure in a wider vicinity of the measurement point $\mathbf{x}^r$ and thus contains some information about the S-wave velocity structure. But how much is it affected and in which regions exactly? To answer this question we proceed as follows:

Let $\mathbf{u}(\mathbf{p}; \mathbf{x}, t)$ be a displacement field recorded over time t at the location $\mathbf{x} = \mathbf{x}^r$ depending on some model parameters $\mathbf{p} = (\rho, \mu, ...)$. We define the *apparent shear wave speed* β_a by the ratio of the rms amplitudes of the velocity and the rotation displacement field:

$$\beta_a(\mathbf{p}; \mathbf{x}^r, t) := \frac{1}{2} \frac{\|\dot{\mathbf{u}}(\mathbf{p}; \mathbf{x}^r, t)\|_2}{\|\boldsymbol{\omega}(\mathbf{p}; \mathbf{x}^r, t)\|_2} \tag{3.3}$$

with $\|\boldsymbol{\omega}(\mathbf{p}; \mathbf{x}^r, t)\|_2 := \|\frac{1}{2}\nabla \times \mathbf{u}(\mathbf{p}; \mathbf{x}^r, t)\|_2$ and $\|(\cdot)\|_2 := \sqrt{\int_{\mathbb{R}} (\cdot)^2 dt}$.

To simplify the notation we introduce the symbols:

$$A_v(\mathbf{x}^r) := \|\dot{\mathbf{u}}(\mathbf{p}; \mathbf{x}^r, t)\|_2 = \sqrt{\int_{\mathbb{R}} \dot{\mathbf{u}}(\mathbf{p}; \mathbf{x}^r, t)^2 dt} \tag{3.4}$$

$$A_\omega(\mathbf{x}^r) := \|\boldsymbol{\omega}(\mathbf{p}; \mathbf{x}^r, t)\|_2 = \sqrt{\int_{\mathbb{R}} \left(\frac{1}{2}\nabla \times \mathbf{u}(\mathbf{p}; \mathbf{x}^r, t)\right)^2 dt} \tag{3.5}$$

and

$$\beta_a(\mathbf{x}^r) = \frac{1}{2} \frac{A_v(\mathbf{x}^r)}{A_\omega(\mathbf{x}^r)}. \tag{3.6}$$

Now we want to know something about the influence of the S-wave velocity structure on the measurement of the apparent shear wave speed β_a. From a mathematical point of view it is obvious to calculate the first derivative of β_a with respect to the model parameter β. The definition 3.6 directly yields the sensitivity kernel, $\delta_\beta \beta_a$, for the apparent shear wave speed, β_a, with respect

to the S-wave speed, β. In symbols this is

$$\frac{1}{\beta_a}\delta_\beta\beta_a = \frac{1}{A_v}\delta_\beta A_v - \frac{1}{A_\omega}\delta_\beta A_\omega. \tag{3.7}$$

The sensitivity densities $\delta_\beta A_v$ and $\delta_\beta A_\omega$ can be calculated by applying a mathematical method called the adjoint method. The rest of chapter 2 is dedicated to explain what the adjoint method is. This concept can also be found in Fichtner et al. (2006).

3.1 Operator derivatives

The adjoint method permits the computation of the exact first derivative of a physical observable (e.g. displacement field) with respect to its parameters (e.g. S-wave speed). The method is based on a general notation for derivatives of an arbitrary operator that shall be introduced in this section.

Definition 3.1.1 (Linear operator)**.**
An operator T mapping the normed space X to the normed space Y is called linear *if it satisfies*

$$\begin{aligned} i) \quad & T(x+y) = T(x) + T(y) \\ ii) \quad & T(\alpha x) = \alpha T(x) \end{aligned}$$

for all $x,y \in X$ and $\alpha \in \mathbb{R}$. The set of all linear operators mapping X to Y is denoted by $[X \to Y]$.

Definition 3.1.2 (First derivative)**.**
Let P be an operator mapping the open set $\Omega_X \subset X$ of the Banach space X to the subset $\Omega_Y \subset Y$ of the Banach space Y. Moreover, assume that for an element $x_0 \in \Omega_X$ there is a linear operator $U \in [X \to Y]$ such that the relation

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[P(x_0 + \varepsilon x) - P(x_0)] = U(x)$$

holds for all $x \in X$. If it exists, the operator U is called the first derivative *of P in $x_0 \in \Omega_X$ and we write*

$$U =: \mathfrak{D}P(x_0), \quad U(x) = \mathfrak{D}P(x_0)(x).$$

The derivative defined above is also called *weak derivative* or *Gâteaux derivative*. If not explicitly stated otherwise, X and Y will, in what follows, denote two Banach spaces and P will be an operator mapping an open subset $\Omega_X \subset X$ to a subset $\Omega_Y \subset Y$.

Definition 3.1.3 (Differentiability, Fréchet derivative)**.**
Let K be the subset of X with $\|x\| = 1$. If

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[P(x_0 + \varepsilon x) - P(x_0)] = U(x) = \mathfrak{D}P(x_0)(x)$$

exists uniformly with respect to $x \in K$, then P is said to be differentiable *in x_0 and $\mathfrak{D}P(x_0)$ is called the* Fréchet derivative *of P in x_0.*

Theorem 3.1.1.
An operator $P : X \to Y$ is differentiable in x_0 if and only if the following properties are true:

$i_)$ $\exists\ U \in [X \to Y]$

$ii_)$ $\forall\ \varepsilon > 0\ \exists\ \delta > 0$ *,so that* $\|\Delta x\| < \delta, \quad \Delta x \in X$ *implies*
$\|P(x_0 + \Delta x) - P(x_0) - U(\Delta x)\| \ \leq\ \varepsilon \|\Delta x\|.$

For linear operators it is straightforward to show that differentiability in $x_0 \in \Omega_X$ implies continuity in $x_0 \in \Omega_X$. If $P \in [X \to Y]$, then the derivative of P is just given by

$$\mathfrak{D}P(x_0)(x) = P(x).$$

This means that the derivative is constant for all $x_0 \in \Omega_X$ and depends only on the direction of differentiation.

Theorem 3.1.2 (Linearity)**.**
Let $P_1, P_2 : (\Omega_X \subset X) \to (\Omega_Y \subset Y)$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. Moreover, let P be defined by $P := \alpha_1 P_1 + \alpha_2 P_2$. If the derivatives $\mathfrak{D}P_1(x_0)$ and $\mathfrak{D}P_2(x_0)$ with $x_0 \in \Omega_X$ exist, then the derivative $\mathfrak{D}P(x_0)$ exists also and is given by

$$\mathfrak{D}P(x_0) = \alpha_1 \mathfrak{D}P_1(x_0) + \alpha_2 \mathfrak{D}P_2(x_0).$$

If P_1 and P_2 are differentiable in x_0, then P is also differentiable in x_0.

Theorem 3.1.3 (Chain rule)**.**
Let $P : (\Omega_X \subset X) \to (\Omega_Y \subset Y)$ and $Q : (\Omega_Y \subset Y) \to Z$, with Ω_Y an open subset of Y and Z a Banach space. The operator Q is assumed to be differentiable in $P(x_0) \in Y$ with $x_0 \in \Omega_X$, and the derivative of P in x_0 is assumed to exist. Then the derivative of the operator $R := QP : (\Omega_X \subset X) \to Z$ in x_0 exists and is given by:

$$\mathfrak{D}R(x_0) = \mathfrak{D}Q(P(x_0))\mathfrak{D}P(x_0).$$

Theorem 3.1.4 (Mean value theorem)**.**
Let $x, x_0 \in \Omega_X$ and $(1-t)x_0 + tx =: x_0 + t\Delta x \in \Omega_X$ for $t \in [0,1]$. Moreover, assume that P is differentiable along the line connecting x and x_0. Then,

$$\|P(x) - P(x_0)\| \leq \sup_{\phi \in (0,1)} \|\mathfrak{D}P(x_0 + \phi \Delta x)\| \|\Delta x\|.$$

Definition 3.1.4 (Bilinear operator)**.**
Let X and Y be two Banach spaces and let $\langle\cdot,\cdot\rangle : X^2 \to Y$ be an operator assigning to each pair of elements $x, x' \in X$ an element $y \in Y$, i,e.,

$$\langle x, x'\rangle = y \in Y.$$

The binary operator $\langle\cdot,\cdot\rangle$ is said to be bilinear *if the relations*

$$\langle \alpha x_1 + \beta x_2, x_1'\rangle = \alpha \langle x_1, x_1'\rangle + \beta \langle x_2, x_1'\rangle, \tag{3.8}$$
$$\langle x_1, \alpha x_1' + \beta x_2'\rangle = \alpha \langle x_1, x_1'\rangle + \beta \langle x_1, x_2'\rangle \tag{3.9}$$

are satisfied for all $x_1, x_2, x_1', x_2' \in X$ and for all $\alpha, \beta \in \mathbb{R}$. Moreover, the bilinear operator $\langle\cdot,\cdot\rangle$ is termed bounded *if there exists $M > 0$ such that*

$$\| \langle x, x'\rangle \| \leq M\|x\|\|x'\|$$

for all $x, x' \in X$. We shall denote the space of all bounded bilinear operators mapping from X^2 to Y by $[X^2 \to Y]$.

An operator P may act on several elements from different spaces at a time, making it necessary to introduce partial derivatives.

Definition 3.1.5 (Partial derivatives)**.**
Let Ω_1 be an open subset of the Banach space X_1 and Ω_2 be an open subset of the Banach space X_2. Moreover, let $P : X_1 \times X_2 \to Y$ be an operator mapping $X_1 \times X_2$ to the Banach space Y. A linear operator $U_1 \in [X_1 \to Y]$ for which the relation

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[P(x_1 + \varepsilon x_1', x_2') - P(x_1, x_2)] = U_1(x_1')$$

holds for all $x_1' \in X_1$, is called the first partial derivative *of P with respect to x_1. Similarly, a linear operator $U_2 \in [X_2 \to Y]$ is called the* first partial derivative *of P with respect to x_2 if*

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}[P(x_1, x_2 + \varepsilon x_2') - P(x_1, x_2)] = U_2(x_2')$$

holds for all $x_2' \in X_2$. If the respective partial derivatives exist, we shall write

$$\delta_{x_1} P(x_1, x_2) = U_1, \quad \delta_{x_2} P(x_1, x_2) = U_2.$$

Concerning mixed double partial derivatives it is possible to change the order of differentiation under certain conditions:

Theorem 3.1.5 (Change of the order of partial differentiation)**.**
Let Ω_1 be an open subset of the Banach space X_1 and Ω_2 be an open subset of the Banach space X_2. Moreover, let $P : X_1 \times X_2 \to Y$ be an operator mapping $X_1 \times X_2$ to the Banach space Y. If P is continuous and if $\delta_{x_1} P,\ \delta_{x_2} P,\ \delta_{x_1}\delta_{x_2} P,\ \delta_{x_2}\delta_{x_1} P$ exist and are continuous, then

$$\delta_{x_1}\delta_{x_2} P(x_1, x_2)(x_2')(x_1') = \delta_{x_2}\delta_{x_1} P(x_1, x_2)(x_1')(x_2').$$

This is a generalisation of Schwarz's theorem from real calculus.

Theorem 3.1.6.
Let $Q : (\Omega_X \subset X) \to (\Omega_Y \subset Y)$ and $P : (\Omega_X \subset X) \times (\Omega_Y \subset Y) \to Z$, with Ω_X and Ω_Y open subsets and X, Y, Z Banach spaces. Moreover, let Q be differentiable in $x_0 \in \Omega_X$ and P be partially differentiable with respect to Q in $Q(x_0)$ and with respect to x in x_0. If the derivatives $\mathfrak{D}Q,\ \delta_Q P$ and $\delta_x P$ are continuous in x_0 and $(Q(x_0), x_0)$, respectively, then

$$\mathfrak{D}P(Q(x_0), x_0)(x) = \delta_Q P(Q(x_0), x_0)\mathfrak{D}Q(x_0)(x) + \delta_x P(Q(x_0), x_0)(x).$$

Based on this mathematical framework we can now explain the adjoint method for the general case where $\mathbf{u} \in U$ is a vector function depending on a spatial variable $\mathbf{x} \in G \subset \mathbb{R}^3$, a time variable $t \in T = [t_0, t_1]$ and some parameters $\mathbf{p} \in P$,

$$\mathbf{u} = \mathbf{u}(\mathbf{p}(\mathbf{x}); \mathbf{x}, t).$$

U and P are adequately chosen function spaces. For an arbitrary operator $\mathbf{L}$, a source term $\mathbf{f}(\mathbf{x}, t)$, and $\mathbf{u} \in \tilde{U} \subset U$ we write again:

$$\mathbf{L}(\mathbf{u}, \mathbf{p}; \mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t)$$

Assuming that both $\mathbf{L}$ and $\mathbf{u}$ are differentiable with respect to $\mathbf{p}$ and $\mathbf{q} \in P$ we have

$$\mathfrak{D}_p\mathbf{L}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathbf{q}) = \delta_u\mathbf{L}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathfrak{D}_p\mathbf{u}(\mathbf{p})(\mathbf{q})) + \ \delta_p\mathbf{L}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathbf{q}) = \mathfrak{D}_p\mathbf{f} = 0$$

Now consider the bilinear forms

$$\begin{aligned} \langle\cdot,\cdot\rangle_U &: (U \times U) \to \mathbb{R} \\ \langle\cdot,\cdot\rangle_P &: (P \times P) \to \mathbb{R} \end{aligned}$$

We are interested in finding a test function $\boldsymbol{\psi}$ such that there exist two linear operators

$$\begin{aligned} \delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p}) &\in [U \to U] \\ \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p}) &\in [U \to P] \end{aligned}$$

satisfying the relations

$$\begin{aligned} \langle \boldsymbol{\psi}, \delta_u \mathbf{L}(\mathbf{u},\mathbf{p})(\mathbf{v}) \rangle_U &= \langle \mathbf{v}, \delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi}) \rangle_U \\ \langle \boldsymbol{\psi}, \delta_p \mathbf{L}(\mathbf{u},\mathbf{p})(\mathbf{q}) \rangle_U &= \langle \mathbf{q}, \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi}) \rangle_P \end{aligned}$$

with $\mathbf{v} := \mathfrak{D}_p \mathbf{u}(\mathbf{p})(\mathbf{q})$.

In general, these two operators do not exist unless $\boldsymbol{\psi}$ is subjected to some conditions which may be summarised by the requirement that $\boldsymbol{\psi}$ be the element of some space $\tilde{U}^* \subset U$. Finding $\tilde{U}^*$ is part of the problem.
The constructions

$$\begin{aligned} \langle ., \delta_u \mathbf{L}(\mathbf{u},\mathbf{p})(.) \rangle_U &\in [U \times U \to \mathbb{R}] \\ \langle ., \delta_p \mathbf{L}(\mathbf{u},\mathbf{p})(.) \rangle_U &\in [U \times P \to \mathbb{R}] \end{aligned}$$

are bilinear functionals and

$$\begin{aligned} \langle ., \delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p})(.) \rangle_U &\in [U \times U \to \mathbb{R}] \\ \langle ., \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p})(.) \rangle_P &\in [P \times U \to \mathbb{R}] \end{aligned}$$

are their respective *transposed operators*. In the case that the bilinear form is a *scalar product*, the transpose is called an *adjoint operator*.

How these considerations contribute to find expressions for $\delta_\beta A_v$ and $\delta_\beta A_\omega$ will be shown in the next section.

3.2 Objective functions

The process of extracting information from the displacement field $\mathbf{u}$ is represented through an objective function $\mathfrak{C}(\mathbf{u}, \mathbf{p})$, with model parameters $\mathbf{p} = (\rho, \lambda, \mu, \beta, ...)$. Evidently, we are interested in the objective functions

$$\begin{aligned} i) \quad & \mathfrak{C}(\mathbf{u}, \mathbf{p}) = \beta_a(\mathbf{u}(\mathbf{p})) \\ ii) \quad & \mathfrak{C}(\mathbf{u}, \mathbf{p}) = \|\dot{\mathbf{u}}(\mathbf{p})\|_2 \\ iii) \quad & \mathfrak{C}(\mathbf{u}, \mathbf{p}) = \|\boldsymbol{\omega}(\mathbf{p})\|_2. \end{aligned}$$

We assume that $\mathfrak{C}(\mathbf{u}, \mathbf{p})$ is expressible in the form of a time integral $\mathfrak{C}(\mathbf{u}, \mathbf{p}) = \int \mathfrak{f}(\mathbf{u}, \mathbf{p}) dt$, with an adequately chosen function $\mathfrak{f}$. Then $\mathfrak{C}$ can be expressed as

$$\mathfrak{C}(\mathbf{u}, \mathbf{p}) = \langle 1, \mathfrak{f}(\mathbf{u}, \mathbf{p}) \rangle_{\mathbb{R}}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{R}} \in [\mathbb{R} \times \mathbb{R} \to \mathbb{R}]$ is a bilinear form in terms of definition 3.1.4.

We apply the chain rule

$$\mathfrak{D}_p \mathfrak{f}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathbf{q}) = \delta_u \mathfrak{f}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathfrak{D}_p \mathbf{u}(\mathbf{p})(\mathbf{q})) + \ \delta_p \mathfrak{f}(\mathbf{u}(\mathbf{p}), \mathbf{p})(\mathbf{q}) \tag{3.10}$$

and combined with the bilinearity of $\langle \cdot, \cdot \rangle_{\mathbb{R}}$ the total derivative of an objective function $\mathfrak{C}(\mathbf{u}, \mathbf{p})$ with respect to the model parameters $\mathbf{p}$ in some direction $\mathbf{q}$ is given by:

$$\mathfrak{D}_p \mathfrak{C}(\mathbf{u}, \mathbf{p})(\mathbf{q}) = \langle 1, \delta_u \mathfrak{f}(\mathbf{u}, \mathbf{p})(\mathbf{v}) \rangle_{\mathbb{R}} + \langle 1, \delta_p \mathfrak{f}(\mathbf{u}, \mathbf{p})(\mathbf{q}) \rangle_{\mathbb{R}}, \tag{3.11}$$

with $\mathbf{v} := \mathfrak{D}_p \mathbf{u}(\mathbf{p})(\mathbf{q})$.

Again, we try to find two linear operators

$$\begin{aligned} \delta_u \mathfrak{f}^*(\mathbf{u}, \mathbf{p}) &\in [\mathbb{R} \to U] \\ \delta_p \mathfrak{f}^*(\mathbf{u}, \mathbf{p}) &\in [\mathbb{R} \to P] \end{aligned}$$

such that

$$\begin{aligned} \langle 1, \delta_u \mathfrak{f}(\mathbf{u}, \mathbf{p})(\mathbf{v}) \rangle_{\mathbb{R}} &= \langle \mathbf{v}, \delta_u \mathfrak{f}^*(\mathbf{u}, \mathbf{p}) \rangle_U \\ \langle 1, \delta_p \mathfrak{f}(\mathbf{u}, \mathbf{p})(\mathbf{q}) \rangle_{\mathbb{R}} &= \langle \mathbf{q}, \delta_p \mathfrak{f}^*(\mathbf{u}, \mathbf{p}) \rangle_P \end{aligned}$$

Given that all the transposed operators exist, the total derivative of $\mathfrak{C}$ with respect to the parameters $\mathbf{p}$ can be written as

$$\begin{aligned}
\mathfrak{D}_p\mathfrak{C}(\mathbf{u},\mathbf{p})(\mathbf{q}) &= \langle \mathbf{v}, \delta_u \mathfrak{f}^*(\mathbf{u},\mathbf{p})\rangle_U + \langle \mathbf{q}, \delta_p \mathfrak{f}^*(\mathbf{u},\mathbf{p})\rangle_P \\
&+ \langle \mathbf{v}, \delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi})\rangle_U + \langle \mathbf{q}, \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi})\rangle_P \\
&= \langle \mathbf{v}, \delta_u \mathfrak{f}^*(\mathbf{u},\mathbf{p}) + \delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi})\rangle_U \\
&+ \langle \mathbf{q}, \delta_p \mathfrak{f}^*(\mathbf{u},\mathbf{p}) + \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi})\rangle_P . \qquad (3.12)
\end{aligned}$$

The presence of $\mathbf{v} = \mathfrak{D}_p\mathbf{u}$ in equation (3.12) is problematic. Numerical approximation methods for $\mathfrak{D}_p\mathbf{u}$ fail in large model spaces. To eliminate $\mathfrak{D}_p\mathbf{u}$ from the expression for $\mathfrak{D}_p\mathfrak{C}$ it is sufficient to require

$$\delta_u \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi}) = -\delta_u \mathfrak{f}^*(\mathbf{u},\mathbf{p}), \quad \boldsymbol{\psi} \in \tilde{U}^*. \qquad (3.13)$$

Equation (3.13) is known as the *transposed problem* of

$$\mathbf{L}(\mathbf{u},\mathbf{p};\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)$$

In the case that the bilinear forms $\langle .,. \rangle$ are scalar products, the transposed problem may also be called the *adjoint problem*. Finally, if the transposed problem can be solved, the total derivative of $\mathfrak{C}$ with respect to the parameters $\mathbf{p}$ can be written as

$$\mathfrak{D}_p\mathfrak{C}(\mathbf{u},\mathbf{p})(\mathbf{q}) = \langle (\mathbf{q}), \delta_p \mathfrak{f}^*(\mathbf{u},\mathbf{p}) + \delta_p \mathbf{L}^*(\mathbf{u},\mathbf{p})(\boldsymbol{\psi})\rangle_P \qquad (3.14)$$

Thus, we can easily calculate the gradient of $\mathfrak{C}$ with respect to the parameters $\mathbf{p}$ by computing only once the transposed (adjoint) field $\boldsymbol{\psi}$.

Remark: Depending on the operator $\mathbf{L}$, it may not be possible to find the transposed operators and the test function $\boldsymbol{\psi}$. So the transposed problem may not have a solution, and if the solution exists, it may not be unique. Fortunately, we can find and uniquely solve the adjoint problems for all linear wave propagation phenomena, i.e., for all types of linear wave equation operators $\mathbf{L}$ complemented by adequate subsidiary conditions.

3.3 Bilinear operators

The adjoint method simplifies if the considered operator $\mathbf{L}$ is bilinear, i.e., if it satisfies the relation

$$\begin{aligned}\mathbf{L}(\alpha\mathbf{u}+\beta\mathbf{v},\gamma\mathbf{p}+\delta\mathbf{q};\mathbf{x},t) &= \alpha\gamma\mathbf{L}(\mathbf{u},\mathbf{p};\mathbf{x},t)+\alpha\delta\mathbf{L}(\mathbf{u},\mathbf{q};\mathbf{x},t)\\ &+ \beta\gamma\mathbf{L}(\mathbf{v},\mathbf{p};\mathbf{x},t)+\beta\delta\mathbf{L}(\mathbf{v},\mathbf{q};\mathbf{x},t)\end{aligned} \quad (3.15)$$

for all $\mathbf{u},\mathbf{v}\in U$, for all $\mathbf{p},\mathbf{q}\in P$ and for all $\alpha,\beta,\gamma,\delta\in\mathbb{R}$. Due to the bilinearity, the derivative of $\mathbf{L}(\mathbf{u}(\mathbf{p}),\mathbf{p})$ with respect to the parameters $\mathbf{p}$ is given by

$$\begin{aligned}\mathfrak{D}_p\mathbf{L}(\mathbf{u}(\mathbf{p}),\mathbf{p})(\mathbf{q}) &= \delta_u\mathbf{L}(\mathbf{u}(\mathbf{p}),\mathbf{p})\mathfrak{D}_p\mathbf{u}(\mathbf{p})(\mathbf{q})+\delta_p\mathbf{L}(\mathbf{u}(\mathbf{p}),\mathbf{p})(\mathbf{q})\\ &= \mathbf{L}(\mathbf{v}(\mathbf{q}),\mathbf{p})+\mathbf{L}(\mathbf{u}(\mathbf{p}),\mathbf{q})\end{aligned} \quad (3.16)$$

with $\mathbf{v}(\mathbf{p}):=\mathfrak{D}_p\mathbf{u}(\mathbf{p})(\mathbf{q})$. Let $\boldsymbol{\psi}\in U$ be an arbitrary function. Assume that, given a specific bilinear form $\langle\cdot,\cdot\rangle$, it is possible to find operators $\mathbf{L}^u$ and $\mathbf{L}^p$ such that the relations

$$\begin{aligned}\langle\boldsymbol{\psi},\mathbf{L}(\mathbf{u}(\mathbf{p})\rangle_U &= \langle\mathbf{u},\mathbf{L}^u(\boldsymbol{\psi},\mathbf{p})\rangle_U+\boldsymbol{\Lambda}_u(\boldsymbol{\psi},\mathbf{u},\mathbf{p}),\\ \langle\boldsymbol{\psi},\mathbf{L}(\mathbf{u}(\mathbf{p})\rangle_P &= \langle\mathbf{p},\mathbf{L}^p(\mathbf{u},\boldsymbol{\psi})\rangle_P+\boldsymbol{\Lambda}_p(\boldsymbol{\psi},\mathbf{u},\mathbf{p})\end{aligned} \quad (3.17)$$

hold. The terms symbolised by $\boldsymbol{\Lambda}_u$ and $\boldsymbol{\Lambda}_p$ result from manipulations necessary to transform the bilinear forms. For all ψ satisfying

$$\boldsymbol{\Lambda}_u(\boldsymbol{\psi},\mathbf{u},\mathbf{p})=\boldsymbol{\Lambda}_p(\boldsymbol{\psi},\mathbf{u},\mathbf{p})=0 \quad (3.18)$$

the operators $\boldsymbol{\Lambda}_u$ and $\boldsymbol{\Lambda}_p$ are transposed operators of $\mathbf{L}$ with respect to u and $\mathbf{p}$. With the assumed bilinearity, we can now write

$$\begin{aligned}&\langle\boldsymbol{\psi},\delta_u\mathbf{L}(\mathbf{u},\mathbf{p}(\mathbf{v}))\rangle_U+\langle\boldsymbol{\psi},\delta_p\mathbf{L}(\mathbf{u},\mathbf{p}(\mathbf{q}))\rangle_U\\ =\;&\langle\boldsymbol{\psi},\mathbf{L}(\mathbf{v},\mathbf{p})\rangle_U+\langle\boldsymbol{\psi},\mathbf{L}(\mathbf{u},\mathbf{q})\rangle_U\\ =\;&\langle\mathbf{v},\mathbf{L}^u(\boldsymbol{\psi},\mathbf{p})\rangle_U+\langle\mathbf{q},\mathbf{L}^p(\mathbf{u},\boldsymbol{\psi})\rangle_P+\boldsymbol{\Lambda}_u(\boldsymbol{\psi},\mathbf{v},\mathbf{p})=\boldsymbol{\Lambda}_p(\boldsymbol{\psi},\mathbf{u},\mathbf{q})\end{aligned} \quad (3.19)$$

If $\boldsymbol{\psi}$ can now be specified such that

$$\boldsymbol{\Lambda}_u(\boldsymbol{\psi},\mathbf{v},\mathbf{p})=\boldsymbol{\Lambda}_p(\boldsymbol{\psi},\mathbf{u},\mathbf{q})=0 \quad (3.20)$$

holds, we find

$$\langle \boldsymbol{\psi}, \mathbf{L}^u(\mathbf{v}, \mathbf{p}) \rangle_U + \langle \boldsymbol{\psi}, \mathbf{L}^p(\mathbf{u}, \mathbf{q}) \rangle_U = \langle \mathbf{v}, \mathbf{L}^u(\boldsymbol{\psi}, \mathbf{p}) \rangle_U + \langle \mathbf{q}, \mathbf{L}^p(\mathbf{u}, \boldsymbol{\psi}) \rangle_P \quad (3.21)$$

Consequently, the transposed problem reduces to

$$\mathbf{L}^u(\boldsymbol{\psi}, \mathbf{p}) = -\delta_u \mathfrak{f}^*(\mathbf{u}, \mathbf{p}), \quad \boldsymbol{\psi} \in \tilde{U}^* \quad (3.22)$$

and the corresponding derivative of the objective function with respect to the parameters is

$$\mathfrak{D}_p \mathfrak{C}(\mathbf{u}, \mathbf{p})(\mathbf{q}) = \langle \mathbf{q}, \delta_p \mathfrak{f}^*(\mathbf{u}, \mathbf{p}) + \mathbf{L}^p(\mathbf{u}, \boldsymbol{\psi}) \rangle_P \quad (3.23)$$

3.4 Application to the isotropic elastic wave equation

It is straightforward to verify that the scalar wave equation operator

$$\mathbf{L}(\mathbf{u}) \equiv \rho\partial_t^2\mathbf{u} - \mu\nabla^2\mathbf{u}$$

is bilinear, i.e., independently linear in u and $\mathbf{p} = (\rho, \mu)$. Moreover, in this specific case, we find for the adjoint operator together with the adjoint wave field the identity

$$\mathbf{L}^u(\boldsymbol{\psi}, \mathbf{p}) = \rho\partial_t^2\boldsymbol{\psi}(\mathbf{x}, t) - \mu\nabla^2\boldsymbol{\psi}(\mathbf{x}, t) = \mathbf{L}(\boldsymbol{\psi}, \mathbf{p}) \tag{3.24}$$

Therefore, the scalar wave equation is *self-adjoint*. Note that the adoint equation is independent of the field $\mathbf{u}$ in the case of a bilinear operator.
Now consider the scalar product

$$\langle \mathbf{u}, \mathbf{v} \rangle_U := \int_{\mathbb{R}} \int_{G\subset\mathbb{R}^3} \mathbf{u}(\mathbf{x}, t) \cdot \mathbf{v}(\mathbf{x}, t)\ dt\ dG \tag{3.25}$$

defined for any two elements $\mathbf{u}$ and $\mathbf{v}$ of U. As $\mathbf{L}$ is self-adjoint it follows that

$$\begin{aligned}
\mathfrak{D}_p\mathfrak{C}(\mathbf{u}, \mathbf{p})(\mathbf{q}) &= \langle (\mathbf{q}), \delta_p\mathbf{L}^u(\mathbf{u}, \mathbf{p})(\boldsymbol{\psi}) \rangle_P \\
&= \langle (\boldsymbol{\psi}), \delta_p\mathbf{L}^u(\mathbf{u}, \mathbf{p})(\mathbf{q}) \rangle_U \\
&= \langle (\boldsymbol{\psi}), \delta_p\mathbf{L}(\mathbf{u}, \mathbf{p})(\mathbf{q}) \rangle_U \\
&= \int_{\mathbb{R}} \int_{G\subset\mathbb{R}^3} \boldsymbol{\psi} \cdot \delta_p\mathbf{L}(\mathbf{u}, \mathbf{p})\ dt\ dG
\end{aligned} \tag{3.26}$$

So the volumetric density of the functional derivative of $\mathfrak{C}$ with respect to the model parameters $\mathbf{p}$ is

$$\delta_p\mathfrak{C} = \int_{\mathbb{R}} \boldsymbol{\psi} \cdot \partial_p\mathbf{L}(\mathbf{u}, \mathbf{p})dt. \tag{3.27}$$

In the literature $\delta_p\mathfrak{C}$ is called Fréchet kernel. We apply Fréchet kernels to the investigation of the sensitivity of the apparent shear wave speed with respect to the true shear wave speed. In this context the terms Fréchet kernel and sensitivity kernel shall be used equivalently.

Chapter 4

Ray method based computation of sensitivity kernels

Chapter 3 ended with a general expression for the Fréchet kernel $\delta_p \mathfrak{C}$, the volumetric density of the functional derivative of $\mathfrak{C}$ with respect to the model parameters $\mathbf{p}$

$$\delta_p \mathfrak{C} = \int_{\mathbb{R}} \boldsymbol{\psi} \cdot \partial_p \mathbf{L}(\mathbf{u}, \mathbf{p}) dt.$$

The variable $\boldsymbol{\psi}$ denotes the adjoint field which is the solution of the adjoint problem. The partial derivative of the wave operator with respect to $\mathbf{p}$ is symbolised by $\partial_p \mathbf{L}$.

The adjoint field $\boldsymbol{\psi}$ is defined through the adjoint wave equation

$$\mathbf{L}(\boldsymbol{\psi}, \mathbf{p}) = \rho \partial_t^2 \boldsymbol{\psi}(\mathbf{x}, t) - \mu \nabla^2 \boldsymbol{\psi}(\mathbf{x}, t) = -\partial_u \ \mathfrak{f}(t) \ \delta(\mathbf{x} - \mathbf{x}^r) \qquad (4.1)$$

and its subsidiary conditions.

In this chapter we apply the adjoint method in terms of equation (3.27) to sensitivity kernels for apparent S-wave speed measurements, $\frac{1}{\beta_a} \delta_\beta \beta_a$. Combined with the ray method we develop an algorithm for the numerical computation of sensitivity kernels.

Let us first recall equation (3.7) from chapter 3

$$\frac{1}{\beta_a} \delta_\beta \beta_a = \frac{1}{A_v} \delta_\beta A_v - \frac{1}{A_\omega} \delta_\beta A_\omega$$

with

$$A_v(\mathbf{x}^r) := \|\dot{\mathbf{u}}(\mathbf{p};\mathbf{x}^r,t)\|_2 = \sqrt{\int_{\mathbb{R}} \dot{\mathbf{u}}(\mathbf{p};\mathbf{x}^r,t)^2 dt}$$

$$A_\omega(\mathbf{x}^r) := \|\boldsymbol{\omega}(\mathbf{p};\mathbf{x}^r,t)\|_2 = \sqrt{\int_{\mathbb{R}} \left(\frac{1}{2}\nabla \times \mathbf{u}(\mathbf{p};\mathbf{x}^r,t)\right)^2 dt}$$

Now consider β_a, A_v and A_ω as objective functions, expressible in the form of a time integral $\mathfrak{C}(\mathbf{u}) = \int \mathfrak{f}(\mathbf{u})dt$. Using equation (3.27) together with equation (3.7) we can write

$$\begin{aligned} \frac{1}{\beta_a}\delta_\beta\beta_a &= \int_{\mathbb{R}} \boldsymbol{\psi}^v \cdot \partial_\beta \mathbf{L}(\mathbf{u},\mathbf{p})dt - \int_{\mathbb{R}} \boldsymbol{\psi}^\omega \cdot \partial_\beta \mathbf{L}(\mathbf{u},\mathbf{p})dt \\ &= \int_{\mathbb{R}} (\boldsymbol{\psi}^v - \boldsymbol{\psi}^\omega) \cdot \partial_\beta \mathbf{L}(\mathbf{u},\mathbf{p})dt \end{aligned} \tag{4.2}$$

where $\boldsymbol{\psi}^v$ and $\boldsymbol{\psi}^\omega$ are the adjoint fields for A_v and A_ω. By definition, the adjoint fields include the scaling factors $1/A_v$ and $1/A_\omega$. Later we will see that the difference $\boldsymbol{\psi}^v - \boldsymbol{\psi}^\omega$ of the adjoint fields is the key element of equation (4.2).
General expressions for the adjoint fields A_v and A_ω with respect to any possible parameter $\mathbf{p}$ can be derived as follows:

4.1 The adjoint field for velocity amplitude measurements

The considerations of this section and section 4.2 are mainly in accordance with Fichtner and Igel (2008): The relative functional derivative of A_v with respect to the model parameters $\mathbf{p}$, denoted by $\frac{1}{A_v}\mathfrak{D}_p A_v$, is

$$\frac{1}{A_v}\mathfrak{D}_p A_v = \frac{1}{{A_v}^2}\int_{\mathbb{R}} \dot{u}_i \; \mathfrak{D}_p \dot{u}_i \; dt. \tag{4.3}$$

In terms of sensitivity densities we may write

$$\frac{1}{A_v}\mathfrak{D}_p A_v = \int_{G\subset\mathbb{R}^3} \frac{1}{A_v}\delta_p A_v \; dG = \frac{1}{{A_v}^2}\int_{\mathbb{R}} \dot{u}_i \; \mathfrak{D}_p \dot{u}_i \; dt = \int_{G\subset\mathbb{R}^3} \frac{1}{{A_v}^2}\int_{\mathbb{R}} \dot{u}_i \; \delta_p \dot{u}_i \; dt dG$$

The symbols G and dG denote the computational domain and the corresponding volume element.

Obviously we need an expression for $\delta_p \dot{u}_i$ at the receiver point $\mathbf{x}^r$. Let $\mathfrak{C}(\mathbf{u})$ be equal to the i component of the displacement field, $u_i(\mathbf{x}^r, \tau)$, that is when $\mathfrak{f}(\mathbf{u}) = \delta(t-\tau)\mathbf{e}_i \cdot \mathbf{u}(\mathbf{x}^r, \tau)$. Then the right-hand side of the adjoint equation (4.1) becomes $-\mathbf{e}_i \; \delta(t-\tau) \; \delta(\mathbf{x}-\mathbf{x}^r)$, implying that the corresponding adjoint field $\mathbf{u}^*$ is the adjoint Green's function with source location $\mathbf{x}^r$ and source time τ, that is $\mathbf{u}^*(\mathbf{x}, t) = \mathbf{g}_i^*(\mathbf{x}^r, t; \mathbf{x}, t')$. This results in

$$\delta_p u_i(\mathbf{x}^r, \tau) = \int_{\mathbb{R}} \mathbf{g}_i^*(\mathbf{x}^r, \tau; \mathbf{x}, t) \cdot \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t)] dt \tag{4.4}$$

With equation(4.4) the sensitivity density $\frac{1}{A_v}\delta_p A_v$ is

$$\begin{aligned} \frac{1}{A_v}\delta_p A_v &= \frac{1}{{A_v}^2}\int_{\mathbb{R}} \dot{u}_i(\mathbf{x}^r, t') \left(\int_{\mathbb{R}} \partial_t \; \mathbf{g}_i^*(\mathbf{x}^r, t; \mathbf{x}, t') \cdot \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')] dt' \right) dt \\ &= \frac{1}{{A_v}^2}\int_{\mathbb{R}^2} \partial_t \; \dot{u}_i(\mathbf{x}^r, t') \; \mathbf{g}_i^*(\mathbf{x}^r, t; \mathbf{x}, t') \cdot \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')] dt' dt \end{aligned}$$

In accordance with the general equation for sensitivity kernels (3.27) we define the adjoint field $\boldsymbol{\psi}^v$ to be

$$\boldsymbol{\psi}^v(\mathbf{x}, t') := \frac{1}{{A_v}^2} \int_{\mathbb{R}} \partial_t \ \dot{u}_i(\mathbf{x}^r, t') \ \mathbf{g}_i^*(\mathbf{x}^r, t; \mathbf{x}, t') dt \tag{4.5}$$

And we can write the sensitivity density $\frac{1}{A_v}\delta_p A_v$ in compact notation

$$\frac{1}{A_v}\delta_p A_v = \int_{\mathbb{R}} \boldsymbol{\psi}^v(\mathbf{x}, t') \ \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')] dt' \tag{4.6}$$

Note that the adjoint field $\boldsymbol{\psi}^v$ can equally be obtained as the solution of the adjoint equation $\mathbf{L}^*(\boldsymbol{\psi}^v, \mathbf{p}) = \mathbf{f}^v$ with the adjoint source $\mathbf{f}^v$

$$\mathbf{f}^v(\mathbf{x}, t') = \frac{1}{{A_v}^2} \ \partial_t^2 \ u_i(\mathbf{x}^r, t') \delta(\mathbf{x} - \mathbf{x}^r). \tag{4.7}$$

4.2 The adjoint field for rotation amplitude measurements

Expressions for the adjoint field for rotation amplitude measurements can be derived in analogy to the procedure for velocity amplitude measurements. The relative functional derivative of A_ω with respect to the model parameters $\mathbf{p}$, denoted by $\frac{1}{A_\omega}\mathfrak{D}_p A_\omega$, is

$$\frac{1}{A_\omega}\mathfrak{D}_p A_\omega = \frac{1}{{A_\omega}^2}\int_{\mathbb{R}} \omega_i \ \mathfrak{D}_p \omega_i \ dt. \tag{4.8}$$

Applying again equation (4.4) leads to

$$\delta_p \omega_i(\mathbf{x}^r, t) = \frac{1}{2}\epsilon_{ijk}\frac{\partial}{\partial x_j^r}\delta_p u_k(\mathbf{x}^r, t) = \frac{1}{2}\epsilon_{ijk}\int_{\mathbb{R}} \frac{\partial}{\partial x_j^r}\mathbf{g}_k^*(\mathbf{x}^r, t; \mathbf{x}, t') \cdot \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')]dt'.$$

Hence

$$\frac{1}{A_\omega}\delta_p A_\omega = \frac{1}{2{A_\omega}^2}\epsilon_{ijk}\int_{\mathbb{R}^2} \omega_i(\mathbf{x}^r, t)\frac{\partial}{\partial x_j^r}\mathbf{g}_k^*(\mathbf{x}^r, t; \mathbf{x}, t') \ \cdot \ \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')]dt'dt. \tag{4.9}$$

Defining the adjoint field $\boldsymbol{\psi}^\omega$ as:

$$\boldsymbol{\psi}^\omega(\mathbf{x}, t') := \frac{1}{2A_\omega^2}\int_{\mathbb{R}} \omega_i(\mathbf{x}^r, t')\frac{\partial}{\partial x_j^r}\mathbf{g}_k^*(\mathbf{x}^r, t; \mathbf{x}, t')dt. \tag{4.10}$$

And we get the sensitivity density $\frac{1}{A_\omega}\delta_p A_\omega$

$$\frac{1}{A_\omega}\delta_p A_\omega = \int_{\mathbb{R}} \boldsymbol{\psi}^\omega(\mathbf{x}, t') \ \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t')]dt' \tag{4.11}$$

Again the adjoint field $\boldsymbol{\psi}^\omega$ can be obtained as the solution of the adjoint equation $\mathbf{L}^*(\boldsymbol{\psi}^\omega, \mathbf{p}) = \mathbf{f}^\omega$ with the adjoint source $\mathbf{f}^\omega$

$$f_k^\omega(\mathbf{x}, t') = \frac{1}{2A_\omega^2}\epsilon_{ijk} \ \omega_i(\mathbf{x}^r, t')\frac{\partial}{\partial x_j^r}\delta(\mathbf{x} - \mathbf{x}^r). \tag{4.12}$$

In view of the next section we add that $\mathbf{f}^{\omega}$ can be written in terms of a moment density $\mathbf{m}$

$$\mathbf{f}^{\omega} = \nabla \cdot \mathbf{m} \tag{4.13}$$

with

$$m_{kj} = \frac{1}{2A_{\omega}^2} \epsilon_{ijk}\ \omega_i(\mathbf{x}^r, t')\delta(\mathbf{x} - \mathbf{x}^r). \tag{4.14}$$

4.3 Sensitivity kernels

In the last two sections we found expressions for the adjoint fields $\boldsymbol{\psi}^v$ and $\boldsymbol{\psi}^\omega$. For the computation of the sensitivity kernel of the apparent shear wave speed, $\frac{1}{\beta_a}\delta_\beta\beta_a = \int_{\mathbb{R}}(\boldsymbol{\psi}^v - \boldsymbol{\psi}^\omega) \cdot \partial_p\mathbf{L}(\mathbf{u},\mathbf{p})dt$, we now look at $\partial_p\mathbf{L}(\mathbf{u},\mathbf{p})$. It is

$$\mathbf{L}(\mathbf{u}) \equiv \rho\partial_t^2\mathbf{u} - \mu\nabla^2\mathbf{u} = \rho\partial_t^2\mathbf{u} - \rho\beta^2\nabla^2\mathbf{u} = \mathbf{f}. \tag{4.15}$$

We are interested in sensitivity densities far away from the source, i.e. $\mathbf{f} = 0$. So $\nabla^2\mathbf{u} = \frac{1}{\beta^2}\partial_t^2\mathbf{u}$ and $\delta_\beta\mathbf{L}(\mathbf{u}) = -2\rho\beta\nabla^2\mathbf{u}$ give

$$\delta_\beta\mathbf{L}(\mathbf{u}) = -2\frac{\rho}{\beta}\partial_t^2\mathbf{u}. \tag{4.16}$$

The expression for the sensitivity kernel is now

$$\frac{1}{\beta_a}\delta_\beta\beta_a = -2\frac{\rho}{\beta}\int_{\mathbb{R}}(\boldsymbol{\psi}^v - \boldsymbol{\psi}^\omega) \cdot \partial_t^2\mathbf{u}\ dt. \tag{4.17}$$

It is basically possible to use exact calculations of the displacement $\mathbf{u}$. The numerical methods are well-known but lead to extremely time-consuming algorithms. Our objective is to implement a completely new approach for structural inversions. To interprete the outcome of our numerical experiments several parameter studies will be necessary. So we need easy manageable and fast running algorithms. For this reason we simplify the forward problem of solving the wave equation by means of a high-frequency approximation of the wave field, i.e. we use ray theory. This simplification will just provide a better insight into the nature of our problem and should not produce misleading outputs.
According to chapter 2 we describe the displacement field $\mathbf{u}(\mathbf{x},t)$ by an amplitude function $\mathbf{U}(\mathbf{x})$ and a time function $T(t-\tau)$

$$\mathbf{u}(\mathbf{x},t) = \mathbf{U}(\mathbf{x}) \cdot T(t-\tau). \tag{4.18}$$

From now on indices $\cdot_S$, $\cdot_{\psi^v}$ *and* $\cdot_{\psi^\omega}$ mark components of the forward S, the adjoint velocity and the adjoint rotation wave field, respectively. As the adjoint wave fields have the same time function we write:

$$\mathbf{u}_S(\mathbf{x},t) = \mathbf{U}_S(\mathbf{x}) \cdot T_S(t-\tau_S) \tag{4.19}$$
$$\mathbf{u}_{\psi^v}(\mathbf{x},t) = \mathbf{U}_{\psi^v}(\mathbf{x}) \cdot T_\psi(t-\tau_\psi) \tag{4.20}$$
$$\mathbf{u}_{\psi^\omega}(\mathbf{x},t) = \mathbf{U}_{\psi^\omega}(\mathbf{x}) \cdot T_\psi(t-\tau_\psi) \tag{4.21}$$

Looking at equation (4.7) and with ray theory the source term of the adjoint velocity wave field is

$$\mathbf{f}_{\psi^v}(\mathbf{x},t) = \frac{1}{{A_v}^2}\mathbf{U}_S(\mathbf{x}^r)\ddot{T}_S(t-\tau_\psi) \tag{4.22}$$

Suggesting that we define

$$T_\psi(t-\tau_\psi) := \ddot{T}_S(t-\tau_\psi) \tag{4.23}$$

Now we are able to rewrite the sensitivity density of the apparent S-wave velocity with ray method

$$\frac{1}{\beta_a}\delta_\beta\beta_a = -2\frac{\rho}{\beta}\Big(\mathbf{U}_{\psi^v}(\mathbf{x}) - \mathbf{U}_{\psi^\omega}(\mathbf{x})\Big)\mathbf{U}_S(\mathbf{x})\int_{\mathbb{R}}\ddot{T}_S(t-\tau_S(\mathbf{x}))\ddot{T}_S(t-\tau_\psi(\mathbf{x}))\;dt. \tag{4.24}$$

At this point it should be underlined that equation (4.24) is a central statement of the diploma thesis, beeing part of nearly all algorithms that were implemented.

Before we care about the amplitudes in equation (4.24) we choose an adequate time function. As an elementary but realistic time signal we take a Gauss function.

$$T(t-\tau) := e^{-(t-\tau)^2\sigma^{-2}} \tag{4.25}$$

In later applications the question will arise what signal frequencies we should use. The parameter σ in equation (4.25) is linked to frequency. A Fourier transform between time- and frequency-domain shows that the smaller we select σ the wider is the frequency band contained in the signal (see figure 4.1). It follows from Fourier transform that the frequency band is given by the interval $[0;\frac{1}{\sigma}]$. As an example for $\sigma = 0.5$ the maximal frequency contained in the signal would be $2Hz$.

Remark: In general there is no maximal frequency in the Gauss signal as its values tend to infinity. In practice we use the σ-dependend width of a Gauss function to define the frequency band of a Gauss signal.

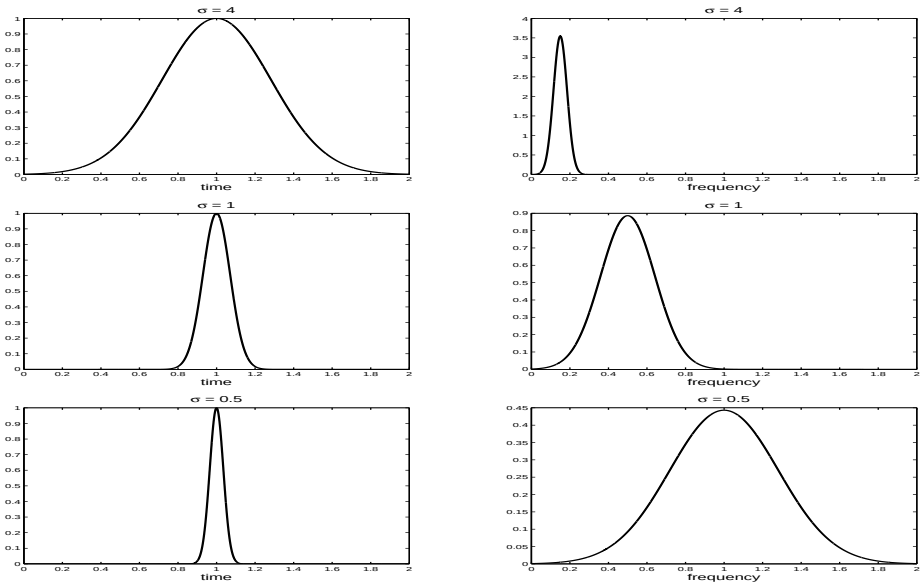


Figure 4.1: *Gauss signal in the time-domain (left column) and frequency-domain (right column) with decreasing σ from top to bottom;*

Unidirectional point source:

Finally we need to calculate the amplitudes of the forward S, the adjoint velocity and the adjoint rotation wave fields. This implies that one has to find a solution of the general equation for elastic displacement, $\mathbf{L}(\mathbf{u}) = \mathbf{f}$. First we have to concretise the body force $\mathbf{f}$. The fundamental case for seismological applications is when $\mathbf{f}$ is excited by a point source $X_0(t)$ pointing in one single direction. Detailed derivation of the solution in homogeneous, isotropic, unbounded medium can be found in standard literature for seismology ,i.g. Aki/Richards (2002), and is omitted here. We just give the result for a point source $X_0(t)$ in the x_j-direction and elastodynamic Green's function G_{ij} (see chapter 1)

$$\begin{aligned} u_i(\mathbf{x},t) &= X_0 * G_{ij} = \frac{1}{4\pi\rho}(3\gamma_i\gamma_j - \delta_{ij})\frac{1}{R^3}\int_{\frac{R}{\beta}}^{\frac{R}{\alpha}} \tau X_0(t-\tau)d\tau \\ &+ \frac{1}{4\pi\rho\alpha^2}\gamma_i\gamma_j\frac{1}{R}X_0(t-\frac{R}{\alpha}) \\ &- \frac{1}{4\pi\rho\beta^2}(\gamma_i\gamma_j - \delta_{ij})\frac{1}{R}X_0(t-\frac{R}{\beta}) \end{aligned} \tag{4.26}$$

with density ρ, P wave velocity α, S wave velocity β, source-receiver distance R and direction cosines $\gamma_i = \delta_i R = \frac{x_i^r - x_i^s}{R}$.

Note that the magnitude of the terms of equation (4.26) depends significantly on the source-receiver distance R. The first term behaves like R^{-2} for sources in which X_0 is nonzero for times that are short compared to $\frac{R}{\beta} - \frac{R}{\alpha}$, whereas the second and the third term behave like R^{-1}. So the second and the third term become dominant over the first as $R \to \infty$, but R^{-2} dominates R^{-1} as $R \to 0$. The terms including $\frac{1}{R}X_0(t - \frac{R}{\alpha})$ and $\frac{1}{R}X_0(t - \frac{R}{\beta})$ are therefore called *far-field* terms and the term including $\frac{1}{R^3}\int \tau X_0(t-\tau)d\tau$ is called *near-field* term.
We want to use far-field S-wave data sets for seismic tomography, so we neglect the first and the second term in equation (4.26), leading to the far-field S-wave displacement

$$u_i(\mathbf{x}, t) = \frac{1}{4\pi\rho\beta^2}(\delta_{ij} - \gamma_i\gamma_j)\frac{1}{R}X_0(t - \frac{R}{\beta}). \tag{4.27}$$

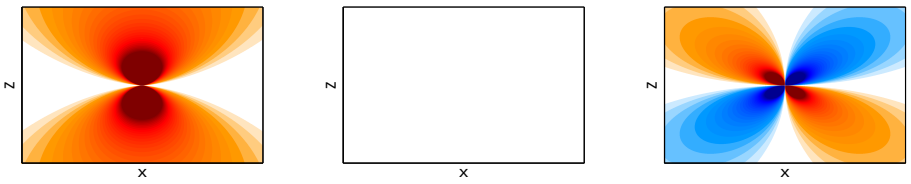
Comparison with equation (4.18) yields for the amplitudes of the displacement field

$$(U_S)_j(\mathbf{x}) = \frac{1}{4\pi\rho\beta^2}(\delta_{ij} - \gamma_i\gamma_j)F_i\frac{1}{R_S}. \tag{4.28}$$

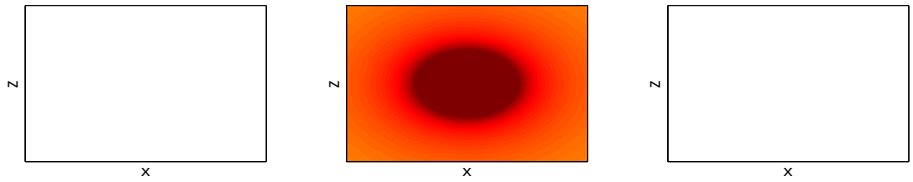
where $\mathbf{F} = (F_1, F_2, F_3)$ is the intensity of the point source and R_S the distance from the source to the grid point $\mathbf{x}$. Later applications deal with relative sensitivities. For that reason we can assume without restriction that $|\mathbf{F}| = 1$.

For an appropriate choice of $\mathbf{F}$ in the outstanding experiments we test amplitudes for three fundamental cases. In the following figures red color simbolises positive amplitudes, blue negative and white zero.

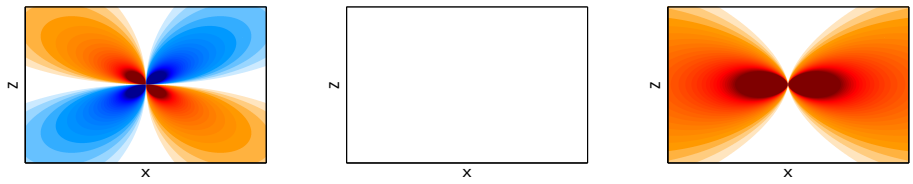
i) unidirectional point source in x_1 direction (i.e.$\mathbf{F}$ = (1,0,0)): vertical slices of the amplitude components U_1, U_2, U_3(from the left to the right);



ii) unidirectional point source in x_2 direction (i.e.$\mathbf{F}$ = (0,1,0)): vertical slices of the amplitude components U_1, U_2, U_3(from the left to the right);



iii) unidirectional point source in x_3 direction (i.e.$\mathbf{F}$ = (0,0,1)): vertical slices of the amplitude components U_1, U_2, U_3(from the left to the right);



Inversion scenarios in this work are constricted to the x_1x_3-plane. So the choice of $\mathbf{F} = (0, 1, 0)$ does not make sense because there are only non-zero contributions in the U_2-direction of the amplitude. For $\mathbf{F} = (0, 0, 1)$ there is no signal for receivers placed straight above the source. To avoid that problem we should remain with $\mathbf{F} = (1, 0, 0)$.

It is quite common in computational seismology to simulate the S-wave source by a point source pointing in one direction. To calculate the adjoint amplitudes it should be outlined that the adjoint wave field does nothing else than sending back the S-wave energy that reaches the receiver. Hence the intensity of the adjoint sources is given by the amplitudes of the S-wave field at the receiver, $\mathbf{F}_\psi = (U_1(\mathbf{x}^r), U_2(\mathbf{x}^r), U_3(\mathbf{x}^r))$.

Passing from the displacement field to the velocity of the displacement field does not affect the wave field source. For describing the adjoint velocity field we can use the same source type as for the S-wave field. So for the amplitudes of the adjoint velocity field we have

$$(U_{\psi^v})_j(\mathbf{x}) = \frac{1}{4{A_v}^2\pi\rho\beta^2}(\delta_{ij} - (\gamma_\psi)_i(\gamma_\psi)_j)(F_\psi)_i\frac{1}{R_\psi} \tag{4.29}$$

with the distance R_ψ from the receiver to the grid point $\mathbf{x}$. For the adjoint rotation field we have to extend our source theory from unidirectional point sources to so-called double-couple point sources.

Double-couple point source:

Experience in seismological research over decades has shown that seismic motions due to earthquakes, explosions etc. are well-described by double-couple point sources. To comprise all possible types a point source has to be characterised by nine different vector couples. They are shown in the picture below.

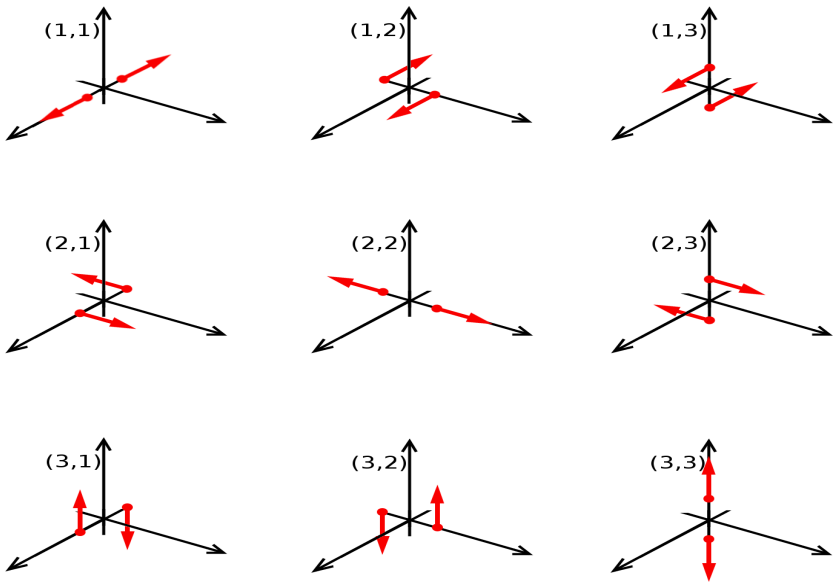


Figure 4.2: *Vector couples for a point source description*

Vector couples are incorporated into displacement fields via the seismic *moment tensor*, $\mathbf{M}$. If we use the elastodynamic Green's function $G_{in}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$ with an unidirectional unit impulse $\delta_{in}\delta(\mathbf{x}-\boldsymbol{\xi})\delta(t-\tau)$ applied at $\mathbf{x} = \boldsymbol{\xi}$, $t = \tau$ and in the n-direction (see chapter 1) then it is a central statement of seismological source theory that we can write

$$u_i(\mathbf{x}, t) = M_{nq} * \frac{\partial G_{in}}{\partial \xi_q}. \tag{4.30}$$

The symbol $*$ is the convolution symbol.

Again refering to Aki/Richards (2002) the double-couple solution in an infinite homogeneous medium is given by

$$\begin{aligned}
u_i(\mathbf{x},t) &= M_{nq} * \frac{\partial G_{in}}{\partial \xi_q} \\
&= \frac{1}{4\pi\rho}(15\gamma_i\gamma_n\gamma_q - 3\gamma_i\delta_{nq} - 3\gamma_n\delta_{iq} - 3\gamma_q\delta_{in})\ \frac{1}{R^4}\int_{\frac{R}{\beta}}^{\frac{R}{\alpha}} \tau M_{nq}(t-\tau)d\tau \\
&+ \frac{1}{4\pi\rho\alpha^2}(6\gamma_i\gamma_n\gamma_q - \gamma_i\delta_{nq} - \gamma_n\delta_{iq} - \gamma_q\delta_{ip})\ \frac{1}{R^2}M_{nq}(t-\frac{R}{\alpha}) \\
&- \frac{1}{4\pi\rho\beta^2}(6\gamma_i\gamma_n\gamma_q - \gamma_i\delta_{nq} - \gamma_n\delta_{iq} - \gamma_q\delta_{ip})\ \frac{1}{R^2}M_{nq}(t-\frac{R}{\beta}) \\
&+ \frac{1}{4\pi\rho\alpha^3}\gamma_i\gamma_n\gamma_q\frac{1}{R}\dot{M}_{nq}(t-\frac{R}{\alpha}) \\
&- \frac{1}{4\pi\rho\beta^3}(\gamma_i\gamma_n - \delta_{in})\gamma_q\frac{1}{R}\dot{M}_{nq}(t-\frac{R}{\beta}). \qquad (4.31)
\end{aligned}$$

Note that $\dot{M}_{nq}$ is equal to the moment density m_{nq} mentioned at the end of the last section. Separating far field and near field terms analogous to the unidirectional point source S-wave associated far field displacement can be expressed by

$$u_i(\mathbf{x},t) = \frac{1}{4\pi\rho\beta^3}(\delta_{in} - \gamma_i\gamma_n)\gamma_q\frac{1}{R}m_{nq}(t-\frac{R}{\beta}). \qquad (4.32)$$

The last step in calculating amplitudes and travel times for the S-wave and the adjoint wave fields is a modification of the moment density $\mathbf{m}$.
First of all we have to think about $\boldsymbol{\omega}$, which was defined by

$$\boldsymbol{\omega}(\mathbf{x},t) := \frac{1}{2}\nabla \times \mathbf{u}(\mathbf{x},t).$$

In terms of ray theory and the assumption that spatial variations of amplitudes are small this reads

$$\begin{aligned}
\frac{1}{2}\nabla \times \mathbf{u}(\mathbf{x},t) &= \frac{1}{2}\epsilon_{ijk}\frac{\partial}{\partial x_j}u_k(\mathbf{x},t) = \frac{1}{2}\epsilon_{ijk}\frac{\partial}{\partial x_j}(U_k(\mathbf{x})T(t-\tau(\mathbf{x}))) \\
&\approx \frac{1}{2}U_k(\mathbf{x})\epsilon_{ijk}\frac{\partial}{\partial x_j}T(t-\tau(\mathbf{x})) \\
&= -\frac{1}{2}U_k(\mathbf{x})\epsilon_{ijk}(\frac{\partial}{\partial x_j}\tau(\mathbf{x}))\dot{T}(t-\tau(\mathbf{x})) \\
&= -\frac{1}{2}(\epsilon_{ijk}p_jU_k(\mathbf{x}))\dot{T}(t-\tau(\mathbf{x})) = -\frac{1}{2}(\mathbf{p}(\mathbf{x})\times\dot{\mathbf{u}}(\mathbf{x},t)).
\end{aligned} \tag{4.33}$$

In accordance with chapter 2 we expressed $\frac{\partial}{\partial x_j}\tau(\mathbf{x})$ by the slowness vector $p_j(\mathbf{x})$ in equation (4.33).

Using the identity $\epsilon_{ijk}\epsilon_{imn} = \delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km}$ and with equation (4.33) we can write for the moment density

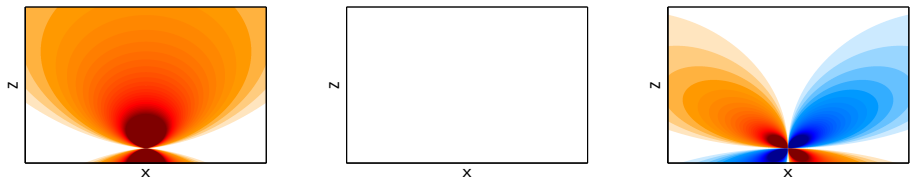
$$\begin{aligned}
m_{kj} &= \frac{1}{2A_\omega^2}\epsilon_{ijk}\ \omega_i\delta(\mathbf{x}-\mathbf{x}^r) \\
&= -\frac{1}{4A_\omega^2}\epsilon_{ijk}\epsilon_{imn}\ p_m\dot{u}_n\delta(\mathbf{x}-\mathbf{x}^r) \\
&= -\frac{1}{4A_\omega^2}(\delta_{jm}\delta_{kn}-\delta_{jn}\delta_{km})p_m\dot{u}_n\delta(\mathbf{x}-\mathbf{x}^r) \\
&= -\frac{1}{4A_\omega^2}(p_j\dot{u}_k - p_k\dot{u}_j)\delta(\mathbf{x}-\mathbf{x}^r).
\end{aligned} \tag{4.34}$$

Inserting recent results in equation(4.32) leads to the amplitudes of the adjoint rotation field

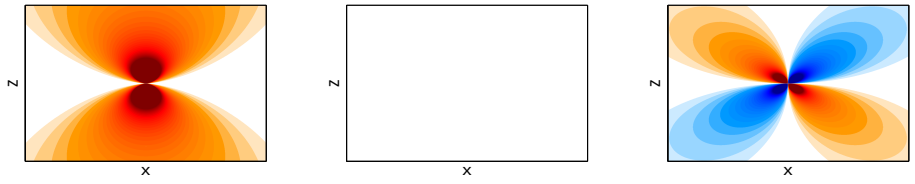
$$(U_{\psi^\omega})_j(\mathbf{x}) = \frac{1}{16{A_\omega}^2\pi\rho\beta^3}((\gamma_\psi)_j(\gamma_\psi)_k - \delta_{jk})(\gamma_\psi)_l(p_l(F_\psi)_k - p_k(F_\psi)_l)\frac{1}{R_\psi}.$$

In the following example we put a unidirectional point source ($F = (1, 0, 0)$) of an S-wave field directly under a receiver of the S-wave field, respectively a source of the adjoint velocity and rotation fields. The figures show:

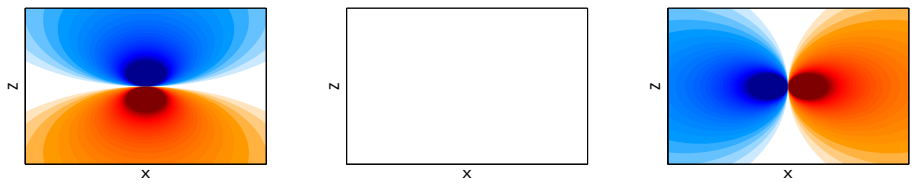
- S-wave amplitude components $(U_S)_1, (U_S)_2, (U_S)_3$ from the left to the right (vertical slices)



- adjoint velocity amplitude components $(U_{\psi^v})_1, (U_{\psi^v})_2, (U_{\psi^v})_3$ from the left to the right (vertical slices)



- adjoint velocity amplitude components $(U_{\psi^\omega})_1, (U_{\psi^\omega})_2, (U_{\psi^\omega})_3$ from the left to the right (vertical slices)



Again red color simbolises positive amplitudes, blue negative and white zero. The discrepancy between the radiation patterns of the adjoint fields will be crucial for the sensitivity kernel of the apparent shear wave speed.

We dedicate the next page to a summary of the most important results obtained so far:

Summary of results

To calculate the sensitivity of the apparent S-wave speed we found the equation

$$\frac{1}{\beta_a}\delta_\beta\beta_a = -2\frac{\rho}{\beta}\Big(\mathbf{U}_{\psi^v}(\mathbf{x}) - \mathbf{U}_{\psi^\omega}(\mathbf{x})\Big)\mathbf{U}_S(\mathbf{x})\int_{\mathbb{R}}\ddot{T}_S(t-\tau_S(\mathbf{x}))\ddot{T}_S(t-\tau_\psi(\mathbf{x}))\;dt.$$

The following expressions are contributed to the individual components of the equation above:

The time function is

$$\begin{aligned} T_S(t-\tau_S(\mathbf{x})) &= e^{-(t-\tau_S)^2\sigma^{-2}} \\ T_S(t-\tau_\psi(\mathbf{x})) &= e^{-(t-\tau_\psi)^2\sigma^{-2}} \end{aligned}$$

with traveltimes τ_S and τ_ψ.

Amplitudes for the S-wave and the adjoint velocity and rotation wave fields are given by

$$\begin{aligned} (U_S)_j(\mathbf{x}) &= \frac{1}{4\pi\rho\beta^2}(\delta_{ij}-\gamma_i\gamma_j)F_i\frac{1}{R_S} \\ (U_{\psi^v})_j(\mathbf{x}) &= \frac{1}{4{A_v}^2\pi\rho\beta^2}(\delta_{ij}-(\gamma_\psi)_i(\gamma_\psi)_j)(F_\psi)_i\frac{1}{R_\psi} \\ (U_{\psi^\omega})_j(\mathbf{x}) &= \frac{1}{16{A_\omega}^2\pi\rho\beta^3}((\gamma_\psi)_j(\gamma_\psi)_k-\delta_{jk})(\gamma_\psi)_l(p_l(F_\psi)_k-p_k(F_\psi)_l)\frac{1}{R_\psi} \end{aligned}$$

with

- slowness vector $p_i(\mathbf{x}^r) = \partial_i\tau_S(\mathbf{x}^r) = \frac{1}{\beta}\frac{x_i^r-x_i^s}{R_S}$
- direction cosines $\gamma_i = \delta_i R_S = \frac{x_i^r-x_i^s}{R_S}$ and $(\gamma_\psi)_i = \delta_i R_\psi = \frac{x_i^r-x_i^s}{R_\psi}$
- $\mathbf{F}_\psi = (U_1(\mathbf{x}^r), U_2(\mathbf{x}^r), U_3(\mathbf{x}^r))$

The expressions of page 60 enable us now to calculate the first derivative of the apparent S-wave speed β_a with respect to the true S-wave speed β, i.e. the sensitivity kernel $\frac{1}{\beta_a}\delta_\beta\beta_a$.
In brevious radiation pattern examples the model size was not relevant. But sensitivity kernels will form the basis for later tomography models. So it is about time to specify a certain model space for imaging sensitivity kernels. As the method is intended for local Earth structure inversions we choose a relatively small model extension by defining a model block with 100 kilometers in each edge length. All mappings of sensitivity kernels are either vertical or horizontal slices through the model block.

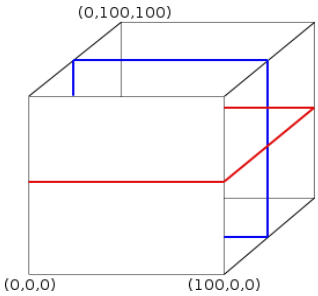


Figure 4.3: *model block, red: horizontal slice; blue: vertical slice*

The following examples show some basic properties of sensitivity kernels and demonstrate in what way they depend on relevant model parameters.
Models are homogeneous with S-wave velocity $\beta = 5km/s$ and density $\rho = 3000kg/m^3$. A more realistic model with a 1-D velocity gradient will be treated in a later section.
The first example will give an answer to the question in which region β_a is sensitive to β. In the second example we see that sensitivity kernels depend significantly on the parameter σ, which is related to the bandwidth of the recorded signal.

Example 1:
The model parameters are defined as: $\beta = 5km/s$, $\rho = 3000kg/m^3$, $\sigma = 0.5$ (i.e. the frequency band is $0 - 2Hz$), $F = (1, 0, 0)$, *source* $S = (20, 50, 20)$, *receiver* $R = (80, 50, 80)$.
In principle the formula for the sensitivity kernel is governed by a time integral and the difference between the adjoint fields. These components are shown in the figures below, that are vertical slices through the source/receiver-plane.

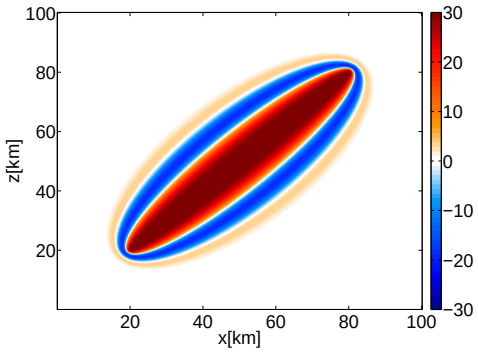


Figure 4.4: *Time integral; vertical slice:* $100 \times 100km$

The time integral gives the kernel its characteristic shape.

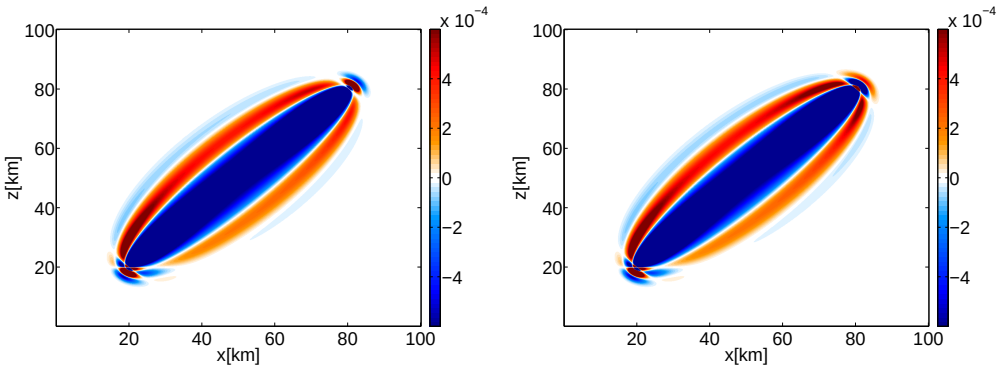


Figure 4.5: *Velocity (left) and rotation kernel (right); vertical slices:* $100 \times 100km$

Both, velocity as rotation kernels have a dominant first Fresnel zone (deep blue). Moreover, the difference between the velocity kernel and the rotation kernel is only near the receiver significant. The consequence is that the

sensitivity kernel $\frac{1}{\beta_a}\delta_\beta\beta_a$ vanishes as we approach the source, i.e. β_a is rather independent of the structure in the source region, which illustrates the next figure

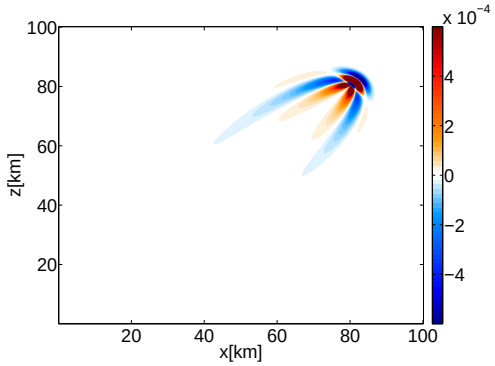


Figure 4.6: *β_a kernel; vertical slice:* $100 \times 100km$

Note that the first Fresnel zone is missing in the β_a kernel. The kernel is characterised by higher-order Fresnel zones producing stronger oscillating values compared to the velocity and the rotation kernels.
The effect of decreasing sensitivity with increasing depth can be reproduced as well by imaging horizontal slices through the model.

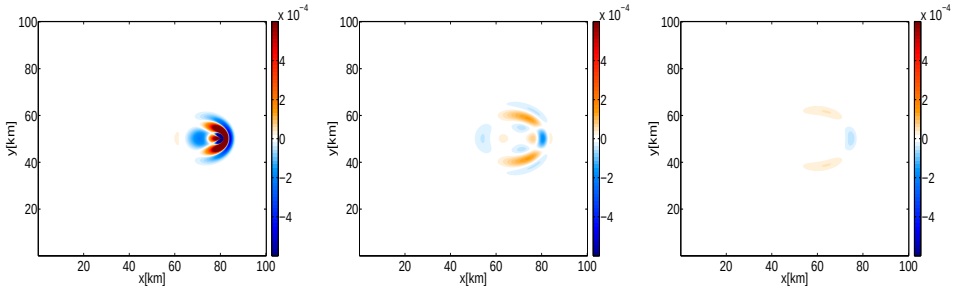


Figure 4.7: *Horizontal slices(*$100 \times 100km$*) through a β_a kernel; depth from the left to the right increasing; left: slice through the receiver plane; middle: 10km below the receiver plane; right: 20km below the receiver plane*

Example 2:
The model parameters are defined as in example 1, except σ. From the left to the right the figures show β_a kernels for $\sigma_1 = 0.25,\ \sigma_2 = 0.5\ and\ \sigma_3 = 1.0$.

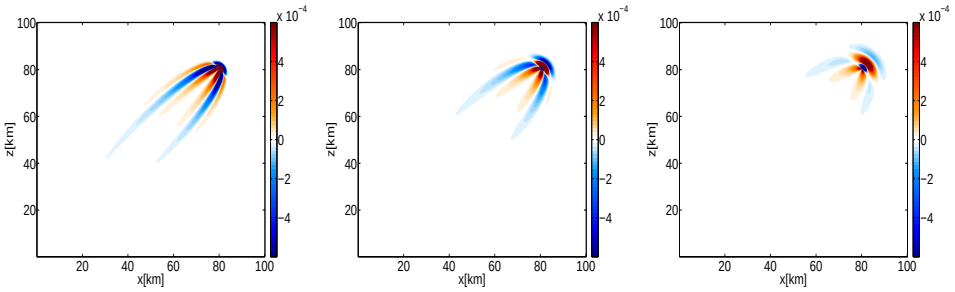
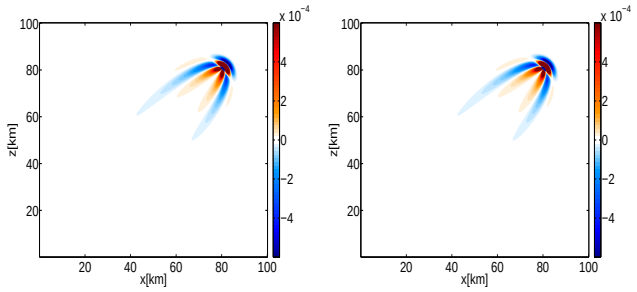


Figure 4.8: *β_a kernels for different frequency bandwidths, left: [0-4]Hz, middle: [0-2]Hz, right: [0-1]Hz; vertical slices:* $100 \times 100km$*;*

Remember: The smaller σ the broader is the frequency band, i.e. the shape of the sensitivity kernel gets slimmer the more (higher) frequencies are contained in the seismic signal. Note also that for kernels with a small frequency band the sensitivity is located to a significant part behind the receiver, whereas kernels with a broad frequency band (with higher frequencies) practically do not extend behind the receiver.

Example 3:
The model parameters are defined as in example 1, except F. We select F to be $F = (1, 0, 0)$ in the first figure and $F = (100, 0, 0)$ in the second figure.



Obviously both figures are identical due to the fact that we defined relative sensitivity kernels, that are independent to source intensities.

At this point we should briefly summarise the main characteristics of sensitivity kernels for the apparent shear wave speed, β_a.

- β_a kernels depend on the spatial direction of the seismic signal but not on its intensity.

- A broad first Fresnel zone is missing what causes strongly oscillating kernel values.

- The sensitivity is located near the receiver and decreases rapidly the closer we get to the source.

- Sensitivity kernels can be regarded as functions of frequency. Smaller frequency bandwidths produce broader β_a kernels and significant sensitivity behind the receiver. For broader frequency bandwidths the effect is revers.

Until now we placed our examples in a homogeneous isotropic unbounded medium. But in practice receiver stations are at the surface of the Earth - for the present we disregard the expensive opportunity of drilling boreholes. So we have to think about what happens if we calculate sensitivity kernels in a half space with the receiver on the free surface of our model. This will be done in the next section.

4.4 Free surface

Consider a scenario with a receiver directly on the free surface of the model. The questions we have to answer is what happens to the incoming S-wave when it reaches the receiver and how does this influence the β_a kernel.

First of all we regard the wavefronts of the incoming S-wave as plane waves. This assumption is true if the receiver is *sufficiently far* away from the source. With *sufficiently far* we refer to the number of wavelengths between the receiver and the source. Thus, even if our model is small, it is possible to use plane wave theory for high frequencies.

Further from basic seismology we know that a S-wave crossing an interface generates a reflected S-wave (SS), a reflected P-wave (SP) and two surface waves (Love- and Rayleigh-waves), as seen below.

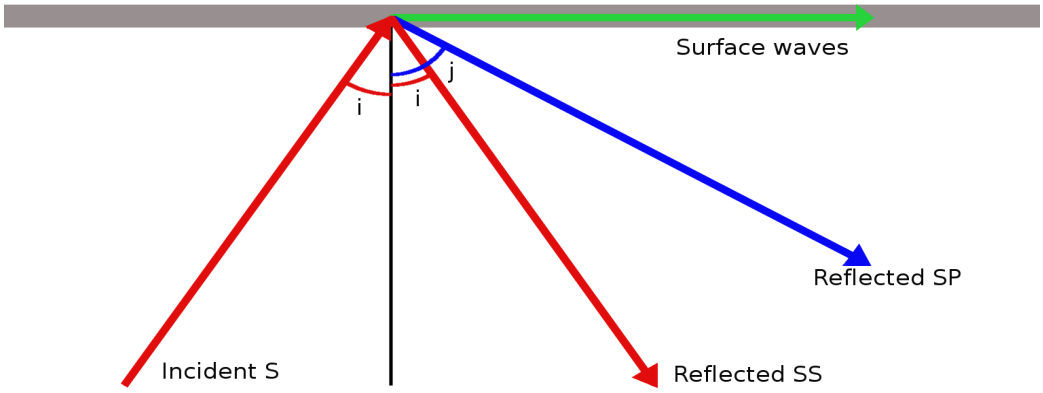
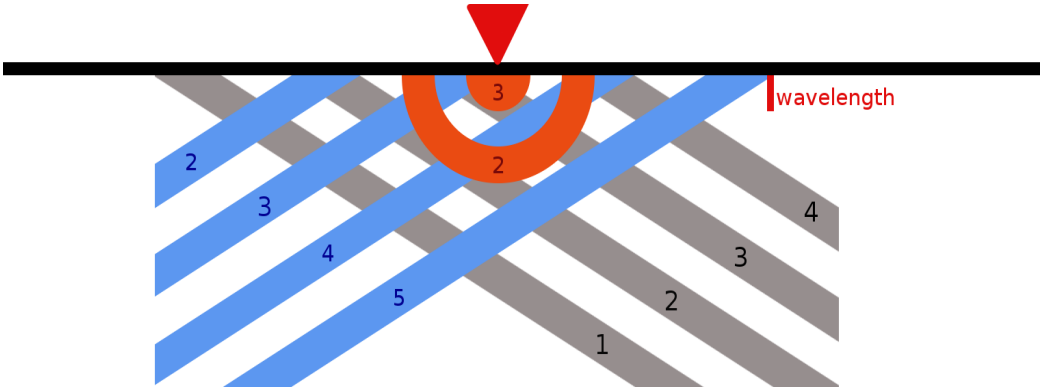


Figure 4.9: *Reflection of a S-wave at a free surface(grey); red: reflected S-wave, blue: reflected P-wave, green: Surface waves; it is* $\frac{sin(j)}{\alpha} = \frac{sin(i)}{\beta}$

Taking Love- and Rayleigh-waves into account for the computation of sensitivity kernels is quite complicated. Surface waves would dominate the amplitudes of the displacement field into a depth which is more or less equal to one wavelength of the incident S-wave. Using high frequencies will keep this zone small. Thus we decided to ignore surface waves in our calculation of sensitivity kernels. In view of numerical efficiency we accept that β_a kernels are not exact very close to the surface.
Besides surface waves another problem is raised by the reflected S- and P-wave. To estimate the area in which the reflected S-wave interacts with the adjoint fields consider the following sketch.



The horizontal black line symbolises the free surface. The receiver of the S-wave, respectively the source of the adjoint wave fields is marked by a red triangle. Incoming S-wavefronts are grey, reflected S-wavefronts are blue. Emitted adjoint wavefronts are drawn in orange colour. Wavefronts represent the displacement field at constant times. In the figure each wavefront is signed by a number that defines a single constant time. Obviously the S-wave field and the adjoint wave field interact only into a depth smaller than the wavelength λ. As P-waves are faster than S-waves this holds even more for P-waves.

Moreover we omitted near-field terms in the brevious section. Hence it seems to be reasonable to calculate sensitivites without surface waves and reflections even if receivers are located at a free surface. The consequence is that senitivity kernels are not exact at depths smaller than one wavelength but as just shown this simplification does not affect the sensitivity kernel at depths larger than the wavelength. The conclusion that higher frequencies are more appropriate is nothing new in our considerations. Applying ray theory even restricted us to high frequencies. Figure 4.8 also proposes high frequencies as otherwise sensitivities extend significantly behind the receiver what is not desirable.

Chapter 4 shall be completed by the computation of a sensitivity kernel for the apparent S-wave speed in a more realistic model with 1-D velocity gradient.

4.5 1-D velocity model

Admittedly one could argue that calculating sensitivity kernels only for a homogeneous medium may be too simplistic to form the basis for tomographic Earth models.
Nevertheless in this section we demonstrate that β_a kernels in more complex models do not differ much from kernels computed for homogeneous media. Thus developing algorithms based on complex models will not change the basic outcome but only increase numerical costs drastically.

A frequently used model in seismological applications is the 1-D velocity gradient model, where the velocity is assumed to increase linearly with depth in the form

$$\beta(x_3) = \beta_0 + bx_3. \tag{4.35}$$

As a first step in constructing a meaningful example we enlarge our model-cube to a length of $600km$ and a height of $300km$. Using the former model ($100km \times 100km$) would not produce any visible difference to the homogeneous case, provided that the velocity gradient stays realistic. Now we put the source at $S(20, 150, 50)$ and the receiver at $R(580, 150, 280)$. We define the velocity gradient b so that the velocity β_S at the source is $7.5km/s$ and $5.2km/s$ at the receiver (1-D velocity model). We calculate the velocity, the rotation and the β_a kernel for the 1-D velocity model and for the homogeneous model. A comparison of the kernels is shown in the following figure.

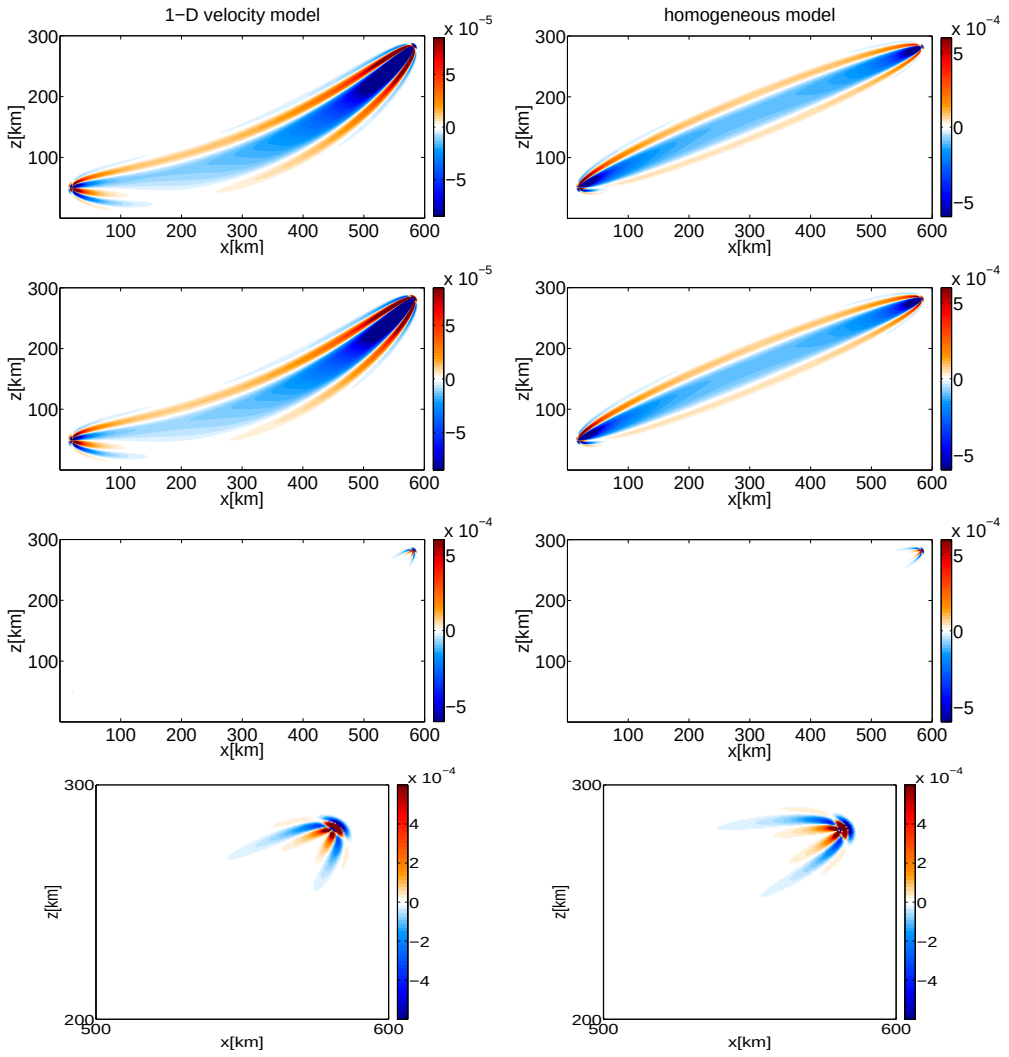


Figure 4.10: *Top to bottom: velocity, rotation and β_a kernel for the 1-D model (left column) and the homogeneous case (right column); vertical slices:* $600 \times 300km$ *(row 1 - 3) and* $100 \times 100km$ *(zoom-in, row 4)*

The effect of rapidly decreasing sensitivity of the β_a kernel towards the source stays the same in the 1-D model and zooming into the figures demonstrates the similarity of the β_a kernel in the 1-D model to the β_a kernel in the homogeneous model (figure 4.10: bottom row). Nevertheless, differences can be observed in the angle of incident of the incoming wave field at the receiver but in inversion applications these effects are managed by variations in source-receiver positions.

The conclusion of this example is that later outputs of inversion algorithms

will not depend much on the complexity of the model in which sensitivity kernels are calculated. We assume that other parameters, like source-receiver positions, number of sources and receivers, choice of frequency bandwidth, etc. are of greater importance for successful inversions. To explore how these parameters affect inversion models with senstitivty kernels for the apparent S-wave speed is the purpose of the next chapter.

Chapter 5

Inversion

There exist many different concepts for the solution of an inverse problem. In practice the process of solving the inverse problem is often equated with the minimisation of the difference between observed and synthetic data. Trying this approach we use an iterative optimisation scheme that is based on the previously computed β_a kernels to solve elementary synthetic inverse problems.

5.1 The non-linear inverse problem

In this chapter the inverse problem is seen as an iterative minimisation of a non-linear misfit function $\mathfrak{E}(\beta_{a,i})$. We choose the quadratic function

$$\mathfrak{E}(\beta_{a,i}) = \frac{1}{2}\sum_{i=1}^{n}(\frac{1}{\beta_{a,i}^{(0)}}(\beta_{a,i}^{(0)} - \beta_{a,i}))^2 \tag{5.1}$$

where $\beta_{a,i}^{(0)}$ denotes the observations, i.e. measurements of the apparent shear wave speed at each receiver station, and $\beta_{a,i}$ denotes the corresponding synthetic values. Index values i symbolise for example different source positions, receiver stations or frequencies.
Now we calculate the gradient, i.e. in our case the first partial derivative of $\mathfrak{E}(\beta_{a,i})$ with respect to the shear wave speed β. Leading to

$$\begin{aligned}\partial_\beta\mathfrak{E}(\beta_{a,i}) &= \sum_{i=1}^{n}\frac{\Delta\beta_{a,i}}{\beta_{a,i}^{(0)}}\frac{1}{\beta_{a,i}^{(0)}}\partial_\beta\beta_{a,i} \\ &\approx \sum_{i=1}^{n}\frac{\Delta\beta_{a,i}}{\beta_{a,i}^{(0)}}\frac{1}{\beta_{a,i}}\partial_\beta\beta_{a,i}\end{aligned} \tag{5.2}$$

with $\Delta\beta_{a,i} = \beta_{a,i}^{(0)} - \beta_{a,i}$. Note that this equation incorporates the sensitivity kernels $\frac{1}{\beta_{a,i}}\partial_\beta \beta_{a,i}$.
In order to invert now for an original model $\beta^{(0)}$ we have to define a starting model β_0. In chapter 4 we stated that a homogeneous model with constant shear wave speed should be appropriate as starting model. The starting model is then updated in arbitrary iteration steps through

$$\beta_{k+1} = \beta_k + s \cdot \partial_\beta \mathfrak{E} \ , \quad k = 0, 1, 2, ... \tag{5.3}$$

The assumption is that after sufficient iterations the updated model represents a satisfying approximation of the original model. The multiplier s has to be determinded by minimising the misfit function (5.1).

Before we try to invert some synthetic models it might be helpful to perform a systematic investigation of synthetic apparent shear wave speed measurements.

5.2 A study of synthetic apparent shear wave speed data

For $\Delta\beta_{a,i} = \beta_{a,i}^{(0)} - \beta_{a,i}$ with apparent shear wave speed measurements in a synthetic model ($\beta_{a,i}^{(0)}$) and a homogeneous model ($\beta_{a,i}$) we regard $\frac{1}{\beta_a^{(0)}}\Delta\beta_a$ as the relative apparent shear wave speed measurement at some receiver station i. We are interested in synthetic apparent shear wave speed data affected by S-wave speed perturbations in the original model. The standard solution to produce the synthetic data consists in using scattering theory. Equally we can calculate the relative apparent shear wave speed via the relation

$$\begin{aligned} \frac{1}{\beta_a^{(0)}}\Delta\beta_a &\doteq \int \Delta\beta(\mathbf{x})\frac{1}{\beta_a^{(0)}}\ \partial_\beta\beta_a(\mathbf{x})\ d^3\mathbf{x} \\ &\approx \int \Delta\beta(\mathbf{x})\frac{1}{\beta_a}\ \partial_\beta\beta_a(\mathbf{x})\ d^3\mathbf{x} \end{aligned} \tag{5.4}$$

where $\Delta\beta$ is the difference of the S-wave speed in the original and the homogeneous model.
The following examples show again vertical slices through a model block ($100 \times 100km$) with one unidirectional point source in $x-$direction, located at $S(50, 50, 20)$. On the surface of the vertical slice we place receiver stations $R_i(i, 50, 100),\ i = 0, ...100$. Into the model block we put a point perturbation $\boldsymbol{\zeta}(\zeta_1, 50, \zeta_3)$, extended by a Gauss function to avoid unrealistic sharp edges. The sketch below shows the scenario for a positive perturbation at $\boldsymbol{\zeta}(50, 50, 80)$, with an extension of $6km$. At the perturbation the S-wave speed increases from $5km/s$ to a maximum of $6.5km/s$.

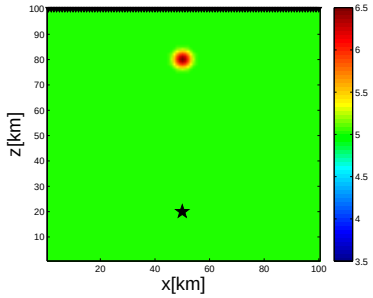


Figure 5.1: *red: perturbation of the S-wave speed; black asterisk: unidirectional point source in $x-$direction; black triangles: receiver positions*

The investigation focuses on what happens with the apparent shear wave speed measurements for different perturbation positions and for variations of frequencies in the seismic signal. Some measurements are compared to the β_a kernel at the receiver located directly above the perturbation position.

Example 1(Testing different perturbation depths)
The situation in the first example is like in the upper model, except that the perturbation depth varies between 60 and $90km$. The frequency band of the seismic signal is constant from $0 - 2Hz$.

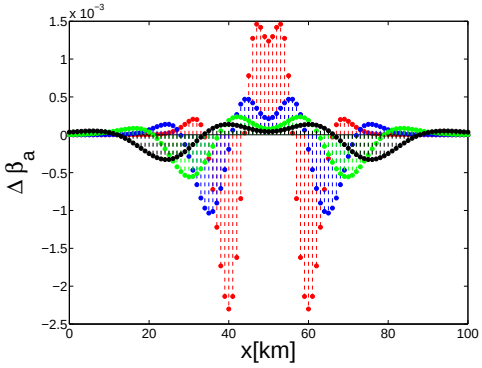


Figure 5.2: *Apparent shear wave speed data for perturbations at* $\boldsymbol{\zeta}(50, 50, 90)[red], \boldsymbol{\zeta}(50, 50, 80)[blue], \boldsymbol{\zeta}(50, 50, 70)[green], \boldsymbol{\zeta}(50, 50, 60)[black]$

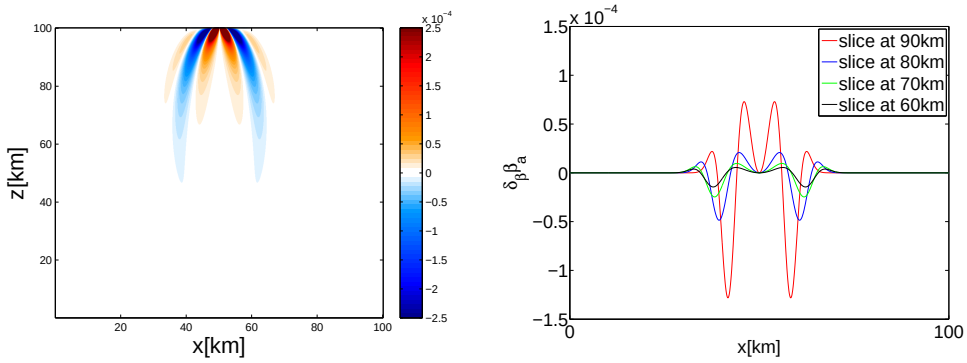


Figure 5.3: *left:* β_a *kernel for a source at* $S(50, 50, 20)$ *and a receiver at* $R(50, 50, 100)$*; right: horizontal slices through the kernel;*

In figure 5.2 different colours represent different perturbation depths. Expectedly the signal gets weaker for increasing perturbation depths. More remarkable is the symmetry in the data reflecting the Fresnel zones of the β_a

kernel in figure 5.3 and the local minimum above the perturbation ($\zeta_1 = 50$), corresponding to the kernel's singularity at the receiver position.

Example 2 deals with variations in the size of the perturbation.

Example 2(Testing different perturbation sizes)
The model set-up is equal to the first example except for the size of the perturbation. In figure 5.4 the perturbation is smaller ($2 \times 2km$), in figure 5.5 it is enlarged to $12 \times 12km$.

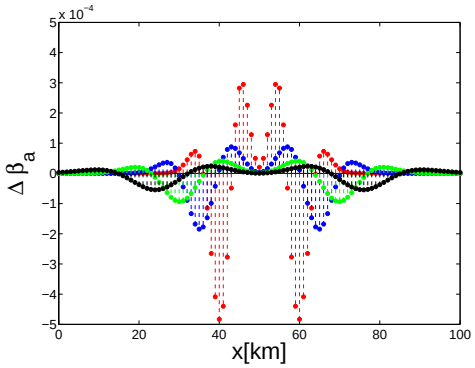


Figure 5.4: *Apparent shear wave speed data for a smaller perturbation: local minimum at the perturbation amplifies*

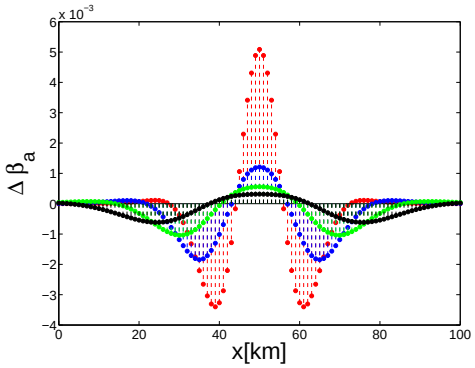


Figure 5.5: *Apparent shear wave speed data for a larger perturbation: local minimum at the perturbation vanishes*

This example shows that the size of the perturbation affects the local minimum above the perturbation. Taking into account how we calculate the data (see equation (5.4)) it is clear why the singularity of the kernel dominates

the data curve for smaller perturbations whereas larger perturbations cancel out the singularity - there is no local minimum at the perturbation anymore.

Another question is what happens for lateral variations of the perturbation position. Therefor example 3 was constructed.

Example 3(Testing lateral perturbation shifts)
In figure 5.6 the perturbation $(2 \times 2km)$ is shifted to the left and in figure 5.8 to the right. Other parameters match with previous examples.

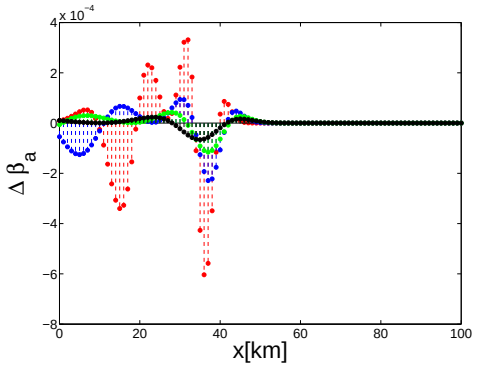


Figure 5.6: *Apparent shear wave speed data for a perturbation at* $\boldsymbol{\zeta}(30, 50, 90)$*: red,* $\boldsymbol{\zeta}(30, 50, 80)$*: blue,* $\boldsymbol{\zeta}(30, 50, 70)$*: green,* $\boldsymbol{\zeta}(30, 50, 60)$*: black*

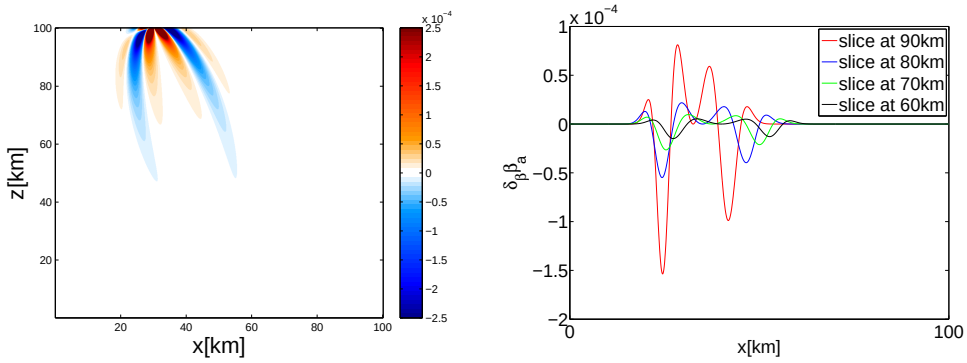


Figure 5.7: *left:* β_a *kernel for a source at* $S(50, 50, 20)$ *and a receiver at* $R(30, 50, 100)$*; right: horizontal slices through the kernel;*

Again the data reflects the Fresnel zones of the β_a kernel. But now due to an inclined angle of incident the curve is asymmetric. Note that the local minimum of each data curve (corresponding to the kernel's singularity) moves to

the left for increasing perturbation depths.

Because of the general symmetry of the problem analogous effects can be observed for the perturbation shifted to the right:

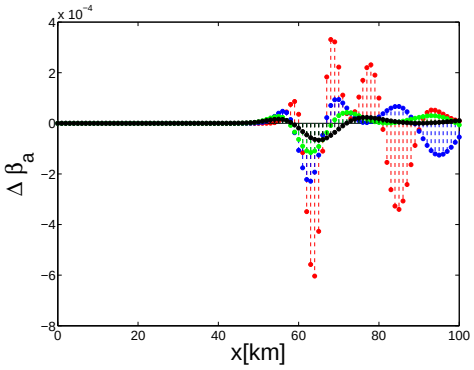


Figure 5.8: *Apparent shear wave speed data for a perturbation at* $\zeta(30, 50, 90)$*: red,* $\zeta(30, 50, 80)$*: blue,* $\zeta(30, 50, 70)$*: green,* $\zeta(30, 50, 60)$*: black*

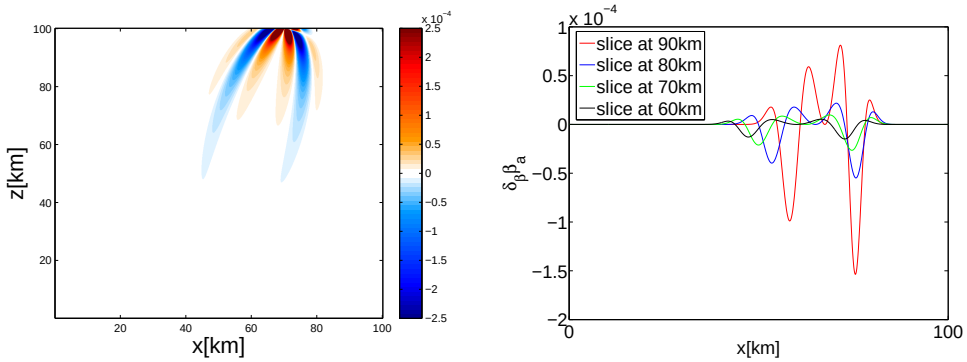


Figure 5.9: *left:* β_a *kernel for a source at* $S(50, 50, 20)$ *and a receiver at* $R(70, 50, 100)$*; right: horizontal slices through the kernel;*

Concerning the perturbation parameters we can further test variations of the intensity of the perturbation.

Example 4(Testing different perturbation intensities and source frequencies) The parameter set-up is the same as in example 1 for the perturbation at $\boldsymbol{\zeta}(50, 50, 90)$ but with variations of the perturbation intensity (left figure: maximal increase of the S-wave speed to $6.5km/s$[red], $5.5km/s$[blue] and $7.5km/s$[green]) and with variations of the source frequency (right figure: frequencies from $0 - 2Hz$[red], $0 - 1Hz$[blue] and $0 - 3Hz$[green]).

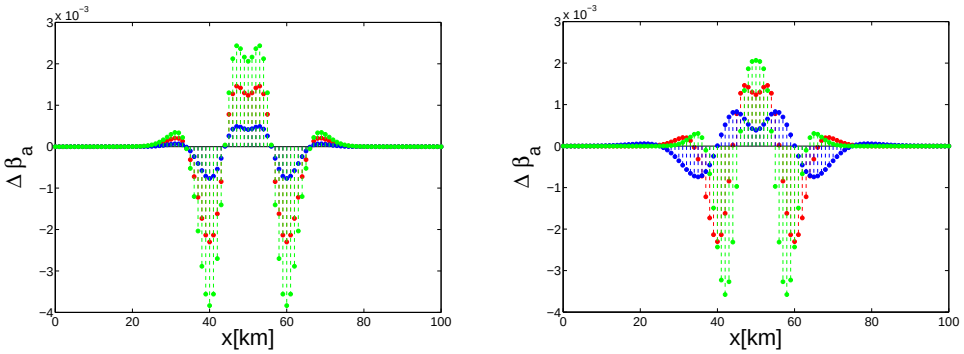


Figure 5.10: *left: apparent shear wave speed data for different perturbation intensities: increase of the S-wave speed up to* $6.5km/s$*(red),* $5.5km/s$*(blue) and* $7.5km/s$*(green); right: apparent shear wave speed data for different source frequencies:* $0 - 2Hz$*(red),* $0 - 1Hz$*(blue) and* $0 - 3Hz$*(green)*

Expectedly stronger perturbations produce only larger measurement values, proportional to the intensity changes (left figure). A similar effect can be demonstrated by changing the frequency band of the seismic signal. Broader frequency bands lead in analogy to sensitivity kernels to stronger and faster oscillating data curves (right figure).

Altogether examples 1-4 proof what we already expected by intuition confirming that the theory outlined so far should be consistent.

Example 5 and 6 differ from the previous ones in the way that we complicate the nature of the perturbations.

Example 5
Here we enlarge step by step the lateral extension of the perturbation. We start with a perturbation of $10\times10km$ and end up at $60\times10km$. The maximal increase of the S-wave speed is in each case from $5km/s$ to $6.5km/s$.

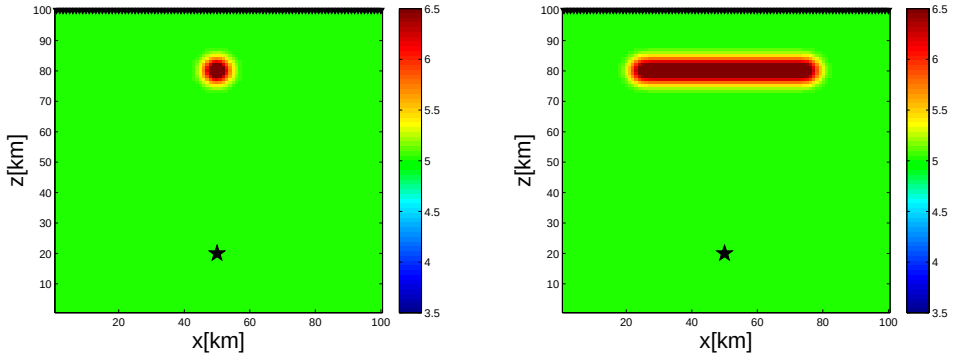


Figure 5.11: *S-wave speed perturbation models, left:* $10 \times 10km$*, right:* $60 \times 10km$

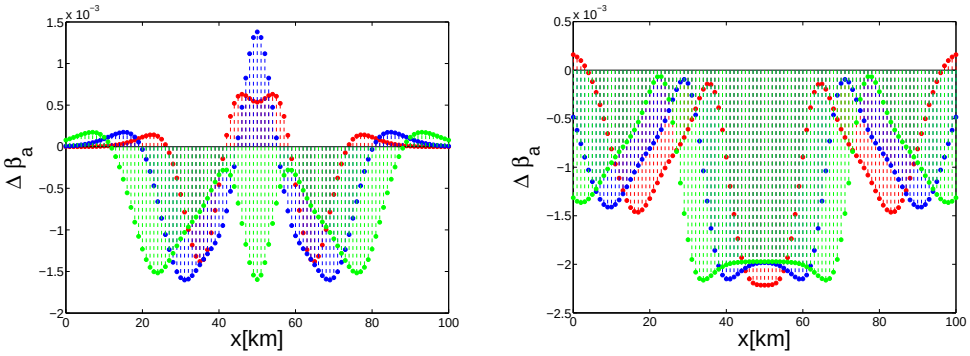


Figure 5.12: *Apparent shear wave speed data for lateral expanding perturbations, left:* $10 \times 10km(red)$*,* $20 \times 10km(blue)$*,* $30 \times 10km(green)$*; right:* $40 \times 10km(red)$*,* $50 \times 10km(blue)$*,* $60 \times 10km(green)$*;*

Remarkable is the phenomenon that large positive perturbations produce negative measurement data. Note the skip from positive to negative values as we pass from $20 \times 10km$ to $30 \times 10km$ perturbation size. Apart from this astonishing behaviour especially the data curves of the right figure capture clearly the lateral extension of the perturbation.

Example 6
The study is completed by a chess board pattern model. We put three positive and three negative point perturbations in the model and simulate three measurement series for small($5 \times 5km$), middle($10 \times 10km$) and large($15 \times 15km$) point perturbations.

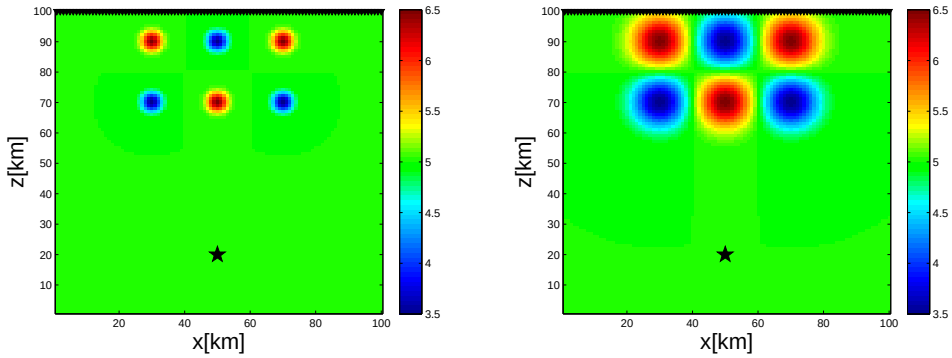


Figure 5.13: *S-wave speed perturbation models, left:* $5 \times 5km$*, right:* $15 \times 15km$

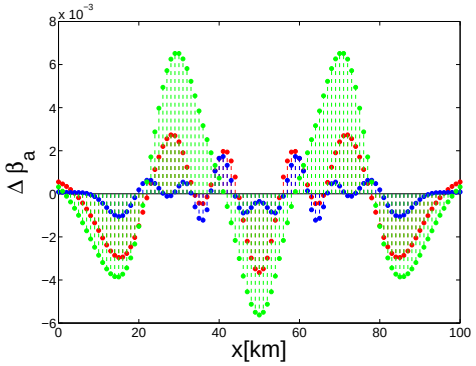


Figure 5.14: *Apparent shear wave speed data for* $5 \times 5km$*(blue),* $10 \times 10km$*(red) and* $15 \times 15km$*(green) point perturbations*

The intention of the example is to show that for more complex perturbation models a direct interpretation of data curves is extremely difficult. In this example we can only state that larger perturbations lead to smoother data curves.

In what follows we try to invert a synthetic model, which is of more demonstrative character than a realistic scenario.

5.3 An inversion experiment

This work ends with an inversion experiment where we apply the developed theory to a simple but insightful perturbation model. The original model (see figure 5.15) we want to invert contains a positive perturbation of the S-wave speed. The perturbation is a point perturbation extended by a Gauss function to a volume of approximately $10 \times 10km$ centered around $\boldsymbol{\zeta}(50, 50, 90)$ and a maximal increase of the S-wave speed from $5km/s$ to $6.5km/s$. In a first trial we use only one source at $\mathbf{S}(50, 50, 20)$ radiating a seismic signal in the frequency band of $0 - 2Hz$, oriented in $x-$direction

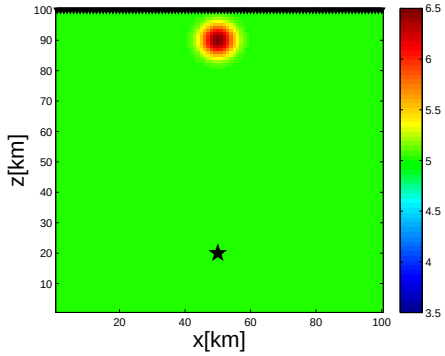


Figure 5.15: *Perturbation model with a positive point perturbation at* $\boldsymbol{\zeta}(50, 50, 90)$ *and a maximal increase of the S-wave speed to* $6.5km/s$*. The seismic source,* $\boldsymbol{S}(50, 50, 20)$*, is marked by a black asterisk*

Before we actually start with the inversion we have to care about the singularity of sensitivity kernels at the receiver. Due to comparatively large values of $\frac{1}{\beta_a}\delta_\beta\beta_a$ in the receiver region it is problematic to use the kernels directly for inversions. In practice sensitivity kernels are often adapted by a so called pre-conditioner. In our case we multiply the kernels by a factor $\mathfrak{p} = 1 - exp(-((z^2 + x^2) \cdot f^2))$, where f is the highest frequency contained in the seismic signal. This leads to a slightly modified kernel (see figure 5.16) without extreme values.

Remark: Because of $\lim_{f\to\infty} \mathfrak{p} = 1$ the manipulation is less significant the broader the frequency band is.

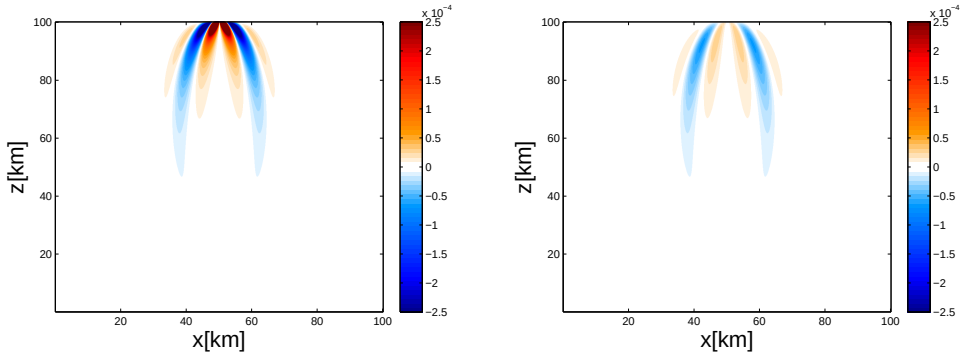


Figure 5.16: *left: β_a kernel without pre-conditioner; right: β_a kernel multiplied by a pre-conditioner*

To demonstrate the utility of a pre-conditioner we just calculate the gradient for the first model update in our inversion example.

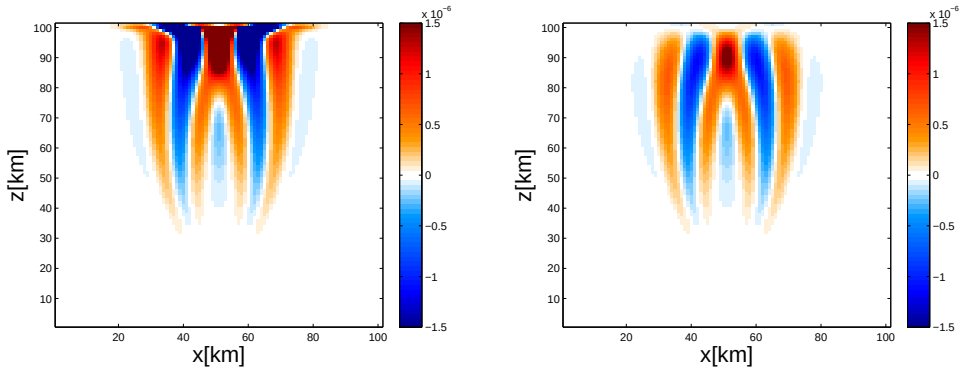


Figure 5.17: *left: gradient for the first model-update without pre-conditioner; right: gradient for the first model-update with pre-conditioner*

In the left figure kernels were used directly. Obviously the extreme kernel values near the surface produce misleading contributions to the model updates. This undesired effect is canceled out in the right figure where kernels were adapted by a pre-conditioner.

Following the inversion theory introduced at the beginning of this chapter we get the first model update by adding a special multiple of the gradient to the homogeneous starting model, so that the misfit between real apparent shear wave speed measurements (original model) and synthetic apparent shear wave speed measurements (model update) is minimised. Figure 5.18

(left) shows apparent shear wave speed data of 101 receiver stations for the original model (black) and the first model update (red) together with the inversion model after the first update (right).

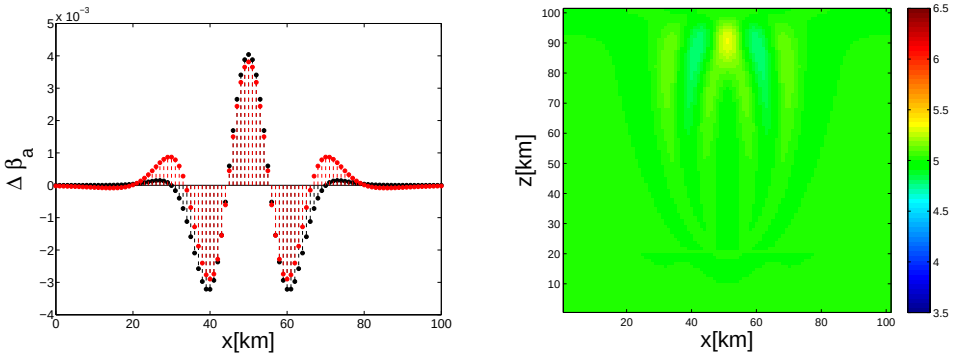


Figure 5.18: *left: apparent shear wave speed data for the original model (black) and for the first model update (red); right: inversion model after the first update*

The inversion result can be improved by iterated model updates. After ten iterations *real* and *synthetic* data match quite well and the inversion model highlights clearly the shape of the perturbation although the amplitudes are too small.

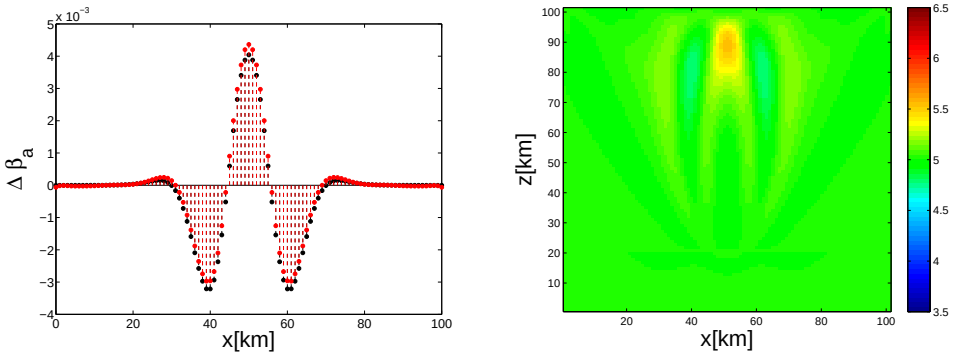


Figure 5.19: *left: apparent shear wave speed data for the original model (black) and for the tenth model update (red); right: inversion model after ten updates, the perturpation is clearly visible*

By iterating up to twenty model updates the result improves again slightly.

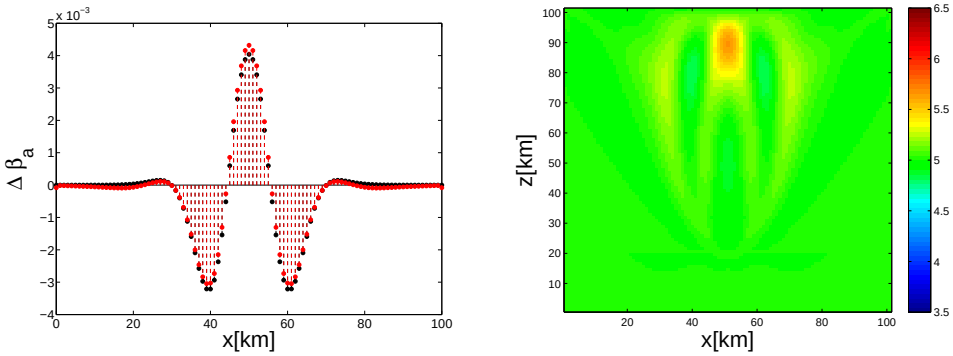


Figure 5.20: *left: apparent shear wave speed data for the original model (black) and for the twentieth model update (red); right: inversion model after twenty updates, the result improves again slightly*

Considering the misfit development for the iteration steps in figure 5.21 explains why there is no further improvement to expect after twenty iterations.

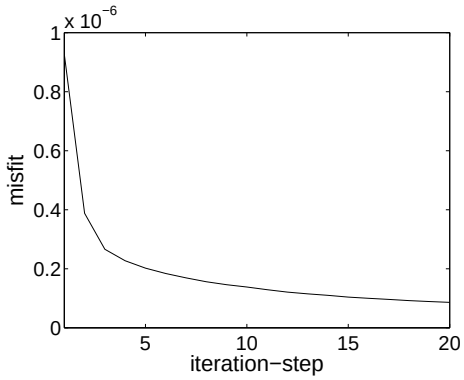


Figure 5.21: *Misfit for twenty iteration steps: There is no significant improvement to expect after twenty iterations.*

Although the result of the previous example is quite respectable, we are interested in what happens if we change certain parameters in the inversion trial. Taking into account that sensitivity kernels can be regarded as functions of frequency we should at first try different frequency bandwidths.

In the following two examples we repeat the inversion trial from above but now with frequency bandwidths from $0 - 1Hz$ and $0 - 3Hz$.

Inversion example for frequencies between 0 and $1Hz$

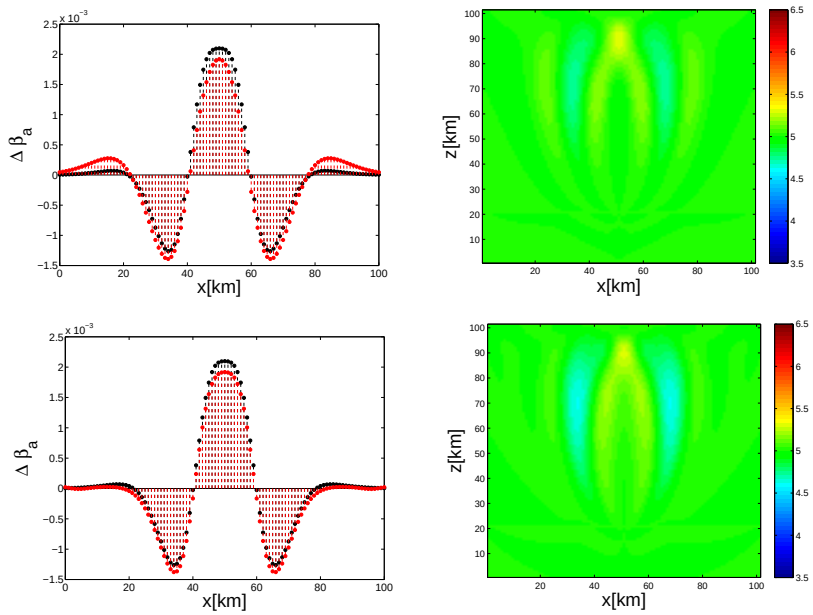


Figure 5.22: *left: apparent shear wave speed data for the original (black) and the updated (red) model [top: first update, bottom: tenth update]; right: inversion model [top: first update, bottom: tenth update]*

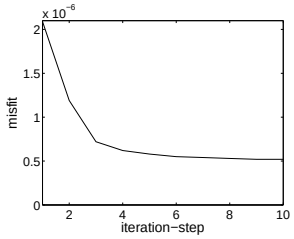


Figure 5.23: *Misfit for ten iteration steps: The iteration is close to its reachable optimum after ten steps*

The misfit-graph (figure 5.23) shows that the inversion process is close to its reachable optimum after ten iteration steps. The discrepancy between *real* (black) and *synthetic* (red) data is acceptable after ten iterations (figure 5.22, bottom-left). Nevertheless the tenth update of the inversion model (figure 5.22, bottom-right) does not really reproduce the original model. Hence in accordance with former postulations we should avoid lower frequencies.

Inversion example for frequencies between 0 and $3Hz$

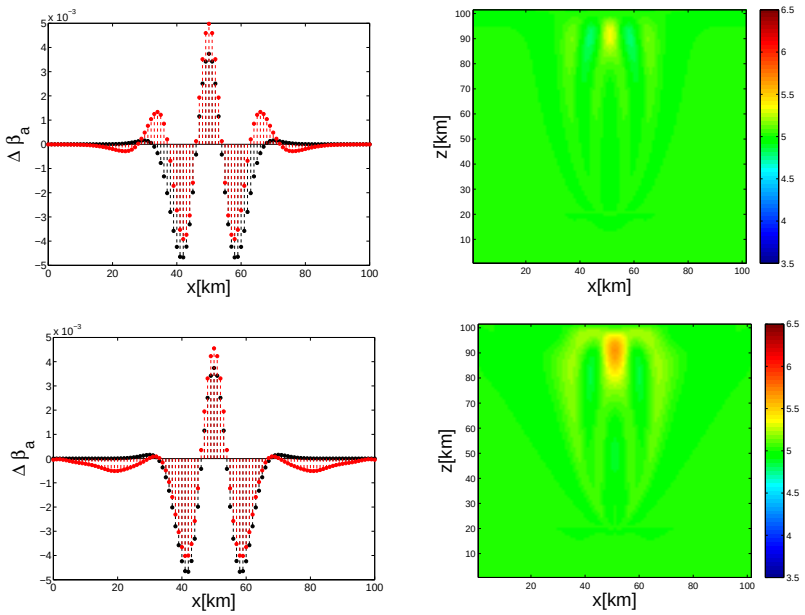


Figure 5.24: *left: apparent shear wave speed data for the original (black) and the updated (red) model [top: first update, bottom: tenth update]; right: inversion model [top: first update, bottom: tenth update]*

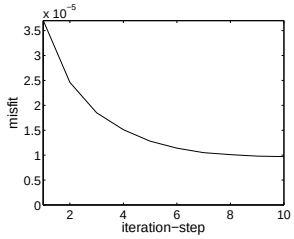


Figure 5.25: *Misfit for ten iteration steps: The iteration is close to its reachable optimum after ten steps*

For a broader frequency band up to $3Hz$ the inversion result (figure 5.24, bottom-right) is quite similar to the first trial with a maximal frequency of $2Hz$. But looking at the apparent shear wave speed data (figure 5.24, bottom-left) especially lateral receiver measurements do not fit very well with the synthetic data. This may be a reason for the obviously too slim reproduction of the perturbation in the inversion model.

Summarising the different frequency trials it seems that for this special example a frequency band from $0 - 2Hz$ is quite optimal. For that reason we expand this experiment with one source to an experiment with three sources, what usually improves results of seismic inversions.

Inversion example for frequencies between 0 and $2Hz$ using three sources

The experimental setup stays the same except that we add two sources at $\mathbf{S}_1(30, 50, 20)$ and $\mathbf{S}_3(70, 50, 20)$.

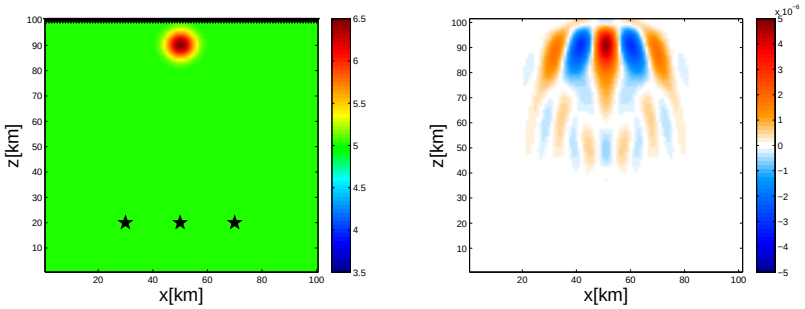


Figure 5.26: *left: The perturbation model is the same as at the beginning of this section, but now three sources,* $\boldsymbol{S}_1(30, 50, 20)$*,* $\boldsymbol{S}_2(50, 50, 20)$*,* $\boldsymbol{S}_3(70, 50, 20)$*, are used for the inversion; right: gradient for the first model update*

Using three sources we have to handle three different data curves. In the following only the diagrams for the left source, $\mathbf{S}_1(30, 50, 20)$, and the middle souce, $\mathbf{S}_2(50, 50, 20)$, are shown. As this is a symmetric problem the diagram for the right source, $\mathbf{S}_3(70, 50, 20)$, is analogous to the left one.

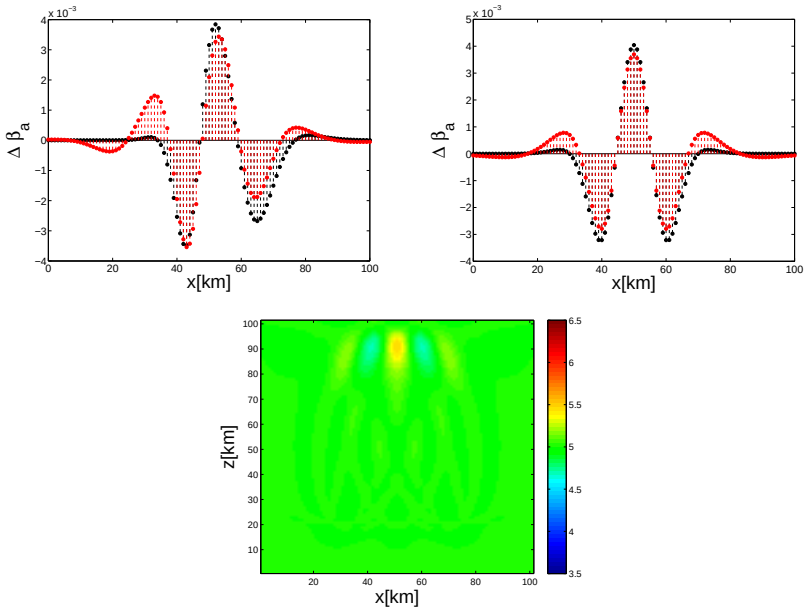


Figure 5.27: *One iteration; top-left: apparent shear wave speed data for* $\boldsymbol{S}_1(30, 50, 20)$*; top-right: apparent shear wave speed data for* $\boldsymbol{S}_2(50, 50, 20)$*; bottom: inversion model*

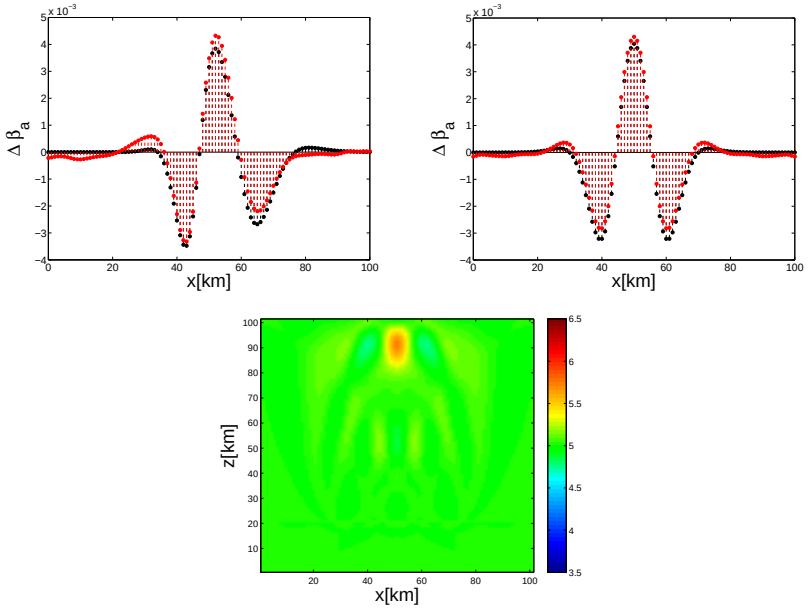


Figure 5.28: *Five iterations; top-left: apparent shear wave speed data for* $\boldsymbol{S}_1(30, 50, 20)$*; top-right: apparent shear wave speed data for* $\boldsymbol{S}_2(50, 50, 20)$*; bottom: inversion model*

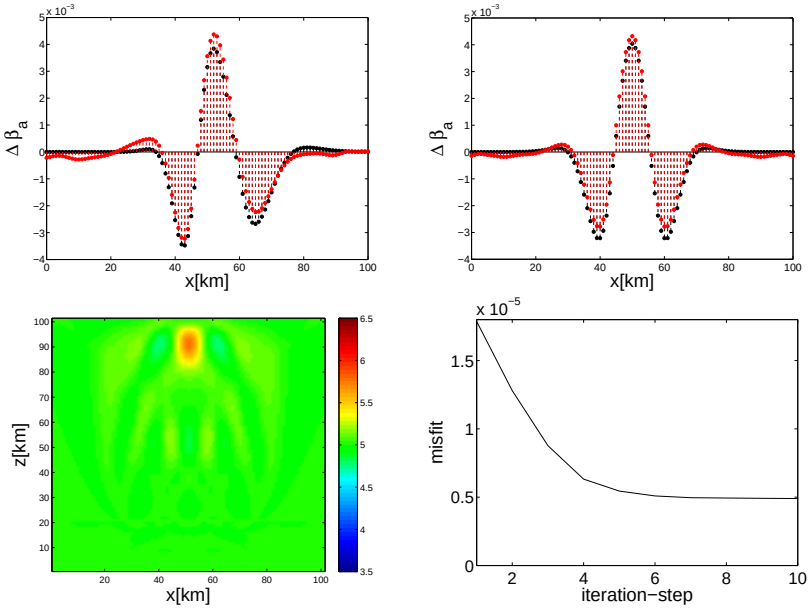


Figure 5.29: *Ten iterations; top-left: apparent shear wave speed data for* $\boldsymbol{S}_1(30, 50, 20)$*; top-right: apparent shear wave speed data for* $\boldsymbol{S}_2(50, 50, 20)$*; bottom-left: inversion model; bottom-right: misfit for ten iteration steps: The iteration is close to its reachable optimum after ten steps*

The inversion result after ten iterations is an improvement to the experiment with only one source. Especially lateral errors decrease. Moreover, with three sources the quality of the reproduction of the original model is already after five iterations remarkable. This suggests that including more sources reduces the non-uniqueness of the inversion problem.

Chapter 6

Conclusions and Outlook

The objective of this work is to contribute to the development of a new method of seismic tomography that combines standard translational and new rotational ground motion measurements - motivated by the recent high-quality and consistent observations of rotational ground motions using ring-laser technology.

The theory is based on the definition of an apparent shear wave speed, β_a, which is the ratio of the rms displacement velocity and the rms displacement rotation. Combining the adjoint method and ray theory for the computation of sensitivity kernels of β_a it is shown that the sensitivity of β_a with respect to β is confined to a volume sorrounding the receiver. Inversion results in chapter 5 indicate that joint measurements of rotational and translational ground motion may indeed be used to obtain images of the subsurface structure near the receiver. The principal advantages of the technique are the independence from deep Earth structure, source magnitude and source timing.

Nevertheless, the quality of the inversion, i.e. the geometry and amplitude of the reconstructed perturbation, is highly sensible to the parametrisation of the inversion model - especially to the choice of frequency. This is nothing new in seismology and moreover not astonishing as sensitivity kernels can be regarded as functions of frequency.

Although β_a kernels are a combination of velocity and rotation kernels they have quite different properties. The most important ones are the vanishing sensitivity in the source region and suppressed contributions in the first-order Fresnel zone but enhanced contributions in higher-order Fresnel zones. The second attribute leads to strongly oscillating values of β_a kernels suggesting that β_a-measurements are sensitive to small-scale structures. This may

explain why it is not possible with our current algorithm to invert for perturbations with lateral extension but only for a single point perturbation. Inversion experiments for more complex models, like the chess board pattern model, failed, too.

So the consistency of the developed theory could be demonstrated in an elementary example but before an application of the method to real data becomes realistic further investigations are necessary in order to make inversions of more complex synthetic models successful.

This investigation should concentrate on following issues:

- In our purpose of adapting pure theory for practical application we had to accept several approximative solutions. Concerning this matter we should primarily mention the ray method. The description of a seismic wave field only along certain ray paths can be problematic (see chapter 2) but in our case the ray method is applied in a homogeneous and a 1-D velocity gradient model. In these extremely smooth models the ray method should not produce any misleading results. Moreover, calculating sensitivity kernels with the ray method is a central issue of the presented work because it leads to fast and flexible algorithms that facilitate extensive parameter studies. Altogether the ray method should be appropriate for further investigations.

- Another discussible issue concerns the starting model that we used for inversions. In principle one should try to keep already the starting model for seismic inversions as realistic as possible. We used a homogeneous starting model that has nothing to do with reality. But the similarity of β_a kernels in a homogeneous medium to β_a kernels in more complex media, due to their concentration to the receiver region, should justify this approach.

- An essential problem is that β_a kernels cannot be used directly for inversions. On the one hand the singularity of the kernel at the receiver suggests highest sensitivity to the structure at that point, so that it seems paradox to exclude exactly this data from inversion. On the other hand the values in that region absolutely dominate the β_a kernel so that inversion trials without pre-conditioner normally fail.
  The choice of a pre-conditioner is quite arbitrary and until now we tested only one possible pre-conditioner. Especially in the context of perturbations with lateral extension the inversion should be successful with an appropriate pre-conditioner.

- Besides many options for a suitable pre-conditioner there exist many different concepts of how to solve an inverse problem, too. Especially when the problem is highly non-linear it may be useful to solve the inversion problem applying stochastic methods. For example Monte Carlo methods are an alternative to the minimisation of the misfit function $\mathfrak{E}$ by means of gradient methods.

- Finally it would be interesting to investigate what happens if we use surface waves instead of S-waves for the inversion.

Altogether I propose to realise a series of additional synthetic experiments primarily concerned with the design of efficient pre-conditioners, the choice of optimal frequency bands and the construction of well-placed source/receiver scenarios. After that the challenge is an application of this new method especially to engineering and exploration problems.

Bibliography

[1] Aki, K., P. G. Richards (2002). *Quantitative seismology*, 2nd Edition, University Science Books.

[2] Brokešová, J. (2006). *Asymptotic ray method in seismology*, Matfyzpress.

[3] Červený, V. (2001). *Seismic ray theory*, Cambridge University Press.

[4] Červený, V. (1972). *Seismic rays and ray intensities in inhomogeneous anisotropic media*, Geophys. J. R. astr. Soc., Vol. 29, pp.1-13.

[5] Dahlen, F. A., A. M. Baig (2002). *Fréchet kernels for body-wave amplitudes*, Geophys. J. Int., Vol.150, pp.440-466.

[6] Fichtner, A., H.-P. Bunge, and H. Igel (2006). *The adjoint method in seismology: I - Theory*, Physics of The Earth and Planetary Interiors, 157(1-2), pp.86-104.

[7] Fichtner, A., H.-P. Bunge, and H. Igel (2006). *The adjoint method in seismology: II - Applications: traveltimes and sensitivity functionals*, Physics of The Earth and Planetary Interiors, 157(1-2), pp.105-123.

[8] Fichtner, A., H. Igel (2008). *Sensitivity densities for rotational ground motion measurements*, Bull. Seism. Soc. Am. (accepted.).

[9] Igel, H., U. Schreiber, A. Flaws, B. Schuberth, A. Velikoseltsev, A. Cochard (2005). *Rotational motions induced by the M8.1 Tokachi-oki earthquake, September 25, 2003*, Geophys. Res. Letters, Vol.32, L08309.

[10] Igel, H., J. Wassermann, A. Cochard, A. Flaws, U. Schreiber, A. Velikoseltsev, N. P. Dinh (2007). *Broad-band observations of earthquake-induced rotational ground motions*, Geophys. J. Int., Vol.168(1), pp.182-197.

[11] Johnson, L. R. (1974). *Green's function for Lamb's problem*, Geophys. J. R. astr. Soc., Vol. 37, pp.99-131.

[12] Jost, J. (1998). *Partielle Differentialgleichungen*, Springer.

[13] Kantorowitsch, L. W., G.P. Akilow (1964).*Funktionalanalysis in normierten Räumen*, Akademie-Verlag, Berlin.

[14] Königsberger, K. (2004). *Analysis 2*, 5.Auflage, Springer.

[15] Lay, T., T. C. Wallace (1995). *Modern global seismology*, Academic Press.

[16] Schwarz, H. R., N. Köckler (2006). *Numerische Mathematik*, 6.Auflage, Teubner.

[17] Tarantola, A. (1984). *Inversion of seismic reflection data in the acoustic approximation*, Geophysics, Vol.49, No. 8, pp.1259-1266.

[18] Tromp, J., C. Tape, Q. Liu (2005). *Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels*, Geophys. J. Int., Vol.160, pp.195-216.

[19] Yomogida, K. (1992). *Fresnel zone inversion for lateral heterogeneities in the Earth*, Pure Appl. Geophys., Vol.138 (3), pp.391-406.

Selbständigkeitserklärung

Hiermit erkläre ich, Moritz Bernauer, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel und Quellen angefertigt habe.

München, den 30. Januar 2009

Moritz Bernauer