GEOLOGICA ULTRAIECTINA

Mededelingen van de
Faculteit Geowetenschappen
Universiteit Utrecht

No. 234

Seismic ray fields and ray field maps: theory and algorithms

Dirk Kraaijpoel

promotor : Prof. Dr R.D. van der Hilst
Faculty of Geosciences
Utrecht University, The Netherlands
and:
Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology, Cambridge MA, USA

copromotoren : Prof. Dr R.K. Snieder
Department of Geophysics
Colorado School of Mines, Golden CO, USA

Dr K. Roy-Chowdhury
Faculty of Geosciences
Utrecht University, The Netherlands

Members of the dissertation committee:

Dr P.M. Bakker	Shell Technology EP Rijswijk The Netherlands
Prof. Dr D. Gajewski	Institut für Geophysik Universität Hamburg Germany
Dr G. Lambaré	Centre de Recherche en Géophysique École des Mines de Paris France
Prof. Dr I. Pšenčík	Geophysical Institute Czech Academy of Sciences, Prague Czech Republic
Prof. Dr Ir C.P.A. Wapenaar	Department of Applied Earth Sciences Delft University of Technology The Netherlands

Seismic ray fields and ray field maps: theory and algorithms

Seismische stralenvelden en stralenveldafbeeldingen: theorie en algoritmen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT UTRECHT, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR W.H. GISPEN, INGEVOLGE HET BESLUIT VAN HET COLLEGE VOOR PROMOTIES IN HET OPENBAAR TE VERDEDIGEN OP DONDERDAG 16 OKTOBER 2003 DES OCHTENDS TE 10:30 UUR

DOOR

Dirk Anton Kraaijpoel

Geboren op 28 april 1974, te Sassenheim

promotor : Prof. Dr R.D. van der Hilst
Faculteit Geowetenschappen
Universiteit Utrecht
and:
Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology, Cambridge MA, USA

copromotoren : Prof. Dr R.K. Snieder
Department of Geophysics
Colorado School of Mines, Golden CO, USA

Dr K. Roy-Chowdhury
Faculteit Geowetenschappen
Universiteit Utrecht

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

ISBN 90-5744-093-8

Contents

Chapter 1

Introduction and outline

1.1 Context

Seismic waves are widely used to study the earth's interior by means of measurements on or close to its surface. Global seismology is concerned with features along the entire depth range of the earth and therefore relies almost exclusively on the high energy waves generated by natural earthquakes. For shallow depths it is feasible to use artificial sources, such as explosives or vibrators, and to set up controlled seismic experiments. Depending on the types of sources and recording equipment, seismic experiments can provide information on subsurface properties at depths ranging from a few metres for engineering purposes, down to the deep crust and upper mantle.

Reflection seismics

The work described in this thesis is relevant for seismic exploration for oil and gas. Hydrocarbon reservoirs are typically found in sedimentary basins, at depths of the order of a few kilometres. The most commonly applied and successful experiment in this geological setting is the *seismic reflection experiment*, in which both sources and receivers are spread out on the surface. The success of the reflection experiment is primarily due to the typically layered structures of sedimentary basins, which reflect (scatter) seismic waves generated at the source back to the surface, where they are recorded by the receivers.

Despite the relative placidity of the environments in which sedimentation usually takes place, sedimentary basins can exhibit very complex structures. These are usually caused by later tectonic events such as folding, faulting and intrusion of salt or basalt. Complex structures are often the most interesting part of such a subsurface because these may contain structural traps that give rise to accumulation of hydrocarbons. The processing and interpretation of seismic data in complex structures are difficult and remain a challenge.

The seismic inverse problem

The general goal of any seismic method is to obtain quantitative information on certain subsurface properties. In reflection seismics, the desired information has for a long time been mostly geometric, in the form of a structural image. Nowadays, however, a certain amount of information regarding physical properties is usually required as well. This information may, often in combination with geological data obtained in boreholes, provide clues regarding rock types and relevant parameters such as porosity, permeability and fluid content.

In mathematical terminology the problem of transforming the seismic measurements, *the data*, into a set of quantitative subsurface properties – *the model* (parameters) – is called an *inverse problem*. As indicated by its name, the inverse problem has a counterpart: the *forward problem*. This problem concerns the expression of data in terms of the model parameters by means of a mathematical model for the physics involved in the experiment. The methodology used for solving the inverse problem relies strongly on the formulation of the forward problem.

For most practical situations in seismics the earth may be considered to be an elastic medium. The elastic wave equation should therefore be an adequate mathematical model for the forward problem. In general, however, the relationship between a solution of the wave equation (the measured data) and its coefficients (the model parameters) is strongly non-linear. This non-linearity makes the inverse problem very hard to be tackled directly.

A common approach to addressing such non-linear inverse problems is by means of perturbations. The model parameters sought are then defined to be small perturbations to a known (reference) model. The forward problem can then be linearised, thereby facilitating the corresponding inverse problem. In reflection seismics a practical approach is to define a smooth reference model and consider sharp features, such as interfaces and faults, as the perturbations to be inverted for. Some common terms that are used for this kind of inversion are *migration*, *imaging*, and *inverse scattering*. The latter name relates to the interpretation that the sharp features act as scatterers of the wave field propagating in the smooth reference model.

The most challenging aspect of the inversion of seismic reflection data is, probably, to determine the smooth reference model for which the linearised inverse scattering procedure may give a reliable result. This procedure is also known as *velocity analysis* and is typically done by means of *tomography*. Whereas the inverse scattering is usually a one step procedure, the tomographic inversion requires iteration. The quality of updates in the reference model is usually estimated by means of a measure of the coherency (focusing) of the result of the inverse scattering.

Both inverse scattering and tomographic inversion rely on many forward calculations of wave propagation in the background medium. These forward calculations can be performed in a number of ways, for example by finite difference or spectral

methods. Some of the most popular forward modelling methods are based on *ray theory*. This theory uses a high frequency approximation of the wave equation and therefore has a limited validity. Nevertheless, its limitations are relatively well understood, and so are those of the inverse methods based on it. Moreover, ray methods are generally very efficient compared to the alternative modelling methods.

References

A number of general references for the context of this thesis are recommended: Lee et al. (2002) and Aki and Richards (2002) on global seismology, Sheriff and Geldart (1995) on reflection seismics, Brouwer and Helbig (1998) on engineering seismics, Duff (1993) on physical geology, Tarantola (1987) on inverse problems, Bleistein et al. (2001) on imaging, and Červený (2001) on seismic ray theory.

1.2 Motivation

One of the characteristic features of ray methods is that the calculations are not performed directly in terms of the spatial coordinates of the medium. Although the seismic ray equations are derived from the elastic wave equation they do not share the same computational domain. The wave equation is a partial differential equation for displacement as a function of the spatial coordinates. The kinematic ray equations, on the other hand, are equations for the ray position as a function of a single parameter such as time or arc length, referred to as the *flow parameter*. The ray described by these equations is a flow line of wave energy in the high frequency approximation.

To study an entire wave field it is necessary to consider an ensemble of rays, or *ray field*, parameterised by the *ensemble parameters*. In the case of a point source, for example, these parameterise the initial ray directions at the source location. Together, the flow parameter and the ensemble parameters form an internal coordinate system for the ray field and are referred to as *ray field coordinates*.

The ray-theoretical wave field is parameterised by travel time and amplitude functions that are determined – along with the spatial coordinates – as a function of the ray field coordinates. To evaluate the travel time and amplitude at a given spatial location one needs to know its corresponding ray field coordinates. In other words, one needs to evaluate the mapping from spatial to ray field coordinates: the *ray field map*.

If the medium is sufficiently complex for the ray field to develop caustics and multi-pathing of rays, the ray field map becomes multi-valued. This multi-valuedness is a source of many practical problems in the application of ray methods to both forward and inverse wave propagation problems. In the forward calculations, for example, it is difficult to get an unambiguous and accurate estimate of all arrivals. This is especially the case in the neighbourhood of caustics, where

the number of arrivals changes abruptly. In inverse methods multi-valued ray field maps are cumbersome to work with. The interpolation of multi-valued maps, for example, is a notorious problem.

In this thesis a number of new approaches to the calculation of ray fields and ray field maps are presented. The central theme is the solution of the practical problems encountered in smooth but *complex media*, i.e., media that give rise to wave front folding and associated multi-pathing. The ultimate aim of the presented material is to enhance the efficiency of seismic inverse methods, by enhancing the efficiency of the forward calculations. Therefore, particular emphasis is placed on the applicability of the results to seismic inverse methods.

1.3 Outline

The essential background in seismic ray theory is provided in **Chapter 2**. The emphasis is on explanation and interpretation of the basic ingredients, rather than on derivation of equations, which are well covered in the existing literature. The concepts of ray fields and ray field maps are introduced as well.

In **Chapter 3** a novel approach to the calculation and representation of ray field maps is introduced that is particularly useful in cases where ray field maps are needed for a dense distribution of sources at an acquisition surface. This is the case in most seismic experiments such as reflection seismics and borehole tomography. For such source distributions it is suggested to construct a single ray field map in an extended space of spatial coordinates and angles, rather than a number of maps in the spatial domain for a range of acquisition coordinates.

A ray field map in the position/angle domain is single-valued, regardless of the complexity of the medium. The ray field information is organised by angles at depth rather than by points of emergence at the surface, which makes the maps particularly suitable for use in modern seismic imaging methods. It is shown that to calculate these maps it is not necessary to trace rays up towards the acquisition surface, which would involve an unacceptable increase in the computational burden. Instead, existing algorithms that trace downwards can be adapted to work in the position/angle domain.

Interpolation is an important tool in both the construction and the application of ray field maps. A new technique for accurate interpolation using derivative information is presented in **Chapter 4**. It is a hybrid of extrapolation to arbitrary order and linear interpolation, and combines the advantages of both methods. Through a modification of the coefficients of the Taylor expansion, extrapolations from a number of locations can be combined to obtain a polynomial order of accuracy that is one higher than that of a single conventional Taylor expansion.

The formulation of the method is very general, and it can be used both with regular and irregular data distributions in arbitrary dimensional spaces. In regular grids it is possible to use finite difference estimates of derivatives if these are not available independently. The interpolation technique is expected to be useful in

many applications and is used at various locations in this thesis.

In **Chapter 5** a ray field construction and mapping algorithm is developed that extends and refines existing wave front construction methods. A modular setup and a hierarchical description of the geometrical structure of the ray field make the algorithm widely applicable. It can be used for the calculation of ray fields and ray field maps in smooth 2-D and 3-D, isotropic and general elastic media.

For ray field mapping in the spatial domain two refinements are proposed that enhance the accuracy and the completeness of the maps by higher order interpolation and improved delineation of caustics. Both refinements are easily included in existing wave front construction methods to enhance the efficiency.

Motivated by the success of wave front construction methods in the spatial domain, the applicability in the position/angle domain is investigated as well. The unfortunate conclusion is that ray field construction in its current form is not suitable for that domain, due to the type of deformation in the geometrical structure of the ray field. In the position/angle domain this deformation is primarily shear-like whereas the algorithm is designed for spreading ray fields in the spatial domain.

A better algorithm for the calculation of ray field maps in the position/angle domain is developed in **Chapter 6**. It is based on the observation that in the position/angle domain the ray field maps are single-valued and that the geometrical spreading is very limited. This implies that the two most important reasons for developing wave front construction methods in the spatial domain are absent in the position/angle domain. Instead, it is possible to use the more primitive – but more efficient – paraxial ray methods.

The one-to-one mapping between position/angle coordinates and ray field coordinates can be exploited in practical applications. Calculations that are typically performed in terms of ray field coordinates can now be performed in terms of position/angle coordinates and the other way around. **Appendix A** shows that this may be advantageous in tomography. If the ray field map is known for a reference model, the cost function gradient can easily be calculated for an arbitrary parameterisation of the model perturbation, using the new concept of a cost function sensitivity kernel.

The change of coordinates from ray field to position/angle coordinates can be exploited even further. In **Appendix B** it is shown how the theory of Chapter 3 may be used to derive equations for the evolution of the ray field coordinates in terms of the position/angle coordinates. These equations may be used as the basis for a finite difference algorithm that calculates the full ray field information for a range of sources directly on a grid in the position/angle domain. This procedure avoids both the explicit mapping step that is usually required after ray tracing and the interpolation of medium properties at arbitrary spatial locations.

Finally, **Appendix C** presents an algorithm for the calculation of ray fields in smooth 2-D media, using a pseudo-spectral expansion of the wave front. This line of research was abandoned in favour of the ray field map methods described

above. Nevertheless, it is presented here because its development provided useful insights for the ray field map approach (e.g., Appendix B) and some of its features may be useful in other applications.

Chapter 2

Seismic ray theory

This chapter provides a background on seismic ray theory. The emphasis is placed on explanation and interpretation of the basic ingredients, rather than on the derivation of equations, which are well covered in existing literature. The most complete account of seismic ray theory available is the monograph by Červený (2001). A compact review of ray theory in general elastic media is given in Chapman (2002).

2.1 Introduction

Seismic ray theory borrows much of its principles and methodology from electromagnetic ray theory, or *geometric optics* as it is usually called (Born and Wolf, 1980; Kline and Kay, 1965). There are two major approaches to the derivation of the ray equations. The first, classical approach is based entirely on heuristic geometric principles and is therefore called *geometric ray theory*. The second approach derives the ray equations from the wave equation by means of asymptotic analysis. This approach is referred to as *asymptotic ray theory*.

Geometric ray theory has played a prominent role in the history of natural science, as a practical theory to explain the propagation of light. Both the rectilinear propagation of light and the law of reflection were described by Euclid already around 300 BC. The most important steps were made in the seventeenth century with Snell's law of refraction and Fermat's principle of least time. The subsequent statement of similar geometrical principles in classical mechanics finally led to an extensive mathematical theory associated with the names of Euler, Lagrange, Jacobi and Hamilton (e.g., Lanczos, 1986; Goldstein, 1980).

Throughout the development of geometric ray theory it was understood that the theory was limited and that it could not explain a large number of phenomena observed in experiments with light. After Maxwell had derived the much more complete wave theory of light, efforts began to reconcile the ray and wave theories.

These efforts led to the development of asymptotic ray theory, in which the ray equations are derived from the wave equation by a high frequency asymptotic analysis.

Apart from establishing ray theory as a high frequency asymptotic approximation, the asymptotic analysis also extended ray theory to explain a wider range of wave phenomena. Whereas geometrical ray theory formally explains only ray paths and travel times, asymptotic ray theory also provides a theoretical basis for the calculation of amplitudes and polarisation. Moreover, ray theory in anisotropic media is much easier to derive from asymptotic analysis than from geometric principles (although it is possible to do so, see Červený, 2002). Seismic ray theory in general elastic media has been based on high frequency asymptotics from its inception (Babich, 1994; Červený, 1972).

The derivation of seismic ray theory from the elastic wave equation relies on three essential analytical techniques: first, the definition of a suitable *ansatz* to the solution of the wave equation; second, a *high frequency asymptotic analysis* to determine the equations for the coefficients of the ansatz; and third, solution of these equations using the *method of characteristics*. These three analytical methods are covered in Sections 2.2 through 2.4. The calculation of paraxial rays and ray perturbations is discussed in Section 2.5; the calculation of amplitudes and Green functions in Section 2.6. Finally, some expressions for simplified ray tracing systems in isotropic and acoustic media are given in Section 2.7.

2.2 Ray theory ansatz

Elastic wave equation

The basic assumption underlying all of the following is that the physical properties of the media under consideration are adequately described in terms of linear elastodynamics. In that case the combination of Newton's equation of motion and Hooke's law of linear elasticity (both generalised to continuous media) yields the elastic wave equation. In Cartesian coordinates and in the frequency domain it is expressed as

$$\rho\omega^2 u_i + \frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial u_k}{\partial x_l}\right) = -f_i, \tag{2.1}$$

with coordinates $\boldsymbol{x} = (x_1, x_2, x_3)$, angular frequency ω, mass density $\rho = \rho(\boldsymbol{x})$, displacement field $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, \omega)$, elastic tensor $c_{ijkl} = c_{ijkl}(\boldsymbol{x})$, and external force field $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x}, \omega)$.

Details of the theory behind linear elastodynamics can be found in the literature (e.g., Červený, 2001; Aki and Richards, 2002). With regard to the elastic tensor it is important to note its symmetries:

$$c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}, \tag{2.2}$$

which are often silently made use of in the derivations.

Solutions to a wave equation with an arbitrary driving force are often expressed as a representation integral in terms of a Green function:

$$\rho\omega^2 G_{in}(\boldsymbol{x},\boldsymbol{y},\omega) + \frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial G_{kn}(\boldsymbol{x},\boldsymbol{y},\omega)}{\partial x_l}\right) = -\delta_{in}\delta(\boldsymbol{x}-\boldsymbol{y}), \tag{2.3}$$

$$u_i(\boldsymbol{x},\omega) = \int_D G_{in}(\boldsymbol{x},\boldsymbol{y},\omega) f_n(\boldsymbol{y},\omega) d\boldsymbol{y}, \tag{2.4}$$

where the Green function G_{in} is a second order tensor due to the vectorial character of both displacement and driving force. Domain D includes at least the support of $\boldsymbol{f}$. The task of ray methods is usually to determine the Green function for a given medium and source location. More general solutions can then be obtained using Equation (2.4).

Asymptotic ray theory ansatz

The first step in the analysis is the definition of an *ansatz* to the solution of the wave equation (2.1). The asymptotic ray theory (ART) ansatz is a polynomial series in inverse powers of ω:

$$\boldsymbol{u}(\boldsymbol{x},\omega) \sim \left(\sum_{n=0}^{\infty} \frac{1}{(i\omega)^n}\boldsymbol{U}^{(n)}(\boldsymbol{x})\right) e^{i\omega T(\boldsymbol{x})}. \tag{2.5}$$

Here, the ansatz is parameterised by a series of vectorial amplitude coefficients $\boldsymbol{U}^{(n)}(\boldsymbol{x})$ and a phase function $\omega T(\boldsymbol{x})$, where the symbol T is introduced in anticipation, as it will stand for travel time later on. This series is a high frequency asymptotic series, which may loosely be interpreted as an (in general non-convergent) Taylor series expansion in ω^{-1} around $\omega = \infty$. In most practical situations only the zeroth-order term of the series is retained (see Červený, 2001, for exceptions):

$$\boldsymbol{u}(\boldsymbol{x},\omega) \sim \boldsymbol{U}(\boldsymbol{x}) e^{i\omega T(\boldsymbol{x})}, \tag{2.6}$$

with amplitude coefficient $\boldsymbol{U} \equiv \boldsymbol{U}^{(0)}$. This is the *zeroth-order asymptotic ray theory* (ZART) ansatz.

It is outside the scope of this thesis to digress on the deeper meaning of asymptotic expansions in general or why the particular series (2.5) is used to develop ray theory (see, e.g., Bender and Orszag, 1978; Bleistein and Handelsman, 1987; Kline and Kay, 1965). It is, however, useful to make the following interpretative remarks.

First, the functional forms of (2.5) and (2.6), i.e. an amplitude times an oscillatory term, are a generalisation of the known exact solutions in homogeneous media, such as plane and spherical waves. The forms are generalised in the sense that both phase and amplitude terms can be arbitrary functions of position. In

the series ansatz (2.5) the amplitude has a dependence on frequency as well. Note that both (2.5) and (2.6) describe a progressing wave field with wave fronts at surfaces of equal $T(\boldsymbol{x})$.

The basic premise behind this choice of functional form is that if the wavelength is relatively short with respect to variations in the medium, the medium locally homogeneous, and the wave field propagation should locally resemble that in a homogeneous medium. The ultimate justification of this premise should follow from the asymptotic analysis of the wave equation described below.

Second, a Fourier transformation of ZART ansatz (2.6) to the time domain yields a Dirac delta pulse: $\boldsymbol{U}(\boldsymbol{x})\delta(t-T(\boldsymbol{x}))$. The additional higher order terms in (2.5) yield progressively higher order integrals of the Dirac delta pulse, and hence increasingly smooth contributions. The ZART ansatz is therefore often said to represent the *most singular* part of the wave field.

Third, according to the online Oxford English Dictionary[1] an ansatz is a "mathematical solution, esp. about the form of an unknown function, which is made in order to facilitate solution of an equation or other problem". In asymptotic ray theory the solution of the wave equation is simplified by a reduction of the degrees of freedom in the solution. In ZART the amplitude and phase functions to be determined do not depend on frequency at all, while in the more general ART at most a small number of amplitude terms are expected to contribute. Moreover, the amplitude and phase functions are expected to be slowly varying quantities compared to the highly oscillatory displacement field. In practical applications this allows for a sparser set of evaluation points.

Finally, it should be mentioned that (2.5) and (2.6) are merely ansätze for elementary solutions of the wave equation. A final solution usually requires multiplication with a frequency dependent factor, say $F(\omega)$, representing a finite bandwidth waveform. Moreover, a general asymptotic wave field may be a summation of several of such contributions, or even an integral, depending on the external force field (e.g., $\boldsymbol{f}$ in Eq. 2.1), to be modelled.

2.3 Asymptotic analysis of the wave equation

By choosing an ansatz with reduced degrees of freedom – to facilitate solution of the wave equation – the prospect of obtaining exact solutions in general circumstances is lost. Instead, the desired solution is to be close to the exact solution, in some specific sense.

In asymptotic ray theory the goal is to make the ansatz fit the wave equation in an asymptotic sense, i.e. in the limit of high frequency. To this end asymptotic analysis is applied to the wave equation, wherein the solution is assumed to attain the form of the ZART ansatz (2.6). This analysis leads to equations for travel time T and amplitude $\boldsymbol{U}$ that, when satisfied, make the ZART ansatz fit the wave equation asymptotically.

[1] `http://dictionary.oed.com`

Initially, solutions are sought that satisfy the wave equation (2.1) away from any external force field, i.e., $\boldsymbol{f} \equiv 0$; explicit sources will be taken into account in Section 2.6. Substituting ansatz (2.6) in (2.1), recombining terms, and division by $\rho(i\omega)^2$ yields:

$$\begin{aligned} &\left[\frac{c_{ijkl}}{\rho}\frac{\partial T}{\partial x_j}\frac{\partial T}{\partial x_l}U_k - U_i\right] + \\ (i\omega)^{-1}&\left[\frac{c_{ijkl}}{\rho}\frac{\partial T}{\partial x_j}\frac{\partial U_k}{\partial x_l} + \frac{1}{\rho}\frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial T}{\partial x_l}U_k\right)\right] + \\ (i\omega)^{-2}&\left[\frac{1}{\rho}\frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial U_k}{\partial x_l}\right)\right] = 0. \end{aligned} \tag{2.7}$$

To satisfy this equation for arbitrary ω each of the three terms within square brackets should vanish. Clearly, in general circumstances this cannot be accomplished, because the total number of equations is nine (three components in each term), while there are only four degrees of freedom in the ansatz (T and three components of $\boldsymbol{U}$). The ordering of terms in inverse powers of frequency in Equation (2.7) shows their relative importance in an asymptotic analysis. The first term in brackets is the most important to be satisfied. As will become clear in the following, this term constrains both T and the direction (polarisation) of $\boldsymbol{U}$. Satisfying this first term, however, is not sufficient to obtain an asymptotic solution for the wave equation (e.g., Bender and Orszag, 1978, Section 10.2). To this end, also the magnitude (scalar amplitude) of $\boldsymbol{U}$ must be determined. A relation for this magnitude will be obtained from the second term of Equation (2.7).

Definitions

In the following it is useful to introduce a number of symbols, viz. the slowness vector $\boldsymbol{p}$, defined as the gradient of travel time:

$$\boldsymbol{p} = \frac{\partial T}{\partial \boldsymbol{x}}, \tag{2.8}$$

the density-normalised elastic tensor a_{ijkl}:

$$a_{ijkl} = \frac{c_{ijkl}}{\rho}, \tag{2.9}$$

and the Christoffel matrix Γ_{ij}:

$$\Gamma_{ik} = a_{ijkl}p_j p_l. \tag{2.10}$$

It will also prove useful to parameterise the amplitude $\boldsymbol{U}$ by a scalar amplitude A and a unit polarisation vector $\boldsymbol{g}$:

$$\boldsymbol{U} = A\boldsymbol{g}, \qquad |\boldsymbol{g}| = 1, \tag{2.11}$$

and, similarly, the slowness vector by (scalar) phase velocity V and wave front normal $\boldsymbol{n}$:

$$\boldsymbol{p} = \frac{1}{V}\boldsymbol{n}, \qquad |\boldsymbol{n}| = 1. \tag{2.12}$$

An alternative Christoffel matrix $\tilde{\Gamma}_{ik}$ may be defined in terms of $\boldsymbol{n}$:

$$\tilde{\Gamma}_{ik} = a_{ijkl} n_j n_l. \tag{2.13}$$

Using (2.12) the relationship between the two Christoffel matrices can be established:

$$\Gamma_{ik} = V^{-2} \tilde{\Gamma}_{ik}. \tag{2.14}$$

The eikonal equation

Using definitions (2.8)-(2.11) and the Kronecker symbol δ_{ik} the first term of (2.7) can be written as

$$(\Gamma_{ik} - \delta_{ik})\, g_k = 0. \tag{2.15}$$

This is an eigenvalue equation subject to the additional constraint that the eigenvalue be equal to 1:

$$(\Gamma_{ik} - G\delta_{ik})\, g_k = 0, \qquad G = 1. \tag{2.16}$$

This eigenvalue constraint basically poses a constraint on the components of Γ_{ik}, which, in turn, for fixed $\boldsymbol{x}$ and $\boldsymbol{n}$ poses a constraint on the phase velocity. Use of the alternative Christoffel matrix (2.13) and relation (2.14) puts (2.16) in the form of a classical eigenvalue problem:

$$\left(\tilde{\Gamma}_{ik} - V^2\delta_{ik}\right) g_k = 0. \tag{2.17}$$

The constraint $G = 1$ of (2.15) is replaced by the constraint $|\boldsymbol{n}| = 1$ in the definition (2.12).

For a fixed $\boldsymbol{x}$ and $\boldsymbol{n}$ the eigenvectors of $\tilde{\Gamma}_{ik}$ correspond to the three possible polarisations. Due to the symmetry of $\tilde{\Gamma}_{ik}$ these polarisations form an orthonormal set. Each corresponds to one of the three wave types, usually called qP ("quasi-P"), qS1 and qS2. The eigenvalue associated with each wave type is equal to the square of its phase velocity. The largest eigenvalue corresponds to qP.

If the phase velocities of the three wave types are distinct it is possible to derive an individual eikonal equation for each wave type. If two phase velocities coincide, as is the case for S-waves in isotropic media, special attention is required. This situation is not of particular interest for this thesis (see, e.g., Červený, 2001; Chapman, 2002).

Eigensystem (2.15) contains information on the travel times and polarisation of all three wave types. The system can be replaced by three independent travel time equations by projecting its terms onto the orthogonal set of polarisations. For a fixed wave type both terms on the left side of (2.15) are parallel to that wave type's own polarisation, and orthogonal to the other two. The projection therefore leaves only one equation that is not trivially solved, the *eikonal equation*:

$$a_{ijkl}p_j p_l g_i g_k - 1 = 0, \tag{2.18}$$

which has the same form for each wave type. It is an equation for the travel time only, as the polarisation is implicitly fixed by the choice of wave type and its corresponding slowness vector at the initial conditions.

The transport equation

Since travel time is governed by the eikonal equation (2.18), and the polarisation is implicit in the choice of wave-type, only one more equation is required to determine the coefficients of ZART ansatz (2.6).

The equation for the scalar amplitude $A(\boldsymbol{x})$, defined in (2.11), must be obtained from the second bracketed term in (2.7). It is not possible to find an $A(\boldsymbol{x})$ that makes the whole term vanish in general circumstances. The remainder has an arbitrary orientation, while variation of the amplitude function can only influence the component of this remainder along the polarisation direction.

The best result is therefore obtained by making the remainder perpendicular to the polarisation, which in this context is also called the *principle direction* (e.g., Chapman, 2002). Taking the inner product of the remainder with the polarisation direction and equating it to zero yields the transport equation. With the help of definitions (2.8-2.11) it can be expressed as

$$a_{ijkl}p_j g_i \frac{\partial(Ag_k)}{\partial x_l} + \frac{1}{\rho}\frac{\partial}{\partial x_j}\left(\rho A a_{ijkl}p_l g_k\right) g_i = 0. \tag{2.19}$$

With some rearrangement of terms (e.g., Červený, 2001) it can be reformulated as

$$\frac{\partial}{\partial x_i}\left(\rho A^2 a_{ijkl}p_l g_j g_k\right) = 0, \tag{2.20}$$

which has the form of a conservation law. Its solution will be analysed in Section 2.6.

Solving the eikonal and transport equations

The eikonal equation (2.18) is a non-linear partial differential equation for travel time. The approaches that are used to solve the equation can generally be categorised as ray methods on the one hand, and eikonal solvers on the other.

The ray methods, based on the method of characteristics, represent the classical approach. They have the general advantage, as will be shown in Section 2.4, that the partial differential equation is replaced by a set of ordinary differential equations, which are generally easier to solve. Moreover, it turns out that the transport equation (2.20) is easily solved in terms of rays as well.

A disadvantage of the ray methods, however, is that they involve a change of coordinates – in fact a central theme of this thesis – which has its computational consequences. One example is the evaluation of the medium properties. In ray methods these have to be evaluated at arbitrary positions in the medium, which, in practice, requires interpolation. As this interpolation has to be smooth, it is relatively expensive, and it turns out that in ray methods a considerable amount of computation time is spent on interpolating medium properties.

The name eikonal solvers is used for a group of methods that evaluate the eikonal equation directly in terms of the spatial coordinates, incorporating finite difference (FD) methods (e.g., Vidale, 1988, 1990; Podvin and Lecomte, 1991; van Trier and Symes, 1991), fast marching (FM) methods (e.g., Sethian, 1999; Sethian and Popovici, 1999), and level set (LS) methods (e.g., Sethian, 1999).

By evaluating the eikonal equation directly in terms of the spatial coordinates these methods bypass the change of coordinates inherent to ray methods. Thus they avoid the interpolation of medium properties and other problems associated with the change of coordinates discussed extensively in this thesis. These methods, however, face a number of other theoretical and practical difficulties, primarily associated with stability (FD) and accuracy (FM, LS). Extension to general anisotropic media is problematic, as is the accurate calculation of amplitudes.

Another important issue in solving the eikonal equation is its non-linearity. Even starting from perfectly regular initial conditions its solution may develop singularities. Depending on the type of solution method these appear in the form of shocks or multi-valuedness. If the equation is solved using the method of characteristics, as in Section 2.4, multi-valued solutions may be obtained. These correspond physically to wave front folding or triplications. Eikonal solvers usually do not allow multi-valued solutions.

Although research is going on challenging all of the above issues (e.g., Kim, 2002; Qian and Symes, 2002; Osher et al., 2002; Qian et al., 2003), it seems that in the context of this research, where both multi-pathing occurs and the evaluation of amplitudes is essential, the eikonal solvers do not yet provide an alternative to ray methods.

2.4 Kinematic ray equations

The third essential step of asymptotic ray theory, after the definition of the ansatz and the asymptotic analysis (Sections 2.2 and 2.3), is the solution of the eikonal equation using the method of characteristics. It is customary to use a formulation in terms of Hamiltonians, to make the connection with the extensive mathematical

theory available for Hamiltonian systems.

Hamiltonian formulation

In the Hamiltonian context the eikonal equation for a single wave type is summarised in terms of the Hamiltonian function $\mathcal{H}$:

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) = 0. \tag{2.21}$$

The eikonal equation may be regarded as a nonlinear partial differential equation for T, or as an algebraic equation in the phase space coordinates $\boldsymbol{x}$ and $\boldsymbol{p}$. The latter view is central to the Hamiltonian formalism.

The kinematic ray equations are derived from the eikonal equation using the method of characteristics (e.g., Bleistein, 1984). The equations thus obtained are the Hamiltonian equations of motion, or the *kinematic ray equations*:

$$\frac{d\boldsymbol{x}}{d\sigma} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} \tag{2.22}$$

$$\frac{d\boldsymbol{p}}{d\sigma} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}} \tag{2.23}$$

$$\frac{dT}{d\sigma} = \boldsymbol{p} \cdot \frac{d\boldsymbol{x}}{d\sigma} \tag{2.24}$$

The parameter σ is referred to as the flow parameter. The explicit dependence of the ray's phase space coordinates $(\boldsymbol{x}, \boldsymbol{p})$ on σ is often suppressed.

In addition it is sometimes useful to define

$$\boldsymbol{v} = \frac{d\boldsymbol{x}}{dT} = \left(\frac{dT}{d\sigma}\right)^{-1} \frac{d\boldsymbol{x}}{d\sigma}, \tag{2.25}$$

the *group velocity* of the ray.

For the travel time equation (2.24) an alternative equation may be used, that stems from geometrical ray theory and its analogy with the variational methods of classical mechanics (e.g., Lanczos, 1986). The travel time in the ray tracing system plays the role of the action in classical mechanics. The travel time derivative along the ray may therefore be identified as the Lagrangian $\mathcal{L}(\boldsymbol{x}, \boldsymbol{p})$. It may be expressed in terms of the ray variables and the Hamiltonian:

$$\frac{dT}{d\sigma} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{p} \cdot \frac{d\boldsymbol{x}}{d\sigma} - \mathcal{H}(\boldsymbol{x}, \boldsymbol{p}). \tag{2.26}$$

In comparison with (2.24) the Hamiltonian term is included. For an exact ray satisfying the eikonal equation this does not make a difference, because its value is zero (2.21). The presence of the Hamiltonian simplifies the derivation of travel time perturbations later on.

Elastic ray equations

The ray equations (2.22-2.24) can be made more explicit for the eikonal equation (2.18). The Hamiltonian is chosen to read

$$\mathcal{H}(\boldsymbol{x},\boldsymbol{p}) = \frac{1}{2}\left(a_{ijkl}p_j p_l g_i g_k - 1\right), \tag{2.27}$$

which leads to the ray equations (2.22)-(2.24):

$$\frac{dx_i}{d\sigma} = a_{ijkl}p_l g_j g_k \tag{2.28}$$

$$\frac{dp_i}{d\sigma} = -\frac{1}{2}\frac{\partial a_{jkln}}{\partial x_i}p_k p_n g_j g_l \tag{2.29}$$

$$\frac{dT}{d\sigma} = a_{ijkl}p_j p_l g_i g_k = 1. \tag{2.30}$$

From the last equation it is clear that, for this choice of Hamiltonian, the flow parameter σ is equal to the travel time ($\sigma \equiv T$). Hence the group velocity (2.25) may be expressed as

$$v_i = a_{ijkl}p_l g_j g_k. \tag{2.31}$$

Despite its correspondence with T in this case, the flow parameter σ is maintained, mainly because it is useful to separate the two in perturbation methods, to be discussed in Section 2.5.

Initial conditions and ray fields

To solve the ray equations they have to be supplied with initial conditions. Mathematically any vector in phase space may serve as the initial condition. To be physically relevant, however, the slowness must be chosen such that the initial condition satisfies the eikonal equation (2.21). Using (2.22) and (2.23) it is easily shown that the Hamiltonian is a constant of motion:

$$\frac{d\mathcal{H}}{d\sigma} = \frac{\partial\mathcal{H}}{\partial\boldsymbol{x}}\cdot\frac{d\boldsymbol{x}}{d\sigma} + \frac{\partial\mathcal{H}}{\partial\boldsymbol{p}}\cdot\frac{d\boldsymbol{p}}{d\sigma} = 0. \tag{2.32}$$

Therefore if the eikonal equation is satisfied for the initial conditions, it is satisfied anywhere along the ray.

To approximate an entire wave field it is necessary to consider an ensemble of rays, or *ray field*, corresponding to a continuous range of initial conditions. The initial conditions are parameterised by the *ensemble parameters*, say $\boldsymbol{\gamma}$. In the case of a point source, $\boldsymbol{\gamma}$ parameterises the initial slowness vectors at the source location. Alternatively, the initial position of the rays may vary as well, for example if a ray field emanating from an interface is modelled. Together, the flow parameter σ and the ensemble parameters form an internal coordinate system of a ray field and may therefore be called *ray field coordinates*: $(\sigma,\boldsymbol{\gamma})$.

In order for the ray field to represent a physical wave, the number of ensemble parameters must be one less than the dimension of the medium, such that the number of ray field coordinates is equal to the number of spatial coordinates. In the current chapter only such *physical ray fields* will be used. In this thesis, however, starting in Chapter 3, ensembles of rays with a higher number of ensemble parameters will be studied as well. The terms ray fields and ray field coordinates will then also be used to address those more general ray ensembles.

An important thing to note is that the travel time and amplitude functions of the ZART ansatz (2.6) are in fact calculated as a function of $(\sigma, \boldsymbol{\gamma})$ rather than as a function of $\boldsymbol{x}$. The latter is calculated as a function of $(\sigma, \boldsymbol{\gamma})$ as well. This does not pose a conceptual problem as long as the mapping between $(\sigma, \boldsymbol{\gamma})$ and $\boldsymbol{x}$ is regular, i.e. as long as the jacobian of transformation $J^{(\sigma)}$, defined as

$$J^{(\sigma)} = \det\left(\frac{\partial \boldsymbol{x}}{\partial(\sigma, \boldsymbol{\gamma})}\right), \tag{2.33}$$

does not vanish. Note that the dimensions of $\boldsymbol{x}$ and $(\sigma, \boldsymbol{\gamma})$ are equal.

In sufficiently complex media, however, $J^{(\sigma)}$ vanishes on so called *caustic* points, curves, or surfaces. There, the ray solution is in conflict with its original ansatz (2.6) and is formally invalid. As shown in Section 2.6 the ray theory amplitude also becomes infinite at caustics, which is another indication that ray theory is locally invalid. After passing through a caustic $J^{(\sigma)}$ usually changes sign but becomes finite again, leading to regular solutions. It is not self-evident that the ray solution can be continued after passing through the caustic singularity. Using extensions of ray theory such as Maslov theory, however, it can be shown that this is correct if a proper phase shift is taken into account, see also Chapman (2002) and Section 2.6.

2.5 Paraxial rays and ray perturbation theory

In many applications one is not only interested in the ray path, but also in the dependence of this ray path on small variations in some parameter. Such a parameter may for example specify an initial condition (paraxial ray theory), or a variation of the medium properties (ray perturbation theory).

Perturbed ray equations

For a certain parameter γ the ray path may be described as a function of two parameters: $\boldsymbol{x} = \boldsymbol{x}(\sigma, \gamma)$, and the same holds for the slowness $\boldsymbol{p}$ and the travel time T. For a small perturbation $\Delta\gamma$ the dependence of the ray path may be linearised around γ:

$$\boldsymbol{x}(\sigma, \gamma + \Delta\gamma) = \boldsymbol{x}(\sigma, \gamma) + \frac{\partial \boldsymbol{x}(\sigma, \gamma)}{\partial \gamma}\Delta\gamma + O(\Delta\gamma^2). \tag{2.34}$$

In the following the symbol δ will be used as a differential operator:

$$\delta = \Delta\gamma \frac{\partial}{\partial \gamma}. \tag{2.35}$$

A differential equation for the first order ray perturbation $(\delta \boldsymbol{x}, \delta \boldsymbol{p})$ may be found from Equations (2.22) and (2.23) through the application of δ:

$$\frac{d}{d\sigma}\begin{pmatrix}\delta \boldsymbol{x} \\ \delta \boldsymbol{p}\end{pmatrix} = \begin{pmatrix}\partial_{\boldsymbol{p}}\partial_{\boldsymbol{x}}\mathcal{H} & \partial_{\boldsymbol{p}}\partial_{\boldsymbol{p}}\mathcal{H} \\ -\partial_{\boldsymbol{x}}\partial_{\boldsymbol{x}}\mathcal{H} & -\partial_{\boldsymbol{x}}\partial_{\boldsymbol{p}}\mathcal{H}\end{pmatrix} \cdot \begin{pmatrix}\delta \boldsymbol{x} \\ \delta \boldsymbol{p}\end{pmatrix} + \begin{pmatrix}\partial_{\boldsymbol{p}}\delta^*\mathcal{H} \\ -\partial_{\boldsymbol{x}}\delta^*\mathcal{H}\end{pmatrix}. \tag{2.36}$$

The operator δ^* is a (partial) differential operator equal to δ, with the exception that it treats $\boldsymbol{x}$ and $\boldsymbol{p}$ as constants. This means that the Hamiltonian vanishes under this operator, unless it depends on γ through the medium properties.

Paraxial rays and the propagator matrix

If the Hamiltonian does not depend on parameter γ explicitly ($\delta^*\mathcal{H} \equiv 0$), the inhomogeneous term of system (2.36) vanishes, and a system of linear ordinary differential equations remains. This system is called the *paraxial* or *dynamic ray tracing* system. The solution to such a homogeneous linear system may be expressed in terms of a propagator matrix $\boldsymbol{P}(\sigma, \sigma_0)$:

$$\begin{pmatrix}\delta \boldsymbol{x} \\ \delta \boldsymbol{p}\end{pmatrix}(\sigma) = \boldsymbol{P}(\sigma, \sigma_0) \cdot \begin{pmatrix}\delta \boldsymbol{x} \\ \delta \boldsymbol{p}\end{pmatrix}(\sigma_0), \tag{2.37}$$

with

$$\boldsymbol{P} = \begin{pmatrix}\boldsymbol{P}_{\boldsymbol{xx}} & \boldsymbol{P}_{\boldsymbol{xp}} \\ \boldsymbol{P}_{\boldsymbol{px}} & \boldsymbol{P}_{\boldsymbol{pp}}\end{pmatrix} = \frac{\partial(\boldsymbol{x}, \boldsymbol{p})}{\partial(\boldsymbol{x}_0, \boldsymbol{p}_0)}, \tag{2.38}$$

and $(\boldsymbol{x}_0, \boldsymbol{p}_0) = (\boldsymbol{x}, \boldsymbol{p})(\sigma_0)$. $\boldsymbol{P}$ is a 6×6 matrix (4×4 in a 2D medium), which satisfies

$$\boldsymbol{P}(\sigma_0, \sigma_0) = \boldsymbol{I}, \tag{2.39}$$

with $\boldsymbol{I}$ the identity matrix, and each column of $\boldsymbol{P}$ satisfies (2.36) with $\delta^*\mathcal{H} \equiv 0$:

$$\frac{d\boldsymbol{P}(\sigma, \sigma_0)}{d\sigma} = \begin{pmatrix}\partial_{\boldsymbol{p}}\partial_{\boldsymbol{x}}\mathcal{H} & \partial_{\boldsymbol{p}}\partial_{\boldsymbol{p}}\mathcal{H} \\ -\partial_{\boldsymbol{x}}\partial_{\boldsymbol{x}}\mathcal{H} & -\partial_{\boldsymbol{x}}\partial_{\boldsymbol{p}}\mathcal{H}\end{pmatrix} \cdot \boldsymbol{P}(\sigma, \sigma_0), \tag{2.40}$$

see, e.g., Farra and Madariaga (1987).

An important property of the propagator matrix is that its determinant is equal to unity:

$$\det \boldsymbol{P}(\sigma, \sigma_0) = 1, \tag{2.41}$$

which shows that the rays in phase space behave similar to flow lines of an incompressible fluid. The matrix also displays symplectic symmetry (e.g., Goldstein,

1980; Červený, 2001). An important practical corollary of this is that it is easily inverted:

$$\boldsymbol{P}(\sigma_0,\sigma) = \boldsymbol{P}^{-1}(\sigma,\sigma_0) = \begin{pmatrix} \boldsymbol{P}_{\boldsymbol{pp}}^T & -\boldsymbol{P}_{\boldsymbol{xp}}^T \\ -\boldsymbol{P}_{\boldsymbol{px}}^T & \boldsymbol{P}_{\boldsymbol{xx}}^T \end{pmatrix}. \tag{2.42}$$

It is usually not necessary to compute the entire propagator matrix. At most four columns are required,because the other two can be obtained analytically from the results of the kinematic ray tracing. The following relations hold for the Hamiltonian (2.27), with $\dot{\boldsymbol{x}} = d\boldsymbol{x}/dT = \boldsymbol{v}$ and $\dot{\boldsymbol{p}} = d\boldsymbol{p}/dT$:

$$\boldsymbol{P} \cdot \begin{pmatrix} \dot{\boldsymbol{x}}_0 \\ \dot{\boldsymbol{p}}_0 \end{pmatrix} = \begin{pmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{p}} \end{pmatrix}, \tag{2.43}$$

$$\boldsymbol{P} \cdot \begin{pmatrix} 0 \\ \boldsymbol{p}_0 \end{pmatrix} = \begin{pmatrix} (T-T_0)\dot{\boldsymbol{x}} \\ \boldsymbol{p} + (T-T_0)\dot{\boldsymbol{p}} \end{pmatrix}, \tag{2.44}$$

see Červený (2001, Section 4.2.1).

Propagator solution for perturbed rays

The solution to the general inhomogeneous system (2.36) may be expressed in terms of the propagator matrix:

$$\begin{pmatrix} \delta\boldsymbol{x} \\ \delta\boldsymbol{p} \end{pmatrix}(\sigma) = \boldsymbol{P}(\sigma,\sigma_0) \cdot \begin{pmatrix} \delta\boldsymbol{x} \\ \delta\boldsymbol{p} \end{pmatrix}(\sigma_0) + \int_{\sigma_0}^{\sigma} \boldsymbol{P}(\sigma,\bar{\sigma}) \cdot \begin{pmatrix} \partial_{\boldsymbol{p}}\delta^*\mathcal{H} \\ -\partial_{\boldsymbol{x}}\delta^*\mathcal{H} \end{pmatrix} d\bar{\sigma}, \tag{2.45}$$

where the first term on the right hand side expresses the propagation of an initial perturbation, and the second term expresses the perturbations induced by perturbations of the medium properties (e.g., Farra and Madariaga, 1987).

Travel time perturbations

A similar analysis for the traveltime Equation (2.24) yields

$$\frac{d\delta T}{d\sigma} = \delta\boldsymbol{p} \cdot \frac{d\boldsymbol{x}}{d\sigma} + \boldsymbol{p} \cdot \frac{d\delta\boldsymbol{x}}{d\sigma}. \tag{2.46}$$

Note that this right hand side must vanish for Hamiltonian systems that have the travel time as flow parameter ($\sigma \equiv T$, hence $\delta T = 0$).

The same analysis applied to the variational version of the travel time equation (2.26) yields:

$$\frac{d\delta T}{d\sigma} = \frac{d}{d\sigma}(\boldsymbol{p} \cdot \delta\boldsymbol{x}) - \delta^*\mathcal{H}(\boldsymbol{x},\boldsymbol{p}). \tag{2.47}$$

This equation is equivalent to (2.46) and can easily be integrated to give the travel time perturbation:

$$\delta T = [\boldsymbol{p} \cdot \delta\boldsymbol{x}]_{\sigma_0}^{\sigma_1} - \int_{\sigma_0}^{\sigma_1} \delta^*\mathcal{H}(\boldsymbol{x},\boldsymbol{p}) d\sigma. \tag{2.48}$$

This equation shows that the first order travel time perturbation δT is independent of the actual first order ray path perturbation $\delta \boldsymbol{x}$ except for the end point perturbation.

In the derivations above, no distinction was made between variations in medium properties, introduced through terms with $\delta^* \mathcal{H}$, and variations in initial conditions. The resulting perturbation equations can handle both.

In practice, it is often useful to consider both types of variations simultaneously. This is the case, for example, in travel time tomography, where the ray end points are fixed (see also Appendix A). Since a variation in the medium causes a change in the ray path, an additional perturbation is necessary to keep the ray end point fixed. Usually this requires a perturbation of the initial slowness vector.

The number of degrees of freedom in the initial slowness perturbation is limited by the requirement that the perturbed slowness vector fit the eikonal equation[2]. This means that another degree of freedom has to be found to counteract the end point perturbation caused by the medium variation. This degree of freedom is provided by a variation in the flow parameter σ (Snieder and Sambridge, 1993). The advantage of expression (2.48) is that it is not necessary to explicitly calculate these initial slowness and flow parameter perturbations. It is sufficient to know that they can be chosen in such a way that the end point perturbations of position vanish.

Restricted end point perturbations

The ray perturbation equations (2.36), and the propagator matrix (2.38) are defined for constant flow parameter σ. In practice it is often desirable to restrict the end point perturbations to a surface, in order to find a perturbed point of emergence. In general this requires a perturbation of the flow parameter σ.

Using $\dot{\boldsymbol{x}} = d\boldsymbol{x}/d\sigma$, perturbed point of emergence $\delta \tilde{\boldsymbol{x}}$ and surface normal $\boldsymbol{\nu}$, the required flow parameter perturbation $\delta\sigma$ can be determined. Perturbation $\delta \tilde{\boldsymbol{x}}$ can be expressed as the sum of an arbitrary perturbation $\delta \boldsymbol{x}$ and a perturbation along the ray direction:

$$\delta \tilde{\boldsymbol{x}} = \delta \boldsymbol{x} + \dot{\boldsymbol{x}} \delta\sigma. \tag{2.49}$$

This perturbation must be perpendicular to the surface normal:

$$\delta \tilde{\boldsymbol{x}} \cdot \boldsymbol{\nu} = 0, \tag{2.50}$$

such that

$$\delta\sigma = -\frac{\delta \boldsymbol{x} \cdot \boldsymbol{\nu}}{\dot{\boldsymbol{x}} \cdot \boldsymbol{\nu}}. \tag{2.51}$$

[2]Here, cases are neglected where additional degrees of freedom are lost due to the presence of a caustic at one of the end points of the ray. In such cases a perturbation approach with fixed end points is not feasible.

The corresponding slowness and travel time perturbations ($\delta\tilde{\boldsymbol{p}}$, $\delta\tilde{T}$) then follow naturally:

$$\delta\tilde{\boldsymbol{p}} = \delta\boldsymbol{p} + \dot{\boldsymbol{p}}\delta\sigma \tag{2.52}$$

$$\delta\tilde{T} = \delta T + \dot{T}\delta\sigma. \tag{2.53}$$

See Farra and LeBégat (1995) for more details.

2.6 Amplitudes and Green functions

Relative amplitude evolution

The evolution of the scalar amplitude A in the ZART ansatz (2.6) is governed by the transport equation (2.20). Although this is a linear ordinary differential equation for A^2 its character is complicated by the fact that the slowness vector $\boldsymbol{p}$ and polarisation vector $\boldsymbol{g}$ appear in its coefficients. Both are determined by the non-linear eikonal equation and may become multi-valued, leading to the same multiplicity in A. It turns out that the transport equation is easily solved in terms of the flow parameter along the ray.

Using (2.31) the transport equation (2.20) can be expressed in terms of the group velocity vector:

$$\nabla \cdot (\rho A^2 \boldsymbol{v}) = 0. \tag{2.54}$$

This equation can be solved in terms of the spreading of a ray field. The volume occupied by a differential element of the ray field is equal to the jacobian $J^{(\sigma)}$ of transformation from ray field coordinates $(\sigma, \boldsymbol{\gamma})$ to spatial coordinates (2.33). The normalised *ray field jacobian* replaces an arbitrary σ by T:

$$J = J^{(T)} = \left(\frac{dT}{d\sigma}\right)^{-1} J^{(\sigma)}. \tag{2.55}$$

As discussed in Section 2.4 J may change sign along a ray, which indicates crossing a caustic.

Using Smirnov's lemma (Thomson and Chapman, 1985) the divergence of the group velocity can be expressed in terms of Jacobian J:

$$\nabla \cdot \boldsymbol{v} = \frac{d}{dT} \ln J, \tag{2.56}$$

and with $\boldsymbol{v} \cdot \nabla = d/dT$ the transport equation (2.54) can be solved up to a constant multiplier $C(\boldsymbol{\gamma})$:

$$A^2 = \frac{C(\boldsymbol{\gamma})}{\rho J}, \tag{2.57}$$

where J is still allowed to have negative sign and A can in general be complex.

Absolute point source amplitudes

In order to find an absolute amplitude the constant $C(\boldsymbol{\gamma})$ in Equation (2.57) has to be determined, typically by means of an initial condition. For a point source this can be achieved by comparison with the analytical solution for the Green function in a homogeneous medium (Kendall et al., 1992; Gajewski, 1993; Pšenčík and Teles, 1996).

Following the analysis of Kendall et al. (1992) as quoted by Chapman (2002) this leads to the following expression for the scalar amplitude A:

$$A = \frac{e^{-i\frac{\pi}{2}\mathrm{sgn}(\omega)\kappa}}{4\pi\sqrt{\rho(\boldsymbol{x})\rho(\boldsymbol{x}_0)|R|}}. \tag{2.58}$$

In this expression A, κ, $\boldsymbol{x}$, and R implicitly depend on the flow parameter, say, T, and the subscript 0 denotes a quantity defined at the source point (i.e., at $T = T_0$). Variable R is defined below, and KMAH-index κ is introduced to choose the right (real or imaginary) root of R, implementing the caustic phase shift. Usually κ increases by 1 upon each zero-crossing of R, although its behaviour can be more complicated, especially in anisotropic media (e.g., Bakker, 1998; Červený, 2001).

The expression for the scalar amplitude (2.58) is valid for point source ray fields in any type of medium and does not include the radiation factor induced by the direction of the source (equal to 1 for scalar wave fields), as discussed below.

For the variable R two expressions are available (Chapman, 2002). The first is an expression in terms of J (2.55) and a radiation term D:

$$R = \frac{J}{D}, \tag{2.59}$$

with

$$D = \frac{1}{|\boldsymbol{v}_0|}\left|\frac{\partial \boldsymbol{p}_0}{\partial \gamma_1} \times \frac{\partial \boldsymbol{p}_0}{\partial \gamma_2}\right|, \tag{2.60}$$

where γ_1 and γ_2 are the ensemble parameters. This expression therefore depends on the specific ray field parameterisation, although the effects of numerator and denominator cancel.

An alternative expression can be derived in terms of the ray propagator matrix (2.38) and the group velocity $\boldsymbol{v}$, quantities that depend only on the medium properties and the source and receiver locations:

$$R = \det(\boldsymbol{P}_{\boldsymbol{xp}})(\boldsymbol{P}_{\boldsymbol{xp}}^{-1} \cdot \boldsymbol{v}) \cdot \boldsymbol{v}_0. \tag{2.61}$$

Using (2.44) and $\boldsymbol{p}_0 \cdot \boldsymbol{v}_0 = 1$ this expression can be simplified to:

$$R = \frac{\det(\boldsymbol{P}_{\boldsymbol{xp}})}{T - T_0}. \tag{2.62}$$

Note that this expression does not depend on the (arbitrary) parameterisation of the ray field.

Green functions and ray field maps

The Green function in an elastic medium for a point source with arbitrary direction is a dyadic tensor. The excitation of each ray is proportional to the projection of the force vector on the polarisation vector of the ray at the source location. The complete expression for the Green function of a single wave type at location $\boldsymbol{x}_1$ due to a point source at $\boldsymbol{x}_0$ reads

$$G_{in}(\boldsymbol{x}_1, \boldsymbol{x}_0, \omega) = \sum_{I=1}^{N} A(\boldsymbol{x}_0, \sigma_I, \boldsymbol{\gamma}_I) g_i(\boldsymbol{x}_0, \sigma_I, \boldsymbol{\gamma}_I) g_{0n}(\boldsymbol{x}_0, \boldsymbol{\gamma}_I) e^{i\omega T(\boldsymbol{x}_0, \sigma_I, \boldsymbol{\gamma}_I)}, \quad (2.63)$$

with N the number arrivals and $(\sigma_I, \boldsymbol{\gamma}_I)$ the ray field coordinates of the rays that arrive at $\boldsymbol{x}_1$:

$$(\sigma_I, \boldsymbol{\gamma}_I) \in \{(\sigma, \boldsymbol{\gamma}) \,|\, \boldsymbol{x}(\boldsymbol{x}_0, \sigma, \boldsymbol{\gamma}) = \boldsymbol{x}_1\}. \quad (2.64)$$

The mapping from spatial coordinates $\boldsymbol{x}_1$ to the possibly multi-valued set of ray field coordinates $(\sigma_I, \boldsymbol{\gamma}_I)$ is called a *ray field map* in this work.

2.7 Expressions for scalar wave fields

In isotropic elastic media the eikonal equations (2.18) for both P-waves and S-waves simplify considerably. The precise analysis is not relevant here (see, e.g., Červený, 2001), but its final form is used at a few places in this thesis:

$$|\boldsymbol{p}|^2 = V^{-2}(\boldsymbol{x}). \quad (2.65)$$

The phase velocity may be that of either P-waves or S-waves and is independent of the wave front normal $\boldsymbol{n}$. The same form of eikonal is obtained also for acoustic waves, after a similar asymptotic analysis of the acoustic wave equation. It is the eikonal equation for scalar waves, and in many practical situations the isotropic elastic waves are treated as scalar waves.

It is sometimes useful to use other forms of the eikonal equation. As suggested by Červený (2001), any solution that satisfies (2.65) also satisfies the more general equation

$$F(|\boldsymbol{p}|^2) = F(V^{-2}(\boldsymbol{x})). \quad (2.66)$$

Here $F(x)$ is a continuous function with continuous first and second derivatives, whose first derivative $F'(x)$ satisfies $xF'(x) > 0$ in the region of interest. In particular, the following choice of Hamiltonian is useful:

$$\mathcal{H}_n(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{n}\left(|\boldsymbol{p}|^n - V^{-n}(\boldsymbol{x})\right) = 0, \quad (2.67)$$

with n an integer. This form may be generalised to include even $n = 0$ by taking the limit $n \downarrow 0$:

$$\mathcal{H}_0(\boldsymbol{x}, \boldsymbol{p}) = \lim_{n \downarrow 0} \mathcal{H}_n(\boldsymbol{x}, \boldsymbol{p}) = \ln(|\boldsymbol{p}|) + \ln(V(\boldsymbol{x})) = 0. \tag{2.68}$$

The ray equations for the eikonal equations (2.67) read

$$\frac{d\boldsymbol{x}}{d\sigma} = |\boldsymbol{p}|^{n-2}\boldsymbol{p} \tag{2.69}$$

$$\frac{d\boldsymbol{p}}{d\sigma} = \frac{1}{n}\frac{\partial V^{-n}(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{2.70}$$

$$\frac{dT}{d\sigma} = |\boldsymbol{p}|^{n}. \tag{2.71}$$

The last equation shows the physical meaning of flow parameter σ for each choice of n. For $n = 0$ it is equal to travel time as in the general elastic case above. For $n = 1$ the flow parameter measures the distance along the ray curve. The choice $n = 2$ is sometimes used because it simplifies Equation (2.69).

Chapter 3

Ray field mapping in position/angle coordinates

In ray theory a propagating wave is parameterised by travel time and amplitude functions that are calculated as a function of the ray field coordinates. The evaluation of a ray-theoretical wave field at a given point in space requires a mapping from spatial coordinates to ray field coordinates. In heterogeneous media this mapping may become locally multi-valued, which leads to practical problems in various applications.

In this chapter a new approach for the calculation and representation of ray field maps is introduced that is particularly useful when ray field maps are needed for a dense distribution of sources at an acquisition surface. This is the case in most seismic imaging experiments such as reflection seismics and borehole tomography. For such source distributions it is suggested to construct a single ray field map in an extended space of spatial coordinates and angles, rather than a number of maps in the spatial domain for a range of acquisition coordinates.

A ray field map in the position/angle domain is single-valued, regardless of the complexity of the medium. The ray field information is organised by angles at depth rather than by the point of emergence at the surface, which makes the maps particularly suitable for use in modern seismic imaging methods. It is shown that to calculate these maps it is not necessary to trace rays up towards the acquisition surface, which would involve an major increase in the computational burden. Instead, existing algorithms that trace downwards can be be adapted to work in the position/angle domain.

3.1 Introduction

Ray fields and ray field maps

A characteristic feature of ray methods is that the calculations are not performed directly in terms of the spatial coordinates of the medium. Although the seismic ray equations are derived from the elastic wave equation, they do not share the same computational domain. The wave equation, on the one hand, is a partial differential equation for displacement as a function of the spatial coordinates of the medium. The kinematic ray equations, on the other hand, are ordinary differential equations for the phase space coordinates of the ray as a function of the *flow parameter* (for details see Chapter 2).

The study of an ensemble of rays, or *ray field*, leads to the introduction of additional *ensemble parameters*, which parameterise a continuous range of initial conditions of the rays in the ensemble. Together, the flow and ensemble parameters constitute an internal coordinate system for the ray field and are referred to as *ray field coordinates*.

The ray-theoretical wave field is parameterised by travel time and amplitude functions that are determined – along with the spatial coordinates – as a function of the ray field coordinates. To evaluate the travel time and amplitude at a given spatial location one needs to know its corresponding ray field coordinates. In other words, one needs to evaluate the mapping from spatial to ray field coordinates: the *ray field map*.

If the medium is sufficiently complex for the ray field to develop caustics and multi-pathing, the ray field map becomes multi-valued. This multi-valuedness is a source of practical problems in the application of ray methods in both forward and inverse wave propagation problems.

Ray field mapping algorithms

Over the years a large range of seismic ray tracing algorithms have been developed. The algorithms are all based on the same ray equations but may differ for example in the way the medium is parameterised and what type of ray information is required.

The algorithms that are used in smooth media may broadly be categorised in bending and shooting algorithms. The bending algorithms (e.g., Julian and Gubbins, 1977; Um and Thurber, 1987) treat ray tracing as a boundary value problem. This corresponds to the traditional geometric description of ray theory based on Fermat's principle. An initial ray estimate between two points is iteratively updated to improve the fit with the ray equations. These methods typically require considerable computing time for every source-receiver pair and are practical only for sparse source and receiver distributions, as in seismology.

Instead, ray tracing to a dense distribution of receivers is more efficiently tackled by shooting algorithms. These algorithms treat ray tracing as an initial value problem and implement the concepts of ray fields and ray field maps described

above. The spatial coordinates $\boldsymbol{x}$ of a ray are calculated as a function of the ray field coordinates $(\sigma, \boldsymbol{\gamma})$, with flow parameter σ, and ensemble parameters $\boldsymbol{\gamma}$. The task of the shooting algorithms is to calculate the ray field map, that is, to find the set of ray field coordinates $\{(\sigma_I, \boldsymbol{\gamma}_I), I = 1 \ldots N(\boldsymbol{x}^*)\}$ that solve equation

$$\boldsymbol{x}(\sigma, \boldsymbol{\gamma}) = \boldsymbol{x}^* \tag{3.1}$$

for every receiver $\boldsymbol{x}^*$. Different types of shooting algorithm differ primarily on how this ray field mapping is performed.

One way to solve (3.1) is by searching the space of $(\sigma, \boldsymbol{\gamma})$ for every $\boldsymbol{x}^*$. This is the classical two-point shooting method, which, like the bending methods, is practical only for a sparse distribution of receivers.

A much more efficient approach is implemented by the paraxial ray methods (Červený et al., 1984; Beydoun and Keho, 1987). In these algorithms a number of rays is shot from the source, covering a range of take-off angles. Paraxial ray theory is then used to extrapolate information from each ray to a volume of grid points in its neighbourhood.

In complex velocity structures paraxial shooting methods do not give satisfactory results, however. Strong variations in geometrical spreading and the possibility of multi-pathing make it difficult to find complete and unambiguous solutions. In these situations the shooting methods have to be enhanced with a mechanism to control accuracy and completeness.

A powerful technique called wave front construction was introduced by Vinje et al. (1993) for 2-D isotropic media. A number of variations and extensions have been published, in 2-D (Sun, 1992; Ettrich and Gajewski, 1996; Lambaré et al., 1996) as well as in 3-D (Vinje et al., 1996a,b; Lucio et al., 1996; Coman and Gajewski, 2002). Implementations in anisotropic media have also been reported (Gibson, 2000; Mispel, 2001).

Wave front construction currently seems to be the algorithm of choice in complex media, because it provides complete and unambiguous results within a certain level of accuracy. The accuracy control mechanism, however, makes WFC computationally much more expensive than the paraxial methods mentioned above.

Ray field maps and imaging

The mapping from medium coordinates to ray parameters plays a central role in ray-based imaging ("Kirchhoff migration") in reflection seismics. From each acquisition point $\tilde{\boldsymbol{x}}$ – source or receiver – the Green function has to be calculated to every point $\boldsymbol{x}$ of a dense subsurface grid. In the classical approach each Green function is assumed to consist only of a single ray arrival, which means that for each combination of $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ only a single travel time $T(\tilde{\boldsymbol{x}}, \boldsymbol{x})$ has to be determined. Depending on the type of imaging the amplitude $A(\tilde{\boldsymbol{x}}, \boldsymbol{x})$ may also be required.

In the common practice of Kirchhoff migration the travel time maps are first calculated on a coarse grid in $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$, to be interpolated to a denser grid during the

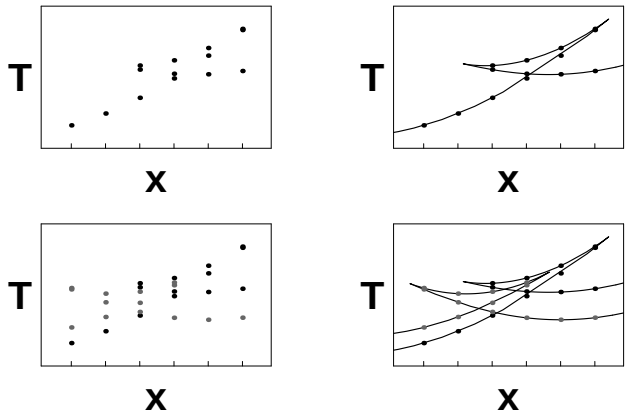


Figure 3.1: The panel on the top left displays the travel time at a number of grid locations in a single spatial dimension (x). The travel time map is multi-valued on this grid (triplication). Interpolation between the grid positions is difficult if the geometry of the branches of the traveltime curve (as shown top right) is unknown. The bottom panels show that the interpolation of travel times for different source locations (bottom left) is practically impossible in the presence of multi-pathing, even if the geometry of the branches (bottom right) is known.

actual migration (e.g., Epili and McMechan, 1996). The main purpose of this two-step procedure is to limit storage space and disk traffic, because in 3-D applications the maps become too voluminous to be kept in working memory.

This classical approach to Kirchhoff migration becomes problematic if the ray fields become multi-valued, because of difficulties in interpolation, as sketched in Figure 3.1. Partly to avoid the interpolation problems and partly because many industrial codes cannot cope with more than one arrival, Kirchhoff migration is often performed using only a single arrival, usually the first or the strongest. It is commonly agreed, however, that for ray-based imaging in complex media the incorporation of all arrivals is essential for satisfactory results (e.g., Geoltrain and Brac, 1993; Operto et al., 2000).

A modern way of dealing with problems of multi-pathing in seismic imaging is to parameterise the imaging integrals in terms of scattering angles and azimuths at depth (e.g., ten Kroode et al., 1998; de Hoop and Brandsberg-Dahl, 2000; Xu et al., 2001). A commonly used approach to both avoid the interpolation problems and perform imaging and velocity analysis in the angle domain, is to trace rays from each subsurface point to be imaged towards the acquisition surface (e.g., Koren et al., 2002; Brandsberg-Dahl et al., 2003). The subsurface grid, however, extends over one dimension (depth) more than the acquisition surface. Hence, this approach increases the computational burden for ray tracing dramatically.

Ray field maps in the position/angle domain

This chapter shows how the construction of ray field maps in an extended space of medium coordinates and angles, or *position/angle coordinates*, rather than in medium coordinates alone, helps to overcome the practical problems associated with the multi-valued maps.

In this introduction the approach is sketched roughly, to be made more precise in the following sections. The approach is based on the fact that in seismic applications ray field maps are not just calculated for a single source locations, but rather for dense source distribution, spread out on the acquisition surface.

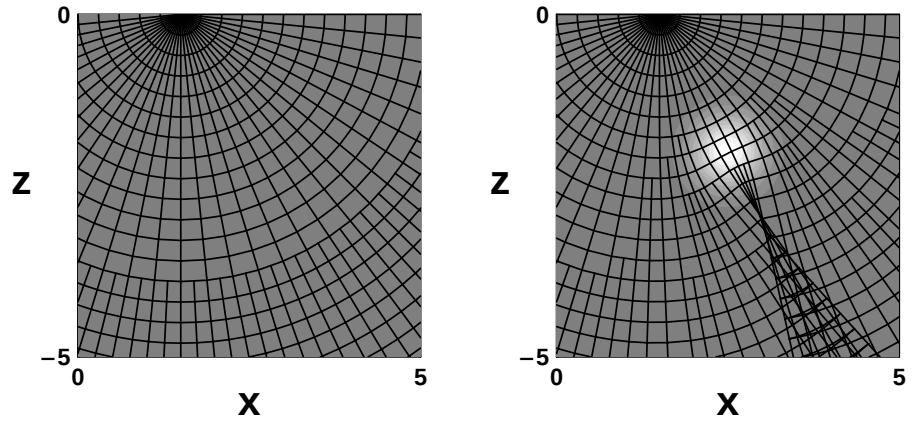
Let the ray field for a single point source be parameterised by travel time T and the slowness angles $\tilde{\boldsymbol{\phi}}_0$. The concept of a ray field can be extended to include rays from a distribution of source locations by adding the coordinates $\tilde{\boldsymbol{x}}_0$ of the acquisition surface to the set of ray field coordinates. Mapping this extended ray field to a specific subsurface position $\boldsymbol{x}$ now yields an infinite number of solutions, because at least one ray is found for every source location. All these rays, however, necessarily have a unique slowness at $\boldsymbol{x}$, which can be deduced from the well-known fact that rays do not cross in phase space. By adding the slowness angles $\boldsymbol{\phi}$ to the map coordinates, a single-valued map is obtained.

Hence, the idea is to calculate maps of the extended set of ray field coordinates $(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)$ on the position/angle coordinates $(\boldsymbol{x}, \boldsymbol{\phi})$. This way, the ray field information is organised by angles at depth immediately, rather than by acquisition coordinates, as is the usual practice. The maps can therefore readily be used in modern imaging/inversion and velocity analysis in the angle domain (de Hoop and Brandsberg-Dahl, 2000; Xu et al., 2001; Brandsberg-Dahl et al., 2003). Using ray field maps in the position/angle domain it is not necessary to trace rays up towards the surface, which represents a considerable improvement of efficiency. Algorithms for the construction of the ray field maps will be discussed in Chapters 5 and 6, and Appendix B.

To illustrate the concept of ray fields in the position/angle domain two examples in simple, isotropic, 2-D media are shown in Section 3.2. The concepts are made more precise in Section 3.3, which treats the definition of a suitable set of slowness angles in 3-D, and Section 3.4, which provides some mathematical relations regarding the coordinate transformations between ray field, position/angle, and phase space coordinates.

3.2 Examples in simple media

As an illustration of the concept of a ray field map in position/angle coordinates a 2-D isotropic medium with a constant phase velocity V is considered. A ray field for a single source position in this medium is displayed in Figure 3.2(a). The range of source positions is chosen at the level $z = 0$. The mapping from ray field coordinates $(\tilde{x}_0, T, \tilde{\phi}_0)$ to position/angle coordinates (x, z, ϕ) may be expressed



(a) Homogeneous medium (b) Gaussian anomaly

Figure 3.2: Ray fields corresponding to a source at $x = 1.5$, calculated using wave front construction. Figure (a) shows the ray field in a homogeneous medium with slowness 1.0, Figure (b) the one for a Gaussian slowness anomaly of magnitude +0.3 superimposed on the homogeneous background of Figure (a). The anomaly causes triplication in the ray field.

analytically:

$$\begin{pmatrix} x \\ z \\ \phi \end{pmatrix} = \begin{pmatrix} \tilde{x}_0 + TV \cos \tilde{\phi}_0 \\ TV \sin \tilde{\phi}_0 \\ \tilde{\phi}_0 \end{pmatrix}. \tag{3.2}$$

Expressions for the ray field map in this configuration are easily found:

$$\begin{pmatrix} \tilde{x}_0 \\ T \\ \tilde{\phi}_0 \end{pmatrix} = \begin{pmatrix} x - z \cot \phi \\ zV^{-1} \csc \phi \\ \phi \end{pmatrix}. \tag{3.3}$$

Geometrical insight into these equations may be obtained by looking at the 3-D contour surface plots in Figure 3.3.

Figure 3.3(a) shows a contour surface for constant $\tilde{x}_0$, or a single source. The contour surface is spanned by all rays emitted from a source at $\tilde{x}_0 = 1.5$, as displayed in Figure 3.2(a). The more or less vertical lines drawn on top of the surface indicate individual ray paths (constant $\tilde{x}_0$ and $\tilde{\phi}_0$); the more or less horizontal lines are individual wave fronts (constant $\tilde{x}_0$ and T). Figure 3.3(b) shows contour surfaces for constant traveltime T. Each contour surface may be interpreted as a generalised wavefront for a continuous distribution of sources, or as a surface spanned by the wavefronts of all source positions $\tilde{x}_0$. The lines drawn on top of the surfaces correspond to individual wave fronts (constant T and $\tilde{x}_0$). Figure 3.3(c)

shows a contour for constant $\tilde{\phi}_0$, which is spanned by rays with a common angle at all source locations. The nearly vertical lines indicate individual rays (constant $\tilde{\phi}_0$ and $\tilde{x}_0$); the horizontal lines indicate constant $\tilde{\phi}_0$ and T. Finally, Figure (d) shows all the contours of Figures 3.3(a), (b), and (c) combined. The intersections of the contours of T with that of $\tilde{x}_0$ represent a wave front from a single source. The intersection of the contours of $\tilde{x}_0$ and $\tilde{\phi}_0$ represents a single ray.

The ray field maps become more complicated for inhomogeneous media. In order to understand the influence of a localised inhomogeneity it is illustrative to look at an example of a medium with a single gaussian-shaped positive slowness anomaly. No analytic expressions are available for either ray field or ray field map for this kind of medium. The ray field for a single source position is depicted in Figure 3.2(b). Figure 3.4 contains the same set of contour plots that were shown for the homogeneous medium. Clearly, the anomaly that causes multi-pathing for a single source in the spatial domain leaves the ray field map in the position/angle domain single-valued. The contours are still smooth.

3.3 Slowness surface parameterisation

The slowness surface is defined as the set of slowness vectors that satisfy the eikonal equation $\mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) = 0$ (2.21) for a fixed $\boldsymbol{x}$. Since the choice of a Hamiltonian like (2.27) implicitly fixes the wave-type in elastic media, the slowness surface determines a single phase velocity V for each wave front normal $\boldsymbol{n}$ (see also Chapter 2). Any parameterisation of the unit circle in 2-D or unit sphere in 3-D may therefore serve as a parameterisation of the slowness surface. Here, a parameterisation of the unit sphere in terms of projection angles is proposed.

Parameterisation of the 2-D unit circle

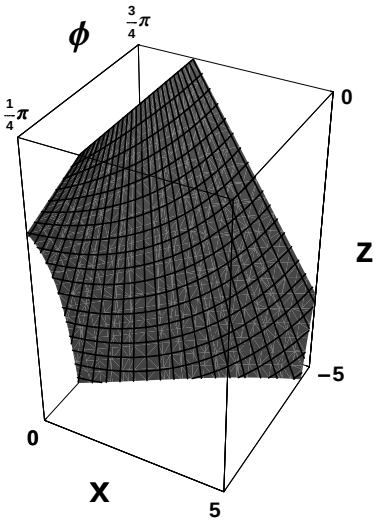
In 2-D, directions are uniquely parameterised by a single angle. If the angle is constrained to a 2π interval, it, in turn, is uniquely determined for any given vector of non-zero length. Let the angle ϕ_{xz} be defined to lie on the interval $[0, 2\pi\rangle$, with a value of 0 on the positive x-axis, and $\frac{1}{2}\pi$ radians at the positive z-axis. The direction of any vector $\boldsymbol{p} = (p_x, p_z)$, or the ratio of its components, is determined by the tangent of this angle:

$$\frac{p_z}{p_x} = \tan \phi_{xz}. \tag{3.4}$$

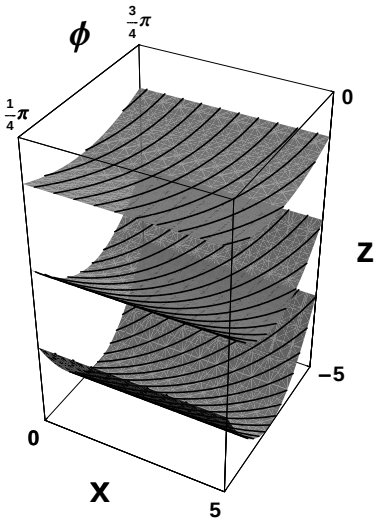
Likewise, for a given vector $\boldsymbol{p}$ the direction angle ϕ_{xz} is uniquely determined on the interval $[0, 2\pi\rangle$:

$$\phi_{xz} = \arctan\left(\frac{p_z}{p_x}\right). \tag{3.5}$$

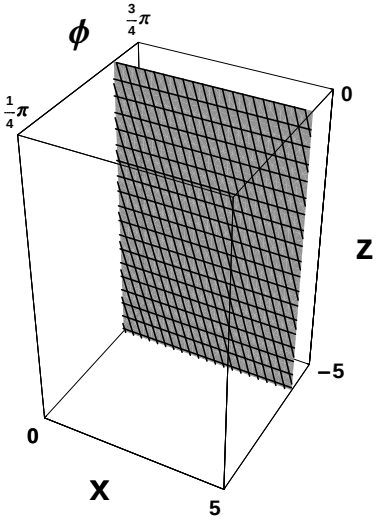
Note that even if p_x vanishes, the arctan is well defined. Hence, in 2-D there is a bijective mapping between angles and unit vectors.



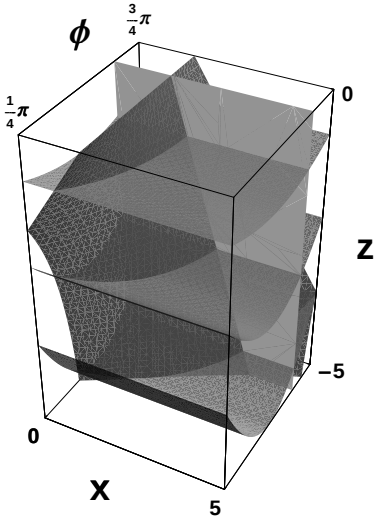
(a) $\tilde{x}_0 = 1.5$



(b) $T = 1.0, 3.0, 5.0$

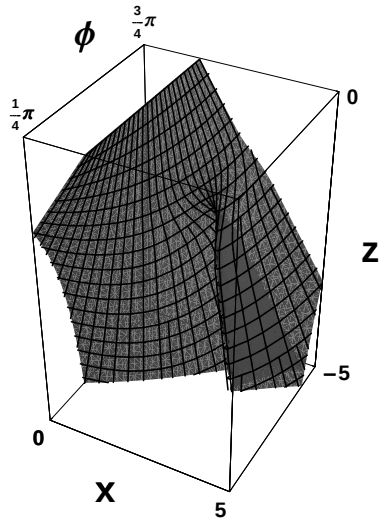


(c) $\tilde{\phi}_0 = 2.0$

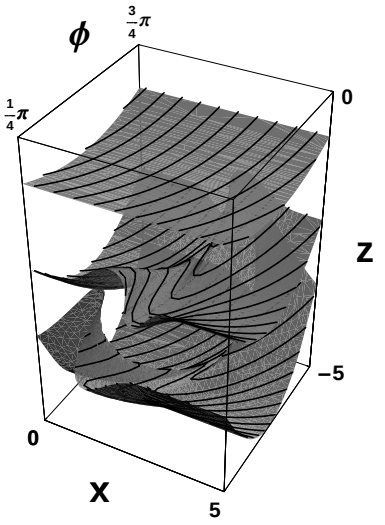


(d) Figs. (a),(b), and (c) combined

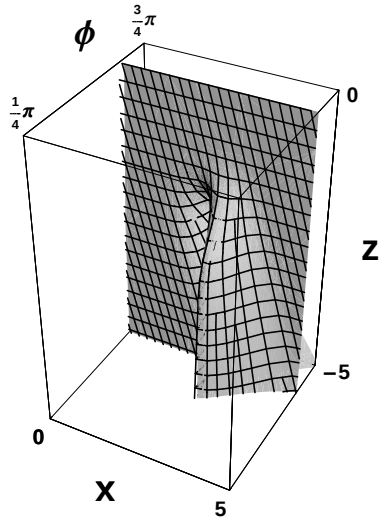
Figure 3.3: 3-D Contours for the ray field map $(\tilde{x}_0, T, \tilde{\phi}_0)$ in (x, z, ϕ)-coordinates for the homogeneous medium of Figure 3.2(a), with a source distribution at $z = 0$.



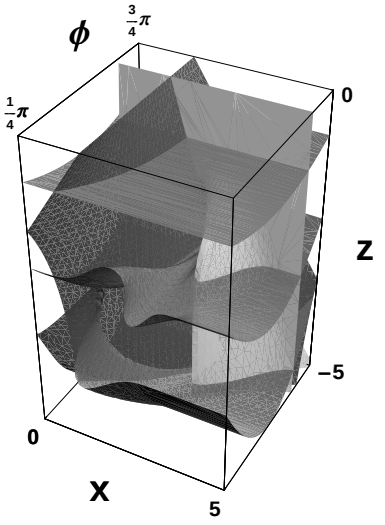
(a) $\tilde{x}_0 = 1.5$



(b) $T = 1.0, 3.0, 5.0$

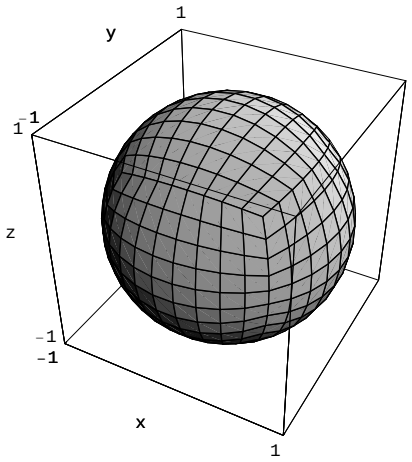


(c) $\tilde{\phi}_0 = 2.0$

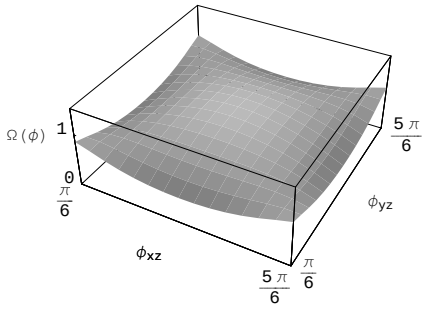


(d) Figs. (a),(b), and (c) combined

Figure 3.4: The contour plots equivalent to those of Figure 3.3, now for the inhomogeneous medium introduced in Figure 3.2(b).



(a) Patchwork for the unit sphere



(b) Differential solid angle function for a vertical patch

Figure 3.5: Figure (a) shows the parameterisation of the unit sphere by means of a patchwork of six projection angle patches, of which only three are visible. Every patch is parameterised by two angles covering a range of $\pi/2$ radians each. Figure (b) shows the solid angle function (3.9) of the patch associated with the positive z-axis, using a range of $2\pi/3$ radians. A uniform distribution of rays in (ϕ_{xz}, ϕ_{yz}) results in a more or less uniform ray density $(1/\Omega)$ on the unit sphere. For the "ideal" parameterisation Ω would be equal to 1 on the entire domain.

Parameterisation of the unit sphere by projection angles

In 3-D it is not possible to define a single non-degenerate parameterisation for the unit sphere. Commonly used coordinate systems such as the spherical and geographical coordinates are degenerate at the poles. The projection of a vector pointing in the direction of one of the poles onto the azimuthal plane vanishes, which leaves the azimuthal angle undetermined. A non-degenerate parameterisation of the unit sphere therefore necessarily consists of a number of patches.

The patchwork to be used for unit sphere parameterisation proposed here is based on the Cartesian rather than the spherical coordinate system, because the former is the coordinate system of choice in seismics.

The parameterisation consists of six patches, each associated with one of the six – positive and negative – coordinate semi-axes. Each patch is is also associated with two *projection angles*. These angles are determined by the projection of a vector onto both coordinate planes that contain the central axis of the patch.

As an example consider the patches associated with the central axis in z-direction. The two coordinate planes that contain this axis are the (x, z) and the (y, z) planes. Projection of the vector $\boldsymbol{p} = (p_x, p_y, p_z)$ on these planes allows the

Projection angles and directions in 3-D

Central axis	Half space	Angle ranges	Direction vector
x	$p_x > 0$	$\phi_{xz} \in \langle -\frac{1}{2}\pi, \frac{1}{2}\pi \rangle$ $\phi_{xy} \in \langle -\frac{1}{2}\pi, \frac{1}{2}\pi \rangle$	$\boldsymbol{p} \propto \begin{pmatrix} \pm 1 \\ \tan\phi_{xy} \\ \tan\phi_{xz} \end{pmatrix}$
	$p_x < 0$	$\phi_{xz} \in \langle \frac{1}{2}\pi, \frac{3}{2}\pi \rangle$ $\phi_{xy} \in \langle \frac{1}{2}\pi, \frac{3}{2}\pi \rangle$	
y	$p_y > 0$	$\phi_{xy} \in \langle 0, \pi \rangle$ $\phi_{yz} \in \langle -\frac{1}{2}\pi, \frac{1}{2}\pi \rangle$	$\boldsymbol{p} \propto \begin{pmatrix} \cot\phi_{xy} \\ \pm 1 \\ \tan\phi_{yz} \end{pmatrix}$
	$p_y < 0$	$\phi_{xy} \in \langle \pi, 2\pi \rangle$ $\phi_{yz} \in \langle \frac{1}{2}\pi, \frac{3}{2}\pi \rangle$	
z	$p_z > 0$	$\phi_{xz} \in \langle 0, \pi \rangle$ $\phi_{yz} \in \langle 0, \pi \rangle$	$\boldsymbol{p} \propto \begin{pmatrix} \cot\phi_{xz} \\ \cot\phi_{yz} \\ \pm 1 \end{pmatrix}$
	$p_z < 0$	$\phi_{xz} \in \langle \pi, 2\pi \rangle$ $\phi_{yz} \in \langle \pi, 2\pi \rangle$	

Table 3.1: For each coordinate axis two parameterisation patches are defined, one for the positive half space, and one for the negative. The relevant angles for each patch are given as well as the corresponding ranges of values and and a direction vector (not normalised) for these angles.

definition of the projection angles ϕ_{xz} and ϕ_{yz}:

$$\phi_{xz} = \arctan\left(\frac{p_z}{p_x}\right), \tag{3.6}$$

$$\phi_{yz} = \arctan\left(\frac{p_z}{p_y}\right). \tag{3.7}$$

Note that this definition is analogous to the 2-D definition (3.5), and in fact the approach can be generalised to any number of dimensions.

As in 2-D, the definition in (3.6-3.7) is well defined even if p_x or p_y goes to zero. However, the angle parameterisation (3.6-3.7) degenerates if p_z vanishes. Since the argument of the arctan is zero for any vector with $p_z = 0$, the projection angles are only determined by the signs of p_x and p_y respectively, and the direction cannot be reconstructed from these projection angles alone. Hence, each parameterisation patch may be used only for vectors with a non-vanishing component along its central axis.

The projection angles for all patches may be defined in a uniform way. Every patch uses two angles out of the set ϕ_{xz}, ϕ_{yz}, and ϕ_{xy}, where the former two are

defined by (3.6-3.7) and the latter by

$$\phi_{xy} = \arctan\left(\frac{p_y}{p_x}\right). \tag{3.8}$$

The angles used for each patch are those that contain the central axis of the patch in its subscript.

In contrast to the 2-D case, all angles are restricted to an interval of length π, where the interval bounds depend on the patch under consideration. Table 3.3 gives a summary of the angles and direction vectors for each patch.

In many applications, all directions of interest may have a non-vanishing component along one coordinate axis. In those cases a single patch is sufficient for the parameterisation of the directions. However, if all possible directions must be accounted for, it is most convenient to use all six patches, each covering one sixth of the unit sphere (Figure 3.5(a)).

Advantages with respect to alternatives

The parameterisation of directions in terms of projection angles has three major advantages with respect to alternative parameterisations. First, it is easy to design a non-degenerate patchwork to cover the entire unit sphere. This is not so for the commonly used spherical and geographical coordinate systems.

Second, the parameter ranges to be represented by a grid – say, in a ray field map – do not depend on the size of the slowness vectors to be parameterised. This means that a fixed grid can be used both in the presence of spatial variations in phase velocity and in a tomographic setting, such that a change in the local phase velocities does not require a redefinition of the grid. The opposite is true if, for example, two out of three slowness components are used to parameterise a patch of the slowness surface.

Finally, the parameterisation of the unit sphere is reasonably uniform, that is, an equidistant grid in terms of projection angles is reasonably efficient in sampling the unit sphere. This can be seen by looking at the differential solid angle function Ω, e.g., for a patch around the z-axis, for a unit vector $\boldsymbol{n}(\phi_{xz}, \phi_{yz})$:

$$\Omega(\phi_{xz}, \phi_{yz}) = \det\left(\frac{\partial \boldsymbol{n}}{\partial \phi_{xz}} \quad \frac{\partial \boldsymbol{n}}{\partial \phi_{yz}} \quad \boldsymbol{n}\right) = \frac{\csc^2 \phi_{xz} \csc^2 \phi_{yz}}{(1 + \cot^2 \phi_{xz} + \cot^2 \phi_{yz})^{3/2}}, \tag{3.9}$$

which is illustrated in Figure 3.5(b).

3.4 Mathematics of the eikonal manifold

In this section some mathematical relations are derived that are useful in the construction and application of ray field maps in the position/angle domain. A central theme will be to show that these maps are indeed single-valued, by showing that

the jacobian of transformation between ray field coordinates and position/angle coordinates does not vanish:

$$\det\left(\left.\frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial(\boldsymbol{x}, \boldsymbol{\phi})}\right|_{\mathcal{H}}\right) \neq 0. \tag{3.10}$$

Other useful relations are derived along the way. In the following derivations it is important to realise that vector $\boldsymbol{x}$ has three components, while vectors $\tilde{\boldsymbol{x}}_0$, $\tilde{\boldsymbol{\phi}}_0$ and $\boldsymbol{\phi}$ have only two.

Note that for the partial derivatives in (3.10) the Hamiltonian is explicitly kept constant:

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{p}) = 0, \tag{3.11}$$

or in other words, only variations are considered that satisfy the eikonal equation. The eikonal equation (3.11) is an algebraic equation in phase space. If N is the spatial dimension of the medium, its corresponding phase space is $2N$-dimensional. The eikonal equation then defines a $2N-1$ dimensional submanifold in phase space, which will be referred to as the *eikonal manifold*. For a fixed position $\boldsymbol{x}$ the eikonal manifold is equal to the slowness surface discussed in Section 3.3. The following analysis will show that both the ray field coordinates and the position/angle coordinates are two independent coordinate systems for the eikonal manifold.

In the following derivations it is assumed that the acquisition surface is a horizontal plane at a constant level of $z = z_0$, with coordinates $\tilde{\boldsymbol{x}}_0 = (x, y)$. Also, the slowness surface is assumed to be parameterised by the projection angles $\boldsymbol{\phi} = (\phi_{xz}, \phi_{yz})$ associated with the central z-axis. Similar relations for other acquisition surfaces and other angle systems may be derived analogously.

The following symbols are introduced for notational convenience:

$$\alpha = \frac{dT}{d\sigma} = \boldsymbol{p} \cdot \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} \tag{3.12}$$

$$\alpha \boldsymbol{v} = \frac{d\boldsymbol{x}}{d\sigma} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} \tag{3.13}$$

$$\alpha \boldsymbol{w} = \frac{d\boldsymbol{p}}{d\sigma} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}}, \tag{3.14}$$

with $\boldsymbol{v}$ the group velocity defined in Section 2.4. Note that usually travel time is used as the flow parameter of the ray tracing system, which renders α equal to unity.

Phase space and position/angle coordinates

The first step is to find the relationship between phase space and position/angle coordinates. The projection angles $\boldsymbol{\phi}$ parameterise the slowness surface for a

fixed $\boldsymbol{x}$ (see Section 3.3). To study the transformations between phase space coordinates and position/angle coordinates it is useful to add to the latter an additional coordinate for the directions normal to the eikonal manifold. A very convenient choice is the Hamiltonian $\mathcal{H}$ itself. The partial derivatives of ϕ and $\mathcal{H}$ with respect to $\boldsymbol{p}$ are known from the definitions (3.6), (3.7), and (3.13):

$$\left.\frac{\partial(\phi,\mathcal{H})}{\partial \boldsymbol{p}}\right|_{\boldsymbol{x}} = \begin{pmatrix} -\frac{p_z}{p_x^2+p_z^2} & 0 & \frac{p_x}{p_x^2+p_z^2} \\ 0 & -\frac{p_z}{p_y^2+p_z^2} & \frac{p_y}{p_y^2+p_z^2} \\ \alpha v_x & \alpha v_y & \alpha v_z \end{pmatrix} \equiv \boldsymbol{A}. \tag{3.15}$$

Which can be inverted to

$$\left.\frac{\partial \boldsymbol{p}}{\partial(\phi,\mathcal{H})}\right|_{\boldsymbol{x}} = \begin{pmatrix} \left.\frac{\partial \boldsymbol{p}}{\partial \phi}\right|_{\boldsymbol{x},\mathcal{H}} & \left.\frac{\partial \boldsymbol{p}}{\partial \mathcal{H}}\right|_{\boldsymbol{x},\phi} \end{pmatrix} \equiv \boldsymbol{A}^{-1}, \tag{3.16}$$

with components

$$\left.\frac{\partial \boldsymbol{p}}{\partial \phi}\right|_{\boldsymbol{x},\mathcal{H}} = \begin{pmatrix} \frac{p_x^2+p_z^2}{p_z} \begin{pmatrix} v_x p_x - 1 \\ v_x p_y \\ v_x p_z \end{pmatrix} & \frac{p_y^2+p_z^2}{p_z} \begin{pmatrix} v_y p_x \\ v_y p_y - 1 \\ v_y p_z \end{pmatrix} \end{pmatrix}, \tag{3.17}$$

and

$$\left.\frac{\partial \boldsymbol{p}}{\partial \mathcal{H}}\right|_{\boldsymbol{x},\phi} = \frac{1}{\alpha}\boldsymbol{p}. \tag{3.18}$$

For the entire position/angle domain, extended by coordinate $\mathcal{H}$, this gives

$$\frac{\partial(\boldsymbol{x},\phi,\mathcal{H})}{\partial(\boldsymbol{x},\boldsymbol{p})} = \begin{pmatrix} & \boldsymbol{I} & \boldsymbol{O} \\ (\boldsymbol{0} & \boldsymbol{0} & -\alpha \boldsymbol{w})^T & \boldsymbol{A} \end{pmatrix}, \tag{3.19}$$

with $\boldsymbol{0}$ a zero vector and $\boldsymbol{O}$ a zero matrix. Its inverse reads

$$\frac{\partial(\boldsymbol{x},\boldsymbol{p})}{\partial(\boldsymbol{x},\phi,\mathcal{H})} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{p}\boldsymbol{w} & \boldsymbol{A}^{-1} \end{pmatrix}. \tag{3.20}$$

Position/angle and ray field coordinates

For a ray field intersecting a surface non-tangentially, the position/angle coordinates may locally be parameterised in terms of the point of emergence at the surface $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\phi}})$ and the travel time to that point T. These may be called *emergence coordinates*. For the surface at $z = z_0$ the emergence coordinates correspond to the ray field coordinates $(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)$.

For a general surface at constant z, the emergence coordinates may be chosen as $\tilde{\boldsymbol{x}} = (x, y)$. If $v_z \neq 0$ this gives the following derivative matrix for the

position/angle domain coordinates in terms of the emergence coordinates:

$$\frac{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})}{\partial(\tilde{\boldsymbol{x}},T,\tilde{\boldsymbol{\phi}},\mathcal{H})} = \begin{pmatrix} 1 & 0 & v_x & 0 & 0 & 0 \\ 0 & 1 & v_y & 0 & 0 & 0 \\ 0 & 0 & v_z & 0 & 0 & 0 \\ 0 & 0 & u_x & 1 & 0 & 0 \\ 0 & 0 & u_y & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{E} & \boldsymbol{O} \\ \boldsymbol{F} & \boldsymbol{I} \end{pmatrix}, \tag{3.21}$$

with vector $\boldsymbol{u}$ defined as

$$\boldsymbol{u} = \frac{d(\boldsymbol{\phi},\mathcal{H})}{dT} = \frac{\partial(\boldsymbol{\phi},\mathcal{H})}{\partial(\boldsymbol{x},\boldsymbol{p})} \cdot \frac{d(\boldsymbol{x},\boldsymbol{p})}{dT} = \begin{pmatrix} u_x \\ u_y \\ 0 \end{pmatrix}, \tag{3.22}$$

and $u_z = 0$ because the Hamiltonian is constant along the ray. The determinant of (3.21) reads

$$\det\left(\frac{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})}{\partial(\tilde{\boldsymbol{x}},T,\tilde{\boldsymbol{\phi}},\mathcal{H})}\right) = v_z, \tag{3.23}$$

and its inverse:

$$\frac{\partial(\tilde{\boldsymbol{x}},T,\tilde{\boldsymbol{\phi}},\mathcal{H})}{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})} = \begin{pmatrix} \boldsymbol{E}^{-1} & \boldsymbol{O} \\ -v_z^{-1}\boldsymbol{F} & \boldsymbol{I} \end{pmatrix}, \tag{3.24}$$

with

$$\boldsymbol{E}^{-1} = \begin{pmatrix} 1 & 0 & -v_x/v_z \\ 0 & 1 & -v_y/v_z \\ 0 & 0 & 1/v_z \end{pmatrix}. \tag{3.25}$$

Note that matrix (3.24) can be used to relate general end point perturbation of rays to perturbations in the emergence coordinates. The same relations can be obtained by construction as discussed at the end of Section 2.5.

Matrices (3.21) and (3.24) can be combined with (3.19) and (3.20), respectively, to obtain derivative matrices in terms of the phase space variables.

Ray field map jacobian

The conditions under which the ray field map jacobian (3.10) does not vanish can now easily be checked by combining some of the relations defined above. First, observe that

$$\det\left(\left.\frac{\partial(\boldsymbol{x},\boldsymbol{\phi})}{\partial(\tilde{\boldsymbol{x}}_0,T,\tilde{\boldsymbol{\phi}}_0)}\right|_{\mathcal{H}}\right) = \det\left(\frac{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})}{\partial(\tilde{\boldsymbol{x}}_0,T,\tilde{\boldsymbol{\phi}}_0,\mathcal{H})}\right). \tag{3.26}$$

Now, the derivative matrix on the right hand side can be written as a superposition of the matrices (3.21), (3.20), and (3.19) derived above:

$$\frac{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})}{\partial(\tilde{\boldsymbol{x}}_0,T,\tilde{\boldsymbol{\phi}}_0,\mathcal{H})} = \frac{\partial(\boldsymbol{x},\boldsymbol{\phi},\mathcal{H})}{\partial(\boldsymbol{x},\boldsymbol{p})} \cdot \boldsymbol{P} \cdot \frac{\partial(\boldsymbol{x}_0,\boldsymbol{p}_0)}{\partial(\boldsymbol{x}_0,\boldsymbol{\phi}_0,\mathcal{H}_0)} \cdot \frac{\partial(\boldsymbol{x}_0,\boldsymbol{\phi}_0,\mathcal{H})}{\partial(\tilde{\boldsymbol{x}}_0,T,\tilde{\boldsymbol{\phi}}_0,\mathcal{H})}, \tag{3.27}$$

and $\boldsymbol{P}$ is the phase space propagator (2.38). Using (2.41) and (3.23) the ray field map jacobian may be written as

$$\det\left(\left.\frac{\partial(\boldsymbol{x},\boldsymbol{\phi})}{\partial(\tilde{\boldsymbol{x}}_0,T,\tilde{\boldsymbol{\phi}}_0)}\right|_{\mathcal{H}}\right) = v_{z0}\frac{\det(\boldsymbol{A})}{\det(\boldsymbol{A}_0)}, \tag{3.28}$$

with

$$\det(\boldsymbol{A}) = \frac{\alpha p_z}{(p_x^2+p_z^2)(p_y^2+p_z^2)} \tag{3.29}$$

$$= \frac{\alpha V^3(\boldsymbol{x},\boldsymbol{\phi})}{\Omega(\boldsymbol{\phi})}, \tag{3.30}$$

where V is the phase velocity ($\boldsymbol{p} = V^{-1}\boldsymbol{n}$), and Ω the differential solid angle defined in (3.9).

The ray field map jacobian does not vanish unless v_{z0} vanishes. These are "grazing" rays at the acquisition surface, which are normally not useful in reflection seismics. The ray field coordinates and the position/angle coordinates therefore provide two independent coordinates systems for the eikonal manifold, the former associated with the acquisition coordinates, the latter with the subsurface coordinates.

Amplitude calculation

To conclude, an example of the kind of ray field information that can be extracted from a ray field map in the position/angle domain. Because the map contains information on all the rays in the medium that intersect the acquisition surface, it also contains the information on paraxial rays. Using the gradients of the ray field map and some algebra all components of the ray propagator (2.38) can be recovered, including the non-eikonal component, because of relations (2.44) and (3.18).

The amplitude may be obtained using relations (2.58) and (2.59). Radiation factor D in the latter may be identified as (3.16):

$$D = \det\left(\frac{\partial \boldsymbol{p}}{\partial(\boldsymbol{\phi},\mathcal{H})}\right), \tag{3.31}$$

while J can be estimated from the ray field map:

$$J = \det\begin{pmatrix}\frac{\partial \boldsymbol{x}_0}{\partial \boldsymbol{\phi}} & \boldsymbol{v}_0\end{pmatrix} = \det\begin{pmatrix}\frac{\partial \boldsymbol{x}_0}{\partial \tilde{\boldsymbol{x}}_0}\cdot\frac{\partial \tilde{\boldsymbol{x}}_0}{\partial \boldsymbol{\phi}} & \boldsymbol{v}_0\end{pmatrix}, \tag{3.32}$$

with

$$\frac{\partial \boldsymbol{x}_0}{\partial \tilde{\boldsymbol{x}}_0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{3.33}$$

The two expressions for J in Equation (3.32) are equal because $\partial \boldsymbol{x}_0/\partial \boldsymbol{\phi}$ differs from $\partial \tilde{\boldsymbol{x}}_0/\partial \boldsymbol{\phi}$ only by a component along $\boldsymbol{v}_0$, which does not contribute to the value of J.

3.5 Discussion and conclusions

The construction of ray field maps in position/angle coordinates, rather than in spatial coordinates alone, helps to overcome the practical difficulties in imaging algorithms related to the handling of multi-valued ray field maps. The map in the position/angle domain contains the ray field information of a range of sources and is single-valued under all circumstances. The map may therefore be easily interpolated.

The approach is particularly useful for modern imaging/inversion algorithms that address the data as a function of scattering angles at the subsurface image point (e.g., de Hoop and Brandsberg-Dahl, 2000; Xu et al., 2001), because the ray field information is already organised by angles at depth. For implementation of angle domain imaging methods it is often considered to be most convenient to trace the rays upwards to the acquisition surface from each subsurface image point rather than the other way around (e.g., Koren et al., 2002; Brandsberg-Dahl et al., 2003). The disadvantage of this approach is a large increase in the computational effort related to the ray tracing. The theory developed in this chapter shows that in order to obtain the ray field information organised by angles at depth it is not necessary to trace the rays upwards. Instead, existing algorithms that trace downward may be adapted to work in the position/angle domain. Two important ray tracing algorithms, viz. wave front construction (e.g., Vinje et al., 1993) and the paraxial ray method (e.g., Beydoun and Keho, 1987), will be adapted and studied for this purpose in Chapters 5 and 6 respectively.

It should be noted that for the implementation of the angle domain imaging the greatest practical problems currently seem to be related to the data flow. In contrast to more classical formulations of Kirchhoff imaging the data – which are organised in terms of acquisition coordinates – cannot be handled sequentially. The evaluation of the imaging integral in terms of angles at depth requires more or less random data access. Since the total volume of data is much too large to be kept in working memory, this may result in a dramatic increase in disk I/O. Solutions to this problem may be found in alternative ways of storing the data, but that is outside the scope of this thesis.

Apart from in reflection imaging, ray field maps in the position/angle domain are expected to be useful in various other applications of ray theory. Whenever

a large and rather dense source distribution is present, and rays have to be calculated through the entire medium, the ray field map approach promises tools to do this efficiently. After a ray field map has been calculated it contains information on practically every ray in the medium, that is, those that intersect the source surface within a limited range positions and angles. The organised way of storing this information in a single-valued map makes this information highly accessible. Paraxial ray information, for example, can easily be obtained from the map through finite difference estimates. The ray field map is therefore particularly useful for modelling wave phenomena by ray theory.

Extensions to ray theory such as Maslov theory (e.g., Chapman and Drummond, 1982; Kendall and Thomson, 1993; Chapman, 2002), Gaussian beams (e.g., Červený et al., 1982; Červený, 2001), and coherent state theory (Klauder, 1987; Thomson, 2001) all extend the validity of zeroth order ray theory by making use of paraxial ray information. These methods may therefore take advantage of the amount of information present in the position/angle domain ray field maps. Maslov waveforms, for example, may be evaluated at a receiver point by an integral over the take-off angles at the source point (Chapman, 2002). The variables needed as a function of these angles in the neighbourhood of the receiver are the slowness vector, the travel time and the ray field jacobians. Due to reciprocity the same calculation can be done using the receiver point as the source, by integration over angles. The ray field map in the position/angle domain contains exactly the information needed: source position, slowness angle at the source, and travel time as a function of angle for a fixed receiver location. The calculation may be done efficiently because all the information is conveniently organised.

The one-to-one mapping between position/angle coordinates and ray field coordinates can be exploited in practical applications. Calculations that are typically performed in terms of ray field coordinates can now be performed in terms of position/angle coordinates and the other way around. Appendix A shows that this may be advantageous in tomography. If the ray field map is known for a reference model, the cost function gradient can easily be calculated for an arbitrary parameterisation of the model perturbation, using the new concept of a cost function sensitivity kernel.

The change of coordinates from ray field to position/angle coordinates can be exploited even further. In Appendix B it is shown how the theory developed in this chapter may be used to derive equations for the evolution of the ray field coordinates in terms of the position/angle coordinates. These equations may be used as the basis for a finite difference algorithm that calculates the full ray field information for a range of sources directly on a grid in the position/angle domain. This procedure avoids both the explicit mapping step that is usually required after ray tracing and the interpolation of medium properties on arbitrary spatial locations.

Chapter 4

Accurate interpolation using derivatives

A new technique for accurate interpolation using derivative information is presented. It is a hybrid of extrapolation to arbitrary order and linear interpolation, and combines the advantages of both methods. Through a modification of the coefficients of the Taylor expansion, extrapolations from a number of locations can be combined to obtain a polynomial order of accuracy that is one higher than that of a single Taylor expansion.

The formulation of the method is very general: it can be used both in regular and irregular data distributions in spaces of arbitrary dimension. The technique is expected to be useful in many applications. It is used at various stages in the algorithms described in the following chapters.

4.1 Introduction

The theory developed in Chapter 3 shows that ray field maps in position/angle coordinates are useful tools for seismic imaging in complex media. To be relevant for practical applications it is important that these maps are calculated efficiently. Therefore, two algorithms for the calculation of ray field maps are investigated in Chapters 5 and 6. In preparation, this chapter introduces a new and accurate local interpolation technique that is used at various stages in these algorithms.

Interpolation methods approximate a function using data that are available at a distribution of data points. The data consist of the function values and possibly some of its derivatives. The data points may be distributed in a space of arbitrary dimension. Both regular and irregular data distributions are relevant for this work, and hence both are discussed here.

Interpolation on a *regular*, rectangular grid is essential for the practical application of the ray field maps. In the common practice of Kirchhoff migration,

for example, travel time maps are first calculated on a coarse grid, but during the actual migration the travel time maps are interpolated to a denser grid (e.g., Epili and McMechan, 1996). The main purpose of this approach is to limit storage space and disk traffic, because in 3-D applications the maps become far too voluminous to keep them in working memory. The more accurately the interpolation can be performed, the more computer time is saved (e.g., Vanelle and Gajewski, 2002).

Interpolation on *irregular* distributions comes into play in the construction of ray field maps. In algorithms such as wave front construction (Vinje et al., 1993; Lambaré et al., 1996) and related methods (Chapters 5 and 6) ray fields are first calculated in terms of ray field coordinates, and subsequently mapped onto spatial coordinates. In the spatial domain the points at which the ray field is evaluated form an irregular distribution, which has to be regularised for practical application. One way to perform this regularisation is by means of interpolation. Another approach, based on averaging integrals, is used in Chapter 6.

This chapter introduces a general approach to enhancing the accuracy of local interpolation by using derivative information. The approach can be applied to both regular and irregular data distributions in arbitrary dimension.

Interpolation methods are widely used in many fields of science. Usually classical techniques as described in textbooks such as Press et al. (1992) and Ralston and Rabinowitz (2001) are used. Recent work has been devoted to the smooth interpolation of irregular data distributions (Sibson, 1981; Sambridge et al., 1995). The use of derivative information is usually limited to Hermite interpolation in a single variable (Ralston and Rabinowitz, 2001) or to bi-cubic interpolation in 2-D rectangular grids (Press et al., 1992).

The method proposed here combines aspects of interpolation and extrapolation. Using the available derivatives the function is extrapolated from the data points to the point of evaluation. Linear interpolation is used to weight the extrapolations from a number of data points. The best known extrapolation formula that uses the function values and derivatives from a single point is the Taylor expansion. Using the Taylor expansion for the extrapolation, however, can be shown to be sub-optimal. Instead, a a modification of the Taylor expansion is proposed, in which the expansion terms are multiplied by a set of coefficients. This makes the accuracy of the interpolation one order higher than that of each individual Taylor expansion. Since the proposed interpolation method is a hybrid of interpolation and extrapolation the term *intrapolation* is coined to refer to it. The modified Taylor expansion is given the name *Dutch Taylor expansion*.

The basic principles of intrapolation and the Dutch Taylor expansion are explained in Section 4.2. Application of intrapolation depends on the availability of the derivatives as independent data. If the derivatives are not available they can, under some accuracy conditions, be estimated from the available data. In regular grids finite difference approximation of derivatives is adequate, as explained in Section 4.3. The accuracy and convergence rates of intrapolation are compared to a number of alternative interpolation methods in Section 4.4.1. An example of

intrapolation in the context of ray field maps is provided in Section 4.4.2.

4.2 Accurate interpolation using derivatives

The objective of a numerical interpolation method is to approximate a function using data known at a set of isolated points. The data may include the function values and various derivatives. In the current context only local interpolation by means of low order polynomials is considered. The adjective local indicates that data is used only from a neighbourhood of the point at which the function value is estimated.

The general approach is to construct an interpolating function, the *interpolant*, which fits the available data perfectly. However, this assumes that the available data are exact, which in many applications is not the case. Methods exist for reducing the effect of errors by exploiting redundancy in the available data, and by relaxing the requirement of perfect fit, but these methods are inherently less local and computationally more expensive.

Even in the presence of errors, local interpolation is, in many cases, the most practical method for approximation. It is then important to choose an interpolant that does not enhance the errors. In case of first order (i.e. linear) interpolation the interpolated error is bounded by the errors in the contributing data. In general, however, the risk of enhancing errors increases with increasing polynomial order of the interpolant, especially if derivatives are used in its construction.

The reason to choose higher order interpolation may either be to enhance the smoothness or to enhance the accuracy (Press et al., 1992). Here, the main interest lies in enhancing the accuracy. This can be accomplished by using independent derivative information or data from a larger neighbourhood. A larger neighbourhood can be used either to define interpolants with a larger support, or to obtain derivatives estimates for the data points within original support, which can subsequently be used to constrain the higher order interpolant. Since it is more practical, the latter approach is followed in Section 4.3. For the remainder of this section it is assumed that the derivatives are known exactly or with sufficient accuracy.

4.2.1 Interpolation of Taylor expansions

The classical approach for incorporating derivative information in interpolation schemes is to use interpolants that fit the derivatives as well as the function values. Construction of such interpolants is relatively easy in 1-D (Hermite interpolation, e.g., Ralston and Rabinowitz, 2001), but becomes more complicated even for 2-D rectangular grids (e.g., Press et al., 1992). Generalisation to irregular data distributions and to rectangular grids in higher dimensions leads to complex calculations.

An alternative approach is to use a hybrid of extrapolation and interpolation. The function values and the available derivatives at each data point are used to construct an extrapolating function. At the point of evaluation the extrapolations from neighbouring data points are then weighted with a standard linear interpolant.

In the following it will first be shown how this hybrid strategy works for extrapolation using the Taylor expansion, which is the most intuitive choice for extrapolation. It turns out that for this choice the interpolation does not improve on the accuracy of the individual extrapolations. A major improvement is achieved by adopting another expansion that renders the order of accuracy of the interpolation equal to its polynomial order. The approach is outlined below in 1-D; generalisations to higher dimensions are discussed in Section 4.2.4.

The n-th order Taylor expansion of a function $f(x)$ around $x = \xi$ is given by

$$\mathcal{T}_n[f(x)](x;\xi) = \sum_{k=0}^{n} \frac{1}{k!}(x-\xi)^k f^{(k)}(\xi), \tag{4.1}$$

where $f^{(k)}$ stands for the k-th derivative of $f(x)$. A two-point interpolation formula based on x_0 and x_1 may then be constructed as follows:

$$\mathcal{I}_n^*[f(x)](x;x_0,x_1) = \frac{(x_1-x)}{(x_1-x_0)}\mathcal{T}_n[f(x)](x;x_0) + \frac{(x-x_0)}{(x_1-x_0)}\mathcal{T}_n[f(x)](x;x_1). \tag{4.2}$$

This hybrid interpolation formula may in some sense be regarded as a generalisation of linear interpolation. If no derivative information is available, the extrapolation is of zeroth order ($n = 0$) and the usual linear interpolation is obtained:

$$\mathcal{I}_0^*[f(x)](x;x_0,x_1) = \frac{(x_1-x)}{(x_1-x_0)}f(x_0) + \frac{(x-x_0)}{(x_1-x_0)}f(x_1). \tag{4.3}$$

Despite its intuitive character, the interpolation of Taylor expansions has a disadvantage that is best illustrated for interpolation using first order derivatives. The Taylor expansion (4.1) is in this case a linear function, or first order polynomial. The interpolation of both Taylor expansions (4.2) therefore produces a second order polynomial. It is interesting to see what second order interpolation formula (4.2) does to an arbitrary second order polynomial sampled at x_0 and x_1. Figure 4.1(a) clearly illustrates that the quadratic function is not reconstructed.

To understand what happens it is instructive to look at the general quadratic function $p_2(x)$:

$$p_2(x) = c_0 + c_1 x + c_2 x^2. \tag{4.4}$$

The first order Taylor expansion of this function around ξ is

$$\begin{aligned}\mathcal{T}_1[p_2(x)](x;\xi) &= p_2(\xi) + (x-\xi)p_2'(\xi)\\ &= c_0 + c_1 x + c_2(2x-\xi)\xi,\end{aligned} \tag{4.5}$$

which, after some algebra, leads to the interpolant

$$\mathcal{I}_1^*[p_2(x)](x;x_0,x_1) = p_2(x) + c_2(x-x_0)(x-x_1). \tag{4.6}$$

The interpolant differs from the function itself by a second order polynomial. This quadratic function has coefficient c_2 as a multiplier, which indicates that if this coefficient is zero, the function is perfectly reconstructed. This is no surprise, because in that case the function is linear, and both Taylor approximations are exact.

For comparison, the result of using a standard linear interpolant is

$$\mathcal{I}_0^*[p_2(x)](x;x_0,x_1) = p_2(x) - c_2(x-x_0)(x-x_1). \tag{4.7}$$

Apart from the sign, the error is exactly the same as in Equation (4.6). Hence, the use of derivative information using Equation (4.2) does not improve the accuracy of the interpolation.

4.2.2 Intrapolation and the Dutch Taylor expansion

Here, a hybrid of extrapolation and interpolation is proposed that is similar to the Taylor-based approach of Equation (4.2), but with a major improvement in accuracy. The interpolants are constructed in such a way that the order of accuracy is equal to the order of interpolation. In order to stress that the resulting interpolation formula is a hybrid of extrapolation and interpolation, the name *intrapolation* is introduced to refer to it. The corresponding interpolant is analogously called the *intrapolant*.

In analogy of Equation (4.2) the intrapolant is defined as

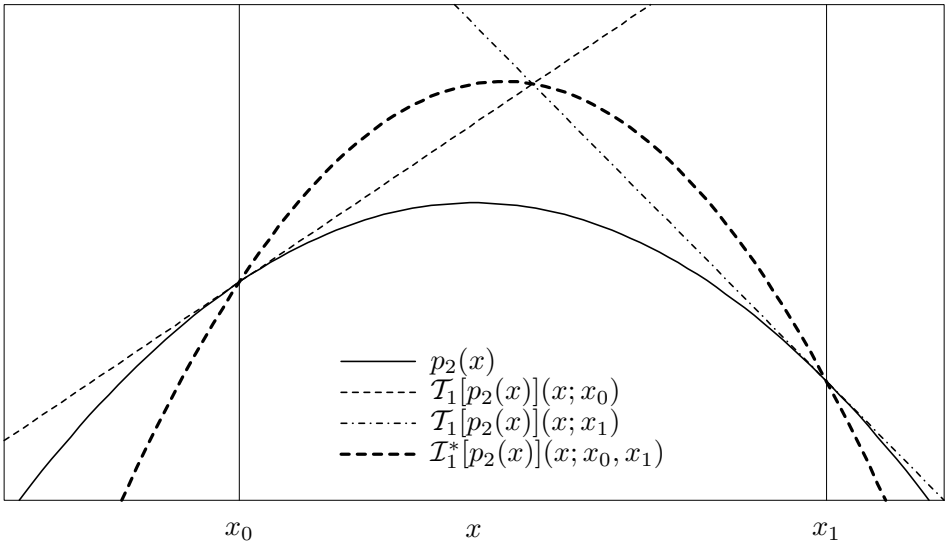
$$\mathcal{I}_n[f(x)](x;x_0,x_1) = \\ \frac{(x_1-x)}{(x_1-x_0)}\mathcal{D}_n[f(x)](x;x_0) + \frac{(x-x_0)}{(x_1-x_0)}\mathcal{D}_n[f(x)](x;x_1), \tag{4.8}$$

with $\mathcal{D}_n$ an extrapolation formula defined as

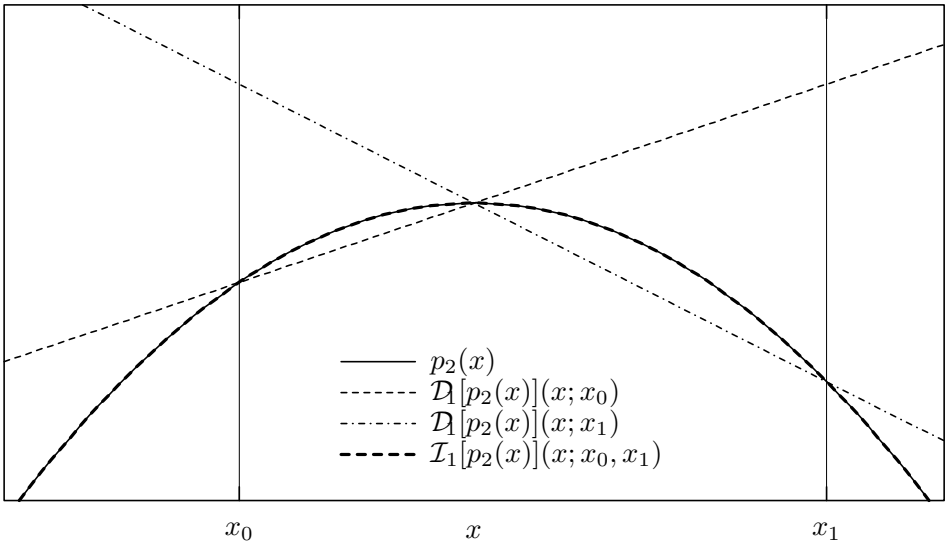
$$\mathcal{D}_n[f(x)](x;\xi) = \sum_{k=0}^{n} \frac{a_k^n}{k!}(x-\xi)^k f^{(k)}(\xi). \tag{4.9}$$

At this point the expansion is expressed as an ansatz in terms of unknown coefficients a_k^n, where the superscript n should be interpreted as an index rather than a power. Note that the parametric form of (4.9) reduces to the Taylor expansion with $a_k^n = 1$.

The coefficients introduced in extrapolation formula $\mathcal{D}_n$ provide the degrees of freedom that can be used to force the intrapolant (4.8) to satisfy some accuracy requirements. Since an n-th order Taylor expansion gives n-th order accuracy, interpolation of n-th order expansions is useful if it can enhance the order of



(a) Quadratic interpolation between points x_0 and x_1 based on the Taylor expansion, as in Equation (4.2). An arbitrary quadratic function is not reproduced by the interpolation.



(b) Quadratic interpolation between points x_0 and x_1 based on the Dutch Taylor expansion, as in Equation (4.21). An arbitrary quadratic function is reproduced exactly.

Figure 4.1: Approximation of a quadratic function $p_2(x)$, based on interpolation of first order extrapolations. For Figure (a) the extrapolation is the conventional Taylor expansion $\mathcal{T}$, as in Equation (4.1), for Figure (b) it is the Dutch Taylor expansion $\mathcal{D}$ (4.9) introduced in this paper.

accuracy to $n+1$. It should be noted here that the polynomial order of the intrapolant $\mathcal{I}_n$ is $n+1$. This means that the goal is to choose coefficients a_k^n in such a way that the order of accuracy of the intrapolant is equal to its polynomial order, which is the highest order of accuracy that can be obtained. For the extrapolation formula thus obtained the term *Dutch Taylor expansion* is coined, for reasons that will become clear later on.

A general algorithm for finding coefficients in linear functionals such as (4.9) is the method of undetermined coefficients (e.g., Ralston and Rabinowitz, 2001). As a first, trivial, example determine the coefficient a_0^0, that makes intrapolant (4.8) exact for first order polynomials. This should result in the well known linear interpolant. Since any first order polynomial in x can be written as a linear combination of the monomials 1 and x, it is sufficient to solve the system of equations

$$\begin{cases} \mathcal{I}_0[1](x;x_0,x_1) = 1 \\ \mathcal{I}_0[x](x;x_0,x_1) = x, \end{cases} \tag{4.10}$$

which, using (4.8), leads to the more explicit system

$$\begin{cases} a_0^0 \left(\dfrac{(x_1-x)}{(x_1-x_0)} + \dfrac{(x-x_0)}{(x_1-x_0)} \right) = 1 \\ a_0^0 \left(\dfrac{(x_1-x)}{(x_1-x_0)} x_0 + \dfrac{(x-x_0)}{(x_1-x_0)} x_1 \right) = x. \end{cases} \tag{4.11}$$

Although there are two equations for one coefficient the solution is consistent:

$$a_0^0 = 1. \tag{4.12}$$

Hence the first order intrapolant is given by

$$\mathcal{I}_0[f(x)](x;x_0,x_1) = \frac{(x_1-x)}{(x_1-x_0)} f(x_0) + \frac{(x-x_0)}{(x_1-x_0)} f(x_1), \tag{4.13}$$

which is indeed equal to the usual linear interpolant.

A similar system of equations can be composed to determine the coefficients a_k^1 for the second order intrapolant:

$$\begin{cases} \mathcal{I}_1[1](x;x_0,x_1) = 1 \\ \mathcal{I}_1[x](x;x_0,x_1) = x \\ \mathcal{I}_1[x^2](x;x_0,x_1) = x^2. \end{cases} \tag{4.14}$$

After some algebra this yields the coefficients

$$a_0^1 = 1, \quad a_1^1 = \frac{1}{2}, \tag{4.15}$$

and the intrapolant

$$\mathcal{I}_1[f(x)](x;x_0,x_1) = \frac{(x_1-x)}{(x_1-x_0)}\left[f(x_0)+\frac{1}{2}(x-x_0)f'(x_0)\right] + \frac{(x-x_0)}{(x_1-x_0)}\left[f(x_1)+\frac{1}{2}(x-x_1)f'(x_1)\right]. \quad (4.16)$$

This intrapolant is exact for second order polynomials, which can be checked by inserting p_2 as defined in Equation (4.4):

$$\mathcal{I}_1[p_2(x)](x;x_0,x_1) = p_2(x). \quad (4.17)$$

Note the difference with Equation (4.6). Figure 4.1(b) shows graphically that intrapolation with first derivatives is exact to second order.

It is not immediately obvious what the coefficients a_k^n for the Dutch Taylor expansion should be for higher values of n. An intrapolant that certainly has practical relevance (e.g. Section 4.4.2) is that for $n = 2$, which is based on both first and second derivatives. Another system of equations, similar to (4.10) and (4.14), yields coefficients

$$a_0^2 = 1, \quad a_1^2 = \frac{2}{3}, \quad \text{and} \quad a_2^2 = \frac{1}{3}. \quad (4.18)$$

Figure 4.2 compares intrapolation of a cubic function using these coefficients with the less accurate interpolation based on the Taylor expansion.

While the solutions for higher n are not very likely to be useful in practice (see Section 4.2.3), it is interesting to see what happens. The system of equations gets increasingly complex and more difficult to solve, but a pattern soon becomes clear. For general n the coefficients form the series

$$a_k^n = \frac{n+1-k}{n+1} = 1 - \frac{k}{n+1} \quad (4.19)$$

where $n+1$ is the polynomial order of the intrapolant, and n the maximum order of derivatives used.

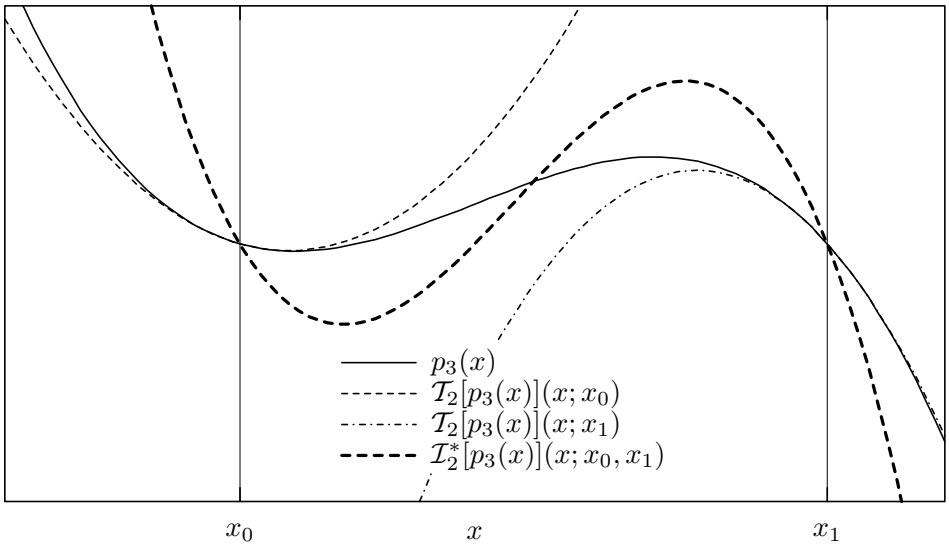
Using coefficients (4.19) the Dutch Taylor expansion (4.9) can be expressed explicitly as

$$\mathcal{D}_n[f(x)](x;\xi) = \sum_{k=0}^{n}\left(1-\frac{k}{n+1}\right)\frac{1}{k!}(x-\xi)^k f^{(k)}(\xi), \quad (4.20)$$

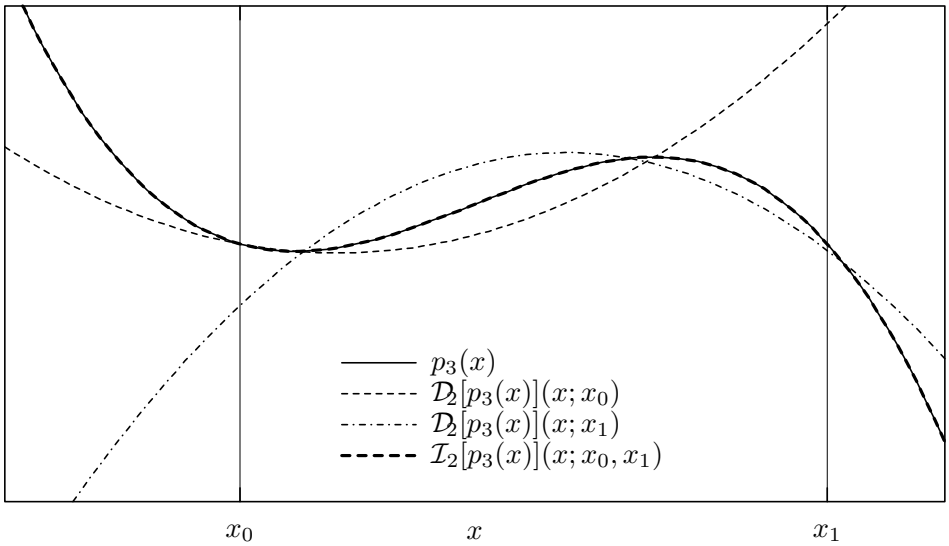
or alternatively as a normal Taylor expansion and a correction term:

$$\mathcal{D}_n[f(x)](x;\xi) = \mathcal{T}_n[f(x)](x,\xi) - \frac{1}{n+1}\sum_{k=1}^{n}\frac{k}{k!}(x-\xi)^k f^{(k)}(\xi). \quad (4.21)$$

The correction term sums over k, starting at $k = 1$ simply because the argument vanishes at $k = 0$. Appendix 4.A.1 proves the correctness of 4.19, by showing that expansion (4.21) makes the intrapolant (4.8) exact to order $n+1$ for arbitrary n.



(a) Cubic interpolation between points x_0 and x_1 based on the Taylor expansion, as in Equation (4.2). A arbitrary cubic function is not reproduced by the interpolation.



(b) Cubic interpolation between points x_0 and x_1 based on the Dutch Taylor expansion, as in Equation (4.21). An arbitrary cubic function is reproduced exactly.

Figure 4.2: The equivalent of Figure 4.1, now for a cubic function $p_3(x)$, with interpolation based on second order extrapolations.

4.2.3 Some interpretative remarks

The high order of accuracy achieved by intrapolation is primarily due to the Dutch Taylor expansion used for the extrapolation. In order to understand why this expansion is so successful it is instructive to look at it in some more detail and compare it with the Taylor expansion.

First, the case of first order expansion is examined. Coefficient a_1^1 in (4.15) is equal to $1/2$. Hence, the first order term in the Dutch Taylor expansion is only half the size of the first order term of the Taylor expansion. Why this is better for the sake of interpolation can be understood qualitatively by realising that the interpolation of the zeroth-order terms alone results already in a linear function. This linear function accommodates the average of the derivative over the interval. Hence, in order to improve the interpolation one should not use the full derivative information in the extrapolation. In the first order expansion using only half the derivative gives the correct result.

From the coefficients of the Dutch Taylor expansion given in Equation (4.19) it can be concluded that the same observation holds more generally. The coefficients of the Taylor expansion are equal to 1, regardless of n and k. All terms ($k > 0$) of the Dutch Taylor expansion are smaller, in an absolute sense, than the equivalent terms in the Taylor expansion.

The Dutch Taylor expansion may thus be regarded as an "economical" [1] alternative to the Taylor expansion. For extrapolation alone, it is less accurate than the Taylor expansion. When used in combination with a linear interpolation, however, it provides greater accuracy. This is possible because the interpolation joins the information coming from two sides. The combination of economy and the fact that in the interpolation each side brings in its own share of information provides ample motivation for the choice of the name *Dutch* Taylor expansion to refer to extrapolation formulas (4.20) and (4.21).

One of the key observations of Appendix 4.A.1 is that the n-th order Dutch Taylor expansion (4.21) is, for a fixed point of evaluation x, a linear function of the data point position ξ for polynomials up to order $n+1$. At $\xi = x$ this linear function passes through the correct value $f(\xi = x)$. Hence, the n-th order Dutch expansion alone gives an exact approximation only for constant functions, but the errors are such that they cancel by linear interpolation for polynomials up to order $n+1$.

In contrast, the n-th order Taylor expansion is exact for polynomials up to order n. For a polynomial of order $n+1$, however, the error is another polynomial of order $n+1$. This type of error is not cancelled by linear interpolation, which explains why an interpolation of Taylor expansions is less accurate than an interpolation of Dutch Taylor expansions.

It is important to note that although intrapolation and the Dutch Taylor expansion are discussed here for general order n, practical applications are usually restricted to low values of n. The range of practical applicability is similar to that

[1] Here "economical" should not be interpreted in the sense of computational efficiency.

of Taylor expansions. High order Taylor expansions are often used in theoretical formulations. In practical applications, however, their use is limited because the convergence with increasing order is typically very slow, and the region of convergence very small. The Dutch Taylor expansion and intrapolation are to be used in small regions where the function to be approximated is well represented by a low order polynomial, i.e. where the Taylor expansion coefficients decrease quickly for increasing order.

4.2.4 Generalisation to higher dimensions

Since extrapolation to arbitrary order and first order interpolation are both well defined in spaces of arbitrary dimension, the intrapolation method can easily be generalised to higher dimensions.

The generalisation of the extrapolation formula is straightforward. It is defined analogous to a multivariate Taylor expansion:

$$\mathcal{T}_n[f(\boldsymbol{x})](\boldsymbol{x};\boldsymbol{\xi}) = \sum_{k=0}^{n} \frac{1}{k!}(\boldsymbol{x}-\boldsymbol{\xi})^{[k]} : \nabla^{[k]} f(\boldsymbol{\xi}), \tag{4.22}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots)^T$ is now a vector variable. Superscript $[k]$ indicates the iterated tensor product with k indices, as in

$$\boldsymbol{x}^{[k]} = x_{i_1} x_{i_2} \ldots x_{i_k}, \tag{4.23}$$

and the double dot (:) indicates contraction (summation) over all k indices. The multivariate analogue of (4.21) is

$$\mathcal{D}_n[f(\boldsymbol{x})](\boldsymbol{x};\boldsymbol{\xi}) = \mathcal{T}_n[f(\boldsymbol{x})](\boldsymbol{x};\boldsymbol{\xi}) - \frac{1}{n+1}\sum_{k=1}^{n} \frac{k}{k!}(\boldsymbol{x}-\boldsymbol{\xi})^{[k]} : \nabla^{[k]} f(\boldsymbol{\xi}). \tag{4.24}$$

As argued in Section 4.2.1 for intrapolation in a single dimension, the accuracy of order $n+1$ can be explained by the fact that the Dutch Taylor expansion (4.21) is a linear function of data point ξ for polynomials up to order $n+1$, passing through the correct value at $\xi = x$. The same argument holds for multiple variables.

For multivariate polynomials up to order $n+1$, the Dutch Taylor expansion (4.24) yields a first order polynomial in terms of the data point variables $\boldsymbol{\xi}$. Hence, in order to do intrapolation with an accuracy of order $n+1$ any type of linear interpolation will do. Natural generalisations of linear interpolation to higher dimensions include the N-linear interpolation on rectangular grids, and barycentric interpolation for general data distributions organised by N-dimensional triangulation. Both types of interpolation are discussed in Appendix D.

4.3 Calculation of derivatives

The interpolation methods introduced in the previous sections assume the presence of derivative information on the data points. The alternative is to calculate

the derivative information from the available data. In essence this means introducing data available at neighbouring data points which do not participate in the interpolation itself.

In order to maintain the accuracy of intrapolation it is important that the accuracy of the derivatives match the order of the interpolating polynomial. That is to say, if the aim is a second order interpolant, which requires only first derivatives, these derivatives should be accurate to second order.

Derivative information on arbitrary data distributions can be obtained using finite difference (FD) methods (e.g., Fornberg, 1996, 1998). The calculation of these FD estimates is another example of the method of undetermined coefficients. The derivatives are expressed as a sum over neighbouring data values, multiplied by undetermined coefficients. The requirement that the derivatives are exact for polynomials up to a given order leads to a system of equations similar to those discussed in Section 4.2.1.

The number of data points used in the summation, and thus the size of the neighbourhood, depends on the number of coefficients needed to parameterise a polynomial of the required order. The combination of the (relative) data locations and the corresponding weights is called the FD stencil.

The complexity of the system of equations leading to the weights in an FD stencil strongly depends on the dimension of the data space, on the order of accuracy that is required, and on the geometry of the data points. A particular advantage of rectangular, equidistant grids is that the stencils only depend on relative data point positions, which means that the same stencil can be used for almost every data point in the grid. Only the grid points near the edges require adapted stencils.

In arbitrary data distributions the situation is a lot more complicated. The definition of a neighbourhood around a data point is much less obvious, and the relative positions of neighbouring data points change from place to place. The result is that each data point requires a dedicated FD stencil to obtain accurate derivatives. For most practical applications, including ours, this is computationally too expensive.

If the objective for the calculation of derivatives is smooth rather than accurate interpolation, one can think of many alternatives for the calculation of reasonable derivative estimates (e.g., Sibson, 1981). The author is, however, not aware of a practical approach that meets the accuracy requirements. For irregular data distributions it is assumed that the derivative information is independently available.

Centred finite differences

In the following the calculation of low order FD estimates in equidistant grids is summarised. The derivations are not be discussed here but can be found for example in Fornberg (1996, 1998). First, consider the one dimensional FD estimate for the derivative $f'(x)$ of a function $f(x)$. A three-point centred FD is exact for

polynomials up to second order:

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2). \tag{4.25}$$

The coefficient multiplying $f(x)$ itself happens to be equal to zero, which effectively makes it a two-point formula.

The second derivative estimate requires all three function values:

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2). \tag{4.26}$$

This formula is actually exact for polynomials up to third order, due to the symmetry of the FD stencil. This suggests the use of three-point centred FD for third order accurate intrapolation. The first order derivatives obtained using (4.25), however, are only accurate to second order. The application of (4.25) to a third order monomial cx^3 yields

$$\frac{c(x+h)^3 - c(x-h)^3}{2h} = 3cx^2 + ch^2. \tag{4.27}$$

This shows that the three-point centred FD estimate of the first derivative of any third order polynomial is wrong by a constant value. Fortunately, the intrapolant (4.8) only depends on the difference of the first derivatives at the two data points. This can be seen by looking at the interpolation of the first order expansion terms only:

$$\begin{aligned} a_1^n \frac{(x_1 - x)}{(x_1 - x_0)}(x - x_0)f'(x_0) + a_1^n \frac{(x - x_0)}{(x_1 - x_0)}(x - x_1)f'(x_1) \\ = a_1^n \frac{(x - x_0)(x - x_1)}{(x_1 - x_0)}(f'(x_1) - f'(x_0)) \end{aligned} \tag{4.28}$$

This means that the constant is cancelled in the interpolant, and it is indeed possible to do third order accurate interpolation using only three-point FD estimates.

One-sided finite difference formulas

At the boundaries of a grid it is necessary to compute the derivatives using one-sided rather than centred FD formulas. Depending on the desired order of accuracy the following formulas are available (from Fornberg, 1996, table 3.1-2):

$$f'(x) = \frac{-f(x) + f(x+h)}{h} + O(h) \tag{4.29}$$

$$= \frac{-3f(x) + 4f(x+h) - f(x+2h)}{2h} + O(h^2) \tag{4.30}$$

$$= \frac{-11f(x) + 18f(x+h) - 9f(x+2h) + 2f(x+3h)}{6h} + O(h^3), \tag{4.31}$$

where the latter is exact for third order polynomials. Similarly, for the second derivatives

$$f''(x) = \frac{f(x) - 2f(x+h) + f(x+2h)}{h^2} + O(h) \tag{4.32}$$

$$= \frac{2f(x) - 5f(x+h) + 4f(x+2h) - f(x+3h)}{2h^2} + O(h^2) \tag{4.33}$$

can be used, the latter being exact for third order polynomials. Note that the errors are different for one-sided and centred FD estimates, and cancellation of errors in the first derivative does not occur when both are combined.

Multivariate finite differences

The one dimensional finite difference formulas are easily generalised to higher dimensions. For the calculation of derivatives in a single variable, the other variables may be regarded as constant, while mixed derivatives can be calculated by repeated differencing. For example, the mixed second derivative of a bivariate function $f(x, y)$ is given by:

$$\begin{aligned}\frac{\partial^2 f(x,y)}{\partial x \partial y} =& \frac{f(x+h, y+h) - f(x+h, y-h)}{4h^2} \\ &- \frac{f(x-h, y+h) - f(x-h, y-h)}{4h^2} + O(h^2),\end{aligned} \tag{4.34}$$

which is exact for bivariate polynomials of second order in both x and y.

4.4 Examples

4.4.1 Comparison of 1D interpolation methods

To assess the accuracy and convergence behaviour of intrapolation, it is compared with a number of alternative interpolation methods. The experiment is performed only in 1-D, but, qualitatively, the result will be the same in higher dimensions.

Experimental setup

Interpolation and extrapolation methods are typically designed to be exact for polynomials up to a given order (see Section 4.2.1). For an assessment of these methods it is useful to look at their performance when applied to a different class of test function. Fourier functions are ideal for this purpose because they form a complete basis for continuous functions, and they relate to familiar concepts as sampling intervals and length scales of variation.

The interpolation of the complete range of real-valued Fourier functions is studied in the form

$$\sin(kx + \phi), \tag{4.35}$$

where x is the variable to be interpolated, k the frequency (or wave number) and ϕ an arbitrary phase shift.

Without loss of generality the first data point may be chosen at $x_0 = 0$ (NB arbitrary phase shifts are accommodated by ϕ), and the second data point at $x_1 = 1$, because variations in the width of the data point interval are mathematically equivalent to variations in k. In this light k may be interpreted as a relative frequency that can be expressed as a product of the absolute frequency $\hat{k}$ and the width of the interpolation interval h:

$$k = \hat{k}h. \tag{4.36}$$

Variations in k may thus be attributed to variations in interval width (h), or to variations in absolute frequency ($\hat{k}$).

For an arbitrary interpolant $\mathcal{I}$, the error E may be expressed as a function of k, x, and ϕ:

$$E(k, x, \phi) = \mathcal{I}\left[\sin(kx + \phi)\right](x, 0, 1) - \sin(kx + \phi). \tag{4.37}$$

For fixed frequency k and fixed position x the error is a periodic function of ϕ. To obtain insight in the behaviour of the error as a function of k and x, the RMS average over the range of ϕ is calculated:

$$\left[E_{\phi}^{\mathrm{rms}}(k, x)\right]^2 = \frac{1}{2\pi} \int_0^{2\pi} \left[E(k, x, \phi)\right]^2 d\phi. \tag{4.38}$$

To isolate the frequency dependence the error is averaged over the range of x:

$$\left[E_{x\phi}^{\mathrm{rms}}(k)\right]^2 = \frac{1}{2\pi} \int_0^1 \int_0^{2\pi} \left[E(k, x, \phi)\right]^2 d\phi dx. \tag{4.39}$$

The error functions are calculated for a number of different interpolation methods.

Interpolation methods

The interpolation methods discussed here are subdivided into three groups. The first group consists of the intrapolation methods $\mathcal{I}_n$ (4.8), for $n = 0, 1, 2$. The derivatives used in the extrapolation may either be exact or obtained by finite difference. For the FD estimators an equidistant grid is assumed, which means that data points at $x = -1$ and $x = 2$ come into play. In total, five different intrapolation methods are tested.

The second group consists of two variations of Hermite interpolation: one using exact and one using FD derivatives. In two-point Hermite interpolation (e.g., Ralston and Rabinowitz, 2001) a third order polynomial interpolant is constructed

that fits both the function values and its first derivatives exactly. The Hermite interpolant for a function $f(x)$ may be expressed as

$$\mathcal{I}^H[f(x)](x,0,1) = f(0)a_0(x) + f(1)a_1(x) + f'(0)b_0(x) + f'(1)b_1(x), \quad (4.40)$$

with basis functions $a_0(x)$, $a_1(x)$, $b_0(x)$, and $b_1(x)$ defined by

$$a_0(x) = 2x^3 - 3x^2 + 1, \quad (4.41)$$
$$a_1(x) = -2x^3 + 3x^2, \quad (4.42)$$
$$b_0(x) = x^3 - 2x^2 + x, \qquad \text{and} \quad (4.43)$$
$$b_1(x) = x^3 - x^2, \quad (4.44)$$

Two-point Hermite interpolation does not have a simple analogue in dimensions higher than one.

The third group are nearest neighbour interpolations $\mathcal{I}_n^N$. In fact, these are Taylor expansions from the data point that is nearest to the point of interpolation:

$$\mathcal{I}_n^N[f(x)](x,0,1) = \begin{cases} \mathcal{T}_n[f(x)](x,0) & \text{for} \quad 0 \le x < \frac{1}{2} \\ \mathcal{T}_n[f(x)](x,1) & \text{for} \quad \frac{1}{2} \le x \le 1. \end{cases} \quad (4.45)$$

Nearest neighbour interpolations are also studied for $n = 0, 1, 2$. The latter two use derivatives that can either be exact or FD estimates, which makes a total of five different nearest neighbour interpolants. Nearest neighbour interpolation with $n = 2$ and FD derivatives is used by Vanelle and Gajewski (2002) for the interpolation of travel time maps (see Section 4.4.2).

Results

For all types of interpolation, the error functions (4.37) to (4.39) can be calculated analytically. The error curves in Figures 4.3- 4.5 are made using high precision arithmetic in *Mathematica*.

Figure 4.3 shows the phase shift averaged error (4.38) for a fixed frequency of 0.1 as a function of x. For each type of interpolation a single line style is chosen; the distinction between different orders for intrapolation and nearest neighbour interpolation is easily made from the height of the curves. (The lowest curve represents the smallest error and corresponds to the highest order.)

Due to the symmetry of the interpolations, all error curves are symmetrical about $x = 0.5$. Most of the curves, viz. those that represent an interpolation that uses derivatives, are coloured differently on the left and the right side of the graph. Black and grey make a distinction between the methods that use exact derivatives (black on the left side, grey on the right), and those that use FD derivatives (grey on the left, black on the right).

For the more accurate interpolations (lower curves) it obvious that using exact derivatives yield a higher accuracy then FD derivatives. If derivatives are available, Hermite interpolation and third order intrapolation ($n = 2$) contend for the

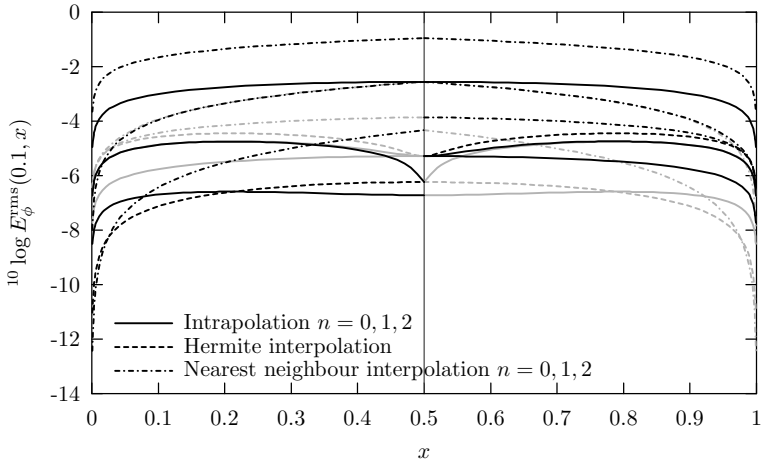


Figure 4.3: The RMS error made in the interpolation of a sine (4.35) with frequency $k = 0.1$, averaged over the full range of possible phase shifts (4.38). All curves are symmetric about $x = 0.5$. Some curves, however, are plotted black on one side, and grey on the other; curves that are black on the left side use either no or exact derivatives, those that are black on the right side use either no or FD derivatives. The line styles of different orders n of an interpolation type are the same, but the lowest curves (least error) correspond to the highest order.

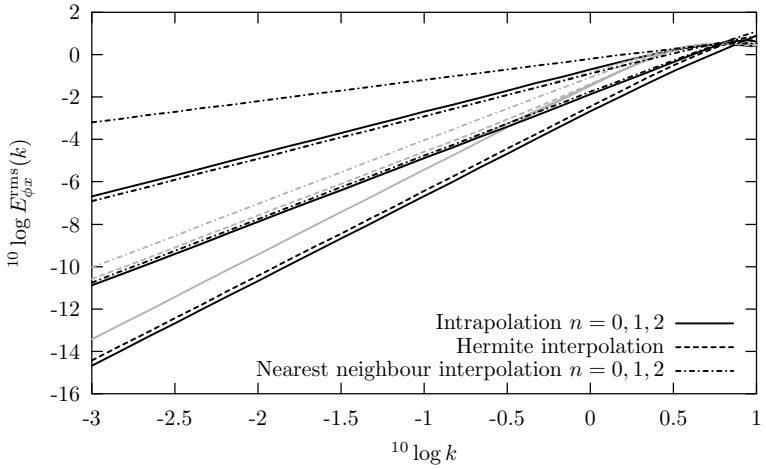


Figure 4.4: The RMS error of interpolating sines (4.35) as a function of frequency, averaged over the full range of possible phase shifts and over the range of x (4.39). The values at $^{10}\log k = -1$ correspond to the RMS averages of the curves of Figure 4.3. Bold black curves correspond to interpolation types that use either no or exact derivatives, while grey curves use FD estimates. The plot clearly indicates the different rates of convergence for the different interpolation types, as well as their relative RMS accuracy.

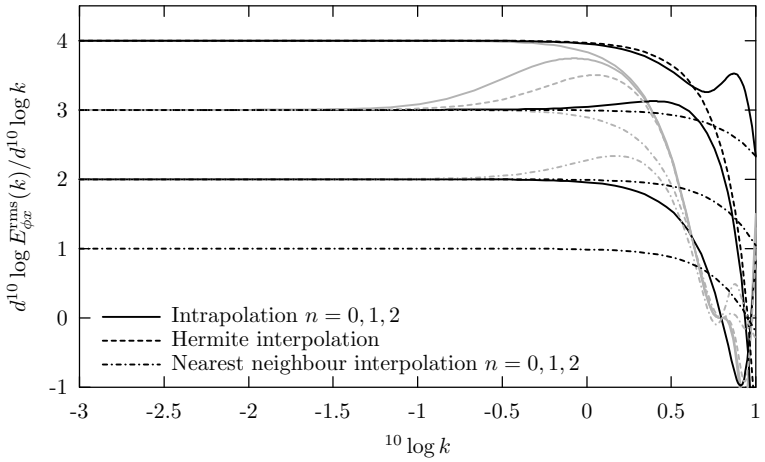


Figure 4.5: The derivatives of the curves in Figure 4.4. The rates of convergence are expressed more clearly. For low frequencies the third order accurate methods have a dominant error term of the fourth order, which gives a logarithmic derivative of 4.0. The only zeroth order method is the nearest neighbour interpolant with $n = 0$.

highest accuracy. Closer to the data points Hermite is more accurate because it fits the derivatives exactly. In the central region, however, intrapolation is more accurate, probably because it uses second order derivatives, while Hermite uses only derivatives of first order. The main disadvantage of Hermite interpolation is that it does not generalise to higher dimension, while intrapolation does.

Looking at the lowest error curves on the right side of the graph, it is clear that in the case of FD derivatives, Hermite interpolation is not as accurate as third order intrapolation. Even if only first derivatives are used, the second order intrapolation is more accurate than the FD Hermite interpolation. From the viewpoint of computational efficiency there is no difference between intrapolation and Hermite interpolation of the same polynomial order. The second order intrapolant is calculated more efficiently, simply because of its lower polynomial order.

The nearest neighbour interpolations are not as accurate as the intrapolations that use the same amount of derivative information. The fact that nearest neighbour interpolation is basically an extrapolation shows from the relatively high error near the centre of the interval.

The convergence behaviour of the interpolation methods is best appreciated by looking at the phase and interval averaged error function (4.39). Figure 4.4 shows the logarithm of this error as a function of the logarithm of the frequency. The linear behaviour of the error functions indicate convergence with polynomial rates. The derivatives of these curves correspond to the order of the error. The order of the error is one higher than the order of accuracy of the interpolations. Figure 4.5 clearly shows that the order of accuracy of the interpolations ranges

from 0 to 3.

In both convergence plots the black lines correspond to the interpolation methods that use exact derivatives, while the grey lines correspond to those that use FD derivative estimates. The intrapolations and the nearest neighbour interpolations show the expected convergence behaviour. In general, the order of accuracy of the nearest neighbour interpolations are equal to the order of the derivatives that are used. For the intrapolations, the order of accuracy is always one higher than the order of derivatives (n) used.

Hermite interpolation with exact derivatives shows the expected result of being accurate to third order. If FD estimates are used for the derivatives, however, Hermite interpolation turns out to be accurate only to second order, because the FD estimates themselves are only accurate to second order (4.25). For intrapolation the latter does not pose a limitation on the order of accuracy because of the fortuitous cancellation of errors discussed in Section 4.3. Even intrapolation based only on first order FD derivatives (n=1) is slightly more accurate than FD Hermite interpolation, while both methods use the same data.

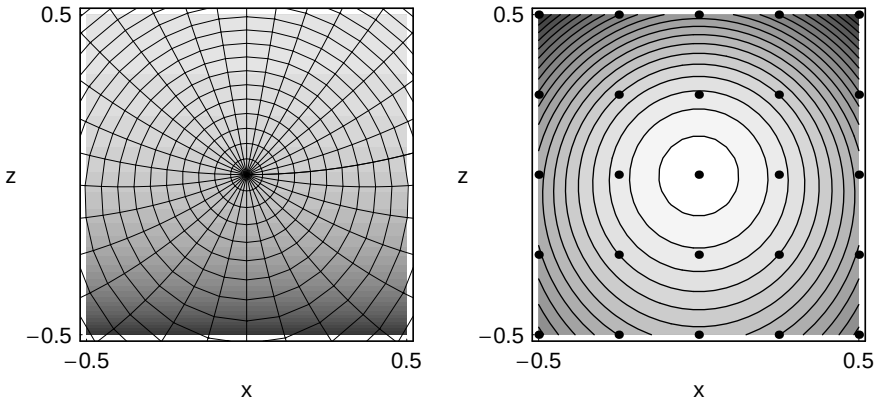
If no exact derivatives are available, intrapolation should be preferred over Hermite interpolation. Also, if both first and second order exact derivatives are available, Figure 4.4 indicates that intrapolation is the method of choice for interpolation, because it gives the lowest error curve in the plot. If only first order exact derivatives are known Hermite interpolation is preferred. In higher dimensions, however, Hermite interpolation is not available. Finally, intrapolation is always to be preferred over nearest neighbour interpolations.

4.4.2 Travel time interpolation

An important application of interpolation in the context of this thesis is the interpolation of travel time maps for Kirchhoff-type imaging. Coarse gridded travel time maps in depth are commonly calculated for the full range of source and receiver locations, prior to the actual imaging. Upon imaging these coarse maps are interpolated to a finer grid. Higher order accuracy in the interpolation allows for coarser maps to start with, which saves storage space and data traffic overhead in practical implementations.

The advantages of travel time interpolation with higher order accuracy were shown by Vanelle and Gajewski (2002). They proposed the second order ($n = 2$) nearest neighbour interpolant discussed in Section 4.4.1, using FD estimates of the derivatives. Here, it is shown that using the same amount of information a greater accuracy can be achieved using intrapolation. The second order method proposed by Vanelle and Gajewski (2002) may therefore be replaced by third order intrapolation, which enhances the advantages at negligible additional cost. Another advantage is that the interpolated travel times are continuous.

To demonstrate that intrapolation works in higher dimensions, a single example of travel time interpolation in a regular grid is shown here. Other examples will be provided in later chapters.



(a) Rays and wave fronts

(b) Contours of squared travel time

Figure 4.6: Figure (a) shows the rays and wave fronts in a medium with V satisfying $V^{-2} = 1 - z$. The grey tones in the background indicates the velocity (darker means faster). Figure (b) shows a contour plot of the corresponding squared travel time. On top a rectangular grid of points is drawn, which contain the data used in the interpolation

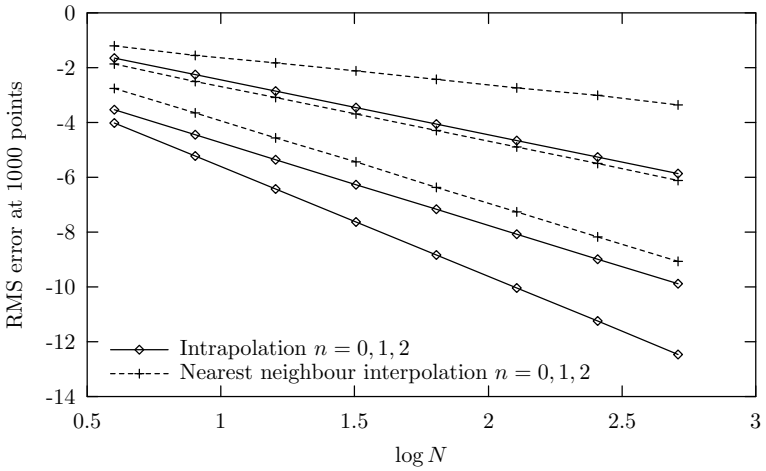


Figure 4.7: Convergence of RMS interpolation error for the square of travel time at 1000 random samples in the domain of figure 4.6. The number N corresponds to the number of grid intervals in both dimensions. Each interpolation type has three curves, arranged in increasing order from top to bottom.

Vanelle and Gajewski (2002) argue that it is better to interpolate the square of the travel time (*hyperbolic interpolation*) than the travel time itself (*parabolic interpolation*), because in a homogeneous medium the square of the travel time is a quadratic function of the position relative to the source. Hence, in a homogeneous medium any interpolation method of second or higher order accuracy will reconstruct the squared travel time perfectly. In inhomogeneous media, the interpolation will not be exact anymore, but close to the source the hyperbolic interpolation will always be superior.

The travel time map to be interpolated is one that can be obtained analytically from the analytic ray solution. The medium, shown in Figure 4.6(a), along with a number of rays and wave fronts emitted by a point source in the origin, has a constant gradient of the square of the slowness (reciprocal velocity) of magnitude 1, and velocity of 1 at the origin. Figure 4.6(b) shows some contours of the squared travel time, as well as an example of a rectangular grid defined in the medium.

Because the travel times on the grid are calculated analytically, the grid spacing determines the accuracy of interpolation. For a range of grid spacings, the RMS interpolation error is calculated from a set of 1000 random locations. For N the number of grid intervals in each dimension the error is plotted in Figure 4.7.

As in Section 4.4.1, the nearest neighbour interpolations as well as the intrapolations show the expected convergence rates. Intrapolation based on first and second order derivatives ($n = 2$) proves to be the most accurate. Surprisingly, even intrapolation with only first order derivatives ($n = 1$), is more accurate than the nearest neighbour interpolation that uses both first and second derivatives.

In practical applications one is usually less interested in the order of convergence of a given interpolation method then in the grid spacing corresponding to a given error. The results of Figure 4.7 can be fitted with linear functions. In this way the number of grid intervals needed for each interpolation method to obtain an acceptable error of, say 10^{-4} can be deduced. Rounded to the first integer above, this gives 60, 6, and 5 for intrapolation with $n = 0, 1, 2$, respectively, and 2393, 46, and 11 for the respective nearest neighbour interpolations. The higher the desired accuracy, the greater the advantage of intrapolation.

4.5 Discussion and conclusions

A new technique for accurate interpolation using derivative information is presented. The technique is called intrapolation, because it is a hybrid of linear interpolation and extrapolation. The extrapolation is done by a modification of the Taylor expansion, referred to as the Dutch Taylor expansion. The order of accuracy of intrapolation is one higher than that of an individual Taylor expansion that uses the same amount of derivative information.

Intrapolation is generally applicable in spaces of arbitrary dimension, in regular as well as irregular data distributions. For interpolation inside an irregular data distribution, the presence of a geometric structure such as an N-dimensional

triangulation is assumed. For interpolation in rectangular grids, intrapolation may also be applied without the presence of independent derivative information. Instead, finite difference estimates for the derivatives may be used. Using only three point FD stencils, an interpolation of third order accuracy is obtained.

Intrapolation does not attempt to fit the available derivative data perfectly. It only uses this data to "steer" the interpolant in the right direction, resulting in an enhanced accuracy. It should be noted, however, that the derivatives are in general discontinuous across the edges of the interpolants. If smoothness rather than accuracy is the reason for higher order interpolation it will be advisable to use a different method (e.g., Press et al., 1992; Sambridge et al., 1995).

An important advantage of not attempting to fit the derivatives perfectly is that it is not necessary to use high polynomial orders for the interpolants. Higher order polynomials are more expensive to evaluate, and are more sensitive to errors in the data. In intrapolation, the influence of the derivatives at all of the data points is (in a way) averaged – note the reduced size of the coefficients in the Dutch Taylor expansion – and therefore the interpolants are less sensitive to errors and more stable than methods that fit the derivatives perfectly.

In order to check the accuracy of intrapolation the method is subjected to two tests. The first test in a 1-D setting shows that the theoretical order of accuracy is indeed obtained in practice. For two-point interpolation, intrapolation outperforms all other interpolation methods. Only if the first derivatives are known exactly, Hermite interpolation yields a third order accuracy, while intrapolation needs second derivatives as well to improve on that. In higher dimensions, however, Hermite interpolation is not available.

The second test involves the interpolation of travel times for a point source in 2-D, where intrapolation is tested against nearest neighbour interpolation, both using FD derivative estimates. It turns out that in this case intrapolation using only a single derivative already outperforms the nearest neighbour interpolation that uses second derivatives as well.

It seems fair to conclude that intrapolation is a very attractive method for accurate interpolation both for regular and – if independent derivative data is available – for irregular data distributions. Implementation is straightforward. The method should be useful in a wide variety of applications. Examples of its use in ray field calculations are provided in Chapter 5, while the applicability of the Dutch Taylor expansion in a broader context is further investigated in Chapter 6.

4.A Appendix

4.A.1 Dutch Taylor expansion coefficients

The intrapolant of polynomial order $n+1$ is constructed using the the function derivatives up to order n. This appendix provides the proof that the Dutch Taylor expansion coefficients a_k^n, as stated in Equation (4.19), render the intrapolant (4.8)

exact for polynomials up to order $n+1$.

Point of departure is the Dutch Taylor expansion 4.21. For a given function $f(x)$, and a data point ξ at which the derivatives up to order n are known, this formula gives the extrapolated value as a function of the point of evaluation x.

Alternatively one can interpret (4.21) as a function of ξ, for a fixed point of evaluation x. This perspective is very useful here, because it clearly shows the conditions under which the intrapolant (4.8) is exact. Since the intrapolant is defined as a linear interpolation, it is exact if and only if extrapolation formula (4.21) is a linear function of data point ξ, passing through the correct value at $\xi = x$. Here it is shown that this is the case for polynomials up to order $n+1$. The proof is divided in two parts, one for polynomials up to order n, and one for the polynomial of order $n+1$. First, however, some binomial identities are shown that are useful in the derivations.

Binomial identities

The binomial 'n over k' is defined as

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}. \tag{4.46}$$

The binomial occurs in the expansion of the n-th power of a sum:

$$(A+B)^n = \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k}. \tag{4.47}$$

This relation will be used in the form

$$\sum_{k=0}^{n} \binom{n}{k} (x-\xi)^k \xi^{n-k} = x^n. \tag{4.48}$$

Finally,

$$k\binom{n}{k} = n\binom{n-1}{k-1} \tag{4.49}$$

is another binomial identity that will be used.

Proof for polynomials of order $m \leq n$.

For a monomial x^m, with $0 \leq m \leq n$, the n-th order Taylor expansion is exact, so the extrapolation formula (4.21) becomes

$$\mathcal{D}_n[x^m](x;\xi) = x^m - \frac{1}{n+1} \sum_{k=1}^{m} k\binom{m}{k} (x-\xi)^k \xi^{m-k}, \tag{4.50}$$

where the limit of the sum has changed from n to m, simply because higher terms are zero.

Using binomial identity (4.49) and the index $K = k - 1$ gives

$$\begin{aligned}\mathcal{D}_n[x^m](x;\xi) &= x^m - \frac{m}{n+1}\sum_{K=0}^{m-1}\binom{m-1}{K}(x-\xi)^{K+1}\xi^{m-1-K} \\ &= x^m - \frac{m}{n+1}(x-\xi)\sum_{K=0}^{m-1}\binom{m-1}{K}(x-\xi)^{K}\xi^{m-1-K}.\end{aligned} \tag{4.51}$$

Application of (4.48) finally leads to

$$\begin{aligned}\mathcal{D}_n[x^m](x;\xi) &= x^m - \frac{m}{n+1}(x-\xi)x^{m-1} \\ &= \left(1 - \frac{m}{n+1}\right)x^m + \frac{m}{n+1}x^{m-1}\xi.\end{aligned} \tag{4.52}$$

The extrapolation formula is apparently a linear function of data point ξ for polynomials up to order n, while it has the exact value x^m for $\xi = x$. Hence, the intrapolant is exact at least up to order n.

Proof for polynomial of order $n+1$

For a monomial x^{n+1} the n-th order Taylor expansion is no longer exact, but the error is easy to find:

$$\mathcal{T}_n[x^{n+1}](x;\xi) = x^{n+1} - (x-\xi)^{n+1}. \tag{4.53}$$

The Dutch Taylor expansion becomes

$$\begin{aligned}\mathcal{D}_n[x^{n+1}](x;\xi) = x^{n+1} - (x-\xi)^{n+1} \\ - \sum_{k=1}^{n}\frac{k}{n+1}\binom{n+1}{k}(x-\xi)^k\xi^{n+1-k}.\end{aligned} \tag{4.54}$$

Using binomial identities (4.48) and (4.49), and the index $K = k-1$ this expression can be simplified to

$$\begin{aligned}\mathcal{D}_n[x^{n+1}](x;\xi) &= x^{n+1} - (x-\xi)^{n+1} - \sum_{K=0}^{n-1}\binom{n}{K}(x-\xi)^{K+1}\xi^{n-K} \\ &= x^{n+1} - (x-\xi)^{n+1} - (x-\xi)\left[x^n - (x-\xi)^n\right] \\ &= x^n\xi.\end{aligned} \tag{4.55}$$

Again, this is an expression that is linear in ξ. In fact, it is of the same form as the expression obtained for lower order polynomials in Equation (4.52), with $m = n+1$. Linear interpolation of this expression using the intrapolant (4.8) indeed returns the exact polynomial x^m for $0 \le m \le n+1$. Hence, the intrapolant is accurate to order $n+1$.

Chapter 5

Ray field construction and mapping algorithm

A ray field construction and mapping algorithm is developed that extends and refines existing wave front construction methods. A modular setup and a hierarchical description of the geometrical structure of the ray field make the algorithm widely applicable. It can be used for the calculation of ray fields in smooth 2-D and 3-D isotropic and general elastic media.

For ray field mapping on the spatial domain two refinements are proposed that enhance the accuracy and the completeness of the maps by higher order interpolation and improved delineation of caustics. Both refinements are easily included in existing wave front construction methods to enhance the efficiency.

Motivated by the success of wave front construction methods in the spatial domain, applicability in the position/angle domain is investigated as well. The unfortunate conclusion is that it is not suitable for that domain, due to the type of deformation in the ray field structure. In the position/angle domain this deformation is primarily shear-like while the algorithm is designed for spreading ray fields in the spatial domain.

5.1 Introduction

In Chapter 3 the advantages of the use of ray field maps in the position/angle coordinates for imaging in complex media have been discussed. An important question, from the viewpoint of practical applicability, is whether these maps can be calculated efficiently. The most straightforward algorithm traces a dedicated ray for each grid point in the position/angle domain. An advantage of this approach is that the implementation is relatively simple. A major disadvantage, however, is that it is inherently inefficient. This can be shown by looking at the computational dimension of the algorithm and the problem: a computation in $2N$

dimensions, i.e. N spatial coordinates, $N-1$ angles, and 1 dimension along each ray, is used to determine the eikonal manifold that has only $2N-1$ degrees of freedom (see Chapter 3).

The search for a more efficient algorithm naturally leads to the more traditional problem of calculating travel time maps for a point source on a spatial grid. As discussed in Chapter 3, the most powerful algorithms at this time seem to be the wave front construction algorithms (Vinje et al., 1993; Lambaré et al., 1996; Mispel, 2001; Coman and Gajewski, 2002).

This chapter presents a number of extensions and refinements of the wave front construction methods. The main characteristic of the proposed algorithm is its general applicability. A modular design and a hierarchical description of the geometrical structure of the ray field make it possible to use a single algorithm for calculations in both 2-D and 3-D media, ranging from acoustic to anisotropic elastic. Refinements are proposed with respect to the accuracy of the mapping to a regular grid, and the ray field sampling near caustics.

Despite the fact that the method is a variation of wave front construction it is referred to as *ray field construction*. Admittedly, choosing a different name for every variation of an existing algorithm is not beneficial for transparency. This new name, however, fits better in the context of this thesis, and it also reflects a slight difference in philosophy. The original wave front construction method (Vinje et al., 1993) was presented as an alternative to the eikonal equation solvers (Vidale, 1988; Podvin and Lecomte, 1991). The advantage of this alternative was a greater numerical stability and the possibility to find multiple arrivals as well as amplitudes. The ray equations were used, basically, to propagate an approximate wave front as the stationary phase resultant of Huygens' principle. Here, the goal is to solve the ray equations for any ensemble of rays that does not necessarily represent a propagating wave field.

Following the theory developed in Chapter 3 this chapter also investigates the applicability of the ray field construction algorithm in the position/angle domain. This investigation is motivated by observation that the only difference, from a practical point of view, is the higher dimensional range of initial conditions and the higher dimensional mapping domain. In fact, the dimension for a position/angle domain calculation in 2-D is equal to the dimension of a regular ray field construction in a 3-D medium.

5.2 Medium description and ray tracer

The ray field construction algorithm is designed in a modular fashion. This means that parts of the algorithm may easily be replaced without drastic changes in other parts. Two central modules are the *model*, i.e. the description of the medium properties, and the *ray tracer*, i.e. the ray equation solver, both are discussed below.

Model parameterisations

The model is described by two types of parameter. The most fundamental characteristics of the model are determined by its *physical parameterisation*. In isotropic elastic media, the propagation of a single wave type may, in the high frequency limit, be described in terms of a single physical parameter, such as the wave propagation velocity. If (relative) amplitudes are to be calculated it is important to know the mass density as well.

In more general – anisotropic elastic – media a larger number of parameters is required. The elastic tensor is described by at most 21 independent parameters (e.g., Červený, 2001), but usually this number is reduced by assuming a certain level of symmetry. Here, only examples of vertical transverse isotropy (VTI) in the Thomsen parameterisation (Thomsen, 1986) will be shown. This type of anisotropy has a vertical axis of symmetry and may be described in terms of 5 independent parameters, which allow the calculation of both quasi-P and quasi-S waves. Lower levels of symmetry such as orthorhombic or triclinic anisotropy may easily be accommodated.

The spatial variation of the physical parameters is determined by the model's *spatial parameterisation*. The spatial variations are required to satisfy a certain smoothness constraint. Depending on the type of ray information desired the spatial derivatives of the model should be continuous to first order for kinematic or second order for dynamic ray tracing. Obviously, this requirement does not allow the explicit specification of interfaces. For applications in reflection imaging this restriction is not a problem, because the background velocity models are usually assumed to be smooth. Vinje et al. (1996a) show how wave front construction methods can be extended to media that include interfaces.

The spatial parameterisation of the model is, technically, not relevant for the rest of the code. The model module must be able to return the desired physical parameters and their spatial derivatives at any position within a certain domain. For kinematic ray tracing only first order derivatives are needed, while for dynamic ray tracing second order derivatives are also required. For testing purposes it is often useful to define analytical models, such as those with constant gradients or Gaussian distributions. For practical applications a cubic spline parameterisation is a popular choice (e.g., Press et al., 1992).

Ray tracing

The type of ray equations to be solved depends on the physical parameterisation of the model, but not on its spatial parameterisation. For isotropic elastic and acoustic media the scalar velocity formulation is used, for more general elastic media the formulation with the full elastic tensor (see Chapter 2). Obviously, the latter is more complicated than the former, and more expensive to calculate. For the ray field construction both types of ray tracing are technically equivalent.

The numerical solution of the ray equations requires an ordinary differential

equation (ODE) solver. Many algorithms are available (e.g., Ralston and Rabinowitz, 2001; Press et al., 1992), but a number of practical considerations facilitate the choice. An important aspect is that the solution of the ray equations is not only desired at the ray's end point, but also at intermediate points along the way. The intervals of these intermediate solutions should be sufficiently small for accurate low order interpolation (see Section 5.4). Therefore, there is no advantage in using sophisticated multi-step or extrapolation methods that allow for large evaluation intervals.

Instead, fourth order Runge-Kutta (RK) integration with fixed step size is used. An implementation with variable step-size may be considered, but is complicated by the fact the rays are grouped in a ray field structure (Section 5.3.1). The entire ray field structure is propagated one step at the time. If the step size changes from ray to ray this leads to problems in the accuracy control (Section 5.3.2). The additional bookkeeping involved probably leads to a decrease of efficiency rather than an increase. A more practical alternative would be to add to the model parameters a local step size parameter, that smoothly decreases the step size towards the more complex parts of the medium. This approach has not been tested so far.

Along with the choice of the Runge-Kutta ODE solver comes the choice of the order of accuracy. The most commonly used RK solver is one with fourth order accuracy (Press et al., 1992). For a step length of size, say, h, this gives a local truncation error of fifth order: $\mathcal{O}(h^5)$. For a series of steps the local truncation errors accumulate to a global error of one order lower: $\mathcal{O}(h^4)$ (Cartwright and Piro, 1992). Hence, in theory, a fourth order RK solver has a fourth order rate of convergence. In practice, however, this is often not the case.

The order conditions only hold if the function to be approximated, in this case a ray variable, is sufficiently smooth. The most commonly used model parameterisation is based on cubic splines. These splines are designed in such a way that the model is continuous up to and including its second order derivatives. The third order derivatives, however, are discontinuous, and cause additional errors in the RK steps. This leads to a global error of $\mathcal{O}(h^3)$ for most of the ray variables, or $\mathcal{O}(h^2)$ for the paraxial perturbations, because the latter depend on the second order derivative of the model parameters.

Although a fourth order RK solver does not achieve a fourth order convergence rate in spline-based models, this does not mean that it should be replaced by one of lower order. In general, fourth order RK gives a lower error than a third order RK, albeit at the same convergence rate. The effect of decreasing step length, however, should not be overestimated.

5.3 Ray field construction

The ray equations and initial conditions described in Chapter 2 provide sufficient information for the forward calculation of ray fields in various types of media. The

ray fields themselves, however, are usually not the type of information required for applications in imaging and inversion. Rather, the desired information is a map of the ray field information, or *ray field map*, on a grid in the domain of interest. This domain of interest can be any combination of ray variables, but for most applications it is the spatial position. Here, the position/angle domain introduced in Chapter 3 will also be considered.

The calculation of a ray field map involves finding the inverse of the forward ray field calculation. Where the ray equations express the domain coordinates as a function of the ray parameters, the ray field map expresses the ray parameters and associated ray field information as a function of the domain coordinates.

Let the domain coordinates be denoted by $\boldsymbol{y}$, the grid points in that domain denoted by $\boldsymbol{y}_I$, the flow parameter by σ and the ensemble parameters by $\boldsymbol{\gamma}$. The mapping may be expressed as the solution of a system of equations:

$$\boldsymbol{y}_I \mapsto \{(\sigma, \boldsymbol{\gamma}) \,|\, \boldsymbol{y}(\sigma, \boldsymbol{\gamma}) = \boldsymbol{y}_I\}, \tag{5.1}$$

where it should be noted that the number of domain coordinates $\boldsymbol{y}$ is always equal to the number of ray field coordinates $(\sigma, \boldsymbol{\gamma})$.

In general circumstances, the mapping (5.1) is non-linear. It may even be multi-valued, such as in the case of a point source travel time map in the presence of multi-pathing. The most practical ray tracing methods approach the mapping in a similar way. Each step in the forward calculation is followed by a local, low order approximation. Using this approximation the mapping is performed for the grid points $\boldsymbol{y}_I$ that lie in a certain neighbourhood.

In the paraxial ray shooting method (Červený et al., 1984; Beydoun and Keho, 1987) the local approximation is based on extrapolation from a single ray, using paraxial perturbations. For applications in complex media this approach has two disadvantages. First, the occurrence of regions of large geometrical spreading makes it difficult to find rays in the neighbourhood of all of the map's grid points. Second, in the presence of multi-pathing the results of extrapolation become highly ambiguous. It is practically impossible to distinguish if two estimates correspond to the same or to two different arrivals.

Wave front construction methods, introduced by Vinje et al. (1993), extend the paraxial ray shooting method by keeping track of the geometrical structure of the wave front spanned by the rays. Knowledge of the structure of the wave front allows a high degree of accuracy control, by inserting new rays whenever the sampling of the wave front becomes too sparse. This solves both problems associated with the paraxial ray shooting method mentioned above. First, the lack of sampling in regions of large geometrical spreading is prevented by inserting a sufficient number of rays dynamically. Second, the ambiguity of multiple arrivals is solved, because the estimation of arrivals is performed by means of interpolation rather than extrapolation. The maintenance and accuracy control of the wave front structure obviously has a price, but the result is a complete and unambiguous, multi-valued ray field map.

Here the same approach is followed as in the wave front construction methods. The current section discusses the forward calculation of the ray field: the definition of the ray field structure and its propagation in Section 5.3.1, and the mechanism for accuracy control in Section 5.3.2. A number of examples of the forward calculation are shown in Section 5.3.3. The subsequent calculation of the ray field maps is discussed in Section 5.4.

5.3.1 Ray field structure and propagation

Ray fronts

For a fixed value of the flow parameter σ, the ray variables are a function of the ensemble parameters $\boldsymbol{\gamma}$ only. If time is used as the flow parameter for the calculation of a point source ray field, then $\boldsymbol{x}(\sigma = T, \boldsymbol{\gamma})$ describes the spatial position a wave front. In the following the term *ray front* is used for a generalisation of this concept: a ray front is a set of ray field variables at a constant value of σ. The calculation of a ray field may therefore be regarded as the propagation of ray fronts, parameterised by the ray ensemble parameters $\boldsymbol{\gamma}$.

The complexity of the geometrical structure of the ray front depends on the number of ensemble parameters. For a point source in a 2-D medium a single ensemble parameter is required, which may be, for example, the take-off angle at the source. In this case, a ray front is a curve, and the geometrical structure consists of small segments of the ray front connecting neighbouring rays (e.g., Vinje et al., 1993; Lambaré et al., 1996).

A ray field emitted by a point source in a 3-D medium is described by two ensemble parameters. Here, take-off angles are a natural choice as well, although there is no parameterisation in terms of only two angles that is non-degenerate over the entire unit sphere. Where necessary a set of three projection angles is used, as described in Chapter 3. The ray front for this situation is a 2-D manifold, and the natural geometric structure is a triangulation (e.g., Vinje et al., 1996a; Lucio et al., 1996).

For calculations in the position/angle domain two or four ensemble parameters must be used, for 2-D and 3-D media, respectively (see also Chapter 3). Although in the examples of this chapter not more than two ensemble parameters will be used, a hierarchical geometric structure for the ray fronts is proposed that may be used to accommodate an arbitrary number of ensemble parameters.

Hierarchical structure

The proposed hierarchical geometrical structure of the ray front is based on simplices. A simplex, or N-simplex, is the N-dimensional equivalent to a triangle in 2-D and a tetrahedron in 3-D (see Figure 5.1). A 1-simplex is a line segment; a 0-simplex is a point. At a ray front each ray is represented by a point, which can be seen as the lowest, zeroth level of geometrical structure. If only a single ray is traced, the geometrical structure of the ray front is limited to that point. A ray

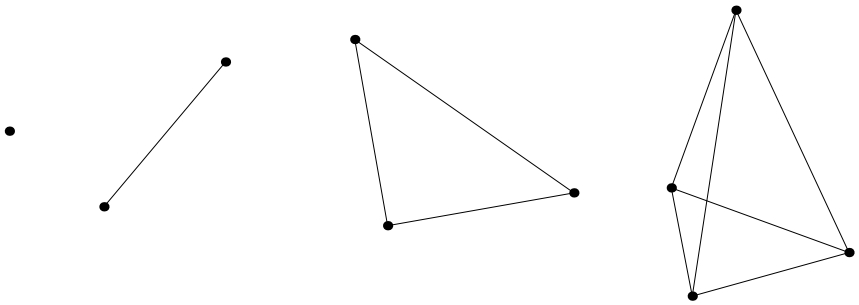


Figure 5.1: From left to right simplices of dimension $N = 0$ to $N = 3$. An N-simplex has exactly $N + 1$ vertices and $N + 1$ sides. Each side is an $(N - 1)$-simplex. For simplices inside a ray front each vertex corresponds to a ray.

field based on a single ensemble parameter has an additional level of structure. This level consists of the segments (1-simplices) that connect the neighbouring rays on the ray front. Adding more parameters leads to successively higher levels of structure. The maximum level that considered here, is the level of triangles, or 2-simplices.

The hierarchical character of the geometrical structure lies in the property that higher level structure "is aware of" the lower level structure, but not the other way around. The main advantage of this approach is that the data structures that implement the geometric structure are simple, and can be maintained with a relatively small amount of bookkeeping. This is due the fact that every N-simplex has exactly $N + 1$ vertices and $N + 1$ sides (see Figure 5.1). In contrast, it is not known a priori for an N-simplex of how many higher level simplices it is a part. For example, a ray in a 2-parameter ray front may be connected to other rays by an arbitrary number of segments.

The data structure of a ray front contains arrays of simplices for every structural level defined. The number of levels is equal to the number of ensemble parameters plus one. The zeroth level, consisting of the rays, contains the crucial ray field information. Higher level simplices mainly contain pointers (indices) to each of its side-simplices at the level beneath, and to each of the rays at its vertices. In this way, every ray and every other structural element needs to be present only once.

Propagation and initialisation

The propagation of a ray front is executed by following a hierarchical chain of commands. The chain starts with a loop over the simplices of the highest structural level. Each of these simplices is propagated by recursive delegation of the propagation command to lower level structures.

The recursive process is explained by the fragment of pseudo-code in Figure

```
function
propagate (simplex, level)
{
    if (level == 0)
        step_ray (simplex);
    else
    {
        for (each side_simplex)
            if (not yet propagated)
                propagate (side_simplex, level-1);
        step_simplex (simplex);
    }

    return;
}
```

Figure 5.2: Pseudo-code for a recursive simplex propagator without accuracy control. The recursion is started at the highest level structure.

5.2. At the zeroth level, a call of `propagate()` is implemented as a single ray step to the new ray front. At higher levels the `propagate()` command is first delegated to the side-simplices, before the main simplex is stepped to the new ray front. An `if`-statement is included to prevent that a simplex is propagated more than once. The reason for this recursive implementation becomes clear in Section 5.3.2, where an accuracy control mechanism is inserted.

It may be noted that in this hierarchical, recursive approach every ray, segment, triangle etc. is treated in exactly the same way, regardless of the structure in which it is embedded.

At the start of the algorithm the ray front structure is initialised. In general, the ranges of initial conditions form a rectangular domain in the space of ensemble parameters. The initial rays can be organised in a rectangular grid. The spacing should be in accordance with the sampling criteria of Section 5.3.2.

5.3.2 Accuracy control

The accuracy control of the ray front structure takes place at the first level of the line segments connecting two rays. If two connected rays diverge, the accuracy of low order interpolation between these two diminishes. In order to maintain sufficient accuracy in the sampling of the ray front, the segment has to be split. This is achieved by inserting a new ray in between, as sketched in Figure 5.3. In the following, three algorithmic ingredients of the accuracy control are discussed: first, the definition of a sampling criterion; second, a ray insertion method; and third, a mechanism that adapts the ray front structure to the presence of the newly inserted ray.

Sampling criteria

The decision to insert a new ray is made on the basis of a sampling criterion. In the wave front construction methods two basic types of criterion have been proposed. The first type is a *distance criterion*, which defines an upper limit on the distance between two rays. Apart from a limit on the spatial distance (Vinje et al., 1993) a limit may also be defined for the angular distance, or the distance in slowness components (Sun, 1992; Lambaré et al., 1996; Vinje et al., 1996a).

Alternatively, a *misfit criterion* can be applied (Lambaré et al., 1996; Lucio et al., 1996). At each of the two rays paraxial derivatives are used to calculate an approximation of the other ray. The misfit of the approximate position and slowness components is a measure for the local curvature of the ray field on the length scale of the segment. If the misfit exceeds a certain limit a new ray has to be inserted.

The two types of criterion correspond to two different philosophies of error management. A distance criterion, on the one hand, aims to prevent errors. This is achieved by looking at the origin of errors. Variations in the ray front are caused by the forcing terms (right hand sides) of the ray equations. These terms contain the medium properties. If the medium properties are adequately sampled along the ray front the errors are small. A distance criterion may therefore be based on the minimum length scale of variations in the medium properties. For ray tracing in smooth media this length scale is commonly known. Additionally, an angular distance criterion, that is, a maximum difference in the angle of the slowness vector, should be defined. In anisotropic media this is essential, because the medium properties vary with angle. In isotropic media, lack of such a criterion leads to insufficient sampling in the neighbourhood of caustics (Lambaré et al., 1996).

A misfit criterion, on the other hand, controls the error by making a direct measurement of the second order error term in a linear approximation. Good results with this type of criterion are reported by Lambaré et al. (1996), who show that the accuracy of sampling in the neighbourhood of caustics is much better than in the case of using a distance criterion. A disadvantage is that the numerical limits of the criterion are much harder to determine a priori. A limit for the distance criterion is easily chosen as a fraction of a minimum length scale of variation, and a maximum angle limit may even be chosen without knowledge of the medium properties. Determining limits for the misfit requires at least some analysis, tuning or experience.

In the ray field construction both types of criterion are implemented. A slight difference with the studies mentioned above is that a metric is used to measure distance or misfit. If $\Delta \boldsymbol{x}$ is the spatial distance or misfit, $\Delta\phi$ the angular distance or misfit, and $\Delta_x^{\max}$ and $\Delta_\phi^{\max}$ the respective upper limits, then the metric distance D is defined as:

$$D = \sqrt{\left(\frac{|\Delta \boldsymbol{x}|}{\Delta_x^{\max}}\right)^2 + \left(\frac{\Delta\phi}{\Delta_\phi^{\max}}\right)^2}. \tag{5.2}$$

If $D > 1.0$ the misfit criterion is violated. For slowness vectors $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, the enclosing angle $\Delta\phi$ is defined as

$$\cos(\Delta\phi) = \frac{\boldsymbol{p}_1 \cdot \boldsymbol{p}_2}{|\boldsymbol{p}_1||\boldsymbol{p}_2|}. \tag{5.3}$$

If the two slowness vectors correspond to two different rays then $\Delta\phi$ refers to the angular distance; if it involves the paraxial estimate of the other $\Delta\phi$ is the angular misfit.

Ray insertion

Once a violation of the sampling criterion occurs, that is $D > 1.0$, a new ray is inserted. In Figure 5.3 the situation is sketched for two rays A and B, with ensemble parameters $\boldsymbol{\gamma}_A$ and $\boldsymbol{\gamma}_B$ respectively. The rays are situated at points A_1 and B_1 at ray front σ_1 and connected by the segment A_1B_1. During the propagation both rays are successfully stepped to points A_2 and B_2 at ray front σ_2, but upon stepping segment A_1B_1 to A_2B_2 at σ_2 the sampling criterion is violated. Therefore, a new ray C is inserted at point C_1 at ray front σ_1, halfway in between A and B in terms of ensemble parameters. The new ray's ensemble parameters are the average of those of A and B:

$$\boldsymbol{\gamma}_C = \frac{1}{2}(\boldsymbol{\gamma}_A + \boldsymbol{\gamma}_B). \tag{5.4}$$

New segments A_1C_1 and B_1C_1 are initiated and in this case both ray C and the new segments are successfully stepped to σ_2.

The most difficult part of the insertion process is the accurate determination of the initial conditions at the point of insertion (C_1 in Figure 5.3). Three approaches to this problem have been proposed. Firstly, Sun (1992) and Ettrich and Gajewski (1996) use a circular approximation to a wave front in 2-D. The midpoint of the circle is the intersection point of the lines tangent to the slowness vectors of both parent rays. A disadvantage of this technique is that it does not generalise to 3-D.

Secondly, Vinje et al. (1993) and Lambaré et al. (1996) use the paraxial derivatives at the parent rays to constrain a third order interpolating polynomial . This results in the following equation for the initial position $\boldsymbol{x}_C$:

$$\boldsymbol{x}_C = \frac{1}{2}\left[(\boldsymbol{x}_A + \boldsymbol{x}_B) - \frac{1}{4}(\boldsymbol{\gamma}_B - \boldsymbol{\gamma}_A) \cdot \left(\left.\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\gamma}}\right|_B - \left.\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\gamma}}\right|_A \right)\right]. \tag{5.5}$$

The initial slowness vector and travel time can be obtained in a similar way by replacing $\boldsymbol{x}$ with $\boldsymbol{p}$ or T in the equation above. The interpolation formula (5.5) is a classical Hermite interpolation (see also Section 4.4.1), evaluated at the centre of the interval. At this location it gives the same result as a second order intrapolation (defined in Chapter 4). Other quantities besides $\boldsymbol{x}$, $\boldsymbol{p}$ and T are interpolated linearly.

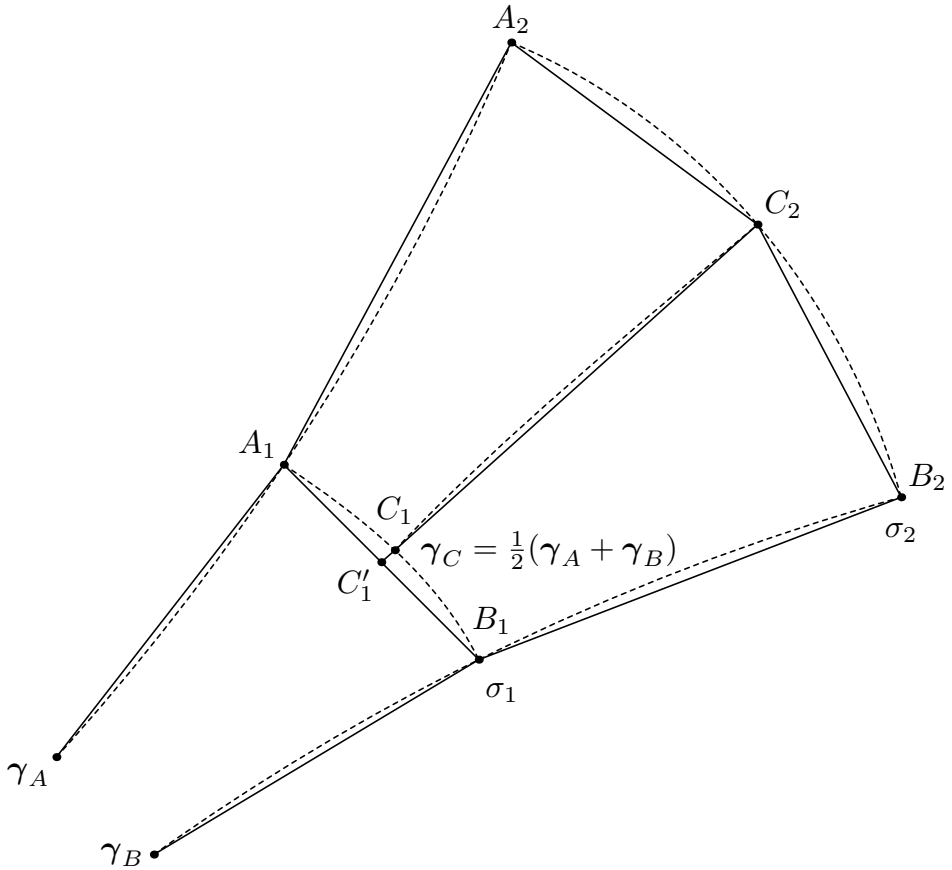


Figure 5.3: Accuracy control mechanism acting on a ray front segment. Dashed lines indicate exact ray and ray front curves, continuous lines indicate linear approximations. Two rays with ensemble parameters γ_a and γ_b may be embedded in a ray field of arbitrary dimension. The rays are connected by a ray front segment at $\sigma = \sigma_i$. Both rays are propagated to $\sigma = \sigma_{i+1}$, where a sampling criterion is evaluated and in this case violated. An infill ray is inserted at point C_1 by accurate interpolation on ray front σ_i. The infill ray is traced to σ_{i+1}. Subsequently the ray point C_1 is replaced by the linearly interpolated C_1' to prevent gaps in the coverage of the mapping domain.

Thirdly, a ray can be traced from the source to the point of insertion. Obviously, tracing an entire ray is more expensive than interpolation, but it removes the ray interpolation as a source of errors. Coman and Gajewski (2002) claim that this approach allows a relaxation of the sampling criterion, resulting in a net gain in efficiency.

While essential for the quality of the ray field, the accuracy of the ray insertion introduces a minor problem for the quality of the mapping discussed below (Section 5.4). The fact that the point of insertion C_1 in Figure 5.3 is not located on the straight line segment A_1B_1, causes a gap in the ray field in the shape of the triangle $A_1B_1C_1$. In order to prevent this, the ray point C_1 has to be replaced by a point C_1' on the segment A_1B_1. This can be done safely after the ray has been propagated to C_2. In principle, all quantities can be interpolated linearly to C_1'. However, since an accurate method for mapping travel times will be proposed in Section 5.4, it is important that both travel time T and slowness vector P are interpolated accurately at this point as well. For the slowness vector this is achieved by interpolating the length of the slowness vector as well as its components. The interpolated components are then scaled to obtain an interpolated slowness vector of the right length. Hermite interpolation using the slowness vectors yields an estimate of second order accuracy for the travel time.

Recursive structure adjustment

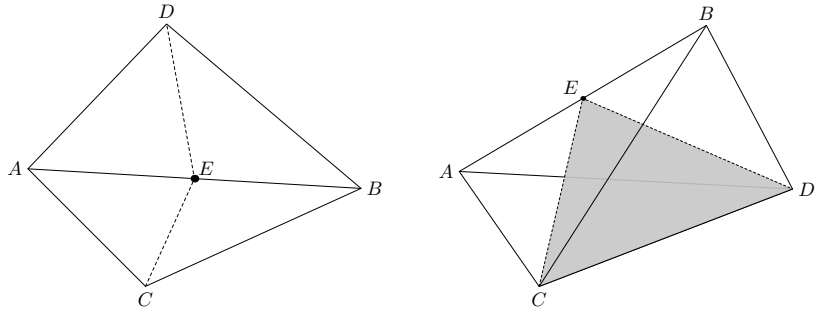
After inserting a new ray on a segment the geometrical structure in the neighbourhood of that segment has to be adjusted. In Figure 5.4 examples are shown of adjustments that have to be made in 2- and 3-parameter structures. If a ray front segment is split in two, every higher level simplex that contains this segment has to be split as well. The adjustments are applied recursively during propagation. The workings of this mechanism are demonstrated using a fragment of pseudo-code in Figure 5.5. It is an extension of the pseudo-code shown in Figure 5.2, with the accuracy control mechanism included. Along with the rays at level zero, also the segments at level one require special treatment. The variable `splitflag` indicates if a ray has been inserted and a structure has been split. If the `splitflag` of one of the sides of a simplex has been set, the simplex itself has to split as well.

5.3.3 Examples

Point source ray fields

Four examples of ray field construction for points sources, the classical wave front construction, are given in Figure 5.6. Issues of accuracy and selection of limits for the sampling criteria are not discussed here. More details on this subject can be found for example in Lambaré et al. (1996). Here, some examples are shown of how the same algorithm works in 2-D and 3-D, isotropic and anisotropic media.

The first example is a ray field in a homogeneous isotropic medium in Figure 5.6(a). Only a distance criterion is used that includes both spatial and angular



(a) 2-D recursive subdivision (b) 3-D recursive subdivision

Figure 5.4: After inserting a new ray E at the segment AB, the geometrical structure of the ray front has to be adapted. Segment AB has to be split, as well as every higher level simplex that contains the segment AB. Figure (a) shows a case of two ensemble parameters. Two triangles contain the segment AB, and both have to be split. This leads to new segments AE, BE, CE, and DE, as well as new triangles ACE, ADE, BCE, and BDE. Figure (b) shows a single tetrahedron in a three parameter case. Two sides, as well as the tetrahedron itself have to be split due to the insertion of the new ray at E. Among the new triangles is CDE.

```
function propagate (simplex, level) returns flag {
    flag splitflag = NO;

    if (level == 1) /* segment: side_simplices are two rays */
    {
        for (both rays)
            if (not yet propagated)
               step_ray (side_simplex, 0);
        splitflag = accuracy_test (both rays);
    }
    else
        for (each side_simplex)
        {
            if (not yet propagated)
                splitflag = propagate (side_simplex, level-1);
            else if (side_simplex was split)
                splitflag = YES;
        }

    if (splitflag == YES)
        split_simplex (simplex);
    else
        step_simplex (simplex);

    return splitflag;
}
```

Figure 5.5: Pseudo-code for a recursive simplex propagator with accuracy control. The recursion is start at the highest level structure. The function is not called for level 0.

limits. At the source the number of rays is determined by the angular distance limit. During propagation the spatial distance between the rays increases and once it exceeds the limit new rays are inserted.

A simple example of ray field construction in 3-D is presented in Figure 5.6(b), where three partial wave fronts are plotted in a VTI medium. More examples of 3-D wave front construction in isotropic media can be found in Vinje et al. (1996a) and Lucio et al. (1996).

The extension to anisotropic media is well illustrated by an example of both quasi-P and quasi-S waves in a 2-D VTI medium, shown in Figures 5.6(c) and (d) respectively. The quasi-P ray field in 5.6(c) clearly shows the rays propagating at oblique angles with the wave front within the VTI anomaly. Where the anomaly is small the rays and wave fronts are perpendicular again. The polarisation of the motion is also slightly out of line with the ray direction. The ray field in Figure 5.6(d) shows the quasi-S waves that a 2-D medium would support, i.e. those with a polarisation in the 2-D plane (equivalent to S2 or SV in a 3-D medium). The ray field shows similar anisotropic behaviour as its quasi-P counterpart, only the variations in polarisation direction are more pronounced.

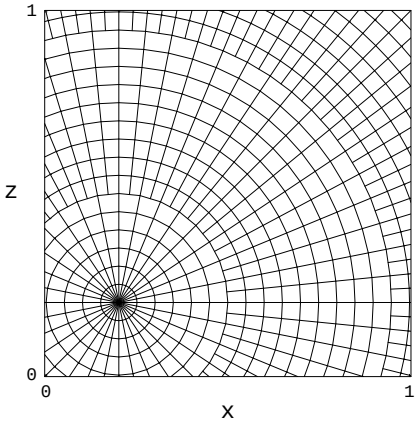
The sharp-sighted reader may notice that the onsets of some inserted rays are slightly displaced. This is the result of the linear interpolation that puts the initial ray point on the ray front segments after a more accurate estimate has been used to calculate the ray at the next ray front (see Section 5.3.2).

Position/angle domain ray fields

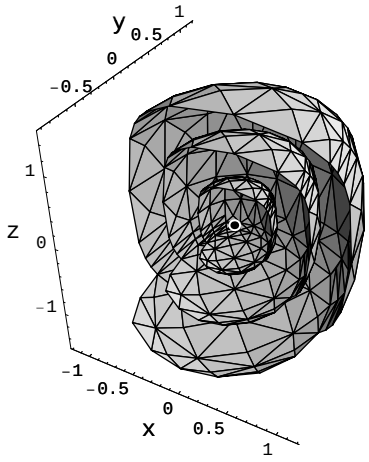
One of the primary purposes of this chapter in the context of this thesis is to investigate the applicability of the ray field construction algorithm in the position/angle domain. Figure 5.7 shows two examples, one in a homogeneous medium and one in a medium with a slow Gaussian anomaly. For both situations three ray fronts are shown that correspond to range of source position at the $z = 1.0$ line, in Figures 5.7(a) and (c) respectively. For reference the wave fronts and rays of a single source at $x = 0.5$ are drawn as well.

At first view there does not seem to be a problem. Although the ray fronts look a bit different from the wave fronts in a normal 3-D wave front construction (e.g. Figure 5.6(b)), the ray fronts are calculated in a similar way. A look at Figure 5.8, however, makes clear that there is a problem. In this Figure the triangulations of the ray fronts of Figure 5.7 are shown in a projection on the (x, ϕ)-plane. Where the deformation of the geometrical structure of a 3-D wave front is mainly characterised by expansion, it is clear that in this case the main type of deformation is shear-like. In hindsight, this was to be expected from the observation made in Chapter 3 that geometrical spreading is very limited in the eikonal manifold.

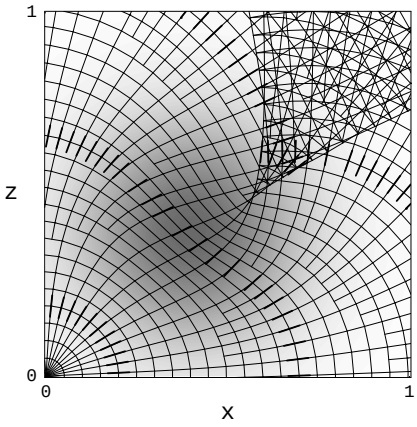
The shearing of the ray front leads to stretching of the segments connecting rays that propagate at different angles. To relax the shear tension new rays are inserted. Since there is little spreading the density of rays quickly rises. Moreover,



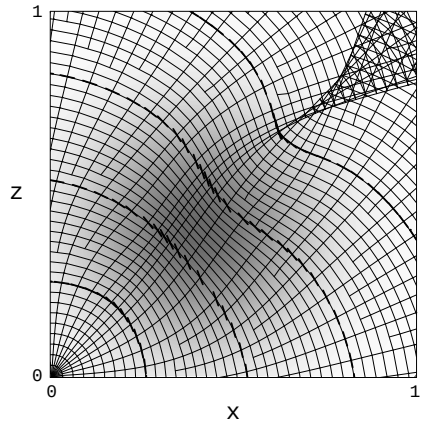
(a) Homogeneous isotropic 2-D medium



(b) Homogeneous anisotropic 3-D medium, partial quasi-P wave fronts



(c) Anisotropic 2-D medium with Gaussian anomaly, quasi-P rays.



(d) Anisotropic 2-D medium with Gaussian anomaly, quasi-S rays.

Figure 5.6: Four examples of the ray field construction algorithm. Figure (a) shows the ray field in a homogeneous medium. Figure (b) shows three partial quasi-P wave fronts in a homogeneous 3-D medium. The medium is anisotropic: VTI with Thomsen parameters $V_P = 1.0$, $V_S = 0.6$, $\epsilon = -0.2$, $\delta = 0.2$, and $\gamma = 0.0$. Figure (c) and (d) show ray fields in a 2-D VTI medium with a Gaussian anomaly. The Thomsen parameters of the isotropic background are $V_{P0} = 1.0$, $V_{S0} = 0.6$, $\epsilon_0 = 0.0$, $\delta_0 = 0.0$, and $\gamma_0 = 0.0$; the anisotropic Gaussian anomaly has parameters $V_{P1} = -0.5$, $V_{S1} = -0.3$, $\epsilon_1 = 0.5$, $\delta_1 = 0.0$, and $\gamma_1 = 0.0$. The spatial variation of the anomaly is proportional to $\exp((-(x-0.4)^2 - (z-0.4)^2)/0.3^2)$. Figure (c) and (d) show the quasi-P and quasi-S ray fields respectively. The line segments plotted on a number of wave fronts indicate the polarisation directions.

the introduction of new rays entails the introduction of new segments, which immediately start to stretch again. The result is a very high number of rays and very narrow triangles. It turns out that ray field construction algorithm, although very efficient for diverging ray fields, is not very suitable for the shear-like deformations in a position/angle domain ray field.

5.4 Ray field mapping

In the previous section it was shown how to calculate a ray field with controlled accuracy by the ray field construction algorithm. To obtain ray field information in a form that is convenient for applications in imaging, the forward ray field calculation must be followed by a ray field mapping. This mapping can be cast as the solution of a non-linear inverse problem, as expressed in Equation 5.1.

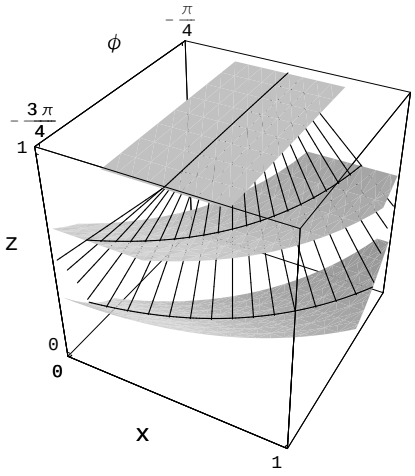
As shown by the examples above, the ray field construction algorithm is not very well suited for applications in the position/angle domain. With regard to the mapping process the discussion can therefore be limited to the classical wave front construction application, that is, the calculation of travel time and amplitude maps in the spatial domain.

In wave front construction methods the mapping is performed locally, by partitioning the ray field into small ray field cells (e.g., Vinje et al., 1993; Lambaré et al., 1996). The cells are formed by the intersections of a group of rays with two consecutive ray fronts. The grouping of the rays is determined by the geometrical structure of the ray front. In 2-D two rays are connected by a ray front segment, in 3-D three rays by a ray front triangle. The local mapping is applied only to the mapping points inside the ray cell. Figures 5.9 and 5.10 show ray field cells in 2-D and 3-D respectively, along with a further decomposition of the ray cells into simplices, to be discussed below.

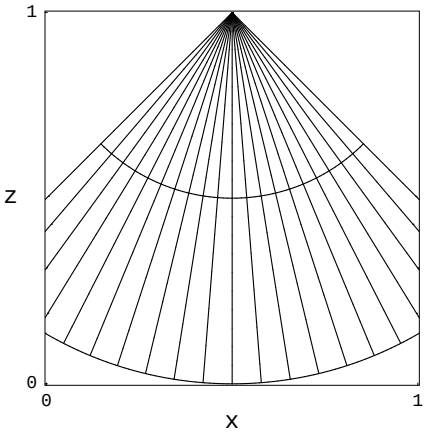
In the following two refinements of the ray field mapping process are proposed. First, in Section 5.4.1, it is demonstrated how *intrapolation*, defined in Chapter 4, and the availability of slowness vectors can be used to enhance the accuracy of the mapping. Second, in Section 5.4.2, it is shown how the partitioning of the ray field in the neighbourhood of caustics can be improved by means of *accurate caustic delineation*, leading to a more complete estimation of the number of arrivals.

5.4.1 Accurate mapping using intrapolation

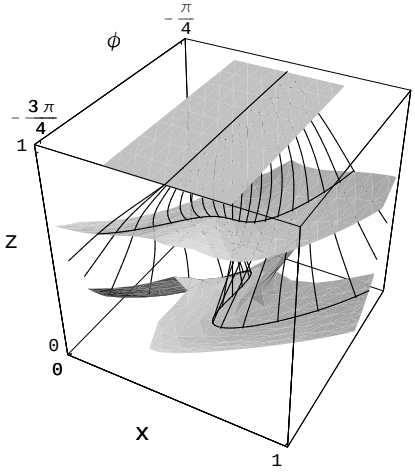
Although the mapping to grid points inside the ray cells is often referred to as interpolation, it is more correctly described as *inverse interpolation*. Since the spatial coordinates of the ray field are calculated as a function of the ray parameters, the process of finding the ray parameters at a given grid point comes down to solving a non-linear system of equations (5.1). If carried out in the neighbourhood of a single solution this process is sometimes referred to as inverse interpolation (e.g., Ralston and Rabinowitz, 2001).



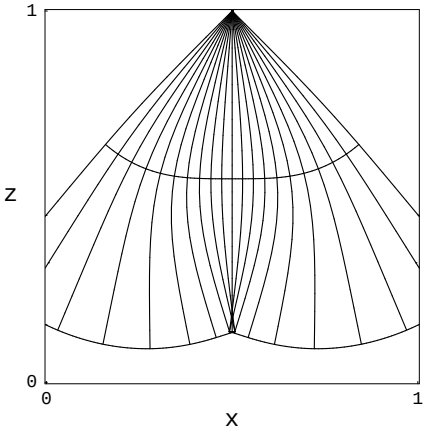
(a) Three position/angle domain ray fronts in homogeneous medium



(b) Wave fronts and rays for a fixed source position in a homogeneous medium



(c) Three position/angle domain ray fronts in a medium with a Gaussian anomaly



(d) Wave fronts and rays for a fixed source position in a medium with Gaussian anomaly

Figure 5.7: Figures (a) and (c) display three ray fronts in position angle domain. The range of initial positions is $x_0 \in [0.25, 0.75]$, the range of initial angles $\phi_0 \in [-\pi/4, -3\pi/4]$. Figure (a) shows the ray fronts in a homogeneous medium, Figure (b) those in the same medium with a slow gaussian anomaly superimposed. Both plots also show the rays and wave fronts for a single source at $x_0 = 0.5$, as displayed in Figures (b) and (d) respectively. The triangulations of the ray fronts are shown in Figure 5.8.

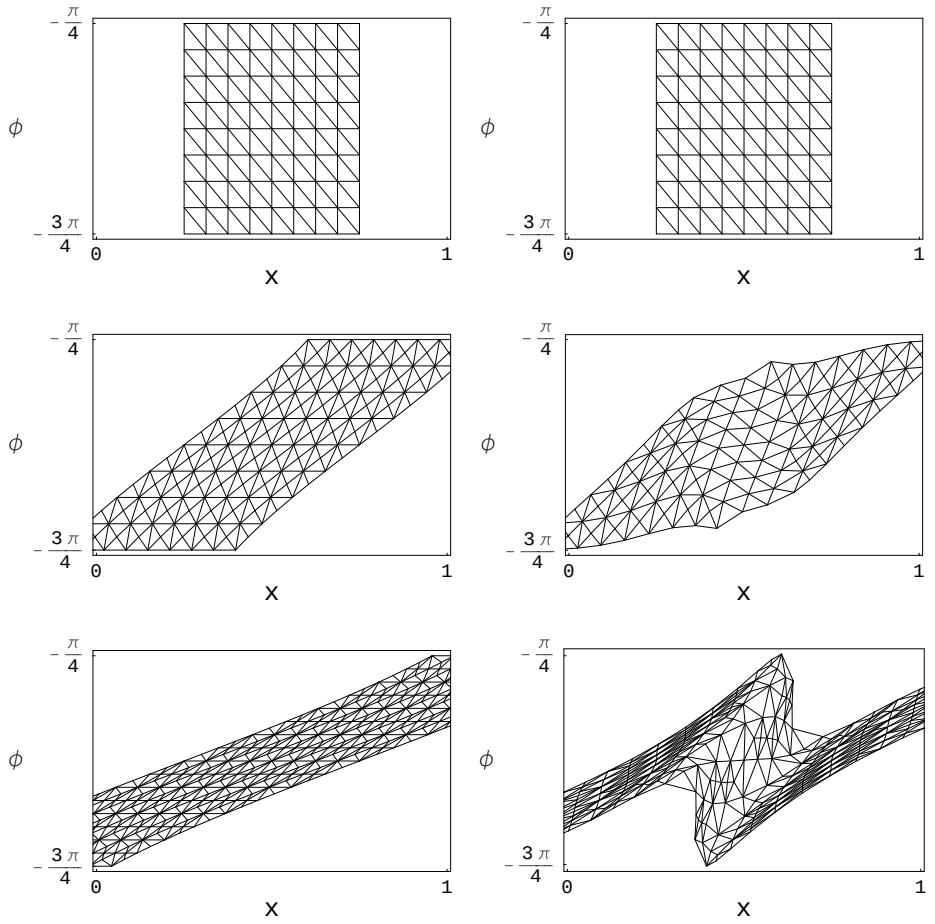


Figure 5.8: Triangulations of the position/angle domain ray fronts displayed in Figure 5.7(a) and (b), viewed in a projection on the (x, ϕ)-plane. The left column corresponds to the homogeneous medium of 5.7(a), the right column to the medium with the Gaussian anomaly of 5.7(b). An important observation is that the main type of deformation in the geometrical structure of the ray fronts is shear-like. The result is a rapidly increasing number of narrow ray front triangles. The effect is stronger in the presence of the Gaussian anomaly. The ray front construction algorithm is not suited to accommodate this type of deformations in the ray front.

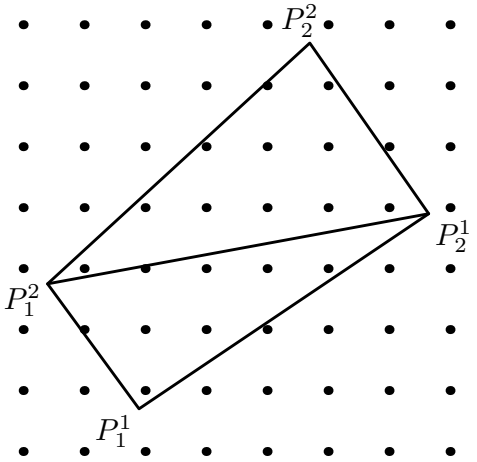


Figure 5.9: Mapping of ray field information on a 2-D grid. Two neighbouring rays and two consecutive ray fronts form a quadrangular ray field cell. The cell is determined by the four points P_i^j, where i is a ray front index, and j a ray index. The quadrangular cell is subdivided into two triangles, to allow a local linear approximation of the ray field. Each triangle is used to calculate the mapping only for the grid points it contains.

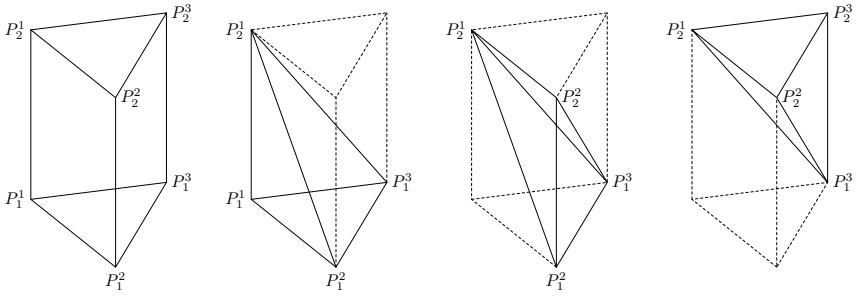


Figure 5.10: A ray field cell in 3-D is determined by three rays and two ray fronts. The vertices of the cells are denoted by P_i^j, where i is a ray front index, and j a ray index. Before interpolation the cell is subdivided in three simplices as shown.

The main advantage of partitioning the ray field into small cells is that locally the ray field may then be approximated by a low order polynomial, which facilitates inverse interpolation. Bulant and Klimeš (1999) show that the ray field within a 3-D ray field cell can be approximated by a single multivariate polynomial and that the corresponding equation can be solved to find up to three solutions for each grid point in the cell.

A simplification of the mapping process is achieved by further subdivision of the ray field cells into simplices (see Figures 5.9 and 5.10), as suggested by Lambaré et al. (1996) and Lucio et al. (1996). Within each of these simplices the ray field can be approximated by a linear function, which turns mapping (5.1) into a linear system of equations. The linearity ensures that there is only one solution for each grid point. Note that since the ray cell is divided into three simplices it is still possible to find three solutions for the entire cell. Nevertheless, in the presence of multi-pathing the linearisation may lead to an inadequate delineation of the caustic and consequently an underestimation of the number of arrivals. An effective way to deal with this problem is introduced in Section 5.4.2.

The distinction between interpolation and inverse interpolation is less pronounced in the case of linear functions. Inverse interpolation in the space of ray parameters can be seen as interpolation in the spatial domain. However, even in the latter case a linear system of equations has to be solved. A convenient way to do this is by barycentric interpolation (Appendix D.2), because it also provides a simple criterion to determine whether a grid point is located inside the simplex. The linear system to be solved in this case is the system that yields the barycentric coordinates (D.13).

For a simplex with vertices at $\boldsymbol{x}_i$ ($i = 0 \dots N$), with N is 2 or 3, the barycentric interpolation of the travel time T at $\boldsymbol{x}$ can be expressed as

$$T(\boldsymbol{x}) = \sum_{i=0}^{N} \alpha_i T(\boldsymbol{x}_i), \tag{5.6}$$

with α_i the barycentric coordinates of point $\boldsymbol{x}$ (see Appendix D.2).

Here the goal is to enhance the accuracy of the mapping in the simplex, because this will allow a relaxation of the ray field sampling criteria, which, in turn, leads to a faster algorithm. The most straightforward approach to achieve this is to approximate the ray field in the simplices by higher order polynomials, constrained by the paraxial derivatives. Obviously, this leads back to a mapping defined by a non-linear system of equations, which is much more difficult to solve. An advantage of this approach, however, would be an improvement of the delineation of the caustics.

A more practical alternative is to use higher order interpolation in the spatial domain, using available spatial derivatives and intrapolation as defined in Chapter 4. This approach is valid only if the ray field mapping inside the simplex is single-valued, i.e. if all vertices of the simplex belong to the same ray field branch, that

is, if the simplex does not contain a caustic. The handling of caustics inside a simplex is treated in Section 5.4.2.

The accuracy of the barycentric interpolation inside a simplex in the spatial domain can be improved using intrapolation. The availability of first order derivatives leads to second order accuracy. For travel times the gradient is available in the form of the slowness vector $\boldsymbol{p}$, and the travel time interpolation (5.6) can be replaced by

$$T(\boldsymbol{x}) = \sum_{i=0}^{N} \alpha_i \left(T(\boldsymbol{x}_i) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_i) \cdot \boldsymbol{p}(\boldsymbol{x}_i) \right), \tag{5.7}$$

which is accurate up to second order, as proven in Chapter 4.

A further enhancement of accuracy, primarily in the neighbourhood of the source can be achieved by interpolating the squared travel time rather than the travel time itself. This is often referred to as *hyperbolic interpolation* (see Vanelle and Gajewski (2002) and Section 4.4.2):

$$T(\boldsymbol{x}) = \sqrt{\sum_{i=0}^{N} \alpha_i T(\boldsymbol{x}_i) \left(T(\boldsymbol{x}_i) + (\boldsymbol{x} - \boldsymbol{x}_i) \cdot \boldsymbol{p}(\boldsymbol{x}_i) \right)}. \tag{5.8}$$

An example of the second order accurate hyperbolic travel time mapping is provided in Figure 5.11. Because of the large curvature of the wave fronts, the source region is a challenge for travel time interpolation. As shown, the errors in the travel time map are strictly caused by finite precision. That the travel time mapping is exact is due to three computational features: (i) the use of hyperbolic interpolation; (ii) the second order accurate intrapolation; and (iii) the accurate interpolation of travel time and slowness vectors inserted at the ray front segments (see Section 5.3.2).

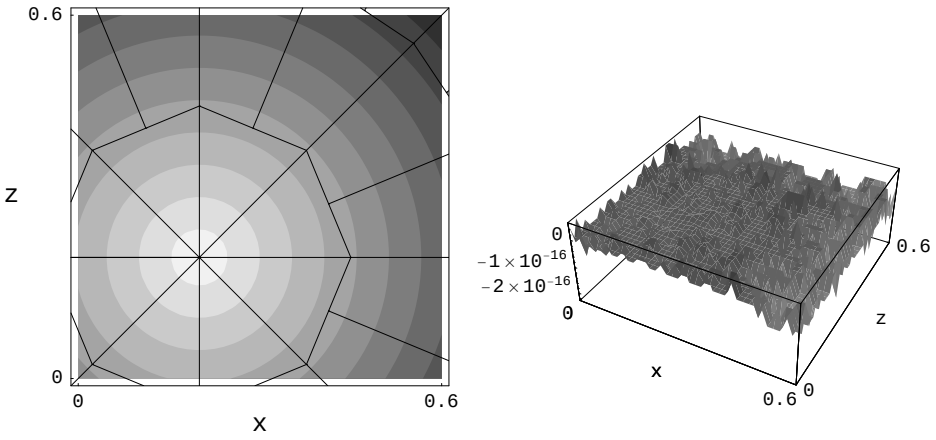
For variables other than travel time, the spatial derivatives are not directly available in the ray field. The availability of paraxial derivatives makes it possible to find these by a simple calculation. The gradient of the slowness vector for example can be found from the paraxial derivatives of both spatial position and slowness:

$$\partial_{\boldsymbol{x}} \boldsymbol{p} = \left(\frac{\partial \boldsymbol{p}}{\partial \boldsymbol{\gamma}} \right) \cdot \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{\gamma}} \right)^{-1}. \tag{5.9}$$

If desired, this slowness gradient, the second order travel time derivative, can also be used to find a third order accurate travel time.

5.4.2 Accurate caustic delineation

The mapping of a ray field in the neighbourhood of a caustic requires special attention. Although local linearisation leads to simplified inverse interpolation, as



(a) Travel time map

(b) Error in travel time map

Figure 5.11: Travel time map in a homogeneous region around the source, the error is of the order of the machine precision only.

discussed in the previous section, it also leads to problems in the delineation of caustics.

An illustration of the problem is provided in Figure 5.12. The mapping of the wave front curve on the spatial coordinates is multi-valued. The turning point of the wave front is the caustic that separates two regions in the mapping domain. On the left of the caustic two rays arrive, on the right none. The approximate wave front curve has a different turning point than the exact curve. After the mapping the rightmost sample acts as an approximate caustic. In the region between the approximate and the exact caustic no arrivals are found, while in reality there are two.

The caustic delineation problem would be alleviated to a great extent if the linear ray field approximation could be replaced by one of higher order. This, however, would lead to a much more difficult mapping, as discussed in Section 5.4.1. The problem is not solved by the second order intrapolation proposed in that section, because that method only works if the ray field branch is locally single-valued.

The solution to the caustic delineation problem is also illustrated in Figure 5.12. If a sampling point is added on or close to the caustic, the caustic delineation is greatly improved. In practice this means that those simplices that contain a caustic must be split. The presence of a caustic can be detected by looking at the ray field map jacobian J:

$$J(\boldsymbol{\gamma}) = \det\left(\frac{\partial \boldsymbol{x}}{\partial(\sigma, \boldsymbol{\gamma})}\right), \tag{5.10}$$

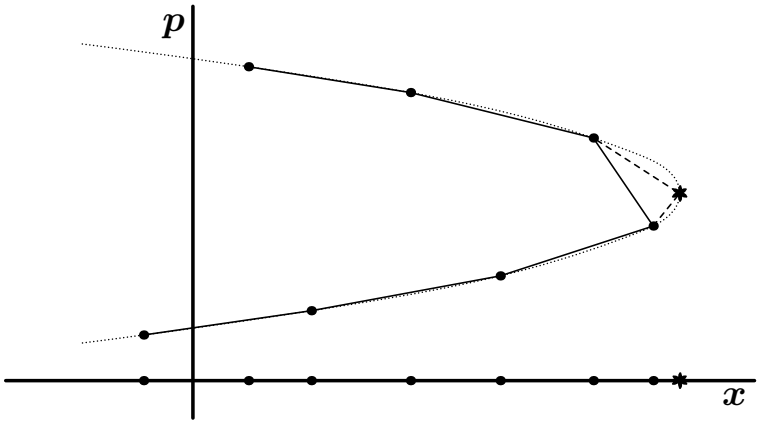


Figure 5.12: Cartoon illustrating the problem of mapping in the neighbourhood of a caustic. The horizontal axis represents the spatial coordinates, the vertical axis a variable that "unfolds" caustics, in this case the slowness. The dotted curve may be interpreted as a wave front, the big dots as the rays by which the wave front is sampled, point out of the drawing. The continuous lines form the piecewise linear approximation of the wave front spanned by the sampled rays. Mapping of the approximate wave front leads to a mislocation of the position of the caustic. The introduction of an extra sampling point (star) on or close to the caustic improves the mapping.

which changes sign upon crossing a caustic.

If two ray field samples at $(\sigma_a, \boldsymbol{\gamma}_a)$ and $(\sigma_b, \boldsymbol{\gamma}_b)$ have a jacobian J of opposite sign a caustic is located in between. The approximate location $(\sigma_c, \boldsymbol{\gamma}_c)$ of the caustic in terms of ray field coordinates can be found by linearisation of J:

$$h = \frac{J(\sigma_a, \boldsymbol{\gamma}_a)}{J(\sigma_a, \boldsymbol{\gamma}_a) - J(\sigma_b, \boldsymbol{\gamma}_b)} \tag{5.11}$$

$$\begin{pmatrix} \sigma_c \\ \boldsymbol{\gamma}_c \end{pmatrix} = (1-h)\begin{pmatrix} \sigma_a \\ \boldsymbol{\gamma}_a \end{pmatrix} + h\begin{pmatrix} \sigma_b \\ \boldsymbol{\gamma}_b \end{pmatrix}. \tag{5.12}$$

The approximate location in the spatial domain can then be found by making a second order intrapolation using the paraxial derivatives:

$$\begin{aligned} \boldsymbol{x}_c(\sigma_c, \boldsymbol{\gamma}_c) = (1-h)\left(\boldsymbol{x}(\sigma_a, \boldsymbol{\gamma}_a) + \frac{1}{2}\frac{\partial \boldsymbol{x}}{\partial(\sigma, \boldsymbol{\gamma})}(\sigma_a, \boldsymbol{\gamma}_a) \cdot \begin{pmatrix} \sigma_c - \sigma_a \\ \boldsymbol{\gamma}_c - \boldsymbol{\gamma}_a \end{pmatrix}\right) \\ + h\left(\boldsymbol{x}(\sigma_b, \boldsymbol{\gamma}_b) + \frac{1}{2}\frac{\partial \boldsymbol{x}}{\partial(\sigma, \boldsymbol{\gamma})}(\sigma_b, \boldsymbol{\gamma}_b) \cdot \begin{pmatrix} \sigma_c - \sigma_b \\ \boldsymbol{\gamma}_c - \boldsymbol{\gamma}_b \end{pmatrix}\right). \end{aligned} \tag{5.13}$$

In this way every simplex can be split in two in such a way that both parts are located on separate ray field branches, each with at least one vertex located at the approximate caustic curve. This *accurate caustic delineation* results in a ray field

partitioning with simplices that are each located within a single ray field branch. This allows the accurate intrapolation scheme proposed in Section 5.4.1 to be used for every simplex.

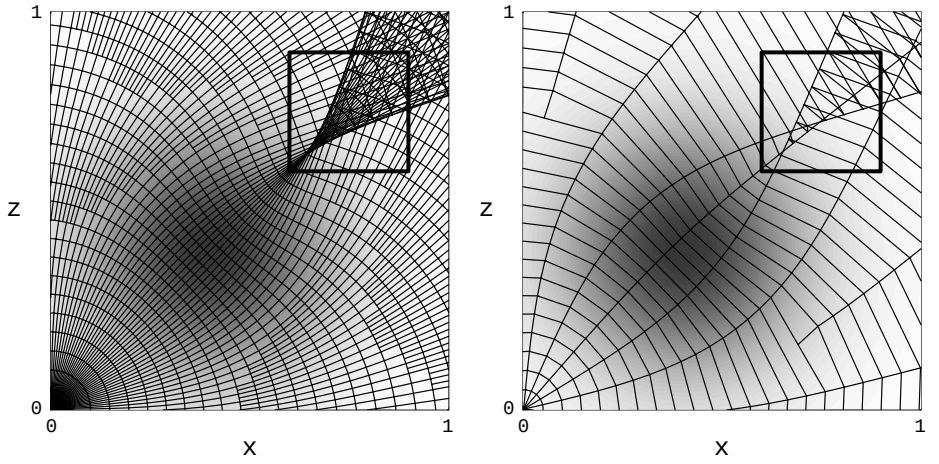
The effects of both the second order accurate travel time mapping by intrapolation and the accurate caustic delineation are illustrated in Figures 5.13 and 5.14. Note that the ray field sampling in Figures 5.13(b) and 5.14(b,c,d) is very sparse. Nevertheless, the multi-valued travel time field in the caustic region is reproduced very accurately, especially in Figure 5.14(d). Only the results for the third arrival are shown, but the second arrival gives comparable results.

The maximum relative error in the travel time of all three arrivals is determined for the ray field mappings illustrated in Figures 5.14(b,c,d) with respect to the map in 5.14(a). The errors are only calculated in the case both maps provide the same number of arrivals. For the calculation in Figure 5.14(b) the errors are 0.025, 0.019, and 0.0062 for arrivals 1 to 3 respectively. The experiment of Figure 5.14(c) gives the numbers 0.0023, 0.0045, and 0.0061. This shows that the second order intrapolation is much better than the linear interpolation, except for the third arrival. The reason is that in the second order interpolation ray field samples are used that do not correspond to the same ray field branch. Some evidence can also be found in Figure 5.14(c), where the travel time contours in neighbourhood of the ray crossing near $(x = 0.75, z = 0.75)$ some odd behaviour. The numbers for the experiment in Figure 5.14(d) read 0.0023, 0.0018, 0.0011. Not only is the caustic much better delineated in this case, the errors are also much smaller due to the fact that second order intrapolation is performed only with ray field samples that belong to the same branch.

5.5 Discussion and conclusions

A ray field construction and mapping algorithm is developed that extends and refines the existing wave front construction methods. A modular setup and a hierarchical description of the geometrical structure of the ray field make the algorithm very flexible and widely applicable. It can be used for the calculation of ray fields in 2-D and 3-D isotropic and general elastic media.

One of the objectives for developing the algorithm was to investigate whether it could be applied for the construction of ray fields and ray field maps in the position/angle domain. It was found that the algorithm is not suitable for that purpose. The accuracy control mechanism inherited from the wave front construction methods is primarily designed for the controlled sampling of diverging ray fields. In the position/angle domain the divergence of rays is limited. Instead, the main deformation in the geometrical structure of the ray field is shear-like. The result is that the ray density quickly rises and the triangulated ray fronts are populated by numerous narrow, nearly degenerate triangles (simplices), which makes the algorithm in its current form unstable and inefficient for application in the position/angle domain.



(a) Dense ray field sampling

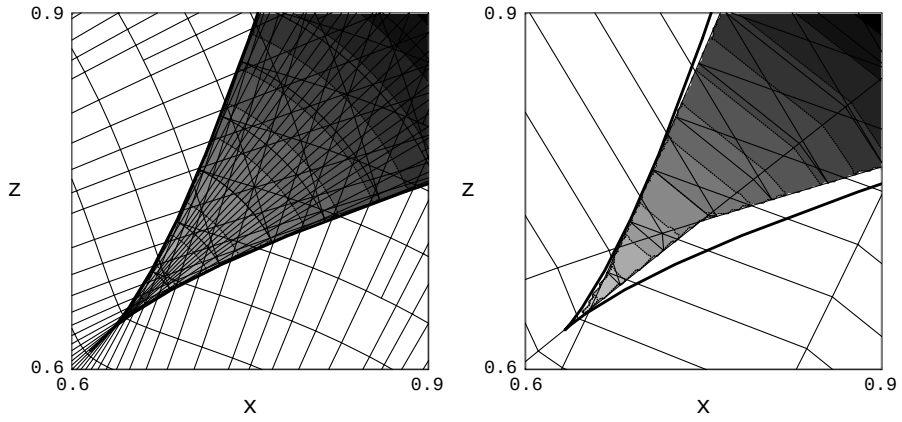
(b) Sparse ray field sampling

Figure 5.13: Ray fields in an isotropic 2-D medium. The velocity V satisfies $V = 1 - 0.4\exp(-((x-0.4)^2 + (z-0.4)^2)/0.3^2)$, that is, a slow Gaussian shaped anomaly. A caustic is formed in the upper right corner. The area within the square is subject to further study in Figure 5.14.

For applications in the spatial domain two refinements of the existing mapping methods are proposed. The use of intrapolation, as defined in Chapter 4 increases the accuracy of travel time mapping within the ray field cells to second order. This allows a coarser sampling of the ray field, which enhances the efficiency of the method. The additional cost per mapping point is just the difference between a second order and a first order polynomial evaluation.

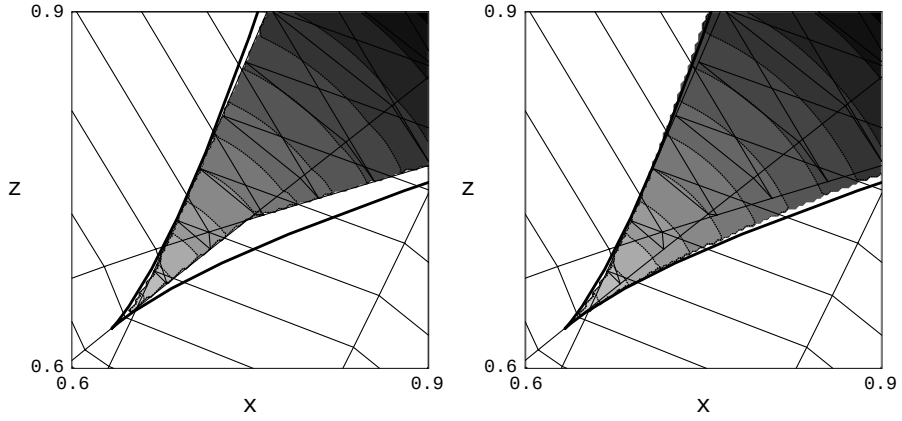
A refinement for the ray field sampling in the neighbourhood of caustics is proposed and called *accurate caustic delineation*. Sparse sampling of ray fields in the neighbourhood of a caustic results in a dubious estimation of the number of arrivals in the caustic region. A solution is provided by placing additional sample points on or close to the caustic, which results in an improved delineation of the latter. This refinement also enables a coarser ray field sampling. The method is tested only for 2-D ray fields. In 3-D, the direct neighbourhood of double caustics will require extra attention.

No attempt was made here to express the expected gain in efficiency in numbers by means of specific experiments. The ray field and wave front construction methods have a large number of degrees of freedom. These degrees of freedom include the choice between various techniques that have been proposed for the sub-parts of the algorithms, and the setting of various parameters. It is not easy to determine which combination of techniques and which setting of parameters provides the best result for a specific situation. Setting up an experiment for a



(a) Reference ray field and map.

(b) First order interpolation, no accurate caustic delineation



(c) Second order interpolation, no accurate caustic delineation

(d) Second order interpolation and accurate caustic delineation

Figure 5.14: Densely sampled travel time map of the third arrival in the square region around the caustic cusp of the ray fields displayed in Figure 5.13. The travel times are indicated by a colour scale and dashed contours, only in the regions where a third arrival is found. Reference map (a) corresponds to the densely sampled ray field of Figure 5.13(a). It clearly shows the curvature of the isochrons. Figures (b), (c), and (d) are based on the sparsely sampled ray field of Figure 5.13(b). In all Figures an accurately determined caustic curve is drawn for reference. Figures (b) to (d) use increasingly better approximations. Figure (d) shows that the combination of second order accurate interpolation and accurate caustic delineation results in a very accurate representation of the multi-valued ray field using only three rays in the caustic cusp.

specific situation and careful tuning of parameters may give an indication of the relative efficiency of different choices, but in the end a conclusion will be obtained only for that specific situation.

In practical applications there is no time for careful tuning of parameters. It is usually much more practical to choose a conservative set of parameters than spending time to find an optimal set of parameters. In that respect there is probably little difference in practical efficiency between different methods of wave front construction. An improvement of some part of the algorithm that in theory allows a relaxation of the parameter settings is useful only if it really changes the choice of parameters in practice.

Therefore, it would be useful to have a rule-of-thumb for the parameter setting, requiring as input a few characteristics of the application, and as output a set of parameters or perhaps a trade-off curve between speed and accuracy. An improvement of the algorithm can then result in a change of the rule-of-thumb, and be translated to an enhanced efficiency.

The characteristics of the application to be used in the rule-of-thumb may include the dimensions of the medium, the minimum length scale of spatial variations, the frequency of the waves, the desired accuracy etc. For each algorithm the rule-of-thumb can be determined rigorously by experiments on a suite of random models. These experiments will also provide data on the relative efficiency of different variations of the algorithm. In this way a considered choice of algorithm and parameter settings can be made.

An analysis of the ray field construction and mapping method as described above is recommended for future research. It is expected that the proposed refinements of the ray field mapping procedure will result in a general increase in efficiency.

Chapter 6

Paraxial ray tracing in the position/angle domain

A ray field in the position/angle domain has only a limited amount of geometrical spreading and its corresponding ray field map is single-valued, regardless of the complexity of the medium. These are exactly the type of circumstances that – in the spatial domain – allow the use of paraxial ray methods for the calculation of ray field maps. This observation suggests the use of paraxial ray methods in the position/angle domain as well.

A paraxial ray method is proposed that can be used for the calculation of ray field maps in the position/angle domain in smooth media of arbitrary complexity. The algorithm is both efficient and simple. The mapping from ray field coordinates to position/angle coordinates is performed by averaging the contributions of rays passing through a small neighbourhood around each of the map's grid points. This is different from the usual approach in paraxial ray tracing in which only the nearest passing ray is used. The advantage of averaging is that the local errors are more smoothly distributed which leads to more stable estimation of derivatives is further applications of the ray field map.

6.1 Introduction

In Chapter 5 the applicability of the proposed ray field construction method in the position/angle domain was tested. The unfortunate conclusion was that this method is not suitable for that domain, due to the type of deformation in the ray field structure. In the position/angle domain this deformation is dominantly shear-like while the ray field construction algorithm is primarily designed for expanding ray fields in the spatial domain.

By looking at the ray field map jacobian (3.28) three important observations can be made. First, the jacobian does not vanish anywhere, so the mapping from

ray field coordinates to position/angle coordinates is regular and single-valued.

Second, the geometrical spreading in the position/angle domain is limited. A look at (3.30) reveals that in 3-D media the geometrical spreading in the position/angle domain is roughly proportional to the third power of the local phase velocity. Given that the ray field map in the position/angle domain is 5-D, the spreading per dimension is limited. Moreover, the local ray density is approximately known a priori, from the values of the local phase velocity.

Third, each ray field emitted from a single point source can exhibit large geometrical spreading in the spatial domain. The combination of spatial divergence of neighbouring rays with the absence of geometrical spreading in the position/angle domain can only be explained if the deformation of the ray field structure in the position/angle domain is shear-like, as was indeed observed in Section 5.3.3.

Two of the most important incentives for the development of the wave front construction methods for ray tracing in complex media were the presence of multi-pathing and the large variations in geometrical spreading. In simpler media, in absence of these complicating features, paraxial ray methods (Červený et al., 1984; Beydoun and Keho, 1987) are generally more efficient, because they do not require the maintenance of a geometrical structure for the ray field. Since in the position/angle domain both multi-pathing and extensive geometrical spreading are absent and the ray field construction method fails due to the shear-like deformations, it is natural to investigate the applicability of paraxial ray methods.

In this chapter a paraxial ray tracing and mapping algorithm is developed in the position/angle domain. In paraxial ray tracing the geometrical structure of the ray field, i.e., the network of connections between neighbouring rays, is not available. Therefore it is not possible to calculate the ray field map by means of (inverse) interpolation as in ray field construction (see Chapter 5). In classical paraxial ray tracing in the spatial domain the mapping is usually performed by extrapolation from the nearest neighbouring ray. In Section 6.2 an alternative mapping method is proposed, which is based on an averaging integral. The paraxial ray method for the position/angle domain itself is introduced in Section 6.3.

6.2 Regularisation by averaging

The task of finding the ray field map on a regular grid can be seen as a problem of regularisation. The points of evaluation of the ray field form an irregular distribution of data points in the mapping domain. One way to perform the regularisation is by means of interpolation.

Unfortunately, interpolation is in general much more complicated in irregular, than in regular data distributions. One of the main difficulties is the determination of a geometric structure on the distribution, which is essential in order to determine the data points that contribute to a given interpolation point (see also Appendix D.2). In the ray field construction method of Chapter 5 the geometrical structure is maintained throughout the ray tracing process. For the paraxial ray method

developed in this chapter no such structure is available. It is then more convenient to perform regularisation by an averaging integral than by interpolation.

6.2.1 Averaging integrals

Continuous formulation

The integral approach to data regularisation is based on convolution of the function to be approximated with an averaging or smoothing kernel. In one dimension a regularisation $\mathcal{R}_0$ for a function $f(x)$ may be expressed as

$$\mathcal{R}_0[f(x)](x) = \int_{-\infty}^{\infty} K(\xi) f(x-\xi) d\xi. \tag{6.1}$$

If kernel $K(x)$ is defined to have a unit integral:

$$\int_{-\infty}^{\infty} K(\xi) d\xi = 1, \tag{6.2}$$

and a vanishing first order moment:

$$\int_{-\infty}^{\infty} K(\xi) \xi d\xi = 0, \tag{6.3}$$

then integral (6.1) may be regarded as an approximate representation integral that is accurate to first order. This can be verified by inserting a first order polynomial:

$$\mathcal{R}_0[c_0 + c_1 x](x) = c_0 + c_1 x. \tag{6.4}$$

Note that this agrees with the common notion that smoothing preserves linear features.

In multiple dimensions an equivalent regularisation may be defined:

$$\mathcal{R}_0[f(\boldsymbol{x})](\boldsymbol{x}) = \iint_{\mathrm{supp}(K)} K(\boldsymbol{\xi}) f(\boldsymbol{x}-\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{6.5}$$

where $\boldsymbol{x}$ and the double integral sign stand for arbitrary dimension. The infinite limits of the integral are replaced by $\mathrm{supp}(K)$ in order to denote that in practice K will have a localised support.

The constraints on kernel K generalise analogously:

$$\iint_{\mathrm{supp}(K)} K(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1, \qquad \text{and} \tag{6.6}$$

$$\iint_{\mathrm{supp}(K)} K(\boldsymbol{\xi}) \boldsymbol{\xi} d\boldsymbol{\xi} = \mathbf{0}. \tag{6.7}$$

Regularisation based on the approximate representation integral (6.5) is performed by evaluating the integral numerically using the available data points. A particular advantage of this approach, as opposed to regularisation by interpolation, is that it does not require as much a priori knowledge about the geometrical arrangement of the data points. A disadvantage, however, is that the numerical evaluation of the integral introduces additional numerical errors.

Discrete formulation

A discrete analogue of (6.5), based on the M data points $\boldsymbol{x}_i$ $(i = 1 \dots M)$ within the support of $K(\boldsymbol{x} - \boldsymbol{x}_i)$, may be expressed as

$$\mathcal{R}_0^D[f(\boldsymbol{x})](\boldsymbol{x}) = \sum_{i=1}^{M} \alpha_i f(\boldsymbol{x}_i), \tag{6.8}$$

where the weights α_i are yet to be determined.

It is obvious that an arbitrary distribution of data points does not offer a high degree of accuracy control in the numerical evaluation of an integral. It should be noted, however, that in order for (6.5) to serve as an approximate representation integral, the actual shape of kernel K is less critical than the validity of accuracy constraints (6.6) and (6.7). These constraints have discrete analogues

$$\sum_{i=1}^{M} \alpha_i = 1, \qquad \text{and} \tag{6.9}$$

$$\sum_{i=1}^{M} \alpha_i \boldsymbol{x}_i = \boldsymbol{x}, \tag{6.10}$$

respectively.

At this point it is interesting to note the similarities between this integral approach to regularisation, and interpolation in terms of barycentric coordinates, as described in Appendix D.2. In fact, barycentric interpolation can be seen as a special case of the method developed here, using only a minimal set of data points. In that case, the weights α_i, the barycentric coordinates, are uniquely determined by the constraints (D.8) and (D.9), which makes the use of an averaging kernel redundant.

Before discussing how to determine weights α_i, it is important to say a word about the practical implementation of this type of regularisation. The numerical evaluation of integral (6.5), through its discrete counterpart (6.8), is most conveniently implemented as a generalised (weighted) binning, by looping over the data points. Because of the local support of kernel K, it is easy to determine the range of grid points that each data point contributes to. Therefore, it is desirable to be able to determine the weight α_i for each contribution before knowing which or how many data points will contribute to the integral in the end. In practice this is

not entirely possible, but it is important that the amount of bookkeeping required for each grid point be kept minimal.

Determination of weights α_i

For a grid point denoted by $\boldsymbol{x}$, the relative contribution $\hat{\alpha}_i$ of data point $\boldsymbol{x}_i$ is determined by kernel K:

$$\hat{\alpha}_i = K(\boldsymbol{x} - \boldsymbol{x}_i). \tag{6.11}$$

If an estimate of a local data point density $\rho(\boldsymbol{x}_i)$ is available it may be included in the relative weight:

$$\hat{\alpha}_i = \rho(\boldsymbol{x}_i) K(\boldsymbol{x} - \boldsymbol{x}_i). \tag{6.12}$$

This is the case, for example, if the averaging integral is calculated on a set of points that form a regular distribution in a domain other than $\boldsymbol{x}$, say $\boldsymbol{y}$. The data point density is then equal to the jacobian of transformation from $\boldsymbol{y}$ to $\boldsymbol{x}$. This is used also in the paraxial ray method to be discussed in Section 6.3.

In the binning process the relative contributions are added, while keeping track of the sum of relative weights. Afterwards the discrete integral (6.8) is obtained using

$$\mathcal{R}_0^D[f(\boldsymbol{x})](\boldsymbol{x}) = \frac{\sum_{i=1}^{M} \hat{\alpha}_i f(\boldsymbol{x}_i)}{\sum_{i=1}^{M} \hat{\alpha}_i}, \tag{6.13}$$

which effectively sets the weights α_i for (6.8):

$$\alpha_i = \frac{\hat{\alpha}_i}{\sum_{i=1}^{M} \hat{\alpha}_i}. \tag{6.14}$$

The constraint (6.9) is automatically satisfied. The bookkeeping for each grid point is limited to a single number that accumulates the sum of relative weights.

The fulfillment of constraint (6.9) guarantees that the regularisation is exact only for constant functions. In order to be first order accurate, it is necessary to satisfy constraint (6.10) as well. Unfortunately this is much more difficult to achieve. Adapting the weights α_i in such a way that both constraints (6.9) and (6.10) are satisfied, requires extensive bookkeeping, which is to be avoided.

Mislocation

The degree of violation of constraint (6.10) can be seen as a measure of how well the continuous regularisation integral (6.5) is approximated by its discrete counterpart (6.8). The error may be interpreted as a *mislocation* of the regularisation estimate.

For a given set of data points $\boldsymbol{x}_i$ and weights α_i that satisfy constraint (6.9), the first order moment $\delta\boldsymbol{x}$ with respect to grid point $\boldsymbol{x}$ may be evaluated:

$$\delta\boldsymbol{x} = \sum_{i=1}^{M} \alpha_i(\boldsymbol{x}_i - \boldsymbol{x}). \tag{6.15}$$

If $\delta\boldsymbol{x}$ vanishes, constraint (6.10) is satisfied and the discrete regularisation integral (6.8) is accurate to first order. For α_i obtained using the procedure outlined above, this will, in general, not be the case. Instead, it is easy to see that the first order moment does vanish if it is calculated with respect to the point $\boldsymbol{x} + \delta\boldsymbol{x}$:

$$\sum_{i=1}^{M} \alpha_i(\boldsymbol{x}_i - \boldsymbol{x} - \delta\boldsymbol{x}) = \boldsymbol{0}. \tag{6.16}$$

In other words, the combination of data points $\boldsymbol{x}_i$ and weights α_i provides a first order accurate estimator for the function value $f(\boldsymbol{x} + \delta\boldsymbol{x})$, rather than for $f(\boldsymbol{x})$.

6.2.2 Kernels and mislocation

Mislocation is an important source of error in the integral regularisation approach, as will be demonstrated in Section 6.2.3. Factors that influence the amount of mislocation are the size and the shape of averaging kernel K, the number of data points within the support of K, and the spatial distribution of the data points. A crucial piece of information is knowledge of the local data point density $\rho(\boldsymbol{x}_i)$ as described above. Here, for two different kernels the mislocation is analysed in 1-D using random samples.

Spline and boxcar kernels

First, two different averaging kernels are introduced that satisfy the zeroth and first order moment constraints (6.2) and (6.3). In practical applications the kernel must also have a finite support. The most important attribute of the shape of a kernel, with respect to mislocation, is its smoothness.

As an example of a smooth kernel is used the cubic spline $K_s(x)$, a piecewise third order polynomial:

$$K_s(x) = \begin{cases} \frac{2}{3} - |x|^2 + \frac{1}{2}|x|^3 & \text{for} \quad |x| \le 1 \\ \frac{4}{3} - 2|x| + |x|^2 - \frac{1}{6}|x|^3 & \text{for} \quad 1 < |x| \le 2 \\ 0 & \text{for} \quad |x| > 2. \end{cases} \tag{6.17}$$

The boxcar kernel $K_b(x)$, serves as an example of a non-smooth kernel:

$$K_b(x) = \begin{cases} \frac{1}{2} & \text{for} \quad |x| \le 1 \\ 0 & \text{for} \quad |x| > 1. \end{cases} \tag{6.18}$$

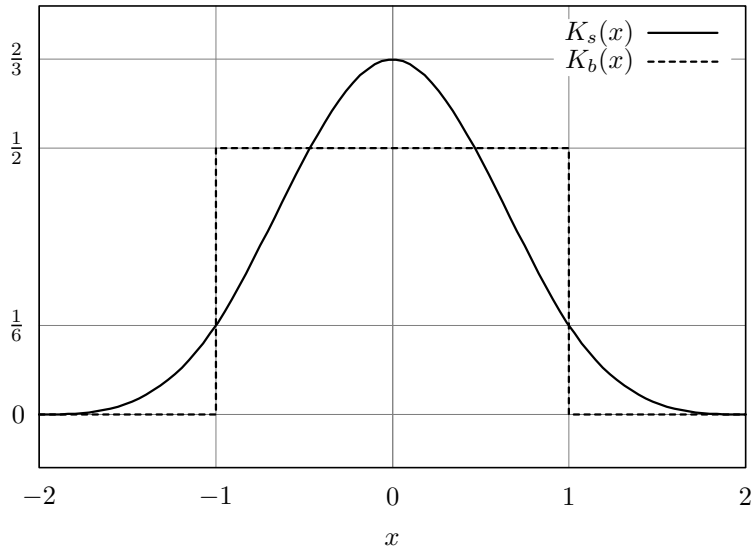


Figure 6.1: Cubic spline averaging kernel $K_s(x)$ (6.17) and boxcar averaging kernel $K_b(x)$ (6.18). The kernels have a different support, but have an equal width according to the second order moment criterion (6.19).

Both kernels are displayed in Figure 6.1.

According to their respective definitions, the spline and boxcar kernels have different regions of support. They have been defined in such a way, that they have equal width according to their second order moment:

$$\int_{-1}^{1} K_b(x)x^2dx = \int_{-2}^{2} K_s(x)x^2dx. \tag{6.19}$$

The second order moment is a plausible measure for the width of an averaging kernel, because it stands for the amount of smoothing it applies to second order polynomials. Two kernels of the same width in terms of second order moment therefore have an equal response up to second order.

Convergence of mislocation for random data points

Figure 6.2 shows the result of an analysis of mislocation as a function of the number of randomly distributed data points within the support of a 1-D kernel. It shows that knowledge of the local data point density – calculated as half the space between the first neighbouring data points to the left and to the right – is crucial. Without this knowledge the convergence is proportional to the square root of the number of data points and is independent of the shape of the kernel. If the data

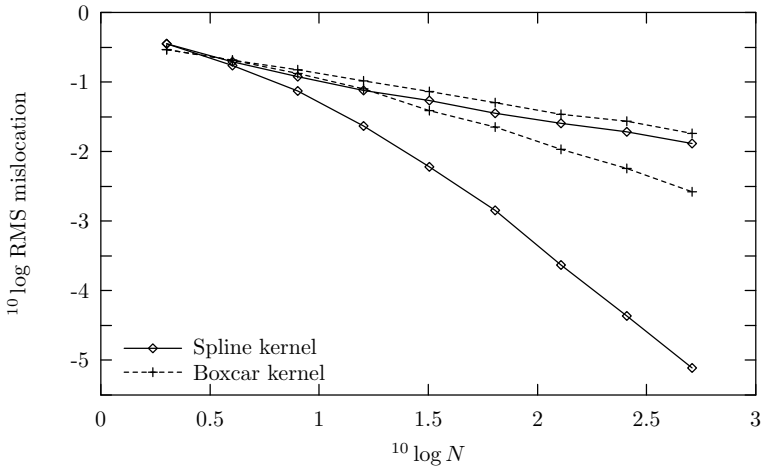


Figure 6.2: Convergence of mislocation as a function of the number N of randomly distributed data points within the support of the kernel. Each data point represent the root-mean-square (RMS) average of 100 random data point distributions. Two curves are shown for both the spline and the boxcar kernels. The lower curve is obtained with knowledge of the local data point density ρ, the upper without. The upper curves for both kernels show an asymptotic gradient of roughly -0.5, the lower curves show asymptotic gradients of roughly -1.0 and -2.5 respectively. For explanation see Section 6.2.2.

point density is known, however, the smooth spline kernel shows a much faster convergence than the discontinuous boxcar kernel.

6.2.3 Regularisation with enhanced accuracy

Higher order moment constraints

As in the case of interpolation, our interest is to enhance the accuracy of the regularisation technique discussed above. The first order accuracy of averaging integral (6.5) is due to the constraints (6.6) and (6.7) on kernel $K(\boldsymbol{x})$. A straightforward approach to enhancing the accuracy of the averaging integral is to increase the number of constraints on $K(\boldsymbol{x})$. The extra constraints will take the form of moment constraints, such as for the k-th order moment:

$$\iint\limits_{\mathrm{supp}(K)} K(\boldsymbol{\xi})(\boldsymbol{x}-\boldsymbol{\xi})^{[k]} d\boldsymbol{\xi} = \boldsymbol{x}^{[k]}, \tag{6.20}$$

where superscript $[k]$ is again the iterated tensor product as defined in Equation (4.23). Note that this general form is compatible with the zeroth and first order moment constraints (6.6) and (6.7).

Hence, in order to obtain an n-th order accurate averaging integral, one may design a kernel $K(\boldsymbol{x})$ that satisfies constraint (6.20) for $k = 0 \ldots n$. Although this makes sense in the continuous formulation, it is not very useful in practice, because in the discrete case it is difficult to satisfy the moment constraints for arbitrary data points. The mislocation discussed above, for example, is the result of not being able to satisfy already the first order moment constraint.

Averaging extrapolations

A more practical method for enhancing the accuracy of the regularisation is available if the data points also contain derivative information. In that case extrapolations from a number of data points can be averaged. The normal averaging of the data can then be seen as the averaging of zeroth order extrapolations. As in Chapter 4 both the Taylor expansion and the Dutch Taylor expansion can be used for the extrapolation. In theory, the Dutch Taylor expansion again promises an accuracy of one order higher than an individual Taylor expansion for an averaging kernel that is exact to first order. This first order accuracy, however, is not obtained in practice due to the mislocation discussed above.

To assess the combined effects of mislocation and the averaging of extrapolations an experiment is performed on the approximation of cosines. The experiment is similar to that discussed in Section 6.2.2, but now for cosines rather than linear functions. The cosines have a wavelength of $\pi/4$ times the support of a spline kernel, and a random phase. Every experiment uses the spline kernel and the estimated local data point density, as described above. Averaging integrals are calculated using both the normal and the Dutch Taylor expansions up to second order. The Dutch Taylor expansion for $n = 0$ is equal the Taylor expansion for $n = 0$ (see also Chapter 4).

Clearly, the accuracy of averaging the Dutch Taylor expansion depends strongly on the mislocation. For sufficiently high N, these expansions lead to a greater accuracy then the corresponding Taylor expansions, but in practical applications the number of data points will be low. Hence, the Dutch Taylor expansion is not as useful in regularisation based on an averaging integral as it is interpolation. The reason is that the averaging integral cannot be made accurate to first order in practice. In the following only the Taylor expansion will be used.

6.3 Paraxial ray tracing

6.3.1 Ray field mapping by averaging

The goal of the paraxial ray method described in this section is to construct a ray field map expressing the ray field coordinates $\boldsymbol{\xi} \equiv (\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)$ as a function of the position/angle coordinates $\boldsymbol{y} \equiv (\boldsymbol{x}, \boldsymbol{\phi})$, as discussed in Chapter 3.

The algorithm basically has two tasks. The first task is to solve the ray equations (Chapter 2) to provide $\boldsymbol{y}(\boldsymbol{\xi})$. The second task is to determine for each grid

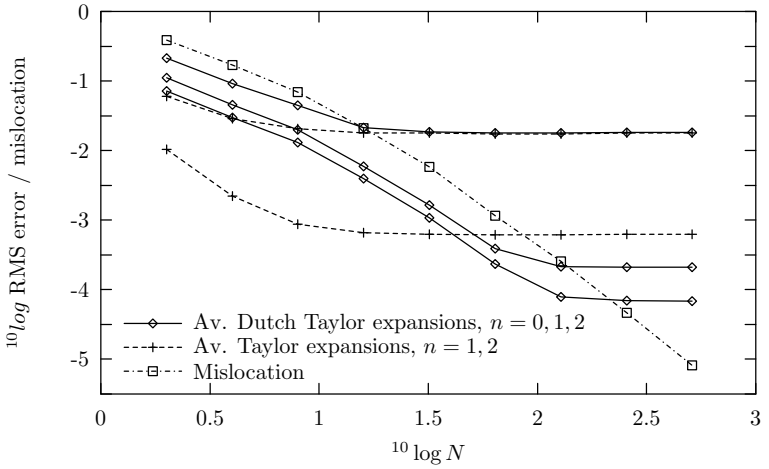


Figure 6.3: Convergence of the approximation of cosines as a function of the number N of randomly distributed data points within the support of the kernel. Each data point represents the root-mean-square (RMS) average of 100 random data point distributions. All curves use spline kernels and estimates of local data point density. For further explanation see Section 6.2.3.

point $\boldsymbol{y}^*$ the $\boldsymbol{\xi} = \boldsymbol{\xi}^*$ that solves

$$\boldsymbol{y}(\boldsymbol{\xi}) = \boldsymbol{y}^*. \tag{6.21}$$

In the ray field construction method (Chapter 5) this second task is performed by determining a local linear approximation to $\boldsymbol{y}(\boldsymbol{\xi})$, which is then solved by means of linear inverse interpolation.

In the paraxial ray method proposed here, system (6.21) is solved approximately using the averaging integrals discussed in the previous section. Formally, the solution $\boldsymbol{\xi}^*$ to (6.21) can be expressed as a representation integral:

$$\begin{aligned} \boldsymbol{\xi}^*(\boldsymbol{y}^*) &= \int \boldsymbol{\xi}(\boldsymbol{y})\delta(\boldsymbol{y}^* - \boldsymbol{y})d\boldsymbol{y} \\ &= \int \boldsymbol{\xi}\delta(\boldsymbol{y}^* - \boldsymbol{y}(\boldsymbol{\xi}))\det\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\xi}}\right)d\boldsymbol{\xi}. \end{aligned} \tag{6.22}$$

Hence, $\boldsymbol{\xi}^*(\boldsymbol{y}^*)$ can be expressed as an integral over $\boldsymbol{\xi}$. For practical applications this integral must be adapted to provide an approximate solution:

$$\boldsymbol{\xi}^*(\boldsymbol{y}^*) \approx \int \boldsymbol{\xi} K(\boldsymbol{y}^* - \boldsymbol{y}(\boldsymbol{\xi}))\det\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\xi}}\right)d\boldsymbol{\xi}, \tag{6.23}$$

with K an averaging kernel. The order of accuracy of (6.23) depends on the properties K, as discussed in Section 6.2).

A simple implementation of a paraxial ray tracing algorithm works by shooting rays from a regular grid in $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\phi}}_0)$ with regular steps of T into the medium. The evaluation points of the ray field then form a regular grid in $\boldsymbol{\xi}$, but an irregular distribution in $\boldsymbol{y}$. Integral (6.23) for grid point $\boldsymbol{y}^*$ can be evaluated approximately using the data points that fall within the support of K. The jacobian $\det(\partial \boldsymbol{y}/\partial \boldsymbol{\xi})$ then serves as the estimator of the local data point density ρ mentioned in Section 6.2.

The defining characteristic of paraxial ray methods is the use of paraxial derivatives to extrapolate information from the rays in the ray field to the grid points in the mapping domain. Rather than averaging the ray field coordinates $\boldsymbol{\xi}$ to find $\boldsymbol{\xi}^*(\boldsymbol{y}^*)$ in (6.23) it is more accurate to average paraxial estimates to the latter from each of the data points. This can be done using the Taylor expansion (4.22):

$$\boldsymbol{\xi}^*(\boldsymbol{y}^*) \approx \int \mathcal{T}_n[\boldsymbol{\xi}](\boldsymbol{y}^*; \boldsymbol{y}) K(\boldsymbol{y}^* - \boldsymbol{y}(\boldsymbol{\xi})) \det\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{\xi}}\right) d\boldsymbol{\xi}. \tag{6.24}$$

In practice only first order paraxial derivatives are calculated:

$$\mathcal{T}_1[\boldsymbol{\xi}](\boldsymbol{y}^*; \boldsymbol{y}) = \boldsymbol{\xi}(\boldsymbol{y}) + (\boldsymbol{y}^* - \boldsymbol{y}) \cdot \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{y}}. \tag{6.25}$$

These paraxial derivatives must be calculated along the rays as explained in the following section.

6.3.2 Paraxial ray information

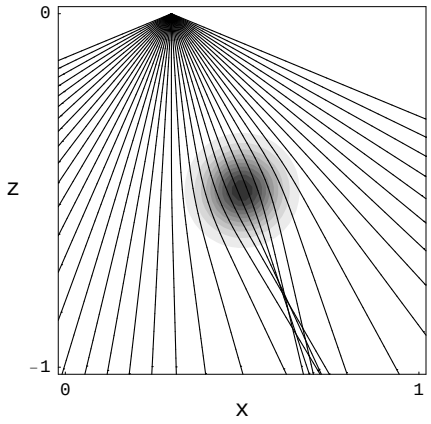
The derivatives $\partial \boldsymbol{\xi}/\partial \boldsymbol{y}$ of Equation (6.25), i.e., the partial derivatives of the position/angle coordinates with respect to the ray field coordinates, can easily be calculated along the rays with the help of the theory of Chapters 2 and 3. The derivatives can be identified as the inverse of a set of paraxial derivatives:

$$\frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial(\boldsymbol{x}, \boldsymbol{\phi})} = \left(\frac{\partial(\boldsymbol{x}, \boldsymbol{\phi})}{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}\right)^{-1}. \tag{6.26}$$

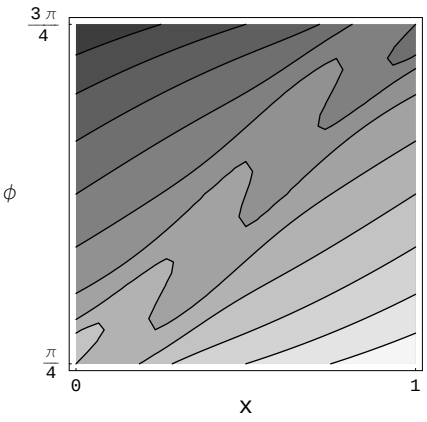
The paraxial derivatives can be calculated directly along the ray using the paraxial ray equations (2.36). However, inverting this paraxial derivative matrix is not a good idea. The shear-like ray field deformations observed in the position/angle domain may make this derivative matrix nearly degenerate, which, in turn, may lead to considerable numerical errors in the matrix inversion.

Instead, it is better to use a formulation in terms of the propagator matrix $\boldsymbol{P} = \partial(\boldsymbol{x}, \boldsymbol{p})/\partial(\boldsymbol{x}_0, \boldsymbol{p}_0)$, which, as shown in Equation (2.42), is easily invertible:

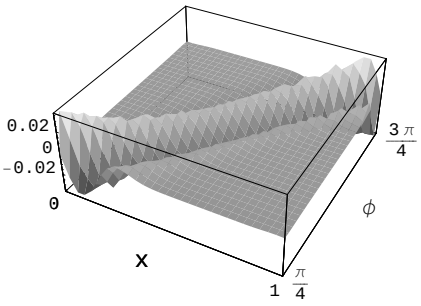
$$\left.\frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial(\boldsymbol{x}, \boldsymbol{\phi})}\right|_{\mathcal{H}} = \frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial(\boldsymbol{x}_0, \boldsymbol{p}_0)} \cdot \boldsymbol{P}^{-1} \cdot \left.\frac{\partial(\boldsymbol{x}, \boldsymbol{p})}{\partial(\boldsymbol{x}, \boldsymbol{\phi})}\right|_{\mathcal{H}}. \tag{6.27}$$



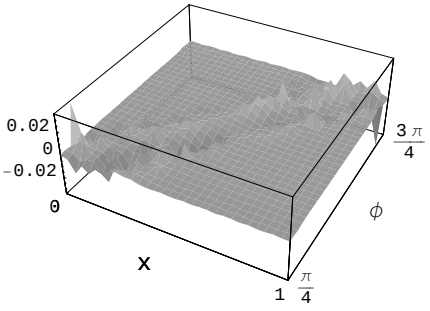
(a) Medium used in paraxial ray tracing example and rays shot for one source location



(b) Contours of $\tilde{x}_0$ evaluated at $z = -1$; contour in the middle represents $\tilde{x}_0 = 0.5$, other contours in steps of 0.25



(c) Error in $\tilde{x}_0$ for paraxial ray method with averaging integral



(d) Error in $\tilde{x}_0$ for paraxial ray method using only nearest ray point

Figure 6.4: Example of the use of paraxial ray tracing in the position/angle domain. For explanation see Section 6.3.3.

Although this formulation requires knowledge of the full ray propagator $\boldsymbol{P}$ it does require the solution of a higher number of paraxial ray equations. Only four columns of $\boldsymbol{P}$ need to be calculated explicitly, because the other two can be obtained from the analytical solutions (2.43) and (2.44). The two other derivative matrices on the right hand side of (6.27) can be calculated using (3.19), (3.20) and (3.24).

6.3.3 Example

As an example of the paraxial ray tracing method in the position/angle domain a ray field map is constructed for an isotropic medium with a Gaussian-shaped slow anomaly on top of a homogeneous background. A contour plot for the medium is shown in Figure 6.4(a). Phase velocity V satisfies $V^{-1} = 1.0 + 0.3\exp(-100.r^2)$, with r the distance to the point $(0.5, 0.5)$.

Rays are shot downwards from the surface $z = 0$, parameterised by $\tilde{x}_0$, $\tilde{\phi}_0$, and T. The rays are calculated using a Runge-Kutta scheme as in Chapter 5 with a time step of 0.025. The ray field map is constructed on a rectangular grid in (x, z, ϕ)-space, with 31 grid points for $x \in [0, 1]$ and $z \in [-1, 0]$, and 31 grid points for $\phi \in [\pi/4, 3\pi/4]$. The rays shot from a single source position are drawn in Figure 6.4(a). The 41 rays cover a range of $\tilde{\phi}_0 \in [1\pi/8, 7\pi/8]$. The shot positions also cover a wider range than the map in order to obtain arrivals at depth $z = -1$ for all angles. The 123 source positions cover a range of $\tilde{x}_0 \in [-1, 2]$.

For comparison a reference calculation is made by tracing rays up towards the surface for every grid point at $z = -1$. The resulting contour plot for $\tilde{x}_0$ is shown in Figure 6.4(b). The folding of contours over the x-range indicates the presence of multi-pathing; the turning points correspond to the caustics.

For the averaging integral (6.24) a spline kernel (6.17) is used with a support of exactly 4 grid intervals in all three directions. The result of the paraxial method with the averaging kernel is evaluated in Figure 6.4(c). It shows the error in the calculation of x_0 with respect to the reference calculation shown in Figure 6.4(b). The errors are largest near the caustics, which is the result of the smoothing caused by the averaging. The overall error is reasonably small, and may be reduced by choosing smaller support for the averaging kernels and increasing the number of rays traced.

For comparison Figure 6.4(d) shows an alternative approach to paraxial ray tracing. At each grid point only one paraxial ray estimate is evaluated from the ray point that is nearest to the grid point. An advantage is that the smoothing is reduced with respect to 6.4(c). A disadvantage is that the errors are more or less randomly distributed. This is a problem for further application of the ray field maps, because this type of error is harmful for the accuracy of finite difference estimates of derivatives.

6.4 Discussion and conclusions

It has been shown that ray field maps in the position/angle domain can be calculated using paraxial ray methods. Rays are traced downwards from the acquisition surface, adding ray field information to those grid points that are contained in a paraxial neighbourhood of the ray. The algorithm is very simple and easy to implement for smooth media of arbitrary complexity; the only ingredients needed are a model description (e.g., grid and interpolator) a data structure for the grid

in the position/angle domain and a paraxial ray tracer.

The algorithm is also very efficient. Because it yields the ray field information organised by angles at depth, it must be compared to other algorithms that yield the same. Currently the only possibility considered by studies regarding imaging in the angle domain seems to be to simply trace the rays upwards from the subsurface image points (e.g., Koren et al., 2002; Brandsberg-Dahl et al., 2003). Obviously, tracing upwards is much more expensive, because the image points extend over one dimension more than the acquisition points. Use of the proposed paraxial ray method instead leads to a gain in speed for the ray tracing roughly proportional to the average length of the rays divided by spatial interval of the image grid.

Note that the paraxial ray algorithm is perfectly applicable to a target-oriented approach as described in Koren et al. (2002). From the boundary of the target rays may first be shot outwards to the acquisition surface to determine the ray field coordinates for each ray. These coordinates may then be propagated inside the target by continuing the rays inward from the boundary.

A general advantage of tracing ray fields in the position/angle domain is that the geometrical spreading, and hence the ray density, does not depend on the length of propagation or some other aspect of the ray path. Apart from the ray density at the source and the specific parameterisation of the slowness surface the ray density only depends on the local phase velocities (see Section 3.4). Hence, the ray density at depth can be predicted reasonably well and the number of rays shot from the acquisition surface can be chosen a priori in such a way that a sufficient number of rays passes in the neighbourhood of each grid point.

An interesting point to note here concerns the occurrence of shadow zones for ray tracing in the spatial domain. Although perfect shadow zones do not exist in smooth media, the geometrical spreading may become very high at places. In wave front construction codes this leads to a large number of ray insertions to maintain a certain accuracy. For ray tracing in the position/angle domain, however, these zones do not have the same effect. The fact that the local ray density in the position/angle domain primarily depends on the local phase velocity promises that a zone that is "left open" by the rays from one source location will be filled in by rays from neighbouring source locations.

The above observation may have very interesting consequences. It seems that the number of rays required to calculate a single ray field map in the position/angle domain can be much smaller than that required for the calculation of a series of ray field maps in the spatial domain for the same range of source positions. If this is true, it may very well lead to the conclusion that even if (possibly multi-valued) ray field maps in the spatial domain are desired, it is advantageous to first construct a ray field map in the position/angle domain, followed by the extraction of spatial ray field maps for the individual source locations. This is left as another area for future research.

The paraxial ray method developed so far is shown to work adequately in the example of Section 6.3.3. A number of issues have not been addressed, however, such as the choice of grid spacing to be used and the number of rays to be shot

and the effect on the accuracy. Although the geometrical spreading in the position/angle domain is independent of the length of propagation, the same does not hold for the number of rays required to maintain sufficient accuracy. In general, the variations in both ray field and ray field map will become smaller in scale along the way. Hence, the number of rays and grid points required to approximate these variations will in general increase during propagation.

An improved version of paraxial ray method for the position/angle domain will therefore probably contain a mechanism to increase the number of rays along the way. It does not have to be as complex, however, as the mechanism used in wave front construction methods (e.g., Chapter 5). It is probably sufficient to use statistical rather than deterministic criteria for the addition of rays, based on the propagation distance and the length scales and sizes of variations in the medium. The ray field may, for example, be re-initialised with enhanced ray density at intermediate surfaces at depth. These criteria should preferably be determined rigourously from experiments on suites of random models, as suggested also in the discussion of Chapter 5.

Chapter 7

General discussion and concluding remarks

The research described in this thesis covers various aspects of the forward calculation of ray fields and ray field maps. The central theme is the solution of problems encountered in smooth but complex media, i.e., media that give rise to wave front folding and associated multi-pathing of rays. The ultimate aim of the presented material is to enhance the efficiency of seismic inverse methods, by enhancing the efficiency of the forward calculations. Particular emphasis has been placed on the applicability of the ray tracing results to seismic inverse methods.

The results can be broadly categorised into four subjects, which will be discussed in the following sections.

7.1 Pseudo-spectral ray tracing

One of the first objectives of this research was to develop an efficient algorithm for the calculation of ray fields in smooth media, and to make this part of a perturbation method that would be able to handle variation in the number of ray arrivals under perturbation of the medium. The algorithm has two distinctive features; first, the wave fronts are approximated by a pseudo-spectral expansion, and second, the parameterisation of the wave fronts is adjusted dynamically to maintain a homogeneous sampling of the ray field throughout the medium.

Although the approach is conceptually attractive, obtaining a stable and efficient algorithm proved to be difficult. The method appears to be more complex and less flexible than, for example, the ray field construction method presented in Chapter 5. Furthermore it cannot easily be extended to 3-D.

Since it is unlikely that pseudo-spectral ray tracing will be able to compete with ray field construction (for speed) it was decided to abandon this line of research in favour of the more practical ray field map methods. However, we have

included the work on the pseudo-spectral ray tracer in this thesis, as Appendix C, because its development provided useful insights for the ray field map approach – e.g., compare the dynamic parameterisation mechanism in Appendix C with the method proposed in Appendix B – and some of its features may be useful in other applications.

7.2 Intrapolation and the Dutch Taylor expansion

Interpolation methods play an important role in various stages of the calculation and application of ray fields and ray field maps. In many cases, not only the values of the function to be interpolated are known at the data points, but also the first or even second order derivatives. This is the case, for example, for the interpolation of travel times in ray field cells, where usually the slowness vector – the gradient of travel time – is available. If paraxial ray tracing is performed, even the second order derivative of travel time – related to the wave front curvature – may be evaluated. Also, for the interpolation of ray field maps on regular grids the derivatives can easily be determined using finite differences.

Many interpolation methods have been developed and applied in many fields of science. However, an extensive search through the literature and the world wide web did not yield a simple and generally applicable method to incorporate derivative information in interpolation to enhance the accuracy. It is possible to construct interpolating polynomials that fit derivatives as well as the function values in arbitrary data point configurations, but this is usually impractical. The analysis required to obtain the interpolants is time consuming and, thus, computationally expensive, and the interpolants are necessarily of higher order, which again means that they are expensive to evaluate, and the risk of enhancing errors in the data is high.

The lack of a suitable method for accurate interpolation using derivative data led to the development of the *intrapolation* technique described in Chapter 4. Intrapolation is a hybrid of extrapolation to arbitrary order and linear interpolation and combines the advantages of both. The extrapolation is done by a modification of the Taylor expansion, for which we coined the term *Dutch Taylor expansion*. The order of accuracy of intrapolation is one higher than that of a single conventional Taylor extrapolation that uses the same amount of derivative information.

Since both extrapolation and linear interpolation are easily performed in arbitrary dimensional spaces and on arbitrary data distributions, intrapolation should be applicable to a wide variety of problems. In the context of ray methods it is immediately applicable to the interpolation of ray field maps on rectangular grids (see Section 4.4.2). It has also proven its value in the forward calculation of travel time maps in the spatial domain: first, as a method for interpolating travel times within ray field cells with second order accuracy, and second, for interpolating additional ray field points to improve the delineation of caustics (see Section 5.4).

The applicability of the Dutch Taylor expansion in a wider context was inves-

tigated in Section 6.2.3. In theory, the expansion can also be used to enhance the accuracy of regularisation by means of averaging integrals. In practice, however, the averaging integrals are evaluated with insufficient accuracy for the Dutch Taylor expansion to have the desired effect.

7.3 Ray field construction in the spatial domain

The calculation of ray-theoretical Green functions on spatial grids is a challenging computational task, especially in complex media, where both strong geometrical spreading and multi-pathing are complicating factors. A very efficient and popular class of algorithm for this type of calculation is called wave front construction. In Chapter 5 a number of extensions and refinements to wave front construction have been presented under the name *ray field construction*.

The proposed algorithm generalises the wave front construction methods by allowing an arbitrary number of ensemble parameters, i.e., degrees of freedom in the initial conditions of the rays in the ray field. This has been accomplished by defining a hierarchical description of the ray field structure and a recursive approach to the ray field propagation. As a result, this algorithm can not only be used for the usual ray tracing in 2-D or 3-D media but also for ray field construction in the position/angle domain, which is discussed in Section 7.4.

A modular setup makes the algorithm highly adaptable to various types of model parameterisation. Both isotropic and general elastic (anisotropic) smooth models are currently supported. Some examples have been shown for media with vertical transverse isotropy (VTI).

For applications in the spatial domain we propose two refinements of the existing ray field mapping methods (Section 5.4). The use of intrapolation, see above, increases the accuracy of travel time mapping within the ray field cells to second order. Also, a refinement for the ray field sampling in the neighbourhood of caustics is proposed under the name *accurate caustic delineation*. Both refinements allow a smaller number of rays in the ray field, which enhances the efficiency of the ray field construction.

7.4 Ray field maps in the position/angle domain

7.4.1 Theory

A large part of this thesis is devoted to the development of theory and algorithms for the construction of ray field maps in the position/angle domain. This new approach for the representation and calculation of ray field information, introduced in Chapter 3, is particularly useful if ray fields have to be calculated for a dense distribution of sources and/or receivers at an acquisition surface. This is a common situation in seismic imaging experiments such as reflection seismics and borehole tomography.

A single ray field map in the position/angle domain contains the ray field information associated with a range of acquisition points. As such, it provides an alternative to a large number of maps in the spatial domain, each containing the ray field information associated with a single acquisition point. The advantages of this approach are particularly pronounced in complex media, where ray field maps in the spatial domain become multi-valued and cumbersome to work with in practical applications. As shown in Chapter 3, a ray field map in the position/angle domain is single-valued, regardless of the complexity of the medium.

Moreover, modern approaches to ray-based imaging in complex media deal with the problems of multi-pathing by parameterising the imaging integrals in terms of scattering angles and azimuths at depth. This requires the ray field information to be organised by angles at depth, exactly as it is provided by a ray field map in the position/angle domain. However, in contrast to what is commonly assumed, obtaining this information does not require the tracing of rays from the image points up towards the acquisition surface. Instead, existing algorithms that trace downwards can be be adapted to work in the position/angle domain, leading to a considerable gain in efficiency. For further discussion on algorithms see Section 7.4.2.

Whether or not ray field maps in the position/angle domain will be able to play an important role in tomorrow's imaging processes not only depends on the availability of practical algorithms to calculate them, but also on the future status of angle domain imaging methods. Although the theoretical advantages of these methods are generally recognised and preliminary case studies show encouraging results, practical application, especially in 3-D, still faces some challenges.

The greatest practical problem for angle domain imaging currently seems to be related to the data flow. The evaluation of imaging integrals in terms of angles at depth requires more or less random data access. Since the total volume of data in 3-D experiments is extremely large this may result in a dramatic increase in disk I/O. Solutions to this problem may perhaps be found in alternative ways of storing the data, another possible future subject of research.

Ray-based imaging methods also face the challenge from competing "wave-equation imaging" methods. Traditionally, these methods, which are based on more accurate wave theories, have not been used as often for production work in 3-D because of their high computational cost. However, continuing advances in both algorithms and computing power have pushed the wave equation methods (almost?) within the range of feasibility. Despite this competition, ray-based imaging techniques will probably remain of interest for some time to come as the preferred tools for target-oriented approaches and migration velocity analysis.

Ray field maps in the position/angle domain are expected to be useful in various other applications of ray theory besides imaging. Within certain practical limits the maps may contain information on every ray in the medium that intersects the acquisition surface. This amount of information may, for example, be utilised by extensions of ray theory such as Maslov, Gaussian beams, and coherent state methods to obtain wave field solutions with a wider range of validity than zeroth-

order ray theory.

Also, the one-to-one mapping between position/angle coordinates and ray field coordinates may be exploited by making a change of variables in practical calculations. As demonstrated in Appendices A and B, this may be advantageous in tomography and the forward calculation of ray fields maps directly on a grid in the position/angle domain. These possibilities are interesting areas for future research.

7.4.2 Algorithms

Adaptations of two well-known ray tracing algorithms to the position/angle domain have been studied. In Chapter 5 it was concluded that the ray field construction algorithm in its current form is not suited for application in the position/angle domain. The main problem is the accommodation of deformations in the ray field structure during propagation. It was found that in the position/angle domain this deformation is predominantly shear-like, whereas the algorithm is designed primarily for diverging ray fields.

In fact, from the findings of Chapter 6 it may be concluded that the ray field construction algorithm is actually overqualified for application in the position/angle domain. The geometrical spreading of ray fields is limited in the position/angle domain, and the corresponding ray field maps are uniformly single-valued. Therefore, two of the most important reasons for developing wave front construction methods in the spatial domain, viz. the occurrence of shadow zones and multi-pathing, are absent in the position/angle domain. As a result, the more primitive but also more efficient paraxial ray methods were adapted successfully to the position/angle domain. In Chapter 6 the paraxial ray method was shown to work adequately in the example provided. The gain in speed with respect to tracing upwards is roughly proportional to the average ray length divided by the spatial grid interval.

Stop press: upon submission of the manuscript of this thesis the author was made aware of the paper by Fomel and Sethian (2002), who propose an algorithm that, basically, has the same objective as the work described here. They also consider an enlarged domain of position and angles, in which they determine what they call the "exit times and positions".

In fact, the algorithm they propose is very similar to the one that was presented in Kraaijpoel et al. (2002), which was not included in this thesis. The essence of both of these algorithms is to trace short ray segments up to a point where the relevant information is available and may be interpolated. The general disadvantage of this approach is the repeated interpolation, which makes the algorithms very sensitive to error propagation. In this respect, the advantage of the paraxial ray method of Chapter 6 is that – apart from the errors in ray tracing that are present in all approaches – an error is introduced only once, in the evaluation of the averaging integral, and hence error propagation is not an issue. Moreover, Fomel and Sethian (2002) require a sorting procedure to determine the order in

which the grid points should be updated. In the paraxial ray method this sorting is implicitly performed by the propagating rays. Nevertheless, a quantitative comparison will be an interesting objective for future research.

Finally, in Appendix B a set of equations is presented for the calculation of ray fields maps directly on a grid in the position/angle domain. In theory, this approach may lead to a very efficient algorithm, because it avoids two computationally expensive steps that are common to the other ray methods: first, the explicit mapping from spatial (or position/angle) coordinates to ray field coordinates, and second, the (sufficiently smooth) interpolation of medium properties at arbitrary spatial locations.

Appendix A

Formalism for tomography in the position/angle domain

The one-to-one mapping between ray field and position/angle coordinates can be exploited in practical applications. Calculations that are typically performed in terms of ray field coordinates can now be performed in terms of position/angle coordinates or the other way around.

This appendix shows that such a change of coordinates may be advantageous in tomography. If the ray field map in position/angle domain is known for a reference model, the cost function gradient can easily be calculated by integration with the cost function sensitivity kernel. Moreover, the parameterisation of the model perturbation may be designed after calculating the ray field map and the accompanying experiment and cost function sensitivity kernels. This provides a greater freedom in the choice of model space regularisation as well as a greater insight in its effects.

A.1 Configuration

Tomography is studied in a borehole configuration. Two boreholes are located in a 2-D medium with spatial coordinates $\boldsymbol{x} = (x, z)$. The source borehole is located at $x = x_s$ and the receiver borehole at $x = x_r$. The medium is assumed to be isotropic with a phase velocity $V(\boldsymbol{x})$. The spatial variations of the velocity are parameterised by a set of basis functions $B_j(\boldsymbol{x})$ and a set of coefficients $\boldsymbol{c} = (c_j)$:

$$\ln V[\boldsymbol{c}](\boldsymbol{x}) = \ln V_0(\boldsymbol{x}) + \sum_j c_j B_j(\boldsymbol{x}), \tag{A.1}$$

where $V_0(\boldsymbol{x})$ is an arbitrary reference model. The basis functions $B_j(\boldsymbol{x})$ may be blocks, splines, Fourier functions, Chebyshev polynomials etc. The parameterisa-

tion in terms of natural logarithms is chosen – without loss of generality – because it leads to simpler equations in the following.

The Hamiltonian ray tracing formulation is explained in Chapter 2, and the Hamiltonian is chosen as in Equation (2.68):

$$\mathcal{H}[\boldsymbol{c}](\boldsymbol{x},\boldsymbol{p}) = \ln|\boldsymbol{p}| + \ln V[\boldsymbol{c}](\boldsymbol{x}), \tag{A.2}$$

with slowness vector $\boldsymbol{p} = (p_x, p_z)$ and an explicit dependence on $\boldsymbol{c}$. The flow parameter for this choice is equivalent to travel time T.

The partial derivative of the Hamiltonian to medium parameter c_j is a simple expression:

$$\frac{\partial \mathcal{H}[\boldsymbol{c}](\boldsymbol{x},\boldsymbol{p})}{\partial c_i} = B_i(\boldsymbol{x}), \tag{A.3}$$

which explains the choices (A.1) and (A.2).

The rays emitted from the source borehole are parameterised by travel time T_s, source depth $\tilde{z}_s$ and initial slowness angle $\tilde{\phi}_s$. The slowness angle ϕ may be defined as, e.g., $\phi = \arctan(p_z/p_x)$.

The corresponding ray field map in position/angle domain (see Chapter 3) consists of the ray field coordinates $(T_s, \tilde{z}_s, \tilde{\phi}_s)$ as a function of the position/angle coordinates $(\boldsymbol{x}, \phi)$.

Similar ray field coordinates and corresponding ray field map can be defined for the ray field emitted from the receiver borehole, using ray field coordinates $(T_r, \tilde{z}_r, \tilde{\phi}_r)$. In practice only one of both maps is required, as shown below.

A.2 Travel time tomography

In travel time tomography the measured data is normally parameterised by the source and receiver depths: $\mathcal{T}^d(\tilde{z}_s, \tilde{z}_r)$. In the following it is useful to reparameterise this data in terms of the source ray field coordinates:

$$\mathcal{T}_s^d[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s) = \mathcal{T}^d(\tilde{z}_s, \tilde{z}_r[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)). \tag{A.4}$$

Because the ray path depends on $\boldsymbol{c}$, both $\tilde{z}_r$ and $\mathcal{T}_s^d$ depend on $\boldsymbol{c}$ if expressed in terms of the source ray field coordinates. The ray-theoretical travel times $\mathcal{T}_s^m$ obtained in the model with parameters $\boldsymbol{c}$ may be expressed as an integral over the ray path using flow parameter T_s. According to (2.26) this gives

$$\mathcal{T}_s^m[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s) = \int \left(\boldsymbol{p} \cdot \frac{d\boldsymbol{x}}{dT_s} - \mathcal{H}[\boldsymbol{c}](\boldsymbol{x},\boldsymbol{p}) \right) dT_s, \tag{A.5}$$

where both $\boldsymbol{x}$ and $\boldsymbol{p}$ depend on $\boldsymbol{c}$ and $(T_s, \tilde{z}_s, \tilde{\phi}_s)$ implicitly. Since the travel time is used as parameter along the ray, the integrand of (A.5) is approximately equal

to 1 for the ray calculated in the reference medium. The travel time misfit $\Delta\mathcal{T}_s$ is defined by

$$\Delta\mathcal{T}_s[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s) = \mathcal{T}_s^m - \mathcal{T}_s^d. \tag{A.6}$$

For least squares travel time tomography the cost function F may be defined as

$$F(\boldsymbol{c}) = \iint W_s \Delta\mathcal{T}_s^2 d\tilde{z}_s d\tilde{\phi}_s, \tag{A.7}$$

where continuous ("sufficiently dense") measurements are assumed.

Weight function W_s may be used for various purposes, such as tapering the data towards the edges of the acquisition domain, or scaling the data by its variance. Another purpose is to remove the arbitrariness of the choice of ray field coordinates. It is useful to define the weight function W in the "neutral" $(\tilde{z}_s, \tilde{z}_r)$-domain and choose

$$W_s[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s) = W(\tilde{z}_s, \tilde{z}_r) \det\left(\frac{\partial(\tilde{z}_s, \tilde{z}_r)}{\partial(\tilde{z}_s, \tilde{\phi}_s)}\right) = W(\tilde{z}_s, \tilde{z}_r)\frac{\partial\tilde{z}_r}{\partial\tilde{\phi}_s}. \tag{A.8}$$

This choice of weight function makes the cost function gradient (A.7) independent of the chosen ray field coordinate system. A similar expression as (A.7), using the receiver ray field coordinate system will then result in exactly the same evaluation of the cost function.

The goal of tomography is to find the set of medium parameters $\boldsymbol{c}$ that minimises cost function (A.7). A general approach to achieve this is the use of gradient methods. These require the calculation of the gradient of the cost function with respect to the medium parameters.

In the following, two formulations for the calculation of cost function gradients in travel time tomography will be treated. The first formulation, in Section A.2.1, assumes the ray end points are fixed, as is the usual approach in travel time tomography. The second formulation, in Section A.2.2, uses fixed initial conditions for the rays and free end points.

A.2.1 Cost function gradient for fixed ray end points

An important result of the ray perturbation theory summarised in Section 2.5, is that the derivative of the travel time with a medium parameter may be calculated along the unperturbed ray by fixing the ray end points. The fact that this is not possible in the presence of caustics is neglected here, because travel time tomography is not defined if the number of arrivals may vary due the medium perturbation.

Fixing the ray's end points means that the partial derivatives of both W_s and $\mathcal{T}_s^d$ with respect to $\boldsymbol{c}$ vanish. Differentiating Equation (A.7) then gives

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = \iint 2W_s \Delta\mathcal{T}_s \frac{\partial\Delta\mathcal{T}_s[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)}{\partial c_i} d\tilde{z}_s d\tilde{\phi}_s, \tag{A.9}$$

and for the derivative of $\Delta\mathcal{T}_s$ Equations (2.48) and (A.3) lead to

$$\frac{\partial \Delta\mathcal{T}_s[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)}{\partial c_i} = \frac{\partial \mathcal{T}_s^m[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)}{\partial c_i} = -\int B_i(\boldsymbol{x}[\boldsymbol{c}](T_s,\tilde{z}_s,\tilde{\phi}_s))dT_s. \tag{A.10}$$

Inserted in equation (A.9) this yields an expression for the cost function gradient as an integral over the source ray field coordinates:

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -\iiint 2W_s\Delta\mathcal{T}_s B_i(\boldsymbol{x}[\boldsymbol{c}](\sigma_s,\tilde{z}_s,\tilde{\phi}_s))dT_s d\tilde{z}_s d\tilde{\phi}_s. \tag{A.11}$$

This expression for the cost function gradient is cumbersome to evaluate. The basis functions by which the medium is parameterised appear in the integral as functions of the ray paths. Since these ray paths may in general be curved, the evaluation of the integral will be rather complicated.

Change of coordinates

At this point the one-to-one mapping between ray field and position/angle coordinates can be exploited by making a change of variables. The integral over the ray field coordinates can be replaced by an integral over position/angle coordinates using the differential relation:

$$dT_s d\tilde{z}_s d\tilde{\phi}_s = \mathcal{J}_s dx dz d\phi, \tag{A.12}$$

in terms of the ray field map jacobian

$$\mathcal{J}_s[\boldsymbol{c}](\boldsymbol{x},\phi) = \det\left(\frac{\partial(T_s,\tilde{z}_s,\tilde{\phi}_s)}{\partial(\boldsymbol{x},\phi)}\right). \tag{A.13}$$

This leads to the following expression for the cost function gradient:

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -\iiint 2W_s\mathcal{J}_s\Delta\mathcal{T}_s B_i(\boldsymbol{x})dx dz d\phi, \tag{A.14}$$

which is rewritten as

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -\iiint 2W\mathcal{J}^*\Delta\mathcal{T} B_i(\boldsymbol{x})dx dz d\phi, \tag{A.15}$$

with

$$\begin{aligned}\Delta\mathcal{T}[\boldsymbol{c}](\boldsymbol{x},\phi) &= \Delta\mathcal{T}_s[\boldsymbol{c}](\tilde{z}_s(\boldsymbol{x},\phi),\tilde{\phi}_s(\boldsymbol{x},\phi)) \\ &= \Delta\mathcal{T}_r[\boldsymbol{c}](\tilde{z}_r(\boldsymbol{x},\phi),\tilde{\phi}_r(\boldsymbol{x},\phi)),\end{aligned} \tag{A.16}$$

and

$$\mathcal{J}^*[\boldsymbol{c}](\boldsymbol{x},\phi) = \frac{\partial\tilde{z}_r}{\partial\tilde{\phi}_s}\mathcal{J}_s = \frac{\partial\tilde{z}_s}{\partial\tilde{\phi}_r}\mathcal{J}_r. \tag{A.17}$$

The latter relation can be proven as follows:

$$\begin{aligned}\frac{\mathcal{J}_r}{\mathcal{J}_s} &= \det\left(\frac{\partial(T_r,\tilde{z}_r,\tilde{\phi}_r)}{\partial(T_s,\tilde{z}_s,\tilde{\phi}_s)}\right)\\ &= \det\left(\frac{\partial(\tilde{z}_r,\tilde{\phi}_r)}{\partial(\tilde{z}_s,\tilde{\phi}_s)}\right)\\ &= -\det\left(\frac{\partial(\tilde{z}_s,\tilde{z}_r)}{\partial(\tilde{z}_r,\tilde{\phi}_r)}\right)^{-1}\det\left(\frac{\partial(\tilde{z}_s,\tilde{z}_r)}{\partial(\tilde{z}_s,\tilde{\phi}_s)}\right)\\ &= \left(\frac{\partial\tilde{z}_s}{\partial\tilde{\phi}_r}\right)^{-1}\frac{\partial\tilde{z}_r}{\partial\tilde{\phi}_s}.\end{aligned} \tag{A.18}$$

In contrast to (A.11), the integral (A.15) is relatively easy to evaluate, once the ray field map for medium parameters $\boldsymbol{c}$ is known, either for the source or the receiver ray field. If the basis functions B_j have only a limited support, as is the case with blocks and splines, integral (A.15) can be limited to within the support of B_i, which makes the evaluation efficient.

If the source ray field is used for the ray field map, the information required for the evaluation of (A.15) is: (i) as a function of (x,z,ϕ): $\tilde{z}_s$, $\tilde{\phi}_s$, and either T_s or $\mathcal{J}_s$, where the latter may be calculated from the former three by finite difference; (ii) as a function of $(\tilde{z}_s,\tilde{\phi}_s)$: $\mathcal{T}^m$, $\tilde{z}_r$, and $\partial\tilde{z}_r/\partial\tilde{\phi}_s$; and (iii) as a function of $(\tilde{z}_s,\tilde{z}_r)$: W, and $\mathcal{T}^d$. The information under (ii) and (iii) may be condensed in a single map.

Sensitivity kernels

Equation (A.15) may be written as

$$\begin{aligned}\frac{\partial F(\boldsymbol{c})}{\partial c_i} &= -\iiint E(\boldsymbol{x},\phi)\Delta\mathcal{T}(\boldsymbol{x},\phi)B_i(\boldsymbol{x})dxdzd\phi\\ &= -\iiint K_A(\boldsymbol{x},\phi)B_i(\boldsymbol{x})dxdzd\phi\\ &= -\iint K_S(\boldsymbol{x})B_i(\boldsymbol{x})dxdz,\end{aligned} \tag{A.19}$$

with E, K_A, and K_S three integral kernels. The first one,

$$E(\boldsymbol{x},\phi) = 2W\mathcal{J}^*, \tag{A.20}$$

may be called the *experiment sensitivity kernel*, as it is determined by the design of the experiment and the variance of the measurement. The other two may be called *cost function sensitivity kernels* (CFSK), one in the position/angle domain:

$$K_A(\boldsymbol{x},\phi) = E(\boldsymbol{x},\phi)\Delta\mathcal{T}(\boldsymbol{x},\phi), \tag{A.21}$$

and one in the spatial domain:

$$K_S(\boldsymbol{x}) = \int K_A(\boldsymbol{x}, \phi) d\phi. \tag{A.22}$$

These kernels express the sensitivity of the cost function to perturbations in phase velocity of the medium and include information on the misfit of the current model and the data.

Since no assumptions were made on the type of basis functions used for the medium parameterisation, $B_i(\boldsymbol{x})$ may represent a perturbation of any shape. Using this approach the parameterisation of the model perturbation, that is, the regularisation in the model space, may be designed after the calculation of the ray field map and its accompanying integral kernels. This provides a greater freedom in the choice of regularisation as well as a greater insight in its effects.

A more general treatment of the theory above will allow for anisotropic medium perturbations as well. The partial derivative of the Hamiltonian with respect to a medium parameter, however, will not be such a simple expression as in Equation (A.3). The basis functions $B_i(\boldsymbol{x})$ in the formulas above will therefore have to be replaced by the appropriate expressions that depend on ϕ as well. This leads to a similar expression for the cost function sensitivity kernel, and hence a similar approach to anisotropic tomography.

A.2.2 Cost function gradient for free ray end points

Ray based tomography for continuous data may also be developed using free ray end points. This means that the perturbation of the ray paths is taken into account explicitly. For travel time tomography it is not clear whether this leads to an improvement. As shown above it is perfectly feasible to express the perturbation of the travel time data in terms of the medium perturbation using fixed ray end points. In that case it is not even necessary to calculate the perturbed ray.

The fixed end point formalism, however, cannot be used in the presence of caustics, that is, when the number of arrivals may change due to a perturbation of the medium. Travel time tomography itself is not well defined then, because it is not possible to extract travel times from data if the number of arrivals is not fixed. Hence, for travel time tomography the fixed ray point formalism is sufficient.

Alternative ray-based tomographic methods that should work in the presence of multi-pathing rely on a formalism that uses free end points. As an example it is instructive to see how travel time tomography may be set up using free end points. In Section A.3 the same formalism is applied to differential semblance tomography.

The first step is to determine the partial derivative of $\Delta\mathcal{T}_s$ (A.8) to c_i:

$$\frac{\partial \Delta\mathcal{T}_s}{\partial c_i} = \frac{\partial \mathcal{T}_s^m}{\partial c_i} - \frac{\partial \mathcal{T}_s^d}{\partial c_i}. \tag{A.23}$$

The partial derivative of the measured travel time $\mathcal{T}_s^d$ (A.4) does not vanish this

time, due to the end point perturbations:

$$\frac{\partial \mathcal{T}_s^d[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)}{\partial c_i} = \frac{\partial \mathcal{T}^d(\tilde{z}_s, \tilde{z}_r)}{\partial \tilde{z}_r} \frac{\partial \tilde{z}_r[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)}{\partial c_i}, \tag{A.24}$$

and the partial derivative of the theoretical travel time $\mathcal{T}_s^m$ may be equated to that of the travel time T_r from the receiver (which may be chosen negative on the way to the receiver):

$$\frac{\partial \mathcal{T}_s^m[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)}{\partial c_i} = \frac{\partial T_r[\boldsymbol{c}](\tilde{z}_s, \tilde{\phi}_s)}{\partial c_i}, \tag{A.25}$$

hence

$$\begin{aligned} \frac{\partial \Delta \mathcal{T}_s}{\partial c_i} &= \frac{\partial \Delta \mathcal{T}_s}{\partial (T_r, \tilde{z}_r)} \cdot \frac{\partial (T_r, \tilde{z}_r)}{\partial c_i} \\ &= \frac{\partial T_r}{\partial\ c_i} - \frac{\partial T_d}{\partial \tilde{z}_r} \frac{\partial \tilde{z}_r}{\partial c_i} \end{aligned} \tag{A.26}$$

The partial derivatives of T_r and $\tilde{z}_r$ with respect to c_i may be expressed in terms of the end point perturbation of the ray for constant flow parameter T_s:

$$\frac{\partial (T_r, \tilde{z}_r)}{\partial c_i} = \frac{\partial (T_r, \tilde{z}_r)}{\partial \boldsymbol{x}_r} \cdot \left. \frac{\partial \boldsymbol{x}_r}{\partial c_i} \right|_{T_s}, \tag{A.27}$$

where $\boldsymbol{x}_r$ is a position at the receiver borehole. The derivatives $\partial (T_r, \tilde{z}_r)/\partial \boldsymbol{x}_r$ implement a change of coordinates from spatial coordinates $\boldsymbol{x}_r$ to emergence coordinates $(T_r, \tilde{z}_r)$, see Chapter 3, especially Equation (3.24).

Now, using Equation (2.45) the end point perturbation can be expressed as an integral along the ray:

$$\frac{\partial \boldsymbol{x}_r}{\partial c_i} = -\int \frac{\partial \boldsymbol{x}_r}{\partial \boldsymbol{p}} \cdot \partial_{\boldsymbol{x}} B_i(\boldsymbol{x}) dT_s, \tag{A.28}$$

and hence

$$\begin{aligned} \frac{\partial \Delta \mathcal{T}_s}{\partial c_i} &= -\int \left[\frac{\partial T_r}{\partial \boldsymbol{p}} - \frac{\partial T_d}{\partial \tilde{z}_r} \frac{\partial \tilde{z}_r}{\partial \boldsymbol{p}} \right] \cdot \partial_{\boldsymbol{x}} B_i(\boldsymbol{x}) dT_s \\ &= -\int \left[\frac{\partial T_r}{\partial (\phi, \mathcal{H})} - \frac{\partial T_d}{\partial \tilde{z}_r} \frac{\partial \tilde{z}_r}{\partial (\phi, \mathcal{H})} \right] \cdot \frac{\partial (\phi, \mathcal{H})}{\partial \boldsymbol{p}} \cdot \partial_{\boldsymbol{x}} B_i(\boldsymbol{x}) dT_s, \end{aligned} \tag{A.29}$$

where a change of coordinates from slowness to angles is made following the theory of Section 3.4. The partial derivatives $\partial (T_r, \tilde{z}_r)/\partial \phi$ can be estimated from the ray field map, and the partial derivatives to $\mathcal{H}$ can be calculated analytically. In fact, using Equations (2.43), (2.44), (3.18) and (3.24) it can be shown that $\partial T_r/\partial \mathcal{H} = T_r$ and $\partial \tilde{z}_r/\partial \mathcal{H} = 0$.

The travel time misfit (A.29) can be inserted in (A.9) and ignoring the partial derivatives of W – assuming W varies slowly – and a change of coordinates to the position/angle domain leads to

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -\iiint \boldsymbol{K}(\boldsymbol{x},\phi)\cdot\partial_{\boldsymbol{x}}B_i(\boldsymbol{x})dxdzd\phi, \tag{A.30}$$

with

$$\boldsymbol{K}(\boldsymbol{x},\phi) = 2W\Delta\mathcal{T}\left[\frac{\partial T_r}{\partial(\phi,\mathcal{H})} - \frac{\partial T_d}{\partial\tilde{z}_r}\frac{\partial\tilde{z}_r}{\partial(\phi,\mathcal{H})}\right]\cdot\frac{\partial(\phi,\mathcal{H})}{\partial\boldsymbol{p}}. \tag{A.31}$$

The cost function sensitivity kernel is now a vector, and the cost function gradient is sensitive to the gradients of the model perturbations. In principle this approach may lead to a better delineation of structures because ray bending is taken into account. Implementation and testing of the formalism, however, are left for future research.

A.3 Differential semblance tomography

This section briefly shows how the ray field map formalism may be applied to differential semblance tomography as well. For details of the tomographic method the reader is referred to other texts such as Symes and Carazzone (1991) and Plessix et al. (2000). Here the formulation of the latter will largely be followed, except that the data will be parameterised in terms of ray field coordinates, to allow for multiple arrivals.

For measured data $s(\tilde{z}_s,\tilde{z}_r,t)$, the back-propagated signal S is defined in terms of the source ray field coordinates:

$$S[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s,\tau) = \frac{s(\tilde{z}_s,\tilde{z}_r[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s),\mathcal{T}_s^m[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)+\tau)}{A[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)}, \tag{A.32}$$

where τ defines a time window, and A is the theoretical amplitude. In the following this amplitude will be considered invariant to first order to the medium perturbation. The differential semblance cost function used is

$$F(\boldsymbol{c}) = \int d\tau\iint\left[w_1\left(\frac{\partial S}{\partial\tilde{z}_s}\right)^2 + w_2\left(\frac{\partial S}{\partial\tilde{\phi}_s}\right)^2\right]d\tilde{z}_s d\tilde{\phi}_s, \tag{A.33}$$

with w_1 and w_2 arbitrary weights.

For the cost function gradient partial integration and discarding boundary terms (Plessix et al., 2000) leads to

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -2\int d\tau\iint\left[w_1\frac{\partial^2 S}{\partial\tilde{z}_s^2} + w_2\frac{\partial^2 S}{\partial\tilde{\phi}_s^2}\right]\left(\frac{\partial S}{\partial c_i}\right)d\tilde{z}_s d\tilde{\phi}_s. \tag{A.34}$$

The perturbation of S can be expressed in terms of the ray's end point perturbations:

$$\frac{\partial S}{\partial c_i} = \frac{1}{A}\left(\frac{\partial s}{\partial \tilde{z}_r}\frac{\partial \tilde{z}_r[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)}{\partial c_i} + \frac{\partial s}{\partial \tau}\frac{\partial T_r[\boldsymbol{c}](\tilde{z}_s,\tilde{\phi}_s)}{\partial c_i}\right). \tag{A.35}$$

The end point perturbations may be calculated in a similar fashion as in Section A.2.2, such that

$$\frac{\partial S}{\partial c_i} = \frac{1}{A}\int\left[\frac{\partial s}{\partial \tilde{z}_r}\frac{\tilde{z}_r}{\boldsymbol{p}} + \frac{\partial s}{\partial \tau}\frac{T_r}{\boldsymbol{p}}\right]\cdot\partial_{\boldsymbol{x}}B_i(\boldsymbol{x})dT_s. \tag{A.36}$$

This leads, after a change of variables to the position/angle domain, to a cost function gradient that is similar in appearance to (A.30)

$$\frac{\partial F(\boldsymbol{c})}{\partial c_i} = -\int d\tau \iiint \mathbf{K}(\boldsymbol{x},\phi,\tau)\cdot\partial_{\boldsymbol{x}}B_i(\boldsymbol{x})dxdzd\phi, \tag{A.37}$$

with the vectorial cost function sensitivity kernel

$$\mathbf{K}(\boldsymbol{x},\phi,\tau) = \frac{2}{A}\left(w_1\frac{\partial^2 S}{\partial \tilde{z}_s^2} + w_2\frac{\partial^2 S}{\partial \tilde{\phi}_s^2}\right)\left[\frac{\partial s}{\partial \tilde{z}_r}\frac{\partial \tilde{z}_r}{\partial(\phi,\mathcal{H})} + \frac{\partial s}{\partial \tau}\frac{\partial T_r}{\partial(\phi,\mathcal{H})}\right]\cdot\frac{\partial(\phi,\mathcal{H})}{\partial \boldsymbol{p}}. \tag{A.38}$$

This formulation of differential semblance tomography allows for efficient evaluation of cost function gradients and may be used in the presence of multi-pathing.

Appendix B

Finite difference calculation of ray field maps

B.1 Introduction

As discussed in Chapters 2 and 3 a distinguishing feature of ray methods is that calculations are performed in a dedicated coordinate system. The ray equations determine the flux lines of wave energy (in the high frequency asymptotic limit) in terms of the ray field coordinates. In order to evaluate the wave field in the spatial domain a mapping between the ray field coordinates and the spatial coordinates is required. Although ray methods are well known for their (relative) efficiency, the necessity of this mapping between ray parameters and spatial coordinates does not work in their advantage.

On the "road" from the initial conditions to the final evaluation of the wave field ray methods take a "detour" in the ray parameter domain. This detour originates in the use of the method of characteristics (Bleistein, 1984) to solve the eikonal equation. The eikonal equation is a non-linear first-order partial differential equation for the travel time of a propagating wave in the spatial domain. With the method of characteristics such an equation can be transformed into a set of ordinary differential equations for the characteristics which are, in general circumstances, easier to solve. Advantages include the possibility to find multi-valued solutions supported by the non-linear eikonal equation (multi-pathing), and easy calculation of additional quantities that are "tied" to the wave fronts, such as amplitudes and polarisation vectors.

Alternatively, the eikonal equation can be solved directly in the spatial domain, by means of finite difference methods (FD) (Vidale, 1988, 1990; Podvin and Lecomte, 1991; van Trier and Symes, 1991; Kim, 2002) or fast marching methods (FM) (Sethian, 1999; Sethian and Popovici, 1999; Alkhalifah and Fomel, 2001; Kim, 2002), both commonly referred to as eikonal solvers. The main disadvantage of these tech-

niques with respect to ray methods is that they yield only first arrivals and no additional information such as amplitudes and polarisations. Generalisation to general elastic media is difficult. Moreover, in complex velocity structures difficulties arise because of instabilities (FD) or low order accuracy (FM). A major advantage of these techniques, however, primarily of interest in less complex velocity structures, is the greater speed.

Both eikonal solvers (except Podvin and Lecomte, 1991) and ray methods solve the eikonal equation by means of expanding surfaces (boxes or "wave fronts"), which means that their computational dimension is roughly the same. A major reason for the computational advantage of the eikonal solvers is that they work directly in the spatial domain, while the ray methods suffer from the mapping between the spatial and ray field coordinates.

The extra computational effort involved with the mapping between the spatial and ray field coordinates manifests in two ways. First, obtaining ray field information such as travel times on a regular spatial grid requires regularisation of the data obtained in the ray field coordinate domain (see Chapters 5 and 6). Second, the rays calculated in terms of the ray field coordinates travel along arbitrary spatial coordinates where the medium properties must be evaluated. If the medium is defined on a regular grid, which is a common approach in complex velocity structure, the grid values must be interpolated. For stability and accuracy reasons this interpolation has to be smooth; a common approach is to use cubic splines. The combination of finding the location in the grid and smooth interpolation poses a considerable computational burden at each point calculated in the ray field.

A considerable improvement in the efficiency of ray methods is to be expected if the complete ray field maps, i.e., not only the travel time maps, can be calculated directly on a regular grid. This will remove the computationally intensive interpolation of medium properties and regularisation of ray information.

This appendix presents the equations that can be used to calculate ray field maps directly on a regular grid in position/angle domain, in a style that is similar to the FD eikonal equation solvers. The advantage of the position/angle domain is that rays in this domain do not cross, regardless of the complexity of the velocity structures. Rays can be regarded as flow lines in the eikonal manifold (see Chapter 3), and the ray field information carried by the rays can be modelled with advective equations on the spatial domain.

B.2 Equations and discussion

Let the ray field map in position angle domain be defined as the the ray field coordinates $(\tilde{\boldsymbol{x}}_0, T, \tilde{\phi}_0)$ as a function of the position angle domain $(\boldsymbol{x}, \phi)$. Further, let the ray field map be known at al level $z = z1$. The ray field map may now be continued in z direction if the vertical component of the local group velocity v_z does not vanish. For $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\phi}}$ defined at level z_1, the vertical derivative of the ray

field map may be expressed as

$$\left.\frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial z}\right|_{(x,y,\boldsymbol{\phi})} = \frac{\partial(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)}{\partial(\tilde{\boldsymbol{x}}, T, \tilde{\boldsymbol{\phi}})} \cdot \left.\frac{\partial(\tilde{\boldsymbol{x}}, T, \tilde{\boldsymbol{\phi}})}{\partial z}\right|_{(x,y,\boldsymbol{\phi})} \tag{B.1}$$

The partial derivatives of $(\tilde{\boldsymbol{x}}_0, T, \tilde{\boldsymbol{\phi}}_0)$ with respect to $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\phi}})$ can be estimated from the available data at z_1, for example by finite difference. The derivatives with respect to T vanish except for $\partial T/\partial T = 1$, because the ensemble parameters $(\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{\phi}}_0)$ do not change along a ray path. The derivative of $(\tilde{\boldsymbol{x}}, T, \tilde{\boldsymbol{\phi}})$ with respect to z is available analytically from Equation (3.24), and depends only on the medium properties.

Similar equations like (B.1) can be used for propagation in x and y directions. Combinations can be used in a finite difference algorithm based on expanding boxes or "wave fronts", similar to the methods described by Vidale (1988, 1990), van Trier and Symes (1991), and Kim (2002). The most complicated aspect of such an algorithm is the treatment of the boundaries of the domain. The development of an algorithm is strongly recommended as an area of future research, because it may have the potential to improve the efficiency of ray tracing in complex media considerably.

A last thing to note is that equation (B.1) basically implements a dynamic reparameterisation of the ray field, similar to the method discussed in Appendix C. The major difference and advantage of the current method is that in the position/angle domain there is a one-to-one mapping between ray field and position/angle coordinates. This allows the position/angle coordinates to be used directly for parameterising the propagating ray field, which avoids the necessity of an additional mapping.

Appendix C

Pseudo-spectral ray tracing with dynamic parameterisation

C.1 Introduction

In this appendix an algorithm is presented for the calculation of seismic ray fields in 2-D smooth media. The method is specifically designed for use in complex but smooth velocity structures where multi-pathing takes place.

The original objective was to make this algorithm part of a perturbation approach that would be able to handle variations in the number of ray arrivals under perturbations of the medium. This line of research was abandoned, however, in favour of the more practical ray field map methods which are central to the main text of this thesis. Nevertheless, the algorithm is interesting enough to be presented here. Its development provided useful insights for the ray field map approach (e.g., Appendix B) and some of its features may be useful in other applications.

The algorithm has two distinctive features related to the numerical representation of the ray field. First, the wave fronts are approximated using a pseudo-spectral expansion. Most ray tracing techniques propagate each ray individually, and sometimes approximate a piece of wave front by local interpolation (e.g., Vinje et al., 1993) or extrapolation (e.g., Beydoun and Keho, 1987). The new algorithm propagates entire wave fronts, represented by their pseudo-spectral expansion coefficients.

Second, it is shown that the parameterisation of the wave front is not necessarily determined by the natural parameterisation of the corresponding ray field, i.e. the parameterisation in terms of initial conditions for the rays. Using the smoothness and continuity of the wave front the parameterisation is adjusted dynamically in

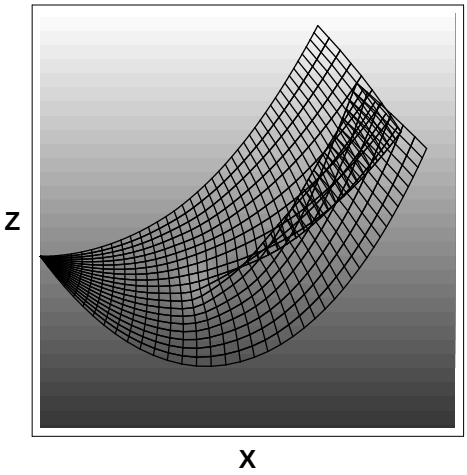


Figure C.1: A ray field emitted from a point source in a simple velocity structure. The velocity structure consists of a constant gradient of the slowness pointing in negative z-direction, and a slow anomaly embedded in it. This anomaly leads to a folding of the ray field and a consequent triplication of the number of arrivals in the region enclosed by the cuspoid caustic.

order to control the accuracy of the expansion.

C.2 Pseudo-spectral ray field calculation

C.2.1 Ray field

Subject of study is a central (point source) ray field in a 2-D medium, parameterised by two ray parameters: a flow parameter varying along the rays and a single ensemble parameter for the initial conditions. If the medium is smooth, each ray variable (position, slowness, traveltime, etc.) is a smooth function of the flow parameter. Similarly, if the initial conditions are smooth functions of the ensemble parameter, the entire ray field can be described by smooth functions of the ray parameters. In spatially varying velocity structures, rays within the ray field eventually cross in physical space and the number of rays crossing a particular point of space may change from place to place (see Figure C.1).

The information of interest in seismic imaging algorithms for example is usually a map of ray information on a grid in physical space. If rays cross in physical space, these maps will be multi-valued and discontinuous. It is important to realise that these irregularities are caused by the projection of the ray information onto physical space, not by discontinuities in the ray field. This can be verified by picturing the ray field in a higher dimensional space. The ray field defines a

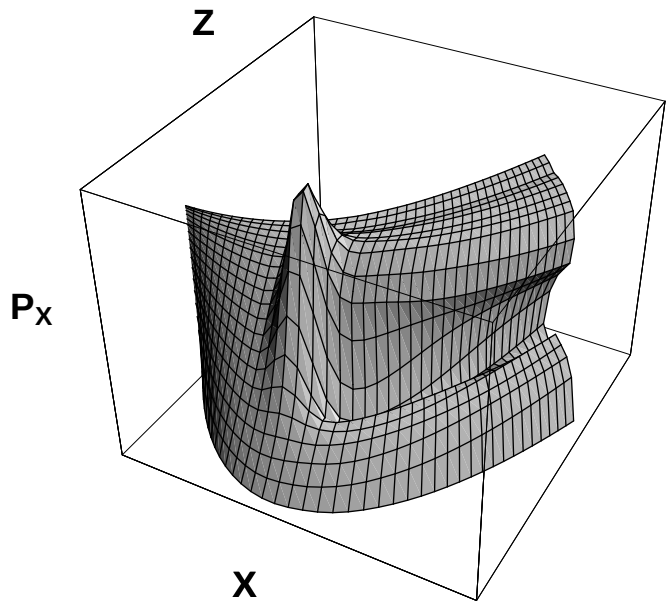


Figure C.2: The same ray field as in Figure C.1, but now in a 3-D projection of the 4D phase space consisting of the physical space at the horizontal plane and the x-component of the slowness along the vertical axis. The rays span a smooth manifold in this subspace, the projection of which onto physical space gives the rays of Figure C.1.

smooth, 2-D manifold in 4-D phase space, a 3-D projection of which can be seen in Figure C.2.

C.2.2 Wave front expansion

In this section and the following the phase space coordinates are written as a single vector $\boldsymbol{y}$, which contains both spatial coordinates and slowness components:

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{p} \end{pmatrix}. \tag{C.1}$$

The system of ray equations is summarised by

$$\frac{d\boldsymbol{y}}{d\sigma} = \boldsymbol{F}(\boldsymbol{y}), \tag{C.2}$$

with $\boldsymbol{F}$ an unspecified right hand side vector (for more details see Chapter 2).

The usual approach in ray tracing algorithms is to solve (C.2) for a discrete set of initial conditions parameterised by ensemble parameter ξ. Here, however,

a continuous estimate of the ray field is calculated. This is accomplished by expanding the ξ-dependence into a set of basis-functions $B_j(\xi)$ that are orthonormal with respect to an associated inner product:

$$\int_{\xi_0}^{\xi_1} B_i(\xi)B_j(\xi)W(\xi)d\xi = \delta_{ij}. \tag{C.3}$$

Here $W(\xi)$ is a weighting function, and δ is the Kronecker delta. The ray field may thus be expressed as a series expansion:

$$\boldsymbol{y}(\sigma,\xi) = \sum_{j=0}^{N} \boldsymbol{c}_j(\sigma)B_j(\xi). \tag{C.4}$$

Each vector of coefficients $\boldsymbol{c}_j$, which depends on σ, may be calculated using the relation

$$\boldsymbol{c}_j(\sigma) = \int_{\xi_0}^{\xi_1} \boldsymbol{y}(\sigma,\xi)B_j(\xi)W(\xi)d\xi. \tag{C.5}$$

In order to trace the coefficients $\boldsymbol{c}_j$, the right hand side of equation (C.2) has to be expanded likewise, with coefficients $\boldsymbol{f}_j(\sigma)$. This yields a differential equation for the coefficients:

$$\frac{d\boldsymbol{c}_j(\sigma)}{d\sigma} = \boldsymbol{f}_j(\sigma). \tag{C.6}$$

If the flow parameter σ corresponds to time, then for a constant σ, $\boldsymbol{x}(\sigma,\xi)$ describes the position of a wave front. Although this does not hold for other types of flow parameter, in the following these "ray fronts" will be referred to as wave fronts as well.

Tracing a progressing wave front with a certain accuracy requires in general a growing number of coefficients. The actual number of coefficients depends on the complexity of the wave front, which in turn is determined by the complexity of the medium traversed so far.

C.2.3 Pseudo-spectral methods

Global expansion methods are popular for the solution of partial differential equations, where they are collectively known as spectral methods. The usual approach is to expand the spatial dependence of the desired solution in a set of suitable basis functions, ending up with a set of ordinary differential equations for the temporal dependence (the method of lines). The current application is slightly different, because the dependence on the initial conditions of an ordinary differential equation

is expanded. Nevertheless, many aspects of both problems are the same, and the existing literature can be used (e.g., Boyd, 2000; Fornberg, 1996). This section provides a summary of the basic principles of spectral methods.

Given an equation

$$Lu = f(x), \tag{C.7}$$

where L is an operator, which is not necessarily linear, the solution is approximated by the series expansion

$$u(x) \approx u_N(x) = \sum_{n=0}^{N} a_n \phi_n(x). \tag{C.8}$$

Substitution of this series in the equation generates a residual function R:

$$R(x; a_n) = Lu_N - f. \tag{C.9}$$

Spectral methods are designed to find a series of coefficients a_n in such a way that the residual function is made as small as possible. The different spectral and pseudo-spectral techniques differ mainly in their way of minimising R.

The basic ingredients of a spectral method are:

- A suitable set of basis functions. The most desirable properties are completeness, rapid evaluation, rapid derivation, rapid convergence.
- An error criterion (norm), needed to quantify the quality of the approximation.
- An algorithm to determine the coefficients.

The most commonly used basis functions are Chebyshev polynomials, the default choice for functions on finite intervals, and trigonometric functions (Fourier) for periodic functions. Different sets are useful in case of special geometries (e.g. spherical harmonics) or unbounded intervals (e.g. Laguerre functions).

Since the quality of the approximation is usually determined by the maximum error on the entire interval, the norm of choice would be the L_∞, or sup-norm. The problem with this norm is that it is difficult to deal with analytically. The most practical norm is the L_2, or least-squares-norm, because it allows fast algorithms for the determination of coefficients, especially if the basis set is orthogonal.

Basically there are two types of algorithms to determine the expansion coefficients. The classical methods (Galerkin, Lanczos "tau") minimise the residual function by making it orthogonal to as many basis functions as possible. The modern collocation methods minimise the residual by making it zero at a finite set of points. The latter methods are usually referred to as pseudo-spectral methods. Choosing the right set of collocation points makes the pseudo-spectral methods equivalent to the classical methods.

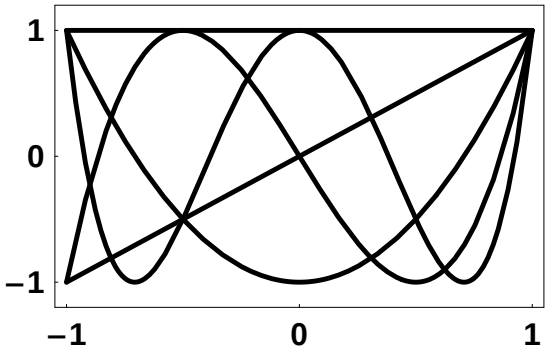


Figure C.3: The first five Chebyshev polynomials. These polynomials oscillate between $+1$ and -1 and form a basis for spectral expansion. For analytic functions the convergence is exponential and uniform over the expansion interval.

In the current application the pseudo-spectral technique is used in combination with Chebyshev polynomials. An advantage of this choice is that it provides uniform exponential convergence in the sup-norm for smooth functions. Algorithms for the Chebyshev expansion may be found in Press et al. (1992). The first five Chebyshev polynomials are displayed in Figure C.3.

C.2.4 Examples

Figure C.4 shows the application of the ray field expansion technique to the same model studied in Figures C.1 and C.2. Time is used as flow parameter and the integration of the coefficients in time was performed using a fourth order Runge-Kutta scheme. The wave fronts are the same as in Figure C.1. The rays, however, are now replaced by small ray segments. Although no individual rays were traced, the segments are plotted to provide insight into the structure of the ray field. The number of ray segments connecting two wave front is equal to the number of terms used in the expansion of the second of the two fronts. The number of coefficients increases during propagation.

Figures C.5 through C.7 show an example for a ray field traced in a low velocity layer. Figure C.5 depicts the ray field for a point source, for a scalar Hamiltonian ray tracing system with the choice $n = 2$ (see Section 2.7). Figure C.6 shows how the number of coefficients used depends on the complexity of the wave front, and may decrease as well as increase. Figure C.7 shows the logarithmic magnitude of the expansion coefficients for the z-coordinate of the last wave front in Figure C.5, and its linear trend reveals exponential convergence. This exponential convergence behaviour was initially considered an asset of the method, but unfortunately it is only observed for simple models like this one.

A problem that may occur is that the length scales of variations of the wave

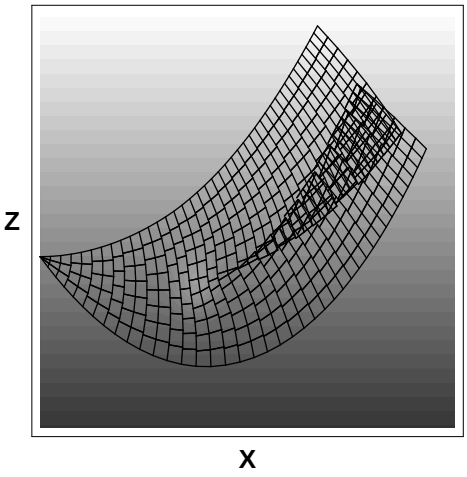


Figure C.4: The same ray field as in Figure C.1, but now traced using the ray field expansion technique.

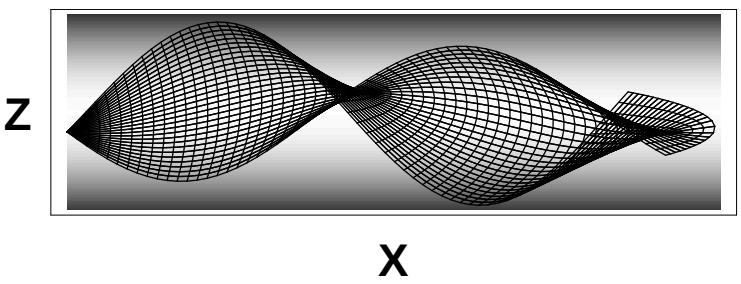


Figure C.5: A fan of rays in a low velocity layer calculated using the expansion approach. The rays are drawn for reference.

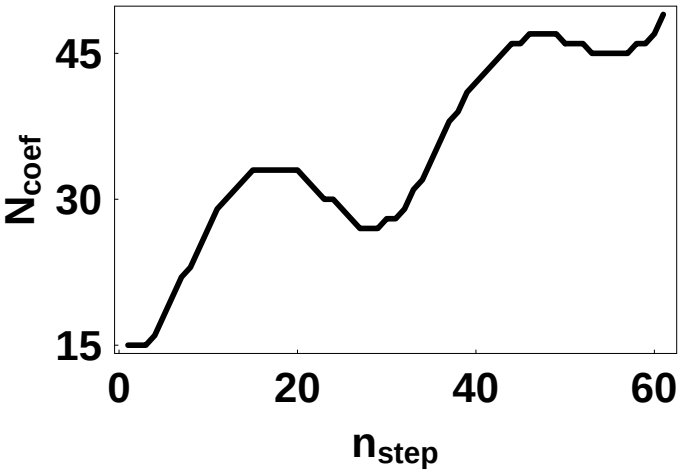


Figure C.6: The number of coefficients is related to the complexity of the wave front. In special circumstances, as in the ray field of Figure C.5, the wave front may get simpler in time which is reflected in the number of coefficients used.

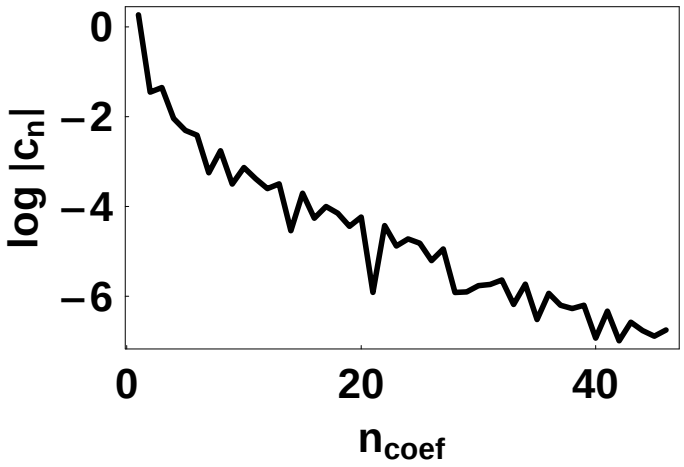


Figure C.7: The Chebyshev spectrum for z in the last wave front of Figure C.5 clearly shows the property of exponential convergence.

front may vary much over the range of ξ. Especially in regions of large geometrical spreading, the variations on a wave front may be large for a small range of take-off angles. The sampling density needed to approximate the wave front adequately is determined by those ranges with the smallest length scales of variations. Since the sampling is homogeneously distributed over the range of ξ — actually the sampling distribution is proportional to the weighting function $W(\xi)$ mentioned in Section C.2.2, as in Gaussian quadrature — this results in oversampling and thus inefficiency in regions of slower variations.

The phenomenon is illustrated by the example of Figures C.8 and C.9. The model consists of two Gaussian shape velocity anomalies. The first anomaly that the wave front meets is fast and causes a shadow zone with very large geometrical spreading. In order to maintain sufficient accuracy to actually see the second anomaly behind the first one, the algorithm needs to use a large amount of expansion terms. This results in severe oversampling in the areas of less geometrical spreading as is seen from the density of ray segments at the sides of the ray ensemble.

Figure C.9 shows the graphs of the ray variables as functions of the ensemble parameter ξ for the last wave front in Figure C.8. It is clear that the variations in these functions are mostly concentrated in a small region of ξ and thus it is this part that determines the number of coefficients to be used in the spectral expansion. This problem is dealt with in the following two sections.

C.3 Dynamic parameterisation

In the examples of Section C.2.4 take-off angles at the source have been used as ensemble parameters. It turns out that this is not a very efficient choice, because the variations of the various ray field variables are not distributed homogeneously over the range of take-off angles. The question is now what parameterisation to use in order to obtain maximal efficiency of the expansion technique.

The highest efficiency is achieved if the number of expansion terms needed for a certain level of accuracy is minimal. Since the function to be approximated is in fact a vector function, the best configuration is such that each vector component needs approximately the same number of terms to achieve its respective accuracy goal. Another way to put this is that all components of the vector should be comparably smooth.

To be able to deal with this analytically, a smoothness or roughness measure for vector functions has to be defined. The roughness of a single-valued function is often measured by an integral over the square of the derivative. An obvious generalisation of this is to define the roughness of a vector function by an integral over a weighted sum of the squared derivatives of all components.

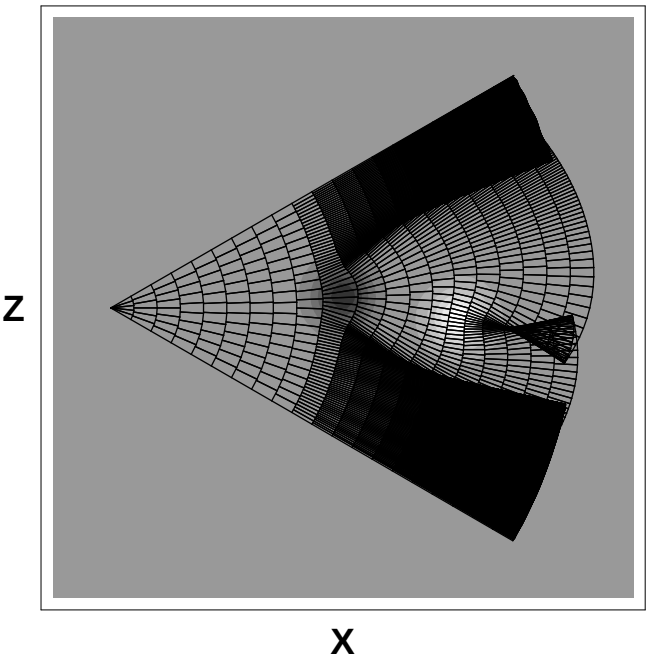


Figure C.8: An example of how large variations in geometrical spreading lead to a non-uniform sampling in the ray field expansion approach.

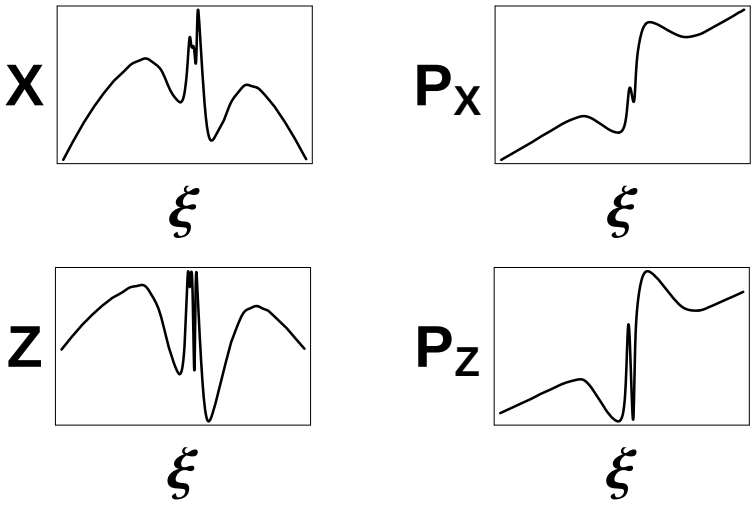


Figure C.9: Graphs of the ray field variables as a function of ensemble parameter ξ. The graphs correspond to the last wave front plotted in Figure C.8

C.3.1 Phase space metric

In a formal sense this is realised by defining a metric, or norm, in the vector space on which the function is defined. In our case, the vector space is the phase space, and the metric may be defined by

$$dl^2 = \langle d\boldsymbol{y}, d\boldsymbol{y} \rangle_{\boldsymbol{A}} = d\boldsymbol{y} \cdot \boldsymbol{A} \cdot d\boldsymbol{y}, \tag{C.10}$$

where $d\boldsymbol{y}$ is a differential in phase space and $\boldsymbol{A}$ is a positive definite matrix containing weights. For a one parameter curve $\boldsymbol{y}(\gamma)$ in phase space this metric determines the length of an infinitesimal segment of the curve by

$$\left(\frac{dl}{d\gamma} \right)^2 = \left\langle \frac{d\boldsymbol{y}}{d\gamma}, \frac{d\boldsymbol{y}}{d\gamma} \right\rangle_{\boldsymbol{A}}. \tag{C.11}$$

The roughness measure R for such a one parameter curve in phase space may then be defined as

$$R = \int_{\gamma_0}^{\gamma_1} \left(\frac{dl}{d\gamma} \right)^2 d\gamma. \tag{C.12}$$

A proper definition of the phase space metric would be such that the parameterisation γ that minimises the roughness R, also minimises the number of expansion terms needed for any component of $\boldsymbol{y}$. Since the optimal metric will probably be different for each situation and cannot be determined without extensive calculations, the most practical thing to do is to define a parametric form for the metric that is adequate for most circumstances, and that may be adapted to specific situations by a rule-of-thumb.

The definition of a metric in phase space implies the summation of coordinate and slowness terms, which have a different physical meaning and hence different ranges of values. In order to compensate this at least one degree of freedom is required in the metric. The weight matrix can be defined as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \alpha \boldsymbol{I} \end{pmatrix}, \tag{C.13}$$

with α, the *phase space geometry factor*, the desired degree of freedom. This yields the metric

$$dl^2 = \langle d\boldsymbol{y}, d\boldsymbol{y} \rangle_{\boldsymbol{A}} = d\boldsymbol{x} \cdot d\boldsymbol{x} + \alpha \, d\boldsymbol{p} \cdot d\boldsymbol{p}. \tag{C.14}$$

The challenge now is to find the parameterisation that minimises the roughness of a wave front. To find it, it is important to recognise the following two facts. First, the smoothness of the Lagrangian manifold in phase space ensures that for any non-degenerate parameterisation γ, the metric derivative $dl/d\gamma$ is strictly positive. Second, the integral of this metric derivative is equal to the length of the wave front

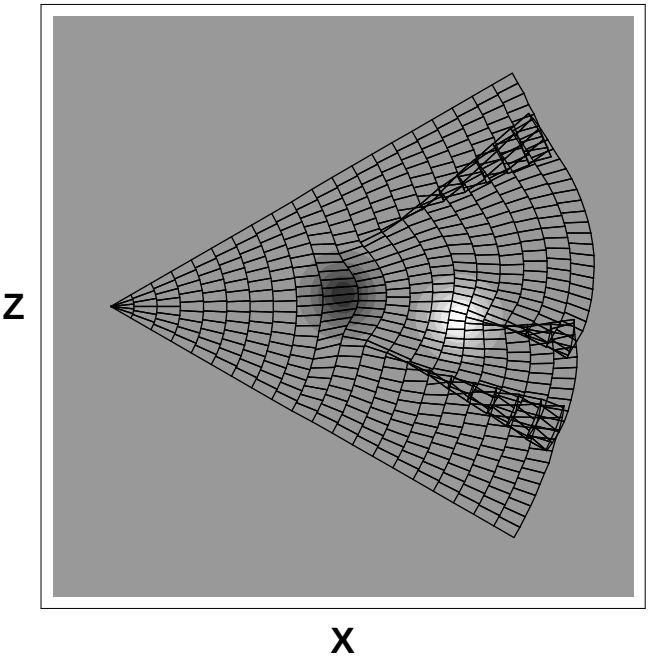


Figure C.10: The same ray field as in Figure C.8, now using the continuous reparameterisation formalism. The ray field is now homogenously sampled.

in phase space corresponding to the chosen metric, and hence is independent of the actual parameterisation. Combination of these facts suggests that the desired parameterisation is the one that renders the length of a differential wave front segment constant:

$$\frac{dl}{d\gamma} = C(\sigma). \tag{C.15}$$

This can be shown by means of the calculus of variations (see Appendix C.A.1). A possible approach to determine a value for α is explained in Appendix C.A.2.

After choosing this optimal parameterisation the next challenge is how to achieve it in practice. Suppose that criterion (C.15) is met for a given wave front, then simply using the system (C.2) to propagate to the next value of σ will violate it instantly. In order to maintain optimal parameterisation throughout the ray field integration, a dynamic reparameterisation formalism is presented in Section C.3.2.

C.3.2 Dynamic wave front parameterisation

Since an ensemble of rays is traced at once, for each value of σ not only the derivative along the ray ($\partial \boldsymbol{y}/\partial\sigma$) is known, but also the derivative along the wave front ($\partial \boldsymbol{y}/\partial\xi$). This makes it possible to integrate the ray field not along the ray paths themselves, but along more general paths. This is accomplished by adding an extra term to the ray equations which is proportional to the derivative along the wave front. In essence this implies a reparameterisation of the ray field, which is represented by the new symbol $\tilde{\boldsymbol{y}}$:

$$\tilde{\boldsymbol{y}}(\lambda,\gamma) = \boldsymbol{y}(\sigma(\lambda,\gamma),\xi(\lambda,\gamma)). \tag{C.16}$$

The new parameters λ and γ may be compared to σ en ξ respectively. Lines of constant λ parameterise the wave fronts, while lines of constant γ do not correspond to rays anymore, since the γ-dependence of $\tilde{\boldsymbol{y}}$ is basically a reparameterisation of the ξ-dependence of $\boldsymbol{y}$. For an arbitrary function $f(\lambda,\gamma)$ the following holds:

$$\frac{\partial\tilde{\boldsymbol{y}}}{\partial\lambda} = \boldsymbol{F}(\tilde{\boldsymbol{y}}) + f(\lambda,\gamma)\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma}, \tag{C.17}$$

in which $\boldsymbol{y}$ now depends on λ and γ as in the right hand side of (C.16). Equation (C.17) may be used as a basis for an expansion approach as in Section C.2. Appendix C.A.3 contains a derivation of (C.17).

The trick is now to choose $f(\lambda,\gamma)$ in (C.17) in such a way that for the ray field solution equation (C.15) holds for every value of λ. To find the proper constraint on $f(\lambda,\gamma)$ it is required that:

$$\frac{d}{d\lambda}\left(\frac{dl}{d\gamma}\right) = D(\lambda), \tag{C.18}$$

where the value of $D(\lambda)$ is the derivative of $C(\lambda)$, which in turn is the equivalent to $C(\sigma)$ from equation (C.15). Hence the meaning of $D(\lambda)$ is the derivative of the length of the wave front with λ. Considering

$$\frac{\partial}{\partial\lambda}\left\langle\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma},\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma}\right\rangle_{\boldsymbol{A}} = 2\left\langle\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma},\frac{\partial^2\tilde{\boldsymbol{y}}}{\partial\gamma\partial\lambda}\right\rangle_{\boldsymbol{A}}, \tag{C.19}$$

for any choice of metric, using (C.11) and (C.17) leads to

$$\frac{\partial}{\partial\lambda}\left(\frac{dl}{d\gamma}\right) = \left(\frac{dl}{d\gamma}\right)^{-1}\left[\left\langle\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma},\frac{\partial\boldsymbol{F}(\tilde{\boldsymbol{y}})}{\partial\gamma}\right\rangle_{\boldsymbol{A}} + f(\lambda,\gamma)\left\langle\frac{\partial\tilde{\boldsymbol{y}}}{\partial\gamma},\frac{\partial^2\tilde{\boldsymbol{y}}}{\partial\gamma^2}\right\rangle_{\boldsymbol{A}} + \left(\frac{\partial f(\lambda,\gamma)}{\partial\gamma}\right)\left(\frac{dl}{d\gamma}\right)^2\right]. \tag{C.20}$$

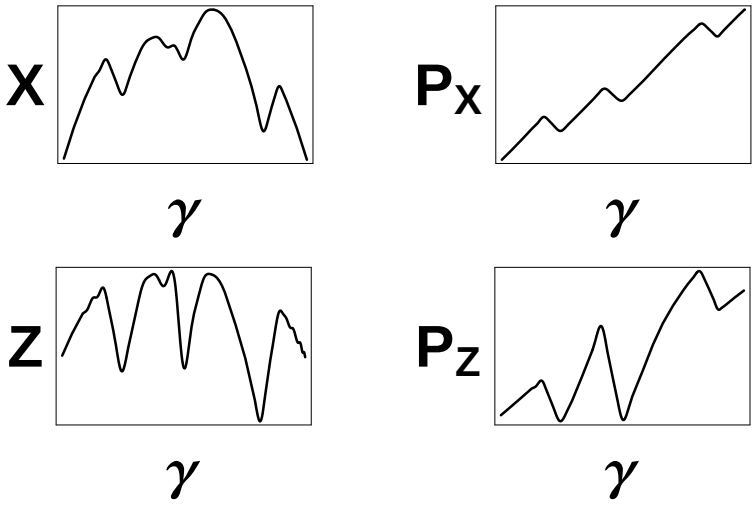


Figure C.11: Graphs of the ray field variables as a function of ensemble parameter γ. The graphs correspond to the last wave front plotted in Figure C.10, and are much smoother than the corresponding ones in Figure C.9

Since the magnitude of $d\tilde{\boldsymbol{y}}/d\gamma$ is constant, the second term in the right hand side vanishes. This is easily verified if by recognising

$$\left\langle \frac{\partial \tilde{\boldsymbol{y}}}{\partial \gamma}, \frac{\partial^2 \tilde{\boldsymbol{y}}}{\partial \gamma^2} \right\rangle_{\boldsymbol{A}} = \left(\frac{dl}{d\gamma} \right) \frac{\partial}{\partial \gamma} \left(\frac{dl}{d\gamma} \right). \tag{C.21}$$

Together, equations (C.18) and (C.20) yield an expression for the derivative of $f(\lambda, \gamma)$, which should be satisfied everywhere to achieve optimal parameterisation:

$$\frac{\partial f(\lambda,\gamma)}{\partial \gamma} = \frac{D(\lambda)}{C(\lambda)} - \left\langle \frac{\partial \tilde{\boldsymbol{y}}}{\partial \gamma}, \frac{\partial \boldsymbol{F}(\tilde{\boldsymbol{y}})}{\partial \gamma} \right\rangle_{\boldsymbol{A}} C(\lambda)^{-2}. \tag{C.22}$$

Hence $f(\lambda, \gamma)$ is determined except for two constants which can be used to make $f(\lambda, \gamma)$ vanish at the boundaries of the ensemble. Although the right hand side of (C.17) is more expensive to calculate than the original one (C.2), the new method is more efficient because the wave fronts require fewer expansion terms.

The dynamic reparameterisation is illustrated in Figures C.10 and C.11. These Figures show the same model as Figures C.8 and C.9 from Section C.2.4, but now computed using the reparameterisation approach. It is obvious both from the ray field plot and from the graphs that the reparameterisation does a good job in reducing the number of expansion terms needed.

C.4 Ray field mapping

The integration of the ray tracing system (C.17) yields ray-related variables such as position, slowness, amplitude, traveltime as a function of the ray parameters (λ, γ). Finding out which rays arrive at a certain grid-position $\boldsymbol{x}^*$ means solving the non-linear system

$$\boldsymbol{x}(\lambda, \gamma) = \boldsymbol{x}^*, \tag{C.23}$$

the ray field map discussed in Chapter 3. There may be multiple solutions, each corresponding to a distinct arrival. The total number of solutions is unknown a priori. In general this means that finding all solutions in such a system is an exhaustive search. The problem may be attacked in three ways.

The first way and conceptually simplest way is to divide the ray field into cells in a way similar to the wave front construction techniques (e.g., Lambaré et al., 1996), followed by inverse interpolation inside the cells, see also Chapter 4.

Another way is to locate the caustics in the ray field. The caustics, expressed in terms of the ray parameters, separate different branches of the ray field, each of which has a single solution to the projection algorithm.

Finally, it is possible to use the fact that the total number of real and complex solutions to a polynomial system of equations is known a priori. The real solutions correspond to the ray-geometric arrivals. The choice of Chebyshev polynomials to expand the ray field allows to find all solutions. For example if one of the coordinates is used as independent parameter λ, each arrival is found by solving numerically a single polynomial equation. If another independent parameter is used, its dependence should also be fitted by a polynomial, which then yields a system of two bivariate polynomial equations. Similarly in 3-D this gives a system of either two or three polynomial equations in as many variables.

Although solving a system of polynomial equations is numerically demanding, the work may be reduced by using the solutions of one grid point as starting values for next. The polynomial system for neighbouring grid points is only slightly different, which allows the use of perturbation methods.

C.5 Discussion and conclusions

A new method for calculating ray-fields in 2-D smooth media was developed. The essence of the method is to consider the complete ray field as a single object. The dependence of the rays on the initial conditions is approximated using a pseudo-spectral expansion in terms of Chebyshev polynomials. The result of a calculation is a continuous analytic representation of the ray-field which may subsequently be used to project ray information onto a grid defined in the medium, to generate tables of possibly multi-valued travel times and amplitudes for use in imaging and source location algorithms.

In order to overcome difficulties of the method related to variations in geometrical spreading over the range of initial conditions a technique for dynamic reparameterisation was developed. It allows a more efficient calculation of the ray-field and its projections by abandoning the ray itself as the basis of the expansion.

The proposed technique is more complex and less flexible than for example the wave front construction methods and it is not easily extended to 3-D. It is not likely that it will be able to compete with wave front construction for speed in the ray field calculation. The primary purpose was to use the resulting analytic description of the ray field as a basis for perturbation techniques. These techniques were not developed, however, because the methods proposed in the following chapters are more powerful.

C.A Appendix

C.A.1 Optimal parameterisation

For a general function $f(x)$, which increases monotonically between x_0 and x_1, a reparameterisation is sought in terms of y ($y_0 \leq y \leq y_1$) such that the roughness condition of equation (C.12) is minimised. Write

$$\tilde{f}(y) = f(x(y)), \tag{C.24}$$

with $\tilde{f}(y_0) = f(x_0)$ and $\tilde{f}(y_1) = f(x_1)$. For the roughness integral this yields

$$R_y = \int_{y_0}^{y_1} \left(\frac{df(x)}{dx} \frac{dx(y)}{dy} \right)^2 dy. \tag{C.25}$$

Using calculus of variations (e.g., Lanczos, 1986) the requirement $\delta R_y = 0$ gives the Euler-Lagrange equation

$$\frac{d}{dy} \left(\frac{df(x)}{dx} \frac{dx(y)}{dy} \right) = \frac{d}{dy} \left(\frac{d\tilde{f}(y)}{dy} \right) = 0. \tag{C.26}$$

This shows that the derivative of the function in the new parameterisation is constant. The constant is determined by the prescribed values of the functions on the boundaries of the domain.

C.A.2 Determination of phase space metric

The purpose of the phase space metric and the phase space geometry factor α (C.3.1) is to be able to relate variations in slowness to variations in spatial coordinates. Comparable smoothness of these quantities along the range of γ may only be achieved if the metric is more or less equally sensitive to variations in both.

The goal is to choose α in such a way that both terms on the right hand side of (C.3.1) are of similar scale. A rule-of-thumb to determine α can be found through scale analysis. The slowness of the medium $u(\boldsymbol{x})$ is approximated by a constant U, and the gradient of the slowness by

$$\nabla u(\boldsymbol{x}) \approx \frac{\delta U}{L_U}, \tag{C.27}$$

where δU is representative for the variations in the slowness, and L_U the corresponding length scale. Now first consider the case where all space coordinates and slowness components are dependent variables. The square of the metric derivative is then given by

$$\left(\frac{dl}{d\gamma}\right)^2 = \frac{d\boldsymbol{x}}{d\gamma}\cdot\frac{d\boldsymbol{x}}{d\gamma} + \alpha\frac{d\boldsymbol{p}}{d\gamma}\cdot\frac{d\boldsymbol{p}}{d\gamma}. \tag{C.28}$$

The slowness may be represented by

$$\boldsymbol{p} = u(\boldsymbol{x})\widehat{\boldsymbol{n}}, \tag{C.29}$$

where the dependence on σ is implicit, and

$$\widehat{\boldsymbol{n}}(\gamma) = \begin{pmatrix}\cos\theta(\gamma)\\ \sin\theta(\gamma)\end{pmatrix}, \tag{C.30}$$

with $\theta(\gamma)$ the angle of $\boldsymbol{p}$ with the x-axis. This gives

$$\frac{d\boldsymbol{p}}{d\gamma}\cdot\frac{d\boldsymbol{p}}{d\gamma} = \left(\nabla u(\boldsymbol{x})\cdot\frac{d\boldsymbol{x}}{d\gamma}\right)^2 + u(\boldsymbol{x})^2\left(\frac{d\theta}{d\gamma}\right)^2. \tag{C.31}$$

For the scale analysis each derivative to γ is substituted by an estimate, as in the case of $\boldsymbol{x}$:

$$\left|\frac{d\boldsymbol{x}}{d\gamma}\right| \approx \frac{\Delta r}{\Delta\gamma}, \tag{C.32}$$

which gives

$$\left(\frac{dl}{d\gamma}\right)^2 \approx \left(\frac{\Delta r}{\Delta\gamma}\right)^2 + \alpha\left[\left(\frac{\delta U}{L_U}\frac{\Delta r}{\Delta\gamma}\right)^2 + U^2\left(\frac{\Delta\theta}{\Delta\gamma}\right)^2\right]. \tag{C.33}$$

A value for α is now found by equating the first and the second term of the right hand side:

$$\begin{aligned}\alpha &\approx \left(\frac{\Delta r}{\Delta\gamma}\right)^2\left[\left(\frac{\delta U}{L_U}\frac{\Delta r}{\Delta\gamma}\right)^2 + U^2\left(\frac{\Delta\theta}{\Delta\gamma}\right)^2\right]^{-1}\\ &= \left(\frac{\delta U^2}{L_U^2} + \frac{U^2}{R^2}\right)^{-1},\end{aligned} \tag{C.34}$$

where R, defined as

$$R = \frac{\Delta r}{\Delta \theta}, \tag{C.35}$$

is a representative value for the radius of curvature of the wave front. It depends on the medium and the initial conditions which of the terms in (C.34) dominates.

If one of the space coordinates is used as independent parameter the case is slightly different. If x is independent for example, and y is the dependent variable with slowness component $u(\boldsymbol{x})\sin\theta$, the metric turns out to be

$$\left(\frac{dl}{d\gamma}\right)^2 = \left(\frac{dy}{d\gamma}\right)^2 + \alpha\left(\frac{du(\boldsymbol{x})}{dy}\frac{dy}{d\gamma}\sin\theta + u(\boldsymbol{x})\cos\theta\frac{d\theta}{d\gamma}\right)^2. \tag{C.36}$$

This yields for α:

$$\alpha \approx \left(\frac{\delta U}{L_U}\sin\theta + \frac{U}{R}\cos\theta\right)^{-2}, \tag{C.37}$$

where it should be noted that θ is often small, to allow x to be used as an independent parameter. The most difficult part of these estimations of α is the a priori determination of a characteristic radius of curvature.

C.A.3 Derivation of Equation (C.17)

Equation (C.17) is derived in a more formal way by considering the change of parameterisation from (σ,ξ) to (λ,γ). The total differential for the new parameterisation $\tilde{\boldsymbol{y}}$ in terms of the old $\boldsymbol{y}$ gives:

$$\begin{aligned} d\tilde{\boldsymbol{y}} = &\left[\left(\frac{\partial \boldsymbol{y}}{\partial \sigma}\right)\left(\frac{\partial \sigma}{\partial \lambda}\right) + \left(\frac{\partial \boldsymbol{y}}{\partial \xi}\right)\left(\frac{\partial \xi}{\partial \lambda}\right)\right] d\lambda \\ &+ \left[\left(\frac{\partial \boldsymbol{y}}{\partial \sigma}\right)\left(\frac{\partial \sigma}{\partial \gamma}\right) + \left(\frac{\partial \boldsymbol{y}}{\partial \xi}\right)\left(\frac{\partial \xi}{\partial \gamma}\right)\right] d\gamma. \end{aligned} \tag{C.38}$$

Since there is no reason to adjust the parameterisation in the direction along the rays, choose

$$\frac{\partial \sigma}{\partial \lambda} = 1, \quad \text{and} \quad \frac{\partial \sigma}{\partial \gamma} = 0, \tag{C.39}$$

such that λ is equivalent to σ, and lines of constant λ will coincide with lines of constant σ. Combining equations (C.38) and (C.2) gives

$$\frac{\partial \tilde{\boldsymbol{y}}}{\partial \lambda} = \boldsymbol{F}(\boldsymbol{y}) + \left(\frac{\partial \boldsymbol{y}}{\partial \xi}\right)\left(\frac{\partial \xi}{\partial \lambda}\right), \quad \text{and} \tag{C.40}$$

$$\frac{\partial \tilde{\boldsymbol{y}}}{\partial \gamma} = \left(\frac{\partial \boldsymbol{y}}{\partial \xi}\right)\left(\frac{\partial \xi}{\partial \gamma}\right). \tag{C.41}$$

To be able to compute the ray field in its new parameterisation the partial derivative to ξ in equation (C.40) is eliminated using (C.41):

$$\frac{\partial \tilde{\boldsymbol{y}}}{\partial \lambda} = \boldsymbol{F}(\boldsymbol{y}) + \left(\frac{\partial \xi}{\partial \gamma}\right)^{-1} \left(\frac{\partial \xi}{\partial \lambda}\right) \frac{\partial \tilde{\boldsymbol{y}}}{\partial \gamma}, \tag{C.42}$$

and since the dependence of ξ on λ and γ is still free, the choice

$$f(\lambda, \gamma) = \left(\frac{\partial \xi}{\partial \gamma}\right)^{-1} \left(\frac{\partial \xi}{\partial \lambda}\right), \tag{C.43}$$

leads to equation (C.17).

Appendix D

First order multivariate interpolation methods

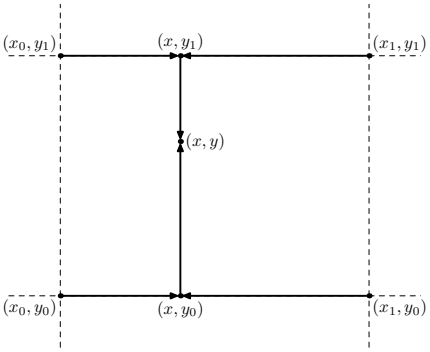
This appendix discusses two interpolation methods with first order accuracy that can be used in arbitrary dimension. For rectangular grids Section D.1 describes N-linear interpolation. For arbitrary data distributions equipped with a geometric structure based on triangulation barycentric interpolation can be used, as discussed in Section D.2.

D.1 Rectangular grids: N-linear interpolation

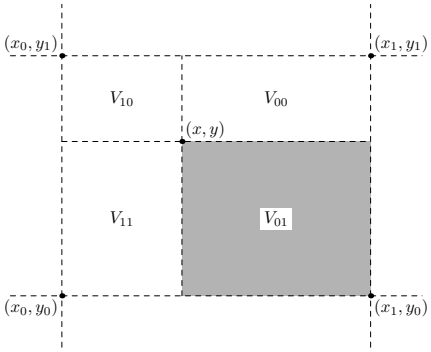
Rectangular grids allow a particularly simple generalisation of linear interpolation that is known as bilinear, trilinear or N-linear interpolation, in 2-D, 3-D, or N-D respectively. The technique is most easily explained in 2-D, but generalisation to N-D is straightforward.

For each interpolation point (x, y) the first step is to locate the cell that contains the point of evaluation, as shown in Figure D.1. If the point is located at a boundary between two cells it is immaterial which one of the two cells is chosen. In that case it is also possible to resort to one-dimensional interpolation on the cell boundary itself. All three options yield the same result.

The data points used in the bilinear interpolation are the vertices of the cell. If the cell is bounded by the grid lines $x = x_0$, $x = x_1$, $y = y_0$, and $y = y_1$, the vertices are located at the coordinates (x_0, y_0), (x_0, y_1), (x_1, y_0), and (x_1, y_1) (see Figure D.1). The bilinear interpolant $\mathcal{I}_0^{\mathrm{bl}}$ for function $f(x, y)$ may now be



(a) Bilinear interpolation: cascade of linear interpolations



(b) Bilinear interpolation: area weighted average

Figure D.1: Two interpretations of bilinear interpolation. Figure (a) shows how bilinear interpolation may be interpreted as the result of three linear interpolations. The first two interpolations are along the axes $y = y_0$ and $y = y_1$ to obtain intermediate data points at (x, y_0) and (x, y_1) respectively. The third interpolation is between these intermediate points to arrive at (x, y). Figure (b) shows how bilinear interpolation may be interpreted as area weighted averaging. The grey rectangle has an area V_{01}. The fraction of the total area of the cell occupied by the area V_{01} is the weight corresponding to grid point (x_0, y_1).

expressed as

$$\begin{aligned}\mathcal{I}_0^{\text{bl}}[f(x,y)](x,y;x_0,y_0,x_1,y_1) = \\ \frac{(y_1-y)}{(y_1-y_0)}\frac{(x_1-x)}{(x_1-x_0)}f(x_0,y_0) + \frac{(y-y_0)}{(y_1-y_0)}\frac{(x_1-x)}{(x_1-x_0)}f(x_0,y_1) + \\ \frac{(y_1-y)}{(y_1-y_0)}\frac{(x-x_0)}{(x_1-x_0)}f(x_1,y_0) + \frac{(y-y_0)}{(y_1-y_0)}\frac{(x-x_0)}{(x_1-x_0)}f(x_1,y_1).\end{aligned} \tag{D.1}$$

This equation can be interpreted in two ways that are both easily generalised to higher dimensions. The first interpretation is that it is an iterated one-dimensional linear interpolation, as in Equation (4.13). This view is clarified by two equivalent representations of (D.1):

$$\mathcal{I}_0^{\text{bl}}[f(x,y)](x,y;x_0,y_0,x_1,y_1) = \mathcal{I}_0\left[\,\mathcal{I}_0[f(x,y)](x;x_0,x_1)\,\right](y;y_0,y_1) \tag{D.2}$$

$$= \mathcal{I}_0\left[\,\mathcal{I}_0[f(x,y)](y;y_0,y_1)\,\right](x;x_0,x_1). \tag{D.3}$$

The sequence of interpolations corresponding to the first line is shown in Figure D.1(a). The fact that bilinear interpolation and N-linear interpolation in general can be seen as a cascade of linear interpolations shows that they are exact for linear functions.

Another way to interpret (D.1) is that the weight factor multiplying the value of each grid point represents a fractional area (see Figure D.1(b)). This is the area

spanned by the point of interpolation and the diametrically opposite grid point. Another alternative representation of (D.1) can thus be written as:

$$\mathcal{I}_0^{\mathrm{bl}}[f(x,y)](x,y;x_0,y_0,x_1,y_1) = \frac{V_{00}}{V}f(x_0,y_0) + \frac{V_{01}}{V}f(x_0,y_1) + \frac{V_{10}}{V}f(x_1,y_0) + \frac{V_{11}}{V}f(x_1,y_1), \quad \text{(D.4)}$$

with

$$V_{00} = (x_1 - x)(y_1 - y), \quad V_{01} = (x_1 - x)(y - y_0), \quad \text{(D.5)}$$

$$V_{10} = (x - x_0)(y_1 - y), \quad V_{11} = (x - x_0)(y - y_0), \quad \text{and} \quad \text{(D.6)}$$

$$V = V_{00} + V_{01} + V_{10} + V_{11}. \quad \text{(D.7)}$$

D.2 Triangulated distributions: barycentric coordinates

In order to interpolate on an irregular data distribution it is necessary to calculate some sort of geometrical structure that determines the connections between the grid points. A common approach is to divide the space into cells using simplices (e.g., Sambridge et al., 1995). In 2-D this is called a triangulation, in higher dimensions one speaks of a tesselation. A simplex, or N-simplex, is the N-D analogue of a triangle in 2-D and a tetrahedron in 3-D. The $N + 1$ vertices of each simplex are the data points.

For each data distribution many tesselations can be constructed. The minimal requirement is that the simplices do not overlap, and that there are no holes. Another important property for a tesselation is that the sides of the simplices are well balanced, that is, there are as few "thin" simplices as possible. The optimal tesselation in that respect is the Delaunay tesselation, which is closely related to the Voronoi diagram, another type of geometrical structure.

The Delaunay tesselation and the Voronoi diagram play an important role in natural neighbour interpolation (Sibson, 1981; Sambridge et al., 1995), an interpolation technique that is specifically designed for irregular data distributions. This type of interpolation is not useful in the context of Chapter 4, because the primary goal of natural neighbour interpolation is smoothness rather than accuracy. Moreover, the calculation of the smooth natural neighbour interpolants is quite expensive, and not easily generalised to higher dimensions.

It should be noted that the construction of tesselations is not the subject of discussion here. References to algorithms for the calculation of Delaunay tesselations are provided in Sambridge et al. (1995). In the ray field construction methods described in Chapter 5 a tesselation is easily constructed from the known ray field. In general, however, calculation of a tesselation from scratch is an expensive procedure.

In order to perform linear interpolation in a tesselated data distribution one must determine which simplex contains the point of evaluation. This is often more challenging than the interpolation itself, especially if the interpolation is performed at an arbitrary isolated point (e.g., Sambridge et al., 1995).

The challenge is somewhat reduced if the interpolation is used as a means for data regularisation, for example in the construction of ray field maps (see also Chapter 5). The points of interpolation are then organised in a rectangular grid. Instead of interpolating each grid point at the time, it is then more efficient to loop over the simplices. For each simplex the interpolations are performed for all the grid points it contains. A problem that remains, however, is to determine whether a given interpolation point lies inside a simplex.

Here, linear interpolation inside a simplex with arbitrary vertices is performed using barycentric coordinates, because these provide a simple criterion to determine whether a given interpolation point lies inside the simplex or not.

A simplex in N dimensions has $N + 1$ vertices. If the N-dimensional volume of the simplex does not vanish, i.e. it is non-degenerate, any point $\boldsymbol{x}$ in the N-dimensional space can be expressed uniquely as a linear combination of the vertices $\boldsymbol{x}_i$ $(i = 0 \ldots N)$:

$$\boldsymbol{x} = \sum_{i=0}^{N} \alpha_i \boldsymbol{x}_i, \tag{D.8}$$

with the constraint

$$\sum_{i=0}^{N} \alpha_i = 1. \tag{D.9}$$

The uniqueness of this representation allows the weights α_i to be interpreted as an alternative set of coordinates for point $\boldsymbol{x}$, the so-called barycentric coordinates.

Before it is shown how to obtain the barycentric coordinates for a given point $\boldsymbol{x}$ and vertices $\boldsymbol{x}_i$, a number of properties of barycentric coordinates are discussed, and how these are used for linear interpolation inside a simplex.

A very useful property of (D.8) and (D.9) is that if $0 \leq \alpha_i \leq 1$ for all i, then $\boldsymbol{x}$ is inside the simplex. This means that if the barycentric coordinates are known for a given point of interpolation it is easy to test whether it is inside the simplex or not.

A interpolation of first order accuracy may also be expressed conveniently in terms of barycentric coordinates and the function values at the vertices of the simplex. For an arbitrary linear function $f(\boldsymbol{x})$:

$$f(\boldsymbol{x}) = \boldsymbol{c}_0 + \boldsymbol{c}_1 \cdot \boldsymbol{x}, \tag{D.10}$$

the first order interpolant is determined using (D.8):

$$\begin{aligned} f(\boldsymbol{x}) &= f\left(\sum_{i=0}^{N} \alpha_i \boldsymbol{x}_i\right) \\ &= \boldsymbol{c}_0 + \boldsymbol{c}_1 \cdot \sum_{i=0}^{N} \alpha_i \boldsymbol{x}_i \\ &= \left(1 - \sum_{i=0}^{N} \alpha_i\right) \boldsymbol{c}_0 + \sum_{i=0}^{N} \alpha_i f(\boldsymbol{x}_i) \end{aligned} \tag{D.11}$$

and using (D.9) this leads to

$$f(\boldsymbol{x}) = \sum_{i=0}^{N} \alpha_i f(\boldsymbol{x}_i), \tag{D.12}$$

where it should be noted that the coordinates α_i implicitly depend on the interpolation point $\boldsymbol{x}$.

In order to determine the barycentric coordinates for a given interpolation point $\boldsymbol{x}$ and simplex vertices $\boldsymbol{x}_i$, it is useful to rewrite (D.8) and (D.9) into a consistent set of equations. This is done by cancelling α_0 from (D.8) using (D.9):

$$\sum_{i=1}^{N} \alpha_i (\boldsymbol{x}_i - \boldsymbol{x}_0) = \boldsymbol{x} - \boldsymbol{x}_0. \tag{D.13}$$

This is a system of N linear equations for N unknowns α_i $(i = 1 \ldots N)$. For any interpolation point $\boldsymbol{x}$ this leads to its barycentric coordinates, where α_0 can be determined using (D.9). If all barycentric coordinates are between 0 and 1, the point is inside the simplex and linear interpolations can be calculated using (D.12).

Bibliography

Aki, K. and P. G. Richards (2002). *Quantitative seismology* (2nd ed.). University Science Books.

Alkhalifah, T. and S. Fomel (2001). Implementing the fast marching eikonal solver: spherical versus cartesian coordinates. *Geophysical Prospecting 49*, 165–178.

Babich, V. M. (1994). Ray method of calculating the intensity of wavefronts in the case of a heterogeneous, anisotropic, elastic medium. *Geophys. J. Int. 118*, 379–383.

Bakker, P. M. (1998). Phase shift at caustics along rays in anisotropic media. *Geophys. J. Int. 134*, 515–518.

Bender, C. M. and S. A. Orszag (1978). *Advanced mathematical methods for scientists and engineers*. International series in pure and applied mathematics. McGraw-Hill.

Beydoun, W. B. and T. H. Keho (1987). The paraxial ray method. *Geophysics 52*, 1639–1653.

Bleistein, N. (1984). *Mathematical methods for wave phenomena*. Academic Press, Inc.

Bleistein, N., J. Cohen, and J. Stockwell, Jr. (2001). *Mathematics of multidimensional seismic imaging, migration, and inversion*. Springer.

Bleistein, N. and R. A. Handelsman (1987). *Asymptotic expansions of integrals*. Dover Publications.

Born, M. and E. Wolf (1980). *Principles of optics* (6th ed.). Cambridge University Press.

Boyd, J. P. (2000). *Chebyshev & Fourier Spectral methods*. Mineola, New York: Dover Publications, Inc.

Brandsberg-Dahl, S., M. V. de Hoop, and B. Ursin (2003). Focusing in dip and AVA compensation on scattering-angle/azimuth common image gathers. *Geophysics 68*, 232–254.

Brandsberg-Dahl, S., B. Ursin, and M. V. de Hoop (2003). Velocity analysis in the common scattering-angle/azimuth domain. *Geophysical Prospecting 51*, 295–314.

Brouwer, J. and K. Helbig (1998). *Shallow High-Resolution Reflection Seismics*, Volume 19 of *Handbook of Geophysical Exploration. Section I, Seismic Exploration*. Elsevier Science.

Bulant, P. and L. Klimeš (1999). Interpolation of ray theory traveltimes within ray cells. *Geophys. J. Int. 139*, 273–282.

Cartwright, J. H. E. and O. Piro (1992). The dynamics of Runge-Kutta methods. *Int. J. Bifurcation and Chaos 2*, 427–449.

Červený, V. (1972). Seismic rays and ray intensities in inhomogeneous anistropic media. *Geoph. J. R. astr. Soc. 29*, 1–13.

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

Červený, V. (2002). Fermat's variational principle for anisotropic inhomogeneous media. *Stud. Geophys. Geod. 46*, 567–588.

Červený, V., L. Klimeš, and I. Pšenčík (1984). Paraxial ray approximations in the computation of seismic wavefields in inhomogeneous media. *Geoph. J. R. astr. Soc. 79*, 89–104.

Červený, V., M. M. Popov, and I. Pšenčík (1982). Computation of wave fields in inhomogeneous media – gaussian beam approach. *Geoph. J. R. astr. Soc. 70*, 109–128.

Chapman, C. H. (2002). Seismic ray theory and finite frequency extensions. In *International Handbook of Earthquake and Engineering Seismology*, Chapter 9. IASPEI.

Chapman, C. H. and R. Drummond (1982). Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory. *Bull. Seism. Soc. Am. 72*(6), S277–S317.

Coman, R. and D. Gajewski (2002). 3-D Wavefront oriented ray tracing: Estimation of traveltimes within cells. In *Expanded abstracts, International Exposition and 72nd Annual Meeting*. SEG.

de Hoop, M. V. and S. Brandsberg-Dahl (2000). Maslov asymptotic extension of generalized Radon transform inversion in anistropic elastic media: a least squares approach. *Inverse problems 16*, 519–562.

Duff, D. (1993). *Holmes' principles of physical geology*. Nelson Thornes.

Epili, D. and G. A. McMechan (1996). Implementation of 3-D prestack Kirchhoff migration, with application to data from the Ouachita frontal thrust zone. *Geophysics 61*(5), 1,400–1,411.

Ettrich, N. and D. Gajewski (1996). Wave front construction in smooth media for prestack depth migration. *PAGEOPH 148*, 481–502.

Farra, V. and S. LeBégat (1995). Sensitivity of qp-wave travel times and polarization vectors to heterogeneity, anisotropy and interfaces. *Geophys. J. Int. 121*, 371–384.

Farra, V. and R. Madariaga (1987). Seismic waveform modeling in heterogeneous media by ray perturbation theory. *J. Geophys. Res. 92*, 2697–2712.

Fomel, S. and J. A. Sethian (2002). Fast phase-space computation of multiple arrivals. *Applied Mathematics 99*, 7329–7334.

Fornberg, B. (1996). *A practical guide to pseudospectral methods*. Cambridge University Press.

Fornberg, B. (1998). Calculation of weights in finite difference formulas. *SIAM Rev. 40*, 685–691.

Gajewski, D. (1993). Radiation from point sources in general anisotropic media. *Geophys. J. Int. 113*, 299–317.

Geoltrain, S. and J. Brac (1993). Can we image complex structures with first-arrival traveltime? *Geophysics 58*(4), 564–575.

Gibson, R. L. (2000). Ray tracing by wavefront construction for anisotropic media. In *Expanded abstracts, International Exposition and 70th Annual Meeting*. SEG.

Goldstein, H. (1980). *Classical Mechanics*. Addison-Wesley.

Julian, B. R. and D. Gubbins (1977). Three-dimensional seismic ray tracing. *J. Geophys. 43*, 95–113.

Kendall, J.-M., W. Guest, and C. J. Thomson (1992). Ray-theory Green's function reciprocity and ray-centred coordinates in anisotropic media. *Geophys. J. Int. 108*, 364–371.

Kendall, J.-M. and C. J. Thomson (1993). Maslov ray summation, pseudo-caustics, Lagrangian equivalence and transient seismic waveforms. *Geophys. J. Int. 113*, 186–214.

Kim, S. (2002). 3D eikonal solvers: first-arrival traveltimes. *Geophysics 67*, 1225–1231.

Klauder, J. R. (1987). Global, uniform asymptotic wave-equation solutions for large wavenumbers. *Anals of physics 180*, 108–151.

Kline, M. and I. W. Kay (1965). *Electromagnetic theory and geometrical optics*. Pure and applied mathematics. Interscience publishers.

Koren, Z., S. Xu, and D. Kosloff (2002). Target-oriented common reflection angle migration. In *Expanded abstracts, International Exposition and 72nd Annual Meeting*. SEG.

Kraaijpoel, D., K. Roy-Chowdhury, and R. Snieder (2002). Ray field maps in position/angle domain: concepts and construction. In *Expanded abstracts, International Exposition and 72nd Annual Meeting*. SEG.

Lambaré, G., P. Lucio, and A. Hanyga (1996). Two-dimensional multivalued traveltime and amplitude maps by uniform sampling of a ray field. *Geophys. J. Int. 125*, 584–598.

Lanczos, C. (1986). *The variational principles of mechanics*. Dover Publications.

Lee, W. H. K., H. Kanamori, P. C. Jennings, and C. Kisslinger (Eds.) (2002). *International Handbook of Earthquake and Engineering Seismology*, Volume 1. IASPEI.

Lucio, P. S., G. Lambaré, and A. Hanyga (1996). 3D multivalued travel time and amplitude maps. *PAGEOPH 148*, 449–479.

Mispel, J. (2001). *Transversely isotropic media: 3-D wavefront construction method and pre-stack depth migration*. Ph. D. thesis, Imperial College of Science, Technology and Medicine.

Operto, M. S., S. Xu, and G. Lambaré (2000). Can we quantitatively image complex structures with rays? *Geophysics 65*, 1223–1238.

Osher, S., L.-T. Cheng, M. Kang, H. Shim, and Y.-H. Tsai (2002). Geometric optics in a phase-space-based level set and Eulerian framework. *J. of Comput. Phys. 179*, 622–648.

Plessix, R.-E., W. Mulder, and A. ten Kroode (2000). Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation. *Geophysical Prospecting 48*, 913–935.

Podvin, P. and I. Lecomte (1991). Finite difference computation of travel times in very contrasted velocity models: a massively parallel approach. *Geophys. J. Int. 105*, 271–284.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C* (2nd ed.). Cambridge University Press.

Pšenčík, I. and T. N. Teles (1996). Point source radiation in inhomogeneous anisotropic structures. *PAGEOPH 148*, 591–623.

Qian, J., L.-T. Cheng, and S. Osher (2003). A level set-based Eulerian approach for anisotropic wave propagation. *Wave Motion 37*, 365–379.

Qian, J. and W. W. Symes (2002). An adaptive finite-difference method for traveltimes and amplitudes. *Geophysics 67*, 167–176.

Ralston, A. and P. Rabinowitz (2001). *A first course in numerical analysis* (2nd ed.). Dover Publications.

Sambridge, M., J. Braun, and H. McQueen (1995). Geophysical parameterization and interpolation of irregular data using natural neighbours. *Geophys. J. Int. 122*, 837–857.

Sethian, J. (1999). *Level set methods and fast marching methods*. Cambridge University Press.

Sethian, J. A. and A. M. Popovici (1999). 3-D traveltime computation using the fast marching method. *Geophysics 64*, 516–523.

Sheriff, R. E. and L. P. Geldart (1995). *Exploration Seismology*. Cambridge University Press.

Sibson, R. (1981). A brief description of natural neighbour interpolation. In V. Barnet (Ed.), *Interpreting multivariate data*, Chapter 2, pp. 21–36. Wiley.

Snieder, R. and M. Sambridge (1993). The ambiguity in ray perturbation theory. *J. Geophys. Res. 98*(B12), 22,021–22,034.

Sun, Y. (1992). Computation of 2-D multiple arrival traveltime fields by an interpolative shooting method. In *Expanded abstracts, International Exposition and 62nd Annual Meeting*. SEG.

Symes, W. and J. Carazzone (1991). Velocity inversion by differential semblance optimization. *Geophysics 56*, 654–663.

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

ten Kroode, A., D. Smit, and A. Verdel (1998). A microlocal analysis of migration. *Wave Motion 28*, 149–172.

Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics 51*, 1954–1966.

Thomson, C. J. (2001). Seismic coherent states and ray geometrical spreading. *Geophys. J. Int. 144*, 320–342.

Thomson, C. J. and C. H. Chapman (1985). An introduction to Maslov's asymptotic method. *Geophys. J. R. Astr. Soc. 83*, 143–168.

Um, J. and C. Thurber (1987). A fast algorithm for two-point ray tracing. *Bull. Seism. Soc. Am. 77*(3), 972–986.

van Trier, J. and W. W. Symes (1991). Upwind finite-difference calculation of traveltimes. *Geophysics 56*, 812–821.

Vanelle, C. and D. Gajewski (2002). Second-order interpolation of travel times. *Geophysical Prospecting 50*, 73–83.

Vidale, J. (1988). Finite-difference calculation of travel times. *Bull. Seism. Soc. Am. 78*(6), 2062–2076.

Vidale, J. E. (1990). Finite-difference calculation of traveltimes in three dimensions. *Geophysics 55*(5), 521–526.

Vinje, V., E. Iversen, K. Åstebøl, and H. Gjøystdal (1996a). Estimation of multivalued arrivals in 3D models using wavefront construction – Part I. *Geophysical Prospecting 44*, 819–842.

Vinje, V., E. Iversen, K. Åstebøl, and H. Gjøystdal (1996b). Part II: Tracing and interpolation. *Geophysical Prospecting 44*, 843–858.

Vinje, V., E. Iversen, and H. Gjøystdal (1993). Traveltime and amplitude estimation using wavefront construction. *Geophysics 58*, 1157–1166.

Xu, S., H. Chauris, G. Lambaré, and M. Noble (2001). Common angle migration: a strategy for imaging complex media. *Geophysics 66*, 1877–1894.

Samenvatting

Seismische experimenten behoren tot de meeste praktische en meest toegepaste methoden voor het verkrijgen van informatie over de ondergrond. Het meest gangbare experiment in de context van de olie- en gaswinning is het *seismische reflectie-experiment*. In dit experiment worden zowel de bronnen (bijvoorbeeld explosieven) als de ontvangers (seismische meetapparatuur) aan het aardoppervlak uitgestald. De trillingen die aan de bronnen worden opgewekt veroorzaken een seismische golfbeweging die zich voortplant in de ondergrond. De seismische golven worden gedeeltelijk weerkaatst (gereflecteerd) door structuren in de ondergrond, zoals overgangen tussen verschillende gesteentelagen en breukvlakken. De weerkaatste golven veroorzaken vervolgens weer trillingen aan het oppervlak die door de onvangers geregistreerd kunnen worden.

Het omzetten van de seismische meetgegevens in kwantitatieve informatie over de ondergrond is een complex probleem dat in wiskundige termen ook wel een *invers probleem* wordt genoemd. Het inverse probleem heeft als tegenhanger het *voorwaartse probleem*. In het geval van seismische experimenten behelst het voorwaartse probleem het berekenen van de seismische golfbeweging voor een gegeven (geschat) model van de ondergrond. Voor het oplossen van het seismische inverse probleem moet het voorwaartse probleem vaak veelvuldig doorgerekend worden. Een populaire theorie voor het uitvoeren van deze berekeningen is de *seismische stralentheorie*. Deze theorie heeft zijn populariteit mede te danken aan het feit dat de voorwaartse berekeningen zeer efficiënt uitgevoerd kunnen worden. Voor toepassing in met name complexe ondergrond modellen zijn er echter nog een aantal problemen in de stralentheorie die nader onderzoek verdienen.

Een karakteristieke eigenschap van straalmethoden is dat de berekeningen niet direct plaats vinden in de ruimtelijke coördinaten van de ondergrond. De straalvergelijkingen beschrijven de stralen – bij benadering de stroomlijnen van golfenergie – als functie van een enkele parameter, zoals bijvoorbeeld tijd. Voor het beschrijven van een bundel van stralen, oftewel een stralenveld, zijn er extra parameters nodig die de beginposities en -richtingen van de stralen bepalen. Samen vormen de parameter langs de straal en de parameters voor de begincondities een intern coördinatensysteem voor het stralenveld, aangeduid als de *stralenveldcoördinaten*.

De variaties in een seismisch golfveld worden in de stralentheorie beschreven door variaties in de looptijd en de amplitude. Beide grootheden worden in eerste

instantie berekend als een functie van de stralenveldcoördinaten. Om een evaluatie van het golfveld op een bepaalde ruimtelijke locatie te verkrijgen moet er een afbeelding van de ruimtelijke coördinaten naar de stralenveldcoördinaten gemaakt worden: de *stralenveldafbeelding*.

Als ondergrondmodellen zo complex worden dat er zich triplicaties in het golfveld gaan voordoen, worden de stralenveldafbeeldingen meervoudig. Deze meervoudigheid leidt tot vele praktische problemen in de toepassing van stralentheorie voor zowel de voorwaartse als inverse problemen.

In dit proefschrift worden een aantal nieuwe technieken voor de berekening van stralenvelden en stralenveldafbeeldingen gepresenteerd. Het centrale thema is het oplossen van de praktische problemen die zich voordoen in complexe ondergronden. Het uiteindelijke doel is de efficiëntie van seismische inverse methoden te verhogen door middel van verbeteringen in de voorwaartse berekeningen. Speciale aandacht wordt besteed aan de toepasbaarheid van de resultaten voor seismische inverse methoden.

De theoretische achtergrond van seismische stralentheorie wordt samengevat in **Hoofdstuk 2**. De nadruk ligt hierbij op de verklaring en interpretatie van verschillende stappen die genomen worden in de ontwikkeling van stralentheorie. Ook worden de concepten stralenveld en stralenveldafbeelding geïntroduceerd.

In **Hoofdstuk 3** wordt een nieuwe aanpak geïntroduceerd voor het berekenen en weergeven van stralenveldgegevens die in het bijzonder geschikt is wanneer er stralenvelden berekend moeten worden voor een groot aantal dicht bijeenstaande bronnen en/of ontvangers. In dat geval wordt voorgesteld om een enkele stralenveldafbeelding te maken in een uitgebreide ruimte van plaatscoördinaten en hoeken, het *plaats/hoek-domein*, in plaats van een afbeelding in de gebruikelijke plaatscoördinaten voor elk van de aanwezige bronnen/ontvangers.

Een voordeel van deze aanpak is dat een stralenveldafbeelding in het plaats/hoek-domein enkelvoudig is, ongeacht de complexiteit van het ondergrondmodel. Bovendien is de stralenveldinformatie geordend naar hoek op diepte, waardoor die bij uitstek bruikbaar is voor moderne seismische beeldvormingstechnieken. Een belangrijk resultaat is dat voor het verkrijgen van deze informatie het niet nodig is om stralen vanuit de diepte omhoog te schieten. In plaats daarvan is het mogelijk om bestaande algoritmen die naar beneden schieten aan te passen zodat deze toegepast kunnen worden in het plaats/hoek-domein. Dit levert een aanzienlijke winst in efficiëntie op.

Interpolatiemethoden spelen een belangrijke rol bij het berekenen en toepassen van stralenveldafbeeldingen. Een nieuwe techniek voor nauwkeurige interpolatie met behulp van gradiëntinformatie wordt gepresenteerd in **Hoofdstuk 4**. De techniek is een hybride van extrapolatie (met een willekeurige orde) en lineaire interpolatie. Door een aanpassing van de coëfficienten van de Taylorbenadering kunnen extrapolaties vanuit een aantal locaties worden gecombineerd, zodanig dat de orde van nauwkeurigheid één hoger wordt dan die van een enkele conventionele Taylorbenadering.

In **Hoofdstuk 5** wordt een stralenveldconstructie- en afbeeldingsalgoritme ont-

wikkeld dat de bestaande golffrontconstructiemethoden uitbreidt en verfijnt. Een modulaire opzet en een hiërarchische beschrijving van de geometrische structuur van het stralenveld maken het algoritme algemeen toepasbaar in 2-D en 3-D, van akoestische tot anisotrope elastische media.

Voor het berekenen van stralenveldafbeeldingen in het ruimtelijke domein worden twee verfijningen voorgesteld die de nauwkeurigheid en volledigheid van de afbeeldingen verhogen door middel van nauwkeuriger interpolatie, en nauwkeuriger afbakening van gebieden met verschillende meerwaardigheid.

Gemotiveerd door het succes van de golffrontconstructiemethoden in het ruimtelijke domein wordt ook de bruikbaarheid van het stralenveldconstructiealgoritme in het plaats/hoek-domein onderzocht. Helaas is de conclusie dat dit algoritme in zijn huidige vorm niet goed werkt in dat domein.

Een succesvol algoritme voor de berekening van stralenveldafbeeldingen in het plaats/hoek-domein wordt ontwikkeld in **Hoofdstuk 6**. Aangezien stralenvelden in het plaats/hoek-domein slechts weinig geometrische spreiding ondergaan en de bijbehorende stralendveldafbeeldingen uniform enkelvoudig zijn, blijkt het mogelijk gebruik te maken van klassieke paraxiale straalmethoden. In verhouding tot het schieten van stralen vanuit de diepte omhoog levert het algoritme een snelheidswinst op die proportioneel is met de gemiddelde straallengte gedeeld door de ruimtelijke roosterafstand.

In **Hoofdstuk 7** wordt een algemene beschouwing gegeven aangaande de resultaten van het onderzoek beschreven in dit proefschrift.

In Appendices A en B worden twee aanzetten gegeven voor toekomstig onderzoek. De één-op-één relatie tussen stralenveldcoordinaten en plaats/hoek-coordinaten kan gebruikt worden voor een verandering van variabelen in praktische toepassingen. Berekeningen die normaal gesproken in het stralenvelddomein plaatsvinden kunnen nu ook in het plaats/hoek-domein worden uitgevoerd. **Appendix A** laat zien dat dit voordelig kan zijn in tomografie. In **Appendix B** wordt getoond dat het op deze manier zelfs mogelijk is om in het plaats/hoek-domein de stralenveldafbeelding direct op een regelmatig rooster uit te rekenen, zonder tussenkomst van de stralenveldcoordinaten.

Tenslotte wordt in **Appendix C** een algoritme beschreven voor het berekenen van 2-D stralenvelden met behulp van een pseudospectrale methode. Het betreft een onderzoeksaanpak die niet is voortgezet, omdat het niet tot de gewenste efficiëntie leidde. Desalniettemin wordt het algoritme hier gepresenteerd vanwege de nuttige inzichten die de ontwikkeling met zich meebracht en de mogelijke bruikbaarheid van de concepten in andere toepassingen.

Dankwoord

Aan de totstandkoming van dit proefschrift hebben vele mensen een bijdrage geleverd. Op deze plek wil ik een aantal van hen graag expliciet bedanken. Ten eerste noem ik mijn beide copromotoren Roel Snieder en Kabir Roy-Chowdhury. Hun gezamenlijke onderzoeksvoorstel heeft de basis gelegd voor een zeer interessant onderzoek in een uitdagend vakgebied. Ik ben hen dankbaar voor het vertrouwen dat zij mij hebben gegeven: eerst, door mij aan te nemen op het onderzoeksproject, en later door mij vrijheid te geven in het kiezen van nieuwe onderzoeksrichtingen.

Kabir wil ik bedanken voor zijn collegiale en vriendschappelijke manier van werken. Zijn deur stond zo niet letterlijk dan toch figuurlijk altijd voor me open. Het feit dat Roel, mijn beoogde promotor, besloot een positie te aanvaarden aan de Colorado School of Mines, was natuurlijk niet in alle opzichten ideaal, maar het was wel de aanleiding voor een tweetal leuke en leerzame reisjes. Hoewel onze interactie noodzakelijkerwijs minder frequent werd, was zijn feedback altijd een bemoedigende stimulans.

In de eindfase van het onderzoek bleek het echter praktischer om dichter bij huis een promotor te vinden. Ik ben Rob van der Hilst enorm dankbaar dat hij deze taak op zich heeft willen nemen. Vanaf het begin was het duidelijk dat hij niet alleen in naam een promotor wilde zijn. Het afgelopen jaar is hij intensief betrokken geweest bij de voltooiing van dit proefschrift.

Bij dit STW-project was een zogeheten gebruikerscommissie betrokken, samengesteld uit Dr P.M. Bakker (Shell Technology EP), Prof. C.H. Chapman (Schlumberger Cambridge Research) en Dr P. van Riel (Jason Geosystems). I want to thank the members of the users' committee for their time and valuable input. The additional visits to Shell Technology EP and Schlumberger Cambridge Research have also been very informative.

The efforts of the members of the dissertation committee (Dr P.M. Bakker, Prof. Dr D. Gajewski, Dr G. Lambaré, Prof. Dr I. Pšenčík, and Prof. Dr Ir C.P.A. Wapenaar) are likewise appreciated. Their detailed comments have helped me to improve the thesis.

Mijn twee bezoeken aan Roel en zijn huidige collega's aan de Colorado School of Mines zullen mij bijblijven als leuke en nuttige ervaringen. I am very grateful to the people at the Center for Wave Phenomena for their hospitality, the use of their facilities and the knowledge they were willing to share with me.

For the work described in Chapter 5 I am very much indebted to Gilles Lambaré and Pascal Podvin from l'École des Mines de Paris. The 2-D and 3-D wave front construction codes that they made available provided an excellent source of solutions and inspiration, which saved me a lot of time in the development of the ray field construction algorithm.

Voor datzelfde Hoofdstuk 5 ben ik ook dank verschuldigd aan Stefan Carpentier, die in het kader van zijn afstudeeronderzoek de uitbreiding van het algoritme naar anisotrope media heeft onderzocht en geïmplementeerd.

Op diverse plaatsen heeft de tekst van dit proefschrift geprofiteerd van de kritische blik van Tim van Zon, Robbert van Vossen, Joost Beltman, Anton Kraaijpoel en Saskia Hendriksen. Alle overgebleven fouten zijn beslist alleen mij aan te rekenenen!

De onontbeerlijke praktische ondersteuning op diverse fronten was tijdens de duur van dit project in goede handen. Voor computerzaken kon ik altijd terecht bij Joop Hoofd, Theo van Zessen en Mark van Alphen, voor papierwerk en dergelijken bij Jacqueline Landsheer en Ildiko Csikos. Jan Jansen was altijd erg behulpzaam in de bibliotheek, en van Paul van Oudenallen zal ik nooit vergeten dat hij me kwam redden uit een zeer penibele situatie met een lekkend inktreservoir aan de vooravond van een SEG meeting.

Een flink aantal mensen hebben bijgedragen aan een goede werksfeer en een fijne tijd op het IvA. Naast de collega's die ik hierboven al genoemd heb wil ik speciaal Sandra Richwalski, Axel Röhm, Henk van der Meer, Tim van Zon, Edith Hafkenscheid, Rob Devilee, Stéphanie Godey, Jesper Spetzler, Caroline Beghein, Joe Resovsky, Hanneke Paulssen, Arie van Wettum, Jeannot Trampert, Karin Vis, Stefan Carpentier, Robbert van Vossen, Lucas Caljé, Walter de Groot, Frédéric Deschamps, Renate Hartog, Menno van der Zedde, Vessa Dimova, Antonio Villaseñor, Jaap Mondt, Ernst Jaarsma, Wim Bredewout, Jojanneke van den Berg – en diegenen die ik nog vergeten ben – bedanken voor vele uren van gezelligheid in diverse vormen. Tevens wil ik Dick Verweij noemen die er helaas niet meer is, maar ook in het rijtje thuishoort.

Maar voor het volbrengen van mijn promotietraject was vooral ook de belangstelling, de steun en het begrip van zowel (schoon)familie als vrienden van essentieel belang. Met name wil ik nog noemen mijn 'dikke' vrienden Joost Beltman en Sijbrand Booij, omdat zij hebben toegezegd tijdens de verdediging mijn paranimfen te willen zijn. Tenslotte gaat mijn grootste dank uit naar Saskia, voor haar geduld, en alle mentale ondersteuning en liefde die ze mij gedurende de hele periode gegeven heeft.

Curriculum Vitae

April 28, 1974	Born in Sassenheim, The Netherlands
1986 - 1992	Gymnasium B, Scholengemeenschap Augustinianum, Eindhoven, The Netherlands
1992 - 1997	Study of Geophysics at the Faculty of Earth Sciences, Utrecht University, The Netherlands
1997 - 2003	Ph.D student at the Faculty of Earth Sciences, Utrecht University, The Netherlands
2003 -	Post-doctoral researcher at NORSAR, Kjeller, Norway