2373-11

Workshop on Geophysical Data Analysis and Assimilation

29 October - 3 November, 2012

Full seismic waveform modelling and inversion

Andreas Fichtner
Department of Earth Sciences, Swiss Federal Institute of Technology (ETH), Zurich
Switzerland
Department of Earth Sciences, Utrecht University Utrecht
The Netherlands

Andreas Fichtner

Department of Earth Sciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland
Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands

Full seismic waveform modelling and inversion

Lecture given at **The Abdus Salam International Centre for Theoretical Physics**

30 and 31 October 2012

Contents

Part I Numerical solution of the elastic wave equation

1 **Introduction** . . . 3
  1.1 Notational conventions . . . 3
  1.2 The elastic wave equation . . . 4
    1.2.1 Governing equations . . . 4
    1.2.2 Formulations of the elastic wave equation . . . 5
  1.3 Discretisation in space . . . 6
  1.4 Discretisation in time or frequency . . . 7
    1.4.1 Time-domain modelling . . . 7
    1.4.2 Frequency-domain modelling . . . 9
  1.5 Summary of numerical methods . . . 9

2 **Finite-difference methods** . . . 13
  2.1 Basic concepts in one dimension . . . 14
    2.1.1 Finite-difference approximations . . . 14
    2.1.2 Discretisation of the one-dimensional wave equation . . . 18
    2.1.3 Von Neumann Analysis: stability and numerical dispersion . . . 20
  2.2 Extension to the three-dimensional cartesian case . . . 24
    2.2.1 The staggered grid . . . 24
  2.3 Accuracy and efficiency . . . 27

3 **Spectral-element methods** . . . 29
  3.1 Basic concepts in one dimension . . . 29
    3.1.1 Weak solution of the wave equation . . . 29
    3.1.2 Spatial discretisation and the Galerkin method . . . 30
  3.2 Extension to the three-dimensional case . . . 34
    3.2.1 Mesh generation . . . 34
    3.2.2 Weak solution of the elastic wave equation . . . 36
    3.2.3 Discretisation of the equations of motion . . . 37
  3.3 Accuracy and efficiency . . . 40

Part II The adjoint method and Fréchet kernels

4 **The time-domain continuous adjoint method** . . . 43
  4.1 Introduction . . . 43
  4.2 General formulation . . . 44
  4.3 Application to the elastic wave equation . . . 47

4.3.1 Derivation of the adjoint equations 47
4.4 First derivatives with respect to selected structural parameters 49
4.4.1 Perfectly elastic and isotropic medium 51
4.4.2 Perfectly elastic medium with radial anisotropy 52

5 Misfit functionals and adjoint sources 55
5.1 Derivative of the pure wave field and the adjoint Greens function 56
5.2 L_2 waveform difference 57
5.3 Cross-correlation time shifts 58

6 Fréchet kernel gallery 61
6.1 Body waves 61
6.2 Surface waves 65
6.2.1 Isotropic Earth models 66
6.2.2 Radial anisotropy 68

Mathematical background for the spectral-element method 71
A.1 Orthogonal polynomials 71
A.2 Function interpolation 72
A.2.1 Interpolating polynomial 72
A.2.2 Lagrange interpolation 73
A.2.3 Lobatto interpolation 74
A.2.4 Fekete points 77
A.2.5 Interpolation error 78
A.3 Numerical integration 79
A.3.1 Exact numerical integration and the Gauss quadrature 79
A.3.2 Gauss-Legendre-Lobatto quadrature 80
Cited literature and further reading material 82

Preface

Full waveform inversion

Full waveform inversion is a novel variant of seismic tomography that is characterised by the numerical solution of the equations of motion, the exploitation of full waveform information and the iterative improvement of the tomographic images that accounts for non-linearity in the relation between model parameters and synthetic data. The numerical solutions ensure the accurate modelling of seismic wave propagation through realistically heterogeneous Earth models, thus making full waveform inversion the tomographic method of choice when the medium properties are highly variable. Thanks to the combination of numerical solutions and adjoint techniques, any type of waveform can be exploited for the benefit of improved tomographic resolution - without the need to identify particular waveforms in terms of classical seismic phases such as P or S.

The purpose of these lecture notes is to present the necessary ingredients for a full waveform inversion applied to real data. It is intended to serve as an accessible introduction to the topic for advanced students. Throughout the text we assume basic knowledge of elastic wave propagation in a seismological context, as it can be found, for instance, in the works of Båth (1979), Bullen & Bolt (1985), Lay & Wallace (1995), Udías (1999), Kennett (2001) or Aki & Richards (2002).

A brief historical overview

The development of full waveform inversion techniques has always been considered a crucial step towards a more detailed understanding of subsurface properties. To see the motivation for and the high expectations in full waveform inversion, we briefly review the history of our seismologically derived knowledge of the Earth's internal structure.

This history probably starts in 1760 when J. Michell (1724-1793) first associated earthquakes with waves that travel through the Earth's crust with a speed of at least 20 miles per minute, that is, roughly 0.5 km/s (Michell, 1760). Michell's observation that waves propagate through the Earth could be explained with the theory of elasticity that was developed in the 18th and 19th centuries. A. L. Cauchy (1789-1857), S. D. Poisson (1781-1840), G. G. Stokes (1819-1903) and many others studied the elastic wave equation, which still forms the basis of modern-day seismological applications. P waves and S waves travelling with different speeds were identified as possible analytical solutions in homogeneous and unbounded media. Solutions for arbitrarily heterogeneous media have not become available until recently, thanks to advances in computational science and numerical mathematics. Between 1852 and 1858, R. Mallet (1810-1881) and his son J. W. Mallet (1832-1912) performed what is likely to be the first active-source experiment by measuring the propagation speed of seismic waves using gun powder explosions. They linked wave speed variations to variations of material properties, thus solving what we would classify today as a seismic inverse problem. In 1889, E. L. A. von Rebeur-Paschwitz (1861-1895) accidentally recorded the first teleseismic earthquake in Potsdam (von Rebeur-Paschwitz, 1889) using a horizontally swinging pendulum that was designed for astronomical measurements. The epicentre was near Tokyo. Teleseismically recorded elastic waves are today's principal source of information on deep Earth structure.

The theoretically predicted P and S waves were first clearly identified by R. D. Oldham (1858-1936) in 1900 (Oldham, 1900). Six years later he discovered the rapid decay of P wave amplitudes at epicentral distances greater than 100°. He correctly inferred the existence of the Earth's outer core (Oldham, 1906), the radius of which was determined by B. Gutenberg (1889-1960) in 1913 (Gutenberg, 1913). K. B. Zoeppritz (1881-1908) compiled traveltime tables for waves observed at teleseismic distances (Zoeppritz, 1907), and he translated them into one-dimensional models of the Earth's mantle. Much of his visionary work was published by his colleagues L. Geiger and B. Gutenberg after he died at the age of 26 (Zoeppritz & Geiger, 1909; Zoeppritz et al., 1912). In 1909 A. Mohorovičić (1857-1936) studied regional earthquakes, and he observed two types of P waves (today's Pn and Pg) and two types of S waves (today's Sn and Sg). He explained their traveltime curves with a discontinuity at 54 km depth − the crust-mantle discontinuity that now bears his name (Mohorovičić, 1910). H. Jeffreys (1891-1989) combined results from seismology and studies of Earth tides to conclude that "there seems to be no reason to deny that the earth's metallic core is truly fluid" (Jeffreys, 1926). Also in 1926, B. Gutenberg provided the first seismological evidence for a low-velocity zone around 100 km depth − the asthenosphere (Gutenberg, 1926). In 1936, I. Lehmann (1888-1993) observed unidentified P waves at large epicentral distances, today's PKIKP, and inferred the existence of the inner core (Lehmann, 1936). Another milestone in the discovery of the Earth's spherical structure was H. Jeffreys' and K. Bullen's (1906-1976) compilation of travel time tables that were used to infer complete radially symmetric Earth models (Jeffreys & Bullen, 1940).

The second half of the 20th century was marked by both the refinement of spherical Earth models (Dziewonski et al., 1975; Dziewonski & Anderson, 1981; Kennett et al., 1995) and the discovery of 3D heterogeneities through systematic tomographic approaches. Aki et al. (1977) were among the first to use teleseismic data in a linearised tomography for regional 3D structure. The analysis of nearly 700,000 P wave arrival time residuals allowed Dziewonski et al. (1977) to derive an early model of large-scale heterogeneities in the deep mantle. Tomographic methods were used also to determine 3D variations of seismic anisotropy (e.g. Montagner, 1985) and visco-elastic dissipation (e.g. Romanowicz, 1990). Increased data quality and data coverage contributed to the continuously improving resolution of tomographic images that could then be linked to mantle convection (e.g. van der Hilst et al., 1997; Ritsema & van Heijst, 2000), lithospheric deformation (e.g. Debayle & Kennett, 2000; Montagner, 2002), chemical heterogeneities (e.g. Jordan, 1978; van der Hilst & Kárason, 1999) and the tectonic evolution of continents (e.g. Zielhuis & Nolet, 1994; Zielhuis & van der Hilst, 1996).

Most seismological inferences concerning the structure of the Earth − including the existence of the inner core, the asthenosphere and the major seismic discontinuities − are based on the simplifying assumption that seismic waves can be represented by rays. This concept is closely related to geometrical optics. Within the ray theoretical framework, the arrival times of seismic waves are sensitive to seismic wave speeds only along a curve connecting source and receiver. The intensive use of ray theory (e.g. Červený, 2001) stems from its simplicity, its intuitive interpretation and from the difficulty of finding solutions of the complete elastic wave equation for realistically heterogeneous Earth models.

The limitations of ray theory in the context of seismic tomography have become a major concern during the past two decades (e.g. Williamson, 1991; Williamson & Worthington, 1993; Spetzler et al., 2001). Ray theory is valid when the length scales of 3D heterogeneities are small compared to the dominant wavelength. This condition imposes an upper limit on the resolution of tomographic images derived from ray theory. Efforts to overcome the limitations of ray theory − and thus to further improve the resolution of tomographic images − include finite-frequency tomography (e.g. Yomogida, 1992; Friederich, 1999, 2003; Dahlen et al., 2000; Yoshizawa & Kennett, 2004, 2005; Zhou et al., 2005; Sigloch et al., 2008) and full waveform inversion (e.g. Bamberger et al., 1982; Tarantola, 1988; Ikelle et al., 1988; Pica et al., 1990; Igel et al., 1996; Pratt, 1999; Djikpéssé & Tarantola, 1999; Dessa et al., 2004; Ravaut et al., 2004; Bleibinhaus et al., 2007; Choi et al., 2008; Fichtner et al., 2009, 2010; Tape et al., 2009, 2010). Finite-frequency tomography is a modification of the ray-theoretical seismic inverse problem that correctly accounts for the spatially extended 3D sensitivity distribution of waves with a finite frequency content. Full waveform inversion goes one step further than finite-frequency tomography by replacing semi-analytical solutions to the wave equation (e.g. ray theoretical solutions or mode summations) by fully numerical solutions. This ensures that the propagation of seismic waves through heterogeneous media is modelled accurately. Furthermore, numerical solutions provide complete seismic waveforms that can be exploited for the benefit of improved tomographic resolution, without the need to identify specific seismic phases. For an extensive review of full waveform inversion, the reader is referred to Virieux & Operto (2009).

Part I
Numerical solution of the elastic wave equation

The numerical solution of the equations of motion is one of the defining characteristics of full waveform inversion and an indispensable tool in seismic ground motion studies. Numerical solutions allow us to accurately model the propagation of seismic waves through strongly heterogeneous media, including, for instance, the Earth's lithosphere on a global scale, sedimentary basins and volcanic edifices.

Seismic wave motion in the Earth is governed by the wave equation, that relates the displacement field to external forces and to the distributions of density and elastic parameters. Since there are no exact analytical solutions to the wave equation in realistically heterogeneous media, we construct approximate solutions by discretising derivatives. The spatial discretisation leads to a system of ordinary differential equations in time that can be solved numerically either in the frequency domain or by iterative time stepping (chapter 1).

Several methods have been developed for the numerical solution of the wave equation, each being particularly well suited for specific types of applications. In the context of full waveform inversion, the finite-difference and the spectral-element methods are most frequently used.

In the finite-difference method (chapter 2), the spatial derivatives are approximated by difference quotients. The approximation leads, as in any other numerical method, to a dispersion error that depends, among other factors, on the number of grid points per wavelength. The feasibility of finite-difference modelling in 3D rests on the definition of a staggered grid where the field variables are evaluated at different grid positions. While increasing the computational efficiency, the staggered grid requires special care in the implementation of anisotropy and the free surface.

The spectral-element method (chapter 3) is based on the weak or variational formulation of the equations of motion. The computational domain is divided into disjoint elements that are mapped onto a reference element. Within the reference element, the dynamic fields are expressed in terms of Lagrange polynomials, collocated at the Gauss-Lobatto-Legendre points. The integrals of the weak form are then approximated by Gauss-Lobatto-Legendre quadrature, which leads to an algebro-differential equation with a diagonal mass matrix.

Chapter 1
Introduction

In this chapter we present the basic elements for the numerical modelling of seismic wave propagation. Following a summary of notational conventions, we introduce the elastic wave equation in its different formulations (section 1.2). While numerical methods differ in the details of the spatio-temporal discretisation, they can still be treated within a unifying framework: The approximation of the spatial derivatives generally leads to a system of ordinary differential equations in time that is commonly referred to as the semi-discrete form of the wave equation. The semi-discrete form can be written in terms of mass and stiffness matrices (section 1.3). Depending on the specifics of an application, the remaining time derivatives can then be approximated using either the Fourier transform or time stepping algorithms such as the Newmark or leapfrog methods (section 1.4).

1.1 Notational conventions

Throughout these notes we have tried to stay close to notations commonly found in the seismological literature. While the meaning of the different symbols is mostly clear from the context, we start with a small collection of conventions that we shall use consistently.
The Fourier transform of a function f is defined as

$$\tilde{f}(\omega) := F[f](\omega) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{-\mathbf{i}\omega t}\, dt\,, \quad \mathbf{i} := \sqrt{-1}\,, \tag{1.1}$$

where the symbol := means that the expression to the left is defined by the expression to the right. The bold-face **i** is intended to distinguish the imaginary unit from the frequently occurring index variable *i*. The inverse Fourier transform corresponding to (1.1) is

$$f(t) = F^{-1}[\tilde{f}](t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \tilde{f}(\omega) e^{\mathbf{i}\omega t}\, d\omega\,. \tag{1.2}$$

With the exception of **i**, we use bold-faced symbols for vectors and tensors. The scalar product of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is denoted by

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i\,. \tag{1.3}$$

We more generally use the dot to signify contraction over adjacent indices. Using this notation, the *i*-component of a matrix-vector product is written as

$$(\mathbf{A} \cdot \mathbf{a})_i = \sum_{i=1}^{n} A_{ij} a_j\,, \quad \mathbf{A} \in \mathbb{R}^{n \times n}\,, \tag{1.4}$$

and the ij-component of a matrix-matrix product is

$$(\mathbf{A}\cdot\mathbf{B})_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}\,. \tag{1.5}$$

Following this scheme, a double dot denotes a contraction over two adjacent indices, for instance

$$\mathbf{A}:\mathbf{B} = \sum_{i,j=1}^{n} A_{ij}B_{ij} \tag{1.6}$$

and

$$(\mathbf{B}:\mathbf{C})_{kl} = \sum_{i,j=1}^{n} B_{ij}C_{ijkl}\,, \quad \mathbf{C}\in\mathbb{R}^{n\times n\times n\times n}\,. \tag{1.7}$$

For the real part of a complex-valued variable $z = x + \mathbf{i}y$ we use the fraktur symbol $\mathfrak{Re}$, i.e.

$$\mathfrak{Re}\,z = x = \frac{1}{2}(z+z^*)\,, \tag{1.8}$$

where $z^* = x - \mathbf{i}\,y$ denotes the complex conjugate of z. The symbol $\mathfrak{Im}$ denotes the imaginary part of z:

$$\mathfrak{Im}\,z = y = \frac{1}{2}(z-z^*)\,. \tag{1.9}$$

Of outstanding importance in any deterministic inverse problem is the definition of a misfit functional, χ, that quantifies the difference between observed and synthetic data. The misfit functional depends on an Earth model $\mathbf{m}\in\mathfrak{M}$, where $\mathfrak{M}$ is the model space. The functional or Fréchet derivative of χ with respect to $\mathbf{m}$ in a direction $\delta\mathbf{m}\in\mathfrak{M}$ is defined by

$$\nabla_m\chi(\mathbf{m})\,\delta\mathbf{m} := \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}[\chi(\mathbf{m}+\varepsilon\,\delta\mathbf{m}) - \chi(\mathbf{m})]\,. \tag{1.10}$$

The derivative $\nabla_m\chi(\mathbf{m})$ is a linear operator acting on the differentiation direction $\delta\mathbf{m}$. In the special case where $\mathbf{m}$ is a vector, the Fréchet derivative $\nabla_m\chi(\mathbf{m})\,\delta\mathbf{m}$ coincides with the directional derivative $\delta\mathbf{m}\cdot\nabla_m\chi(\mathbf{m})$. For convenience, we will mostly use the term 'derivative' instead of 'Fréchet derivative'. The symbol ∇, without subscript, signifies the regular gradient with respect to the position vector $\mathbf{x}$.

1.2 The elastic wave equation

1.2.1 Governing equations

Full waveform inversion is founded on the solution of the *forward problem*, which consists in the simulation of seismic wave propagation through an Earth model, $\mathbf{m}$, and the computation of synthetic seismograms. The propagation of seismic waves in the Earth can be modelled with the *elastic wave equation*

$$\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla\cdot\sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)\,, \quad \mathbf{x}\in G\subset\mathbb{R}^3\,, \quad t\in[t_0,t_1]\subset\mathbb{R} \tag{1.11}$$

that relates the displacement field $\mathbf{u}$ in the Earth $G\subset\mathbb{R}^3$ to its mass density ρ, the stress tensor σ and an external force density $\mathbf{f}$. A truly marvellous matter of fact! Equation (1.11) is the linearised version of Newton's second law that balances the momentum of the particle displacement, $\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t)$, forces resulting from internal stresses, $\nabla\cdot\sigma(\mathbf{x},t)$, and external forces, $\mathbf{f}(\mathbf{x},t)$, that represent the sources of seismic wave motion. For detailed derivations of equation (1.11) the reader is referred to Dahlen & Tromp (1998), Kennett (2001) or Aki & Richards (2002). At the surface, ∂G, of the Earth, the normal components of the stress tensor σ vanish, i.e.

$$\sigma\cdot\mathbf{n}|_{\mathbf{x}\in\partial G} = \mathbf{0}\,, \tag{1.12}$$

where $\mathbf{n}$ is the unit normal on ∂G. Equation (1.12) is the *free surface boundary condition*. Both the displacement field $\mathbf{u}$ and the velocity field $\mathbf{v} = \dot{\mathbf{u}}$ are required to satisfy the *initial condition* of being equal to zero prior to $t = t_0$ when the external force $\mathbf{f}$ starts to act:

$$\mathbf{u}|_{t \leq t_0} = \mathbf{v}|_{t \leq t_0} = \mathbf{0}\,. \tag{1.13}$$

For convenience we will mostly choose $t_0 = 0$. To obtain a complete set of equations, the stress tensor, σ, must be related to the displacement field, $\mathbf{u}$. For this we assume that σ depends linearly on the history of the strain tensor $\varepsilon := \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T)$:

$$\sigma(\mathbf{x},t) = \int_{-\infty}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-t') : \varepsilon(\mathbf{x},t')\,dt'\,. \tag{1.14}$$

Equation (1.14) defines a *linear visco-elastic rheology*. The fourth-order tensor $\mathbf{C}$ is the *elastic tensor*. Since the current stress can not depend on future strain, we require the elastic tensor to be causal:

$$\mathbf{C}(t)|_{t<t_0} = \mathbf{0}\,. \tag{1.15}$$

The symmetry of ε, the conservation of angular momentum and the relation of $\mathbf{C}$ to the internal energy (e.g. Aki & Richards, 2002) require that the components of $\mathbf{C}$ satisfy the following symmetry relations:

$$C_{ijkl} = C_{klij} = C_{jikl}\,. \tag{1.16}$$

The symmetries of the elastic tensor reduce the number of its independent components to 21, and they allow us to write (1.14) directly in terms of the displacement gradient, $\nabla\mathbf{u}$:

$$\sigma(\mathbf{x},t) = \int_{-\infty}^{\infty} \dot{\mathbf{C}}(\mathbf{x},t-t') : \nabla\mathbf{u}(\mathbf{x},t')\,dt'\,. \tag{1.17}$$

The number of non-zero independent elastic tensor components - also referred to as elastic parameters or elastic moduli - determines the *anisotropic properties* of the medium. For instance, a triclinic crystal such as plagioclase requires all 21 independent parameters for its complete description, and 3 elastic parameters are needed to describe crystals with cubic symmetry such as garnet (Babuska & Cara, 1991). On a macroscopic scale the Earth can often be described sufficiently well as an isotropic body where the elastic tensor components are linear combinations of only two elastic moduli: the *Lamé parameters* λ and μ:

$$C_{ijkl} = \lambda\,\delta_{ij}\delta_{kl} + \mu\,\delta_{ik}\delta_{jl} + \mu\,\delta_{il}\delta_{jk}\,. \tag{1.18}$$

The parameter μ relates strain to shear stresses, and is therefore called *shear modulus*. Since λ has no intuitive physical meaning, it is commonly replaced by the *bulk modulus*, $\kappa = \lambda + \frac{2}{3}$, that relates strain to the scalar pressure, defined as $p =: -\kappa\nabla\cdot\mathbf{u}$.

The time dependence of the elastic tensor is responsible for visco-elastic dissipation, that is the process of transforming elastic energy into heat. In the case of a non-dissipative medium, the time dependence of $\mathbf{C}$ takes the form of a unit step or Heaviside function, $H(t)$:

$$\mathbf{C}(\mathbf{x},t) = \mathbf{C}(\mathbf{x})\,H(t)\,. \tag{1.19}$$

The constitutive relation (1.17) then takes the form

$$\sigma(\mathbf{x},t) = \mathbf{C}(\mathbf{x}) : \nabla\mathbf{u}(\mathbf{x},t)\,. \tag{1.20}$$

For simplicity we will assume a non-dissipative medium throughout most of these notes.

1.2.2 Formulations of the elastic wave equation

Equations (1.11) and (1.20) constitute the *displacement-stress formulation* of the elastic wave equation in the absence of dissipation. Together with the initial and boundary conditions they uniquely specify the displacement field, $\mathbf{u}(\mathbf{x},t)$.

Different but fully equivalent formulations are possible and sometimes required by a specific numerical method. We may, for instance, eliminate the stress tensor, σ, by combining equations (1.11) and (1.20). This results in the *displacement formulation* of the elastic wave equation:

$$\rho(\mathbf{x})\,\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla\cdot[\mathbf{C}(\mathbf{x}):\nabla\mathbf{u}(\mathbf{x},t)] = \mathbf{f}(\mathbf{x},t)\,. \tag{1.21}$$

Of particular relevance in numerical modelling is the *velocity-stress formulation* where the wave equation is written as a first-order system in both time and space. We find the velocity-stress formulation by simply differentiating equation (1.20) with respect to time and then substituting $\mathbf{v}$ for $\dot{\mathbf{u}}$:

$$\rho(\mathbf{x})\,\dot{\mathbf{v}}(\mathbf{x},t) - \nabla\cdot\sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)\,, \tag{1.22a}$$

$$\dot{\sigma}(\mathbf{x},t) - \mathbf{C}(\mathbf{x}):\nabla\mathbf{v}(\mathbf{x},t) = \mathbf{0}\,. \tag{1.22b}$$

The elastic wave equation in its different formulations has been studied extensively. Analytical solutions exist for numerous classes of models including, for instance, the isotropic and homogeneous half space (e.g. Lamb, 1904; de Hoop, 1958; Aki & Richards, 2002), stratified media (e.g. Kennett, 1979, 1980, 1981) and spherically symmetric globes with radial anisotropy (e.g. Takeuchi & Saito, 1978; Woodhouse, 1988; Friederich & Dalkolmo, 1995). Ray theory (e.g. Červený, 2001) and perturbation methods (e.g. Woodhouse & Dahlen, 1978; Maupin, 2001) can be used to approximate solutions to the elastic wave equation for mildly heterogeneous media.

In full waveform inversion, the focus is on strongly heterogeneous regions of the Earth where analytical or perturbation methods are not applicable. This is the domain of numerical methods that are the subject of the first part of these notes.

1.3 Discretisation in space

Analytic solutions to the elastic wave equation exist only for comparatively simple models that often do not reflect the structural complexities of the Earth. This deficiency motivates the development of numerical methods for the simulation of seismic wave propagation through almost arbitrarily heterogeneous Earth models.

While being different in the technical details, all numerical methods have one point in common: the discrete spatial approximation of the continuously defined wave field, $\mathbf{u}(\mathbf{x},t)$. This means that $\mathbf{u}(\mathbf{x},t)$ is approximated by a finite number of time-dependent coefficients, $\bar{u}_1(t),...,\bar{u}_N(t)$, that can be summarised in an N-dimensional vector, $\bar{\mathbf{u}}(t)$. Depending on the specifics of the numerical method used, the coefficients, $\bar{u}_i(t)$ $(i=1,...,N)$, may represent, for instance, discrete values of $\mathbf{u}(\mathbf{x},t)$ sampled at a finite number of points, or polynomial coefficients when $\mathbf{u}(\mathbf{x},t)$ is approximated by a polynomial.

Following spatial discretisation, the displacement formulation of the elastic wave equation, given in (1.21), turns into an algebro-differential equation that can always be written in the following canonical form:

$$\mathbf{M}\cdot\ddot{\bar{\mathbf{u}}}(t) + \mathbf{K}\cdot\bar{\mathbf{u}}(t) = \bar{\mathbf{f}}(t)\,. \tag{1.23}$$

The matrices $\mathbf{M}$ and $\mathbf{K}$ are referred to as the *mass matrix* and the *stiffness matrix*, respectively. The vector $\bar{\mathbf{f}}$ represents a discrete version of the force density, $\mathbf{f}$. Both $\mathbf{M}$ and $\mathbf{K}$ tend to be sparse. In practice, the mass and the stiffness matrices are rarely computed explicitly because only the vector-matrix products are needed in actual computations. They are, nevertheless, useful tools in theoretical developments, as we will soon discover.

Depending on the numerical method, it may be more advantageous to discretise the displacement-stress formulation (equations 1.11 and 1.20) or the velocity-stress formulation (equation 1.22) of the elastic wave equation. The corresponding space-discrete systems are then

$$\mathbf{M}\cdot\ddot{\bar{\mathbf{u}}}(t) + \mathbf{K}_1\cdot\bar{\mathbf{s}}(t) = \bar{\mathbf{f}}(t)\,, \tag{1.24a}$$

$$\bar{\mathbf{s}}(t) = \mathbf{K}_2\cdot\bar{\mathbf{u}}(t)\,, \tag{1.24b}$$

for the displacement-stress formulation and

$$\mathbf{M}\cdot\dot{\bar{\mathbf{v}}}(t)+\mathbf{K}_1\cdot\bar{\mathbf{s}}(t)=\bar{\mathbf{f}}(t)\,, \tag{1.25a}$$

$$\dot{\bar{\mathbf{s}}}(t)-\mathbf{K}_2\cdot\bar{\mathbf{v}}(t)=\mathbf{0}\,, \tag{1.25b}$$

for the velocity-stress formulation. The matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are stiffness matrices in a broader sense and $\bar{\mathbf{s}}$ represents a discrete approximation to the stress tensor, σ.

The transition from the continuous wave equation in its various formulations to one of the space-discrete systems (1.23), (1.24) or (1.25), reduces the forward problem to the solution of a large algebraic system and an ordinary differential equation in time.

1.4 Discretisation in time or frequency

The ordinary differential equations in time that arise from the spatial discretisation of the equations of motion can be solved either in the time domain or in the frequency domain. Both approaches have advantages and disadvantages that must be weighted depending on the particular application and the available computational resources.

1.4.1 Time-domain modelling

The time-domain modelling of wave propagation is mostly based on the replacement of the time derivatives in the space-discrete equations of motion by suitable finite-difference approximations. These allow us to advance the wave field in discrete time steps, Δt. The choice of a particular finite-difference scheme depends on the formulation of the wave equation.

Throughout the following paragraphs we assume that the mass matrix, $\mathbf{M}$, can be inverted; noting, however, that the inversion of $\mathbf{M}$ may present a formidable numerical challenge.

1.4.1.1 Displacement and displacement-stress formulation

The space-discrete version of the displacement formulation (1.23) involves the second time derivative of the discrete displacement field, $\ddot{\bar{\mathbf{u}}}$, which is explicitly given by

$$\ddot{\bar{\mathbf{u}}}(t)=\mathbf{M}^{-1}\cdot[\bar{\mathbf{f}}(t)-\mathbf{K}\cdot\bar{\mathbf{u}}(t)]\,. \tag{1.26}$$

Approximating $\ddot{\bar{\mathbf{u}}}(t)$ by the second-order finite-difference

$$\ddot{\bar{\mathbf{u}}}(t)\approx\frac{1}{\Delta t^2}\,[\bar{\mathbf{u}}(t+\Delta t)-2\bar{\mathbf{u}}(t)+\bar{\mathbf{u}}(t-\Delta t)]\,, \tag{1.27}$$

leads to an explicit time-stepping scheme that allows us to compute the displacement at time $t+\Delta t$ from the displacement at times t and $t-\Delta$:

$$\bar{\mathbf{u}}(t+\Delta t)=2\bar{\mathbf{u}}(t)-\bar{\mathbf{u}}(t-\Delta t)+\Delta t^2\,\mathbf{M}^{-1}\cdot[\bar{\mathbf{f}}(t)-\mathbf{K}\cdot\bar{\mathbf{u}}(t)]\,. \tag{1.28}$$

For notational clarity we replaced $\approx$ by $=$ in equation (1.28), keeping in mind that this is an approximation.

A frequently used alternative to (1.28) is the *Newmark scheme* (Newmark, 1959; Chaljub et al., 2007), defined by

$$\bar{\mathbf{u}}(t+\Delta t)=\bar{\mathbf{u}}(t)+\Delta t\,\bar{\mathbf{v}}(t)+\Delta t^2\left[\left(\frac{1}{2}-\beta\right)\ddot{\bar{\mathbf{u}}}(t)+\beta\,\ddot{\bar{\mathbf{u}}}(t)\right]\,, \tag{1.29a}$$

$$\bar{\mathbf{v}}(t+\Delta t)=\bar{\mathbf{v}}(t)+\Delta t\,[(1-\gamma)\,\ddot{\bar{\mathbf{u}}}(t)+\gamma\ddot{\bar{\mathbf{u}}}(t+\Delta t)]\,, \tag{1.29b}$$

with the parameters $\gamma \in [0,1]$ and $\beta \in [0,\frac{1}{2}]$. Second-order accuracy instead of first-order accuracy is achieved if and only if $\gamma = \frac{1}{2}$. In the special case $\gamma = \frac{1}{2}$ and $\beta = 0$, we obtain an explicit central-difference scheme:

$$\bar{\mathbf{u}}(t+\Delta t) = \bar{\mathbf{u}}(t) + \Delta t\, \bar{\mathbf{v}}(t) + \frac{1}{2}\Delta t^2\, \ddot{\bar{\mathbf{u}}}(t)\,, \tag{1.30a}$$

$$\bar{\mathbf{v}}(t+\Delta t) = \bar{\mathbf{v}}(t) + \frac{1}{2}\Delta t\, [\ddot{\bar{\mathbf{u}}}(t) + \ddot{\bar{\mathbf{u}}}(t+\Delta t)]\,. \tag{1.30b}$$

The order of operations in the iterative advancement of $\bar{\mathbf{u}}$ is then as follows: 1.) Compute $\bar{\mathbf{u}}$ at time $t+\Delta t$ from the discrete field variables at time t, using equation (1.30a). 2.) With the help of equation (1.26), compute the acceleration $\ddot{\bar{\mathbf{u}}}(t+\Delta t)$ from $\bar{\mathbf{u}}(t+\Delta t)$. 3.) Advance the velocity, $\bar{\mathbf{v}}$, from time t to $t+\Delta t$, using equation (1.30b). 4.) Go back to 1.) and repeat as often as needed. While being more complicated than (1.28), the Newmark scheme has the advantageous property of conserving linear and angular momentum (e.g. Kane et al., 2003).

We note that the time-discretisation schemes for the displacement formulation are immediately applicable to the displacement-stress formulation from equation (1.24).

1.4.1.2 Velocity-stress formulation

The most commonly used time-discretisation scheme for the velocity-stress formulation (1.25) is the *leapfrog method*. This is based on alternating updates of the discrete velocity and stress fields: Starting from $\bar{\mathbf{s}}(t-\Delta t/2)$ and $\bar{\mathbf{v}}(t)$, we obtain $\bar{\mathbf{s}}(t+\Delta t/2)$ via a second-order finite-difference approximation of equation (1.25b):

$$\bar{\mathbf{s}}(t+\Delta t/2) = \bar{\mathbf{s}}(t-\Delta t/2) + \Delta t\, \mathbf{K}_2 \cdot \bar{\mathbf{v}}(t)\,. \tag{1.31a}$$

With the help of the discrete stress field $\bar{\mathbf{s}}(t+\Delta t/2)$ we can then advance the velocity field from time t to time $t+\Delta t$, using the same second-order approximation applied to equation (1.25a):

$$\bar{\mathbf{v}}(t+\Delta t) = \bar{\mathbf{v}}(t) + \Delta t\, \mathbf{M}^{-1} \cdot [\bar{\mathbf{f}}(t+\Delta t/2) - \mathbf{K}_1 \cdot \bar{\mathbf{s}}(t+\Delta t/2)]\,. \tag{1.31b}$$

Again, for notational convenience, we replaced $\approx$ by $=$ in equations (1.31a) and (1.31b).

1.4.1.3 Stability

All of the above time-stepping algorithms are explicit in the sense that the dynamic fields at time $t+\Delta t$ only depend on the dynamic fields at times *prior* to $t+\Delta t$. This implies that the algorithms are only conditionally stable. The stability criterion, named *CFL condition* in honour of R. Courant, K. Friedrichs and H. Lewy (Courant et al., 1928), typically takes the form

$$\Delta t \leq \text{const.}\, \frac{\min h}{\max v}\,, \tag{1.32}$$

where h is the width of the numerical grid and v is the propagation speed of the fastest wave, that is, the P wave. The constant on the right-hand side of equation (1.32) depends on the methods used for the space and time discretisation. Its order of magnitude is 1. The CFL condition limits the maximum possible time increment and therefore the efficiency of any explicit time stepping. In chapter 2.1 we derive the CFL condition for a finite-difference approximation of the one-dimensional scalar wave equation.

1.4.1.4 Alternative methods

The Newmark and the leapfrog schemes are by far the most frequently used methods for the time discretisation of the space-discrete equations of motion. This may appear surprising given the availability of numerous alternatives such as predictor-corrector variants of implicit multi-step methods or higher-order Runge-Kutta methods (e.g. Quarteroni et al., 2000).

The restriction to conditionally stable methods of comparatively low order can be explained by the larger computational requirements of implicit and higher-order methods, but also by the good performance of the explicit second-order schemes. Empirical studies show that the numerical error is often dominated by the inaccuracies of the *spatial* discretisation. This is true even when the number of time steps is large, that is, on the order of several thousand. Whether the conventional Newmark and leapfrog schemes are sufficient for future applications, such as global-scale wave propagation at periods around 2 s, is still an open question.

1.4.2 Frequency-domain modelling

Frequency-domain modelling is based on the Fourier transformed version of the space-discrete displacement formulation (1.23):

$$-\omega^2 \mathbf{M} \cdot \bar{\mathbf{u}}(\omega) + \mathbf{K} \cdot \bar{\mathbf{u}}(\omega) = \bar{\mathbf{f}}(\omega) \,. \tag{1.33}$$

Upon defining the *impedance matrix*, **L**,

$$\mathbf{L}(\omega) := -\omega^2 \mathbf{M} + \mathbf{K} \,, \tag{1.34}$$

equation (1.33) simplifies to the linear system

$$\mathbf{L}(\omega) \cdot \bar{\mathbf{u}}(\omega) = \bar{\mathbf{f}}(\omega) \,. \tag{1.35}$$

The solution of (1.35) is the discrete wave field, $\bar{\mathbf{u}}$, at one frequency, ω.

It is particularly attractive to solve the system (1.35) using direct matrix factorisation methods, such as LU decomposition (e.g. Press et al., 2007). Once the matrix factors are known, they can be reused to solve the forward problem for any new source, $\bar{\mathbf{f}}$, at very low computational cost (e.g. Pratt et al., 1998; Pratt, 1999). However, in realistic 3D applications, the memory requirements of direct methods can become prohibitive, so that iterative solvers must be used (e.g. Quarteroni et al., 2000).

Frequency-domain modelling is the method of choice when solutions for a few dominant or well-chosen frequencies are needed, or when the problem is two-dimensional thus permitting the application of direct linear system solvers.

One of the outstanding advantages of this approach is the easy implementation of visco-elastic dissipation. This is because the convolution in equation (1.14) translates to a simple product in the frequency domain.

1.5 Summary of numerical methods

In the course of the past decades a large number of numerical methods for the solution of the seismic wave equation have been developed - often for the purposes of ground motion prediction or waveform inversion. Each method comes with advantages and disadvantages that need to be weighted carefully in the light of a specific application.

The most significant distinction between different approaches concerns the spatial discretisation, that is the transformation of the exact spatial derivatives in the wave equation into an algebraic system. The following is a list of some methods that tries to highlight their major similarities and differences.

Finite-difference methods: Numerical modelling in nearly all physical sciences started with the finite-difference method, and seismology is no exception. Early applications can be found in Alterman & Karal (1968), Boore (1970, 1972), Alford et al. (1974) and Kelly et al. (1976). To illustrate the finite-difference concept, we consider a generic function, $f(x)$, that represents the dynamic fields that appear in the wave equation (e.g. stress, strain, displacement, ...).

The fundamental idea is to consider only a finite number of evenly spaced grid points, x_i (i=1, ..., N), and to replace the derivative $\partial_x f(x_i)$ at grid point x_i by a *finite-difference approximation* that involves f evaluated at neighbouring grid points. The best-known example is the second-order central finite-difference approximation

$$\partial_x f(x_i) = \frac{1}{2\Delta x}\left[f(x_i+\Delta x) - f(x_i-\Delta x)\right] + \mathscr{O}(\Delta x^2)\,, \tag{1.36}$$

where Δx is the grid spacing. The feasibility of finite-difference modelling in three dimensions rests on the definition of a *staggered grid* where the dynamic fields are defined at different grid positions (e.g. Madariaga, 1976; Virieux, 1984, 1986, Igel et al., 1995). This is in contrast to the *conventional grid* where all field variables are defined at coincident grid positions. The staggered grid results in a reduced average grid spacing that greatly reduces the *numerical dispersion*, i.e., the artificial dispersion introduced by the discretisation of the original equations of motion.
The popularity of finite-difference modelling is largely due to the comparatively low computational costs and the accuracy especially in the modelling of body wave propagation. In chapter 2 we treat the finite-difference method in detail.

Optimal operators: The discretisation of the equations of motion introduces numerical errors that are particularly prominent near the eigenfrequencies of the elastic medium. This observation led Geller & Takeuchi (1995) to the construction of optimal operators that are designed to minimise the discretisation error first of all in the vicinity of the eigenfrequencies.
The very general criterion for an operator to be optimal is that the inner product of an exact eigenfunction with the net error of the discrete equations of motion at the corresponding eigenfrequency is approximately 0. This criterion is independent of the actual space discretisation scheme. In the particular framework of finite-difference methods, optimality means that the lowest-order errors of the time and the space discretisations cancel, thus leading to highly accurate numerical schemes (Geller & Takeuchi, 1998; Takeuchi & Geller, 2000).
Since time domain optimal operators are inherently implicit, predictor-corrector algorithms must be used to avoid the solution of large algebraic systems in the time stepping. The increased solution accuracy clearly compensates the additional computational costs of the predictor-corrector scheme (Mizutani et al., 2000; Kristek & Moczo, 2006).

Pseudospectral methods: Like finite-difference methods, pseudospectral methods directly discretise the spatial derivatives in the equations of motion (e.g. Kosloff & Baysal, 1982; Furumura et al., 1998). The discretisation proceeds in three steps: First, the wave field sampled at a finite number of grid points is transformed to the wave number domain using the Fast Fourier Transform. The transformed wave field is then multiplied by $\mathbf{i}k$, where k is the wave number. This multiplication corresponds to a space derivative. Finally, using the inverse Fast Fourier Transform, the wave field is transformed back to the space domain. Since the derivative is exact up to the Nyquist wave number, as few as 2 grid points per wavelength are theoretically sufficient for the spatial sampling of the wave field.
Kosloff et al. (1990) proposed a variant of the Fourier pseudospectral method where a Chebyshev transform is used in the vertical direction in order to account more accurately for the free surface condition. An extension to the three-dimensional case with surface topography can be found in Tessmer & Kosloff (1994). The application of the Chebyshev pseudospectral method to wave propagation on the scale of the mantle is presented in Igel (1999).
Pseudospectral methods outperform finite-difference methods regarding the very small amount of numerical dispersion. However, due to the global nature of the derivative approximation, they are restricted to comparatively smooth media (Mizutani et al., 2000). Furthermore, pseudospectral methods face issues of parallelisation because global memory access is required for the computation of the Fourier and Chebyshev transforms.

Finite-element methods: Finite-element methods are based on the *weak* or *variational form* of the wave equation that we describe in section 3.1.1. The computational domain is decomposed into disjoint subdomains, called the elements. Within each element the dynamical fields are approximated by polynomials of low order (e.g. piecewise linear functions, Bao et al., 1998), and continuity between the elements is imposed explicitly. The elastic wave equation then reduces to a space-discrete system for the polynomial coefficients.
Despite its capability to correctly account for irregular geometries and the free surface, applications of the pure finite-element method to elastic wave propagation are comparatively rare (e.g. Lysmer & Drake, 1972; Toshinawa & Ohmachi, 1992; Bao et al., 1998). This is mostly due to the comparatively large numerical dispersion that results from the low-order polynomial approximations. Moreover, the mass matrix in finite-element methods is not diagonal,

which renders its inversion computationally expensive.

While the pure finite-element method does not appear to be well-suited for wave propagation, hybrid schemes have been used very successfully. Moczo et al. (1997, 2007), for instance, combined the finite-element and finite-difference methods for the simulation of wave propagation along irregular surface topography. They discretised the equations of motion in the interior of the computational domain using finite-differences. A rim of finite-elements was then used to mesh the topography.

Spectral-element methods: Spectral-element methods are half-way between finite-element and pseudospectral methods, combining the advantages of both approaches while avoiding many of their drawbacks. Like in finite-element methods, the computational domain is subdivided into non-overlapping elements that can be adapted to irregular geometries. Inside the elements a high-order spectral approximation is used for the dynamic fields. The spectral-element method as originally developed for fluid mechanics (Patera, 1984) and seismic wave modelling (Priolo et al., 1994; Seriani et al., 1995; Seriani, 1998) indeed uses Chebyshev polynomials as basis functions, thus establishing a direct link to the Chebyshev pseudospectral methods.

In a widely used spectral-element variant the Chebyshev polynomials are replaced by Lagrange polynomials collocated at the Gauss-Lobatto-Legendre (GLL) points. The combination with GLL quadrature leads to a diagonal mass matrix that can be trivially inverted. In chapter 3 and appendix A we provide a detailed introduction to the spectral-element method. Applications on a variety of scales can be found in Faccioli et al. (1997), Komatitsch (1997), Komatitsch & Vilotte (1998), Komatitsch & Tromp (2002), Chaljub et al. (2003), Chaljub & Valette (2004) and Nissen-Meyer et al. (2007,2008)

Direct solution method: The direct solution method was introduced in a series of papers by Geller & Ohminato (1994) and Cummins et al. (1994a,b). As finite-element and spectral element methods, it is founded on the weak form of the equations of motion. What distinguishes the direct solution method is the choice of basis functions: linear splines in the radial direction and spherical harmonics in the horizontal directions. The space-discrete equations are then solved in the frequency domain.

Takeuchi et al. (2000) applied the optimal operator formalism of Geller & Takeuchi (1995) to the direct solution method and investigated its applicability to waveform inversion. The method was extended to radially anisotropic media by Kawai et al. (2006) and then applied to waveform inversion for localised heterogeneities by Konishi et al. (2009) and Kawai & Geller (2010).

Discontinuous Galerkin methods: Discontinuous Galerkin methods for seismic wave propagation have been developed only recently (e.g. Dumbser & Käser, 2006; Käser et al., 2007; de la Puente et al., 2007, 2008). They represent a class of finite-element methods where neighbouring elements are linked by numerical fluxes and not by continuity constraints. This allows for solutions that are discontinuous across element boundaries. Discontinuous Galerkin methods are therefore particularly well suited for the modelling of earthquake rupture processes (de la Puente et al., 2009). An application of a discontinuous Galerkin method in the context of full waveform inversion can be found in Brossier et al. (2009).

In chapters 2 and 3 two of the above mentioned discretisation approaches are described in more detail: the finite-element and the spectral element method.

Chapter 2

Finite-difference methods

The finite-difference method can be considered *the* classical and most frequently applied method for the numerical simulation of seismic wave propagation. It is based on the approximation of an exact derivative, $\partial_x f(x_i)$, at a grid position x_i, in terms of the function f evaluated at a finite number of neighbouring grid points.

Early implementations of the finite-difference method in a seismological context used a *conventional grid* where all field variables (e.g. displacement, stress, strain, etc.) are defined at the same grid positions. Examples may be found in Alterman & Karal (1968), Boore (1970, 1972), Alford et al. (1974) and Kelly et al. (1976).

The breakthrough in finite-difference modelling was the application of the *staggered-grid* approach (Madariaga, 1976; Virieux, 1984, 1986). In the staggered grid, field variables are defined at different grid positions, which reduces the effective grid spacing compared to the conventional grid. Further developments in finite-difference wave propagation focussed, for instance, on the modelling of the free surface (e.g. Graves, 1996; Kristek et al., 2002), the incorporation of general anisotropy (Igel et al., 1995) and the correct implementation of material discontinuities (e.g. Moczo et al., 2002). Applications of finite-difference modelling include studies of seismic ground motion in densely populated areas (e.g. Frankel & Vidale, 1992; Wang et al., 2008; Moczo et al., 2007; see figure 2.1), the simulation of wave propagation through random media (e.g. Frankel & Clayton, 1984, 1986) and full waveform inversion (e.g. Igel et al., 1996; Pratt, 1999; Dessa et al., 2004; Bleibinhaus et al., 2007). For an excellent review of finite-difference methods for wave propagation the reader is referred to Moczo et al. (2007).

To introduce the basic finite-difference concepts, we start in section 2.1 with the one-dimensional scalar wave equation. This simple example also allows us to study the stability and the numerical dispersion of the discretised equations. Based on the staggered-grid approach we then make the transition to the three-dimensional elastic case.

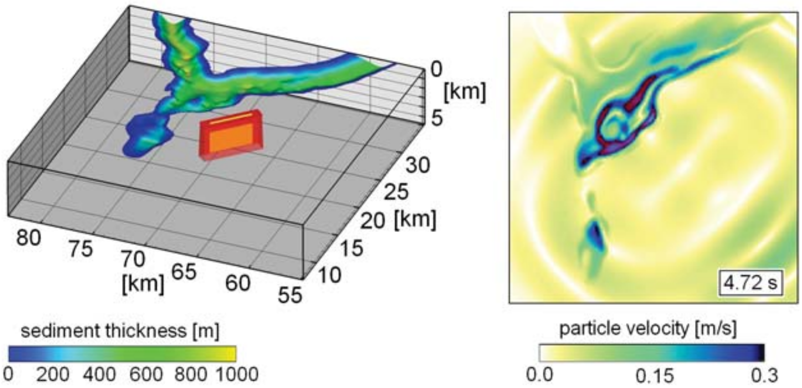

Fig. 2.1 **Left**: Geometrical configuration of the Grenoble (France) sedimentary basin. The fault zone is marked by a red box. **Right**: Snapshot of the absolute value of the horizontal-component particle velocity, 4.72 s after the rupture initiation. The simulation is based on a 4^{th}-order finite-difference discretisation of the elastic wave equation, as described in section 2.2.1. The horizontal displacement velocity is largest above the densely populated sedimentary basin. (Figure modified from Moczo et al. (2007), with permission from the authors.)

2.1 Basic concepts in one dimension

The one-dimensional scalar wave equation is particularly well suited not only for an introduction to the finite-difference method itself but also for the illustration of fundamental concepts in numerical analysis, including stability and grid dispersion. We start our development with the description of two methods for the construction of finite-difference approximations. These are then used to replace the exact derivatives in the wave equation. The result of this procedure is an iterative scheme that allows us to advance a discrete representation of the wave field in time. Both numerical experiments and a rigorous analysis reveal the properties of the iterative scheme in general and its stability requirements in particular.

2.1.1 Finite-difference approximations

2.1.1.1 Truncated Taylor expansions

The most straightforward procedure for the construction of finite-difference approximations to the derivative $\partial_x f(x)$ of a generic function f is based on truncated Taylor expansions. To find an approximation that is correct to second order in the grid spacing Δx, we choose the ansatz

$$\partial_x f(x) \approx g\,[f(x+\Delta x) - f(x-\Delta x)]\,, \tag{2.1}$$

with a scalar coefficient g that we seek to determine such that the right-hand side of (2.1) converges to $\partial_x f(x)$ as $\Delta x \to 0$. Expanding $f(x+\Delta x)$ and $f(x-\Delta x)$ into Taylor series centred around x, yields

$$g\,[f(x+\Delta x) - f(x-\Delta x)] = g\left[2\,\partial_x f(x)\,\Delta x + \frac{2}{3!}\,\partial_x^3 f(x)\,\Delta x^3 + ...\right]. \tag{2.2}$$

Choosing

$$g = \frac{1}{2\Delta x}\,, \tag{2.3}$$

results in the well-known second-order finite-difference stencil

$$\partial_x f(x) = \frac{1}{2\Delta x}\,[f(x+\Delta x) - f(x-\Delta x)] + \mathscr{O}(\Delta x^2)\,. \tag{2.4}$$

To derive approximations that are of arbitrary order in Δx, we generalise (2.1) such that it involves f evaluated at $2N$ grid points $x_n = x \pm n\,\Delta x$, with $n = 1, ..., N$:

$$\partial_x f(x) \approx \sum_{n=1}^{N} g_n\,[f(x+n\,\Delta x) - f(x-n\,\Delta x)]\,. \tag{2.5}$$

To determine the finite-difference coefficients g_n, we replace $f(x \pm n\Delta x)$ by a Taylor series,

$$\sum_{n=1}^{N} g_n\,[f(x+n\,\Delta x) - f(x-n\,\Delta x)] = \sum_{n=1}^{N} g_n\left[\sum_{k=0}^{\infty}\frac{1}{k!}\partial_x^k f(x)(n\Delta x)^k - \sum_{n=0}^{\infty}\frac{1}{k!}\partial_x^k f(x)(-n\Delta x)^k\right]. \tag{2.6}$$

The terms involving even powers of Δx cancel, so that we are left with

$$\sum_{n=1}^{N} g_n\,[f(x+n\,\Delta x) - f(x-n\,\Delta x)] = \sum_{n=1}^{N} g_n\left[\sum_{k=0}^{\infty}\frac{2}{(2k+1)!}\,\partial_x^{2k+1} f(x)(n\Delta x)^{2k+1}\right]. \tag{2.7}$$

Isolating the summands corresponding to $k = 0$, yields

$$\sum_{n=1}^{N} g_n\,[f(x+n\Delta x)-f(x-n\Delta x)] = 2\Delta x\,\partial_x f(x)\sum_{n=1}^{N} n g_n + \sum_{k=1}^{\infty}\frac{2}{(2k+1)!}\partial_x^{2k+1} f(x)\sum_{n=1}^{N} g_n\,(n\Delta x)^{2k+1}\,. \tag{2.8}$$

To ensure that the right-hand side of (2.8) converges to $\partial_x f(x)$ as $\Delta x \to 0$, we impose the condition

$$1 = 2\Delta x \sum_{n=1}^{N} n g_n\,. \tag{2.9}$$

Equation (2.9) leaves us with $N-1$ degrees of freedom for the coefficients g_n that we may now use to eliminate the monomials proportional to $\Delta x^3, ..., \Delta x^{2N-1}$ from (2.8). For this we furthermore require that the finite-difference coefficients satisfy the following system of $N-1$ linear equations:

$$0 = \sum_{n=1}^{N} g_n (n\Delta x)^{2k+1}\,, \quad k = 1,...,N-1\,. \tag{2.10}$$

The resulting finite-difference approximation is then of order $2N$ in Δx:

$$\sum_{n=1}^{N} g_n\,[f(x+n\Delta x)-f(x-n\Delta x)] = \partial_x f(x) + \mathscr{O}(\Delta x^{2N})\,. \tag{2.11}$$

By far the most frequently used finite-difference stencils are those of order 4 to 8. The associated coefficients are the solutions of the linear system given by equations (2.9) and (2.10):

$$\begin{aligned}
&\text{order 4:} \quad g_1 = \frac{2}{3\Delta x}\,, \quad g_2 = -\frac{1}{12\Delta x}\,,\\
&\text{order 6:} \quad g_1 = \frac{3}{4\Delta x}\,, \quad g_2 = -\frac{3}{20\Delta x}\,, \quad g_3 = \frac{1}{60\Delta x}\,,\\
&\text{order 8:} \quad g_1 = \frac{4}{5\Delta x}\,, \quad g_2 = -\frac{1}{5\Delta x}\,, \quad g_3 = \frac{4}{105\Delta x}\,, \quad g_4 = -\frac{1}{280\Delta x}\,.
\end{aligned} \tag{2.12}$$

By construction, higher-order approximations converge faster to the exact derivative as $\Delta x \to 0$. Also for (unavoidably) finite grid spacings, Δx, higher-order approximations generally yield more accurate solutions.

The finite-difference approximations considered so far are based on grid points that are evenly spaced and symmetric with respect to the grid point where the derivative is approximated. While this approach appears most natural, there are useful variations: In the staggered-grid discretisation, for instance (see section 2.2.1), the first derivative, $\partial_x f(x)$, is approximated correct to 4th order *between* the grid points and not *at* the grid points:

$$\begin{aligned}
\partial_x f(x) = &\frac{9}{8\Delta x}\,[f(x+\Delta x/2) - f(x-\Delta x/2)]\\
&- \frac{1}{24\Delta x}\,[f(x+3\Delta x/2) - f(x-3\Delta x/2)] + \mathscr{O}(\Delta x^4)\,.
\end{aligned} \tag{2.13}$$

The *effective grid spacing*, that is the distance between the grid points and the position where the derivative is approximated, is therefore $\pm\Delta x/2$ and $\pm 2\Delta/3$.

2.1.1.2 Sampling of band-limited derivative operators

An alternative way of constructing finite-difference schemes is based on the sampling of a band-limited version of the derivative operator (e.g. Mora, 1986; Igel et al., 1995). For this we note that the exact differentiation of $f(x)$ in the space domain corresponds to a multiplication of its Fourier transform, $\tilde{f}(k)$, by $\mathbf{i}k$, where k denotes the spatial wave number.

The discretisation implies that any numerical derivative can be accurate only within a limited wave number range,

$[-k_{\max}, k_{\max}]$. The *Nyquist wave number*, $k_{\max} = \frac{\pi}{\Delta x}$, is the maximum wave number that can be represented when the grid spacing is Δx. It follows that the wave number domain derivative operator in the range $[-k_{\max}, k_{\max}]$ is given by

$$\tilde{g}(k) = \mathbf{i}k \left[H(k + k_{\max}) - H(k - k_{\max}) \right], \tag{2.14}$$

with H being the Heaviside function. The multiplication $\tilde{g}(k)\tilde{f}(k)$ represents a band-limited differentiation that corresponds to a convolution in the space domain:

$$\widetilde{\partial_x f}(k) \approx \tilde{g}(k)\tilde{f}(k) \quad \Longleftrightarrow \quad \partial_x f(x) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(y) f(x-y)\, dy \,. \tag{2.15}$$

The space domain version of the approximate derivative operator, $\tilde{g}$, is proportional to the derivative of the sinc function,

$$g(x) = \frac{2k_{\max}}{\sqrt{2\pi}} \frac{d}{dx} \operatorname{sinc}(k_{\max} x) = \frac{2}{\sqrt{2\pi} x^2} \left[k_{\max} x \cos(k_{\max} x) - \sin(k_{\max} x) \right]. \tag{2.16}$$

To approximate the convolution in (2.15) we replace the integral by a Riemann sum that only involves f and g sampled at a finite number of grid points,

$$y_n = n\Delta x \,, \quad n = -N, ..., 0, ..., N \,. \tag{2.17}$$

The continuous integral then turns into a discrete convolution sum,

$$\partial_x f(x) \approx \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} g(y_n) f(x - y_n) \Delta x = \sum_{n=-N}^{N} g_n f(x - n\Delta x) \,. \tag{2.18}$$

The finite-difference coefficients, g_n, are given by

$$g_n = \begin{cases} 0\,, & \text{for} \quad n = 0\,, \\ \frac{(-1)^n}{n\Delta x}\,, & \text{for} \quad n \neq 0\,. \end{cases} \tag{2.19}$$

The discrete convolution in equation (2.18) constitutes a finite-difference approximation of the derivative $\partial_x f(x)$. The uncontestable beauty of this approach lies in its ability to effortlessly generate finite-difference coefficients for any distribution of grid points.

Our detour via the wave number domain suggests to quantify the accuracy of the discrete convolution differentiator in terms of its spectrum. First, we note that the coefficients (2.19) are anti-symmetric with respect to the index n. The spectrum of g_n is therefore purely imaginary. As illustrated in the left panel of figure 2.2, the spectrum of the approximate differentiator approaches the spectrum of the exact differentiation, $\mathbf{i}k$, as the number of grid points, N, that contribute to the discrete convolution increases. However, even for impractically large N we observe strong oscillations that indicate an insufficiently good approximation of the exact derivative.

The extent of this failure becomes most apparent when the discrete convolution from equations (2.18) and (2.19) is compared to the differentiators derived from truncated Taylor expansions. As shown in the right panel of figure 2.2, the Taylor method yields finite-difference coefficients that approximate the exact derivative very well for wave numbers smaller than $\approx 50\,\%$ of the Nyquist wave number.

This result does not mean that the sampling of the band-limited derivative operator (2.16) fails per se, because the result depends strongly on the location of the grid points. Sampling (2.16) at the grid points $n\Delta x$ leads, as we have seen, to an inaccurate approximation of the continuous convolution by the discrete Riemann sum (2.18). However, in the *staggered-grid approach* described in chapter 2.2.1, the derivative of f is evaluated halfway between the grid points. This means that we approximate $\partial_x f(x)$ in terms of f given at the grid positions $x - (n + 1/2)\Delta x$ for $n = -N, ..., 0, ..., N-1$. The discrete convolution then takes the form

$$\partial_x f(x) \approx \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N-1} g(y_n) f(x - y_n) \Delta x = \sum_{n=-N}^{N-1} g_n f[x - (n + 1/2)\Delta x] \,, \tag{2.20}$$

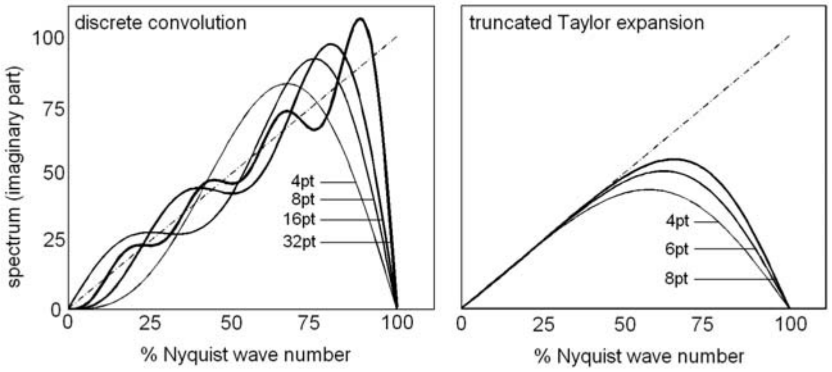

Fig. 2.2 **Left**: Imaginary part of the wave number spectrum corresponding to the discrete convolution differentiator defined by equations (2.18) and (2.19) for variable numbers grid points. (4 points $\leftrightarrow N = 2$, 8 points $\leftrightarrow N = 4$, ...) The dashed line represents the exact differentiation operator, $\mathbf{i}k$. **Right**: The same as on the left but for the finite-difference coefficients from equations (2.12) that we obtained from the truncation of Taylor expansions.

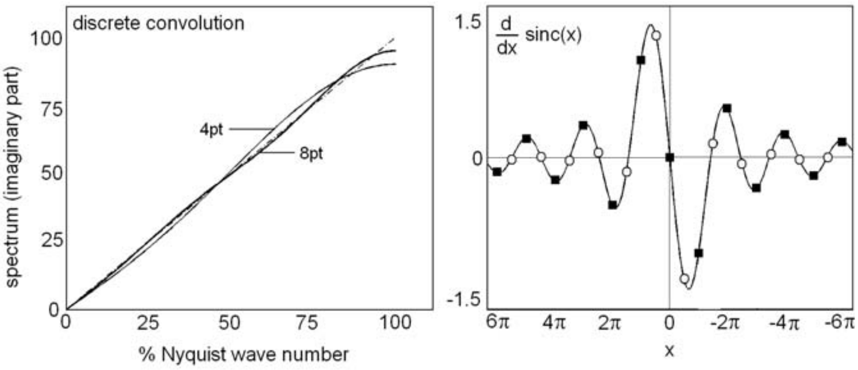

Fig. 2.3 **Left**: Imaginary part of the wave number spectrum corresponding to the 4-point and 8-point discrete convolution differentiator for the staggered grid. **Right**: The derivative of the sinc function and the sampling points in the regular grid (■) and in the staggered grid (○).

and the finite-difference coefficients, g_n, are defined through

$$g_n = \frac{(-1)^{n+1}}{\pi(n+1/2)^2\Delta x}\,. \tag{2.21}$$

The corresponding spectrum for the 4-point and 8-point convolutions is shown in the left panel of figure 2.3. The coefficients from equation (2.21) clearly outperform the coefficients in (2.12) that we derived from the Taylor expansion. In fact, the 8-point operator is hardly distinguishable from the exact derivative operator, even for wave numbers close to the Nyquist wave number. Examples where the remarkable properties of the discrete convolution defined in (2.20) and (2.21) have been used for 3D elastic wave propagation on a staggered grid can, for instance, be found in Igel et al. (1995).

An intuitive explanation for the large differences in accuracy of the discrete convolutions is provided in the right panel of figure (2.3) where we compare the sampling of the derivative of the sinc function (2.16) on the regular grid, $x_n = n\Delta x$, and the staggered grid, $x_n = (n+1/2)\,\Delta x$. The finite-difference coefficients for the staggered grid decay as n^{-2}, meaning that only coefficients with small n effectively contribute to the discrete convolution. On the regular grid, however, the coefficients decay slowly, as n^{-1}. Many more coefficients are therefore needed for an accurate approximation of the continuous convolution by the discrete Riemann sum.

It is, in principle, possible to approximate the exact derivative arbitrarily well with finite-difference operators that involve a large number of grid points. In practice one hardly uses more than 8 points in order to balance accuracy and computational costs.

2.1.2 Discretisation of the one-dimensional wave equation

To introduce the basic concepts of the finite-difference method in the context of wave propagation, we consider the one-dimensional wave equation

$$\rho(x)\,\ddot{u}(x,t) - \partial_x[\mu(x)\,\partial_x u(x,t)] = 0\,, \tag{2.22}$$

with the line density ρ and the elastic parameter μ. In the interest of simplicity, we disregard external forces, and we impose the non-zero initial condition

$$u(x,t)|_{t=0} = u_0(x)\,, \quad \dot{u}(x,t)|_{t=0} = 0\,, \tag{2.23}$$

with the initial displacement field u_0. The computational domain is the interval $[-L,L]$, and as boundary conditions we require zero displacement at $x = \pm L$:

$$u(x,t)|_{x=-L} = u(x,t)|_{x=L} = 0\,. \tag{2.24}$$

Other boundary conditions are possible, but their implementation usually requires additional work that we avoid in this introductory example. The discretisation process starts with the definition of $2N+1$ grid points x_i that are evenly distributed over the interval $[-L,L]$:

$$x_i = i\,\Delta x\,, \quad i = -N, ..., 0, ..., N\,. \tag{2.25}$$

The grid spacing, Δx, is determined by the requirement $N\Delta x = L$. We now approximate the first derivative $\partial_x u$ evaluated at the grid point x_i by the wave field evaluated at neighbouring grid points. Choosing, for instance, the second-order finite-difference approximation from equation 2.4, we find

$$\partial_x u(x_i,t) \approx \frac{1}{2\Delta x}\,[u(x_i+\Delta x,t) - u(x_i-\Delta x,t)] = \frac{1}{2\Delta x}\,[\bar{u}^{i+1}(t) - \bar{u}^{i-1}(t)]\,. \tag{2.26}$$

The time-dependent scalars $\bar{u}^{i\pm 1}$ are the wave field, sampled at the grid positions $x_i \pm \Delta x$. Repeating this procedure for the second spatial derivative in equation (2.22), yields

$$\partial_x[\mu(x)\,\partial_x u(x,t)] \approx \frac{1}{4\Delta x^2}\,\{\mu^{i+1}[\bar{u}^{i+2}(t) - \bar{u}^{i}(t)] - \mu^{i-1}[\bar{u}^{i}(t) - \bar{u}^{i-2}(t)]\}\,, \tag{2.27}$$

where we introduced the variable $\mu^i := \mu(x_i)$ which is the elastic parameter μ evaluated at the grid position x_i. The boundary conditions (2.24) are imposed explicitly by setting

$$\bar{u}^{-N-2} = \bar{u}^{-N-1} = \bar{u}^{-N} = \bar{u}^{N} = \bar{u}^{N+1} = \bar{u}^{N+2} = 0\,. \tag{2.28}$$

We can now assemble the semi-discrete version of the wave equation (2.22):

$$\rho^i \ddot{\bar{u}}^i(t) - \frac{1}{4\Delta x^2}\,\{\mu^{i+1}[\bar{u}^{i+2}(t) - \bar{u}^{i}(t)] - \mu^{i-1}[\bar{u}^{i}(t) - \bar{u}^{i-2}(t)]\} = 0\,, \tag{2.29}$$

with $\rho^i := \rho(x_i)$. It is important to note that equation (2.29) is meaningful only under the assumption that both ρ and μ are continuous with variations that can be represented reasonably well with a finite grid spacing. In the presence of material discontinuities averaging schemes must be applied to ensure accurate numerical solutions (e.g. Moczo et al., 2002; Kristek & Moczo, 2006).

Formally, we can write equation (2.29) in terms of a *mass matrix*, **M**, and a *stiffness matrix*, **K** (see chapter 1):

$$\mathbf{M}\cdot\ddot{\bar{\mathbf{u}}}(t) + \mathbf{K}\cdot\bar{\mathbf{u}}(t) = \mathbf{0}\,. \tag{2.30}$$

The vector $\bar{\mathbf{u}}$ is composed of the $2N+1$ coefficients $\bar{u}^i$. Since **M** is diagonal, it can be trivially inverted. In practice, the mass and the stiffness matrices are never computed explicitly because only the vector-matrix products $\mathbf{M}\cdot\ddot{\bar{\mathbf{u}}}$ and $\mathbf{K}\cdot\bar{\mathbf{u}}$ are needed. We will nevertheless use **M** and **K** for notational convenience.

It now remains to discretise the second time derivative in (2.30). In order to construct an explicit scheme, we use the second-order finite-difference approximation

$$\ddot{\bar{\mathbf{u}}}(t) \approx \frac{1}{\Delta t^2}\left[\bar{\mathbf{u}}(t+\Delta t) - 2\bar{\mathbf{u}}(t) + \bar{\mathbf{u}}(t-\Delta t)\right], \tag{2.31}$$

with a suitably chosen time increment Δt, that we will study later in more detail. Inserting (2.31) into (2.30) then yields a fully discrete version of the scalar wave equation (2.22):

$$\bar{\mathbf{u}}(t+\Delta t) = 2\bar{\mathbf{u}}(t) - \bar{\mathbf{u}}(t-\Delta t) - \Delta t^2 \mathbf{M}^{-1}\mathbf{K}\cdot\bar{\mathbf{u}}(t). \tag{2.32}$$

Equation (2.32) suggests the following recipe for the iterative finite-difference solution of the one-dimensional wave equation: Starting with $\bar{\mathbf{u}}(0)$ and $\bar{\mathbf{u}}(-\Delta t)$ determined by the initial condition (2.23), we compute $\bar{\mathbf{u}}(\Delta t)$. Then with the help of $\bar{\mathbf{u}}(0)$ and $\bar{\mathbf{u}}(\Delta t)$ we find $\bar{\mathbf{u}}(2\Delta t)$. This is repeated as long as required.

The strategy that we followed in the derivation of (2.32) is very general. It can, in particular, be used together with higher-order finite-difference operators. The achieved accuracy will generally depend on the time increment, the grid spacing, the properties of the finite-difference approximations but also on the material parameters.

To test the performance of the algorithm, we consider a homogeneous medium where the exact analytical solution, $u(x,t)$, is well-known:

$$u(x,t) = \frac{1}{2}\left[u_0(x-vt) + u_0(x+vt)\right]. \tag{2.33}$$

The wave field consists of two wave packages propagating in opposite directions with the velocity $v = \sqrt{\mu/\rho}$. The solution (2.33) is valid as long as the wave field does not reach the boundaries at $x = \pm L$. For our numerical experiment we choose $v = 5$ km/s, $\Delta x = 1$ km and $\Delta t = 0.2$ s. The initial wave field is the derivative of a Gaussian with variance σ:

$$u_0(x) = \frac{d}{dx}e^{-x^2/\sigma^2}. \tag{2.34}$$

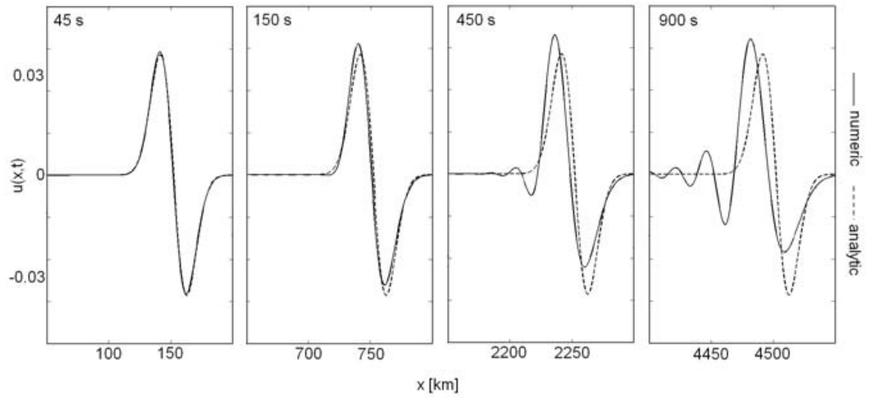

Fig. 2.4 Snapshots of the wave package travelling in positive x-direction for different times. The analytic solution is shown for reference in the form of the dashed curves. The numerical error, that is the discrepancy between the exact analytic solution and the numerical solution increases steadily with increasing propagation distance. Details of the simulation parameters are provided in the text.

Choosing $\sigma = 15$ km results in a dominant wavelength of approximately 30 km. Figure 2.4 shows the results of the numerical simulation and compares them to the analytical solution. For propagation distances up to 200 km, roughly equivalent to 7 wavelengths, the numeric solution is hardly distinguishable from the analytic reference. However, as the simulation proceeds, the numerical errors increase steadily. Clearly, after 450 s (2250 time steps) the numerical solution becomes almost useless. The snapshot at 900 s (4500 time steps) reveals that the numeric solution tends to disperse, meaning that higher-frequency components appear to propagate slower than lower-frequency components. This is despite the non-dispersive character of the analytic solution where all frequencies travel at the same wave speed, v. We refer to this undesirable phenomenon as *numerical dispersion* or *grid dispersion*.

To better quantify the discrepancy between analytical and numerical solutions, we use the following measure of the numerical error:

$$E(t) := \frac{\sqrt{\sum_{i=-N}^{N} [\bar{u}^i(t) - u(i\Delta x, t)]^2}}{\sqrt{\sum_{i=-N}^{N} u^2(i\Delta x, t)}} . \tag{2.35}$$

More elaborate measures of misfit can be defined using, for instance, wavelet transforms (Kristekova et al., 2006, 2009), but $E(t)$ as defined in (2.35) is sufficient for our purposes. There are three major factors that control the numerical error: the dominant wavelength relative to the grid spacing Δx, the order of the finite-difference operator and the number of iterations or time steps.

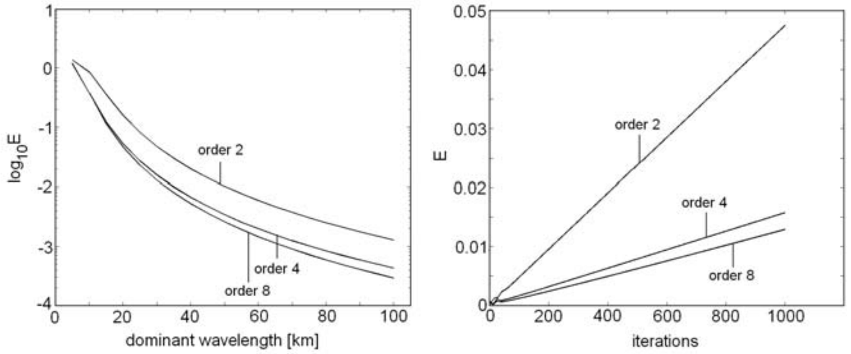

Fig. 2.5 **Left**: Numerical error after 1000 iterations as a function of the dominant wavelength and the finite-difference order. **Right**: Numerical error as a function of the number of iterations and the finite-difference order.

The left part of figure 2.5 illustrates the role played by the finite-difference order and the dominant wavelength. Shown is the error, $E(t)$, after 1000 iterations, that is for $t = 200$ s. Clearly, the error drops rapidly as the dominant wavelength increases. To achieve an error below 1 %, the dominant wavelength should not be less than 20 km for the 4th- and 8th-order operators, or 40 km for the 2nd-order operator. Increasing orders generally lead to more accurate results, but the effect is largest when going from order 2 to order 4.

The right panel of figure 2.5 shows the dependence of the error, $E(t)$, on the number of iterations. It demonstrates that numerical inaccuracies are cumulative, meaning that they increase monotonically - and in this case almost exactly linearly - with the propagation distance of the waves.

2.1.3 Von Neumann Analysis: stability and numerical dispersion

Each method used to approximate the solution of a differential equation requires a detailed analysis in order to learn about errors and stability without relying on specific numerical experiments that may not be representative. *Von Neumann analysis* - first applied by Crank & Nicolson (1947) and Charney et al. (1950) - is a powerful tool for the evaluation of finite-difference approximations to hyperbolic partial differential equations. It allows us to study stability and the numerical dispersion that we already observed qualitatively in the previous paragraph.

For our analysis we assume a homogeneous medium defined on the interval $[-\pi, \pi]$, where the wave field, $u(x,t)$, can be represented in terms of a Fourier series

$$u(x,t) = \sum_{k=-\infty}^{\infty} \psi_k(t)\, e^{\mathbf{i}kx} , \tag{2.36}$$

with time-dependent Fourier coefficients, $\psi_k(t)$. The corresponding representation for the discrete approximation to the wave field is

$$\bar{u}^n(t) = \sum_{k=-N/2}^{N/2} \bar{\psi}_k(t)\, e^{\mathbf{i}kn\Delta x}\,. \tag{2.37}$$

Note that the spatial discretisation with grid spacing $\Delta x = 2\pi/N$ restricts the wave number range to plus/minus the Nyquist wave number, $\pi/\Delta x$, that is to $k \in [-\pi/\Delta x, \pi/\Delta x] = [-N/2, N/2]$. Introducing (2.37) into the semi-discrete wave equation (2.29) gives

$$\ddot{\bar{u}}^n(t) = \frac{v^2}{4\Delta x^2} \sum_{k=-N/2}^{N/2} \left[e^{2\mathbf{i}k\Delta x} - 2 + e^{-2\mathbf{i}k\Delta x}\right] \bar{\psi}_k(t)\, e^{\mathbf{i}kn\Delta x} = -\frac{v^2}{\Delta x^2} \sum_{k=-N/2}^{N/2} \sin^2(k\Delta x)\, \bar{\psi}_k(t)\, e^{\mathbf{i}kn\Delta x}\,. \tag{2.38}$$

Replacing $\ddot{\bar{u}}^n(t)$ by the second-order finite-difference approximation (2.31) and substituting the Fourier series (2.37) yields an equation for the coefficients $\bar{\psi}_k$:

$$\bar{\psi}_k(t+\Delta t) - 2\bar{\psi}_k(t) + \bar{\psi}_k(t-\Delta t) = -\frac{v^2\Delta t^2}{\Delta x^2} \sin^2(k\Delta x)\, \bar{\psi}_k(t)\,. \tag{2.39}$$

To eliminate the explicit dependence on $\bar{\psi}_k$ at time $t - \Delta t$, we define the auxiliary variables

$$\bar{\phi}_k(t+\Delta t) := \bar{\psi}_k(t+\Delta t) - \bar{\psi}_k(t)\,. \tag{2.40}$$

With the help of $\bar{\phi}_k$ we can write equation (2.39) in the form of a linear system of equations that only depends on coefficients at times t and $t + \Delta t$:

$$\begin{aligned} \bar{\phi}_k(t+\Delta t) - \bar{\phi}_k(t) &= -\frac{v^2\Delta t^2}{\Delta x^2} \sin^2(k\Delta x)\, \bar{\psi}_k(t)\,, \\ \bar{\psi}_k(t+\Delta t) - \bar{\psi}_k(t) &= \bar{\phi}_k(t+\Delta t)\,, \end{aligned} \tag{2.41}$$

or using matrix notation,

$$\begin{pmatrix} \bar{\phi}_k(t+\Delta t) \\ \bar{\psi}_k(t+\Delta t) \end{pmatrix} = \mathbf{A}_k \cdot \begin{pmatrix} \bar{\phi}_k(t) \\ \bar{\psi}_k(t) \end{pmatrix}\,. \tag{2.42}$$

The 2×2-matrix $\mathbf{A}_k$ is given by

$$\mathbf{A}_k = \begin{pmatrix} 1 & -\frac{v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \\ 1 & 1 - \frac{v^2\Delta t^2}{\Delta x^2}\sin^2(k\Delta x) \end{pmatrix}\,. \tag{2.43}$$

It follows that we can advance the Fourier coefficients, $\bar{\psi}_k$, in time by the repeated application of $\mathbf{A}_k$ to the initial values of the coefficients:

$$\begin{pmatrix} \bar{\phi}_k(j\Delta t) \\ \bar{\psi}_k(j\Delta t) \end{pmatrix} = \mathbf{A}_k^j \cdot \begin{pmatrix} \bar{\phi}_k(0) \\ \bar{\psi}_k(0) \end{pmatrix}\,. \tag{2.44}$$

The following stability and dispersion analysis is founded on the iteration defined by equation (2.44).

2.1.3.1 Numerical stability

All relevant properties of the algorithm, and its stability in particular, are now encapsulated in the eigenvalues, λ_k, of the matrix $\mathbf{A}_k$. In fact, for the iteration to be stable, the absolute values of the eigenvalues, $|\lambda_k|$, must be smaller than or equal to 1. Otherwise, that is for $|\lambda_k| > 1$, the absolute values of $\bar{\psi}_k$ and $\bar{\phi}_k$ grow indefinitely. To facilitate the eigenvalue analysis, we define

$$\gamma_k := \sin^2(k\Delta x)\,, \tag{2.45}$$

and the *Courant number*

$$c := \frac{v\Delta t}{\Delta x}\,. \tag{2.46}$$

In terms of γ_k and c, the eigenvalues, λ_k, of $\mathbf{A}_k$ are given by

$$\lambda_k = 1 - \frac{1}{2}c^2\gamma_k^2 \pm \sqrt{\frac{1}{4}c^4\gamma_k^4 - c^2\gamma_k^2}\,. \tag{2.47}$$

For $c^2\gamma_k^2 > 4$ the square root is always real, and we can have $|\lambda_k| > 1$. For the algorithm to be stable we therefore require at least $c^2\gamma_k^2 \le 4$. In this case, the square root becomes imaginary, and the absolute value of λ_k is then

$$|\lambda_k| = \sqrt{\lambda_k \lambda_k^*} = 1\,, \tag{2.48}$$

which implies stability. Since γ_k ranges between 0 and 1, it follows that we need to impose

$$\frac{v\Delta t}{\Delta x} = c \le 2\,, \tag{2.49}$$

in order to ensure the stability of the iterative solution to the one-dimensional wave equation discretised by 2^{nd}-order finite-difference approximations, as explained in the previous paragraph. The stability condition (2.49) is commonly referred to as the *CFL condition*, named after R. Courant, K. Friedrichs and H. Lewy (Courant et al., 1928). For our example from figure 2.4, with $\Delta x = 1$ km and $v = 5$ km/s, the CFL condition imposes a maximum time increment of 0.4 s. Smaller time increments are predicted to yield stable solutions, whereas larger time increments are expected to lead to solutions that grow indefinitely as the iteration proceeds. Figure 2.6 proves that the prediction of the CFL condition is remarkably accurate. The solution is stable for $\Delta x = 0.39$ s, but it "explodes" for $\Delta t = 0.41$ s.

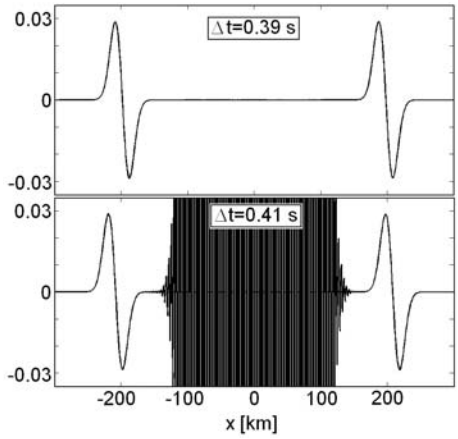

Fig. 2.6 Numerical (in)stability. **Top**: The time increment, $\Delta t = 0.39$ s, is slightly below the limit of 0.4 s imposed by the CFL condition. The solution is stable. **Bottom**: The time increment, $\Delta t = 0.41$, is slightly above the limit of 0.4, leading to instabilities. After few iterations the numerical solution "explodes".

Our analysis was based on a specific finite-difference discretisation of the one-dimensional wave equation with constant parameters. More generally, CFL conditions of the form

$$\Delta t \le \text{const.}\,\frac{\min h}{\max v}\,, \tag{2.50}$$

with a grid spacing parameter, h, and a wave speed, v, are valid for all discretised wave equations where the time-stepping is explicit. The constant on the right-hand side depends on the methods used for the space and time discretisation.

The CFL condition strongly limits the efficiency of numerical methods because it imposes an upper bound for the time increment, Δt. Reducing the minimum grid spacing by a factor m in order to achieve more accurate solutions automatically implies a reduction of the time increment from Δt to $\Delta t/m$.

It is, in principle, possible to circumvent the CFL condition using implicit time-stepping schemes. However, their numerical costs usually compensate the benefit of using a larger time increment.

2.1.3.2 Numerical dispersion

To quantify the numerical dispersion that we observed in figure 2.4, we assume, without loss of generality, that the initial state $(\bar{\phi}_k(0), \bar{\psi}_k(0))$ from equation (2.44) is an eigenvector of the matrix $\mathbf{A}_k$. The iteration that advances the Fourier coefficients in time then simplifies to

$$\begin{pmatrix} \bar{\phi}_k(j\Delta t) \\ \bar{\psi}_k(j\Delta t) \end{pmatrix} = \lambda_k^j \begin{pmatrix} \bar{\phi}_k(0) \\ \bar{\psi}_k(0) \end{pmatrix} . \tag{2.51}$$

Since λ_k is complex-valued within the stable range of time increments, Δt, it can be rewritten in Eulerian form

$$\lambda_k = |\lambda_k| \, e^{\mathbf{i}\alpha_k} , \tag{2.52}$$

where the phase α_k is generally non-zero. To attach physical meaning to the phase, we write α_k in terms of a numerical wave speed, $\bar{v}_k$:

$$\alpha_k =: \Delta t \, k \, \bar{v}_k . \tag{2.53}$$

Introducing equations (2.51) to (2.53) into the Fourier series representation of the finite-difference coefficients $\bar{u}^n$ (equation 2.37) gives

$$\bar{u}^n(j\Delta t) = \sum_{k=-N/2}^{N/2} \bar{\psi}_k(0) \, |\lambda_k|^j \, e^{\mathbf{i}k(j\Delta t \, \bar{v}_k + n\Delta x)} . \tag{2.54}$$

Equation (2.54) reveals that $\bar{v}_k$ indeed plays the role of a wave speed. If the numerical solution were exact, $\bar{v}_k$ would be equal to $v = \sqrt{\mu/\rho}$ and independent of the wave number k. As a result of the discretisation, however, $\bar{v}_k$ is not generally equal to v. Moreover, $\bar{v}_k$ depends on k, meaning that the numerical solution is dispersive, in contrast to the exact solution. Figure 2.7 illustrates the effect of numerical dispersion for the setup that we already used for the example in figure 2.4. There we observed qualitatively that shorter-wavelength components travel at a lower speed than longer-wavelength components, and that the numerical solution in general seems to be slower than the analytical solution. This is confirmed by our qualitative dispersion analysis. Only components with a wave number close to zero propagate at the correct speed, $v = 5$ km/s. As the wave numbers approach the Nyquist wave number, the propagation speed tends to zero.

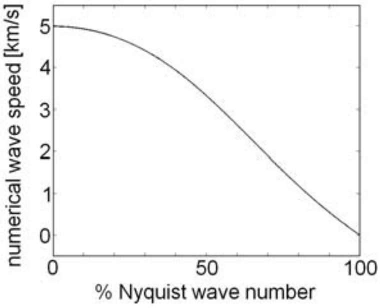

Fig. 2.7 Numerical dispersion. Shown is the numerical wave speed, $\bar{v}_k$, as a function of the wave number, k. The wave speed of the exact solution is $v = 5$ km/s, independent of k. Due to numerical dispersion, $\bar{v}_k$ decreases with increasing wave number. Shorter-wavelength components therefore travel at reduced speed. Wavelengths corresponding to the Nyquist wave number do not propagate at all. Only components with a wave number close to zero travel with the correct speed.

Numerical dispersion is, of course, not a special property of particular example. All numerical solutions are dispersive to some degree, regardless of the method used. In many cases, numerical dispersion analysis is not as straightforward as in the case of low-order finite-difference methods. Its effect should therefore, if possible, be assessed by comparison with analytical solutions.

2.2 Extension to the three-dimensional cartesian case

The application of the finite-difference method to the three-dimensional elastic wave equation is conceptually similar to the one-dimensional scalar case that we studied in the previous section: Exact derivatives are replaced by finite-difference approximations, and this leads to a discrete scheme that can be advanced in time iteratively.
Yet, the simple scheme applied in one dimension proves inefficient in three dimensions because of the large number of grid points per wavelength needed to achieve accurate solutions. Fortunately, a modification of the spatial discretisation - described in section 2.2.1 and referred to as the *staggered grid* - allows us to effectively reduce the grid spacing by 50 % without increasing the number of discrete field variables.

2.2.1 The staggered grid

To translate the finite-difference method to the three-dimensional cartesian case, we consider the velocity-stress formulation of the elastic wave equation in the half-space $G = \mathbb{R}^2 \times (-\infty, z_0]$

$$\rho(\mathbf{x})\,\dot{\mathbf{v}}(\mathbf{x},t) - \nabla\cdot\sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t)\,, \tag{2.55}$$

$$\dot{\sigma}(\mathbf{x},t) = \mathbf{C}(\mathbf{x}) : \dot{\varepsilon}(\mathbf{x},t)\,, \tag{2.56}$$

subject to the free-surface boundary condition

$$\mathbf{e}_z \cdot \sigma|_{z=z_0} = \mathbf{0}\,, \tag{2.57}$$

and the initial conditions

$$\mathbf{u}|_{t\leq t_0} = \mathbf{v}|_{t\leq t_0} = \mathbf{0}\,. \tag{2.58}$$

The symbol $\mathbf{e}_z$ denotes the unit vector in vertical direction. The first step towards the finite-difference approximation of equations (2.55) and (2.56) is to write the stress divergence $\nabla\cdot\sigma$ and the strain rate tensor $\dot{\varepsilon}$ in explicit form:

$$(\nabla\cdot\sigma)_x = \partial_x\sigma_{xx} + \partial_y\sigma_{xy} + \partial_z\sigma_{xz}\,, \quad (\nabla\cdot\sigma)_y = \partial_x\sigma_{yx} + \partial_y\sigma_{yy} + \partial_z\sigma_{yz}\,, \quad (\nabla\cdot\sigma)_z = \partial_x\sigma_{zx} + \partial_y\sigma_{zy} + \partial_z\sigma_{zz}\,, \tag{2.59}$$

$$\dot{\varepsilon}_{xx} = \partial_x v_x\,,\, \dot{\varepsilon}_{yy} = \partial_y v_y\,,\, \dot{\varepsilon}_{zz} = \partial_z v_z\,,\, \dot{\varepsilon}_{xy} = \frac{1}{2}(\partial_x v_y + \partial_y v_x)\,, \quad \dot{\varepsilon}_{xz} = \frac{1}{2}(\partial_x v_z + \partial_z v_x)\,, \quad \dot{\varepsilon}_{zy} = \frac{1}{2}(\partial_z v_y + \partial_y v_z)\,. \tag{2.60}$$

Our goal is to replace the derivatives in the above equations by finite-difference approximations. The practicality of this approach in three dimensions rests on the definition of a *staggered grid* (e.g. Virieux, 1984, 1986; Levander, 1988; Igel et al., 1995; Graves, 1996). To illustrate the staggered-grid concept we start by defining the discrete versions of the velocity components v_x, v_y and v_z on three different grids:

$$\bar{v}_x^{i,j,k} := v_x(x_i + \Delta x/2, y_j, z_k)\,, \tag{2.61}$$

$$\bar{v}_y^{i,j,k} := v_y(x_i, y_j + \Delta y/2, z_k)\,, \tag{2.62}$$

$$\bar{v}_z^{i,j,k} := v_z(x_i, y_j, z_k + \Delta z/2)\,, \tag{2.63}$$

where $\mathbf{x}_{ijk} = (x_i, y_j, z_k)$ is a generic grid point. Based on the above discretisations we can approximate the derivatives of the velocity field, as they appear in equation (2.60). For instance, the 4th-order approximation of $\partial_y v_x(x_i + \Delta x/2, y_j + \Delta y/2, z_k)$ in terms of v_x evaluated at the positions $(x_i + \Delta x/2, y_j + n\Delta y, z_k)$ with $n = -1, 0, 1, 2$, is given by (see equation 2.13)

$$\begin{aligned}
&\partial_y v_x(x_i+\Delta x/2, y_j+\Delta y/2, z_k)\\
&\quad\approx \frac{9}{8\Delta y}\left[v_x(x_i+\Delta x/2, y_j+\Delta y, z_k)-v_x(x_i+\Delta x/2, y_j, z_k)\right]\\
&\quad-\frac{1}{24\Delta y}\left[v_x(x_i+\Delta x/2, y_j+2\Delta y, z_k)-v_x(x_i+\Delta x/2, y_j-\Delta y, z_k)\right]\\
&\quad=\frac{9}{8\Delta y}\left[\bar{v}_x^{i,j+1,k}-\bar{v}_x^{i,j,k}\right]-\frac{1}{24\Delta y}\left[\bar{v}_x^{i,j+2,k}-\bar{v}_x^{i,j-1,k}\right]. \qquad (2.64)
\end{aligned}$$

The corresponding approximation for $\partial_x v_y$ is

$$\partial_x v_y(x_i+\Delta x/2, y_j+\Delta y/2, z_k)\approx\frac{9}{8\Delta x}\left[\bar{v}_y^{i+1,j,k}-\bar{v}_y^{i,j,k}\right]-\frac{1}{24\Delta x}\left[\bar{v}_y^{i+2,j,k}-\bar{v}_y^{i-1,j,k}\right]. \qquad (2.65)$$

At this point we note that the approximations of $\partial_y v_x$ and $\partial_x v_y$ are available at the same grid point, namely at $(x_i+\Delta x/2, y_j+\Delta y/2, z_k)$. We may thus combine (2.64) and (2.65) into the 4th-order approximation of the strain rate component $\dot{\varepsilon}_{xy}$:

$$\begin{aligned}
\dot{\bar{\varepsilon}}_{xy}^{i,j,k}&=\dot{\varepsilon}_{xy}(x_i+\Delta x/2, y_j+\Delta y/2, z_k)\\
&=\frac{1}{2}\left[\partial_y v_x(x_i+\Delta x/2, y_j+\Delta y/2, z_k)+\partial_x v_y(x_i+\Delta x/2, y_j+\Delta y/2, z_k)\right]\\
&\approx\frac{9}{16\Delta x}\left[\bar{v}_x^{i,j+1,k}-\bar{v}_x^{i,j,k}+\bar{v}_y^{i+1,j,k}-\bar{v}_y^{i,j,k}\right]-\frac{1}{48\Delta y}\left[\bar{v}_x^{i,j+2,k}-\bar{v}_x^{i,j-1,k}+\bar{v}_y^{i+2,j,k}-\bar{v}_y^{i-1,j,k}\right]. \qquad (2.66)
\end{aligned}$$

Following this example we find the grid points where the finite-difference approximations of the remaining strain rate components are defined:

$$\begin{aligned}
&\dot{\bar{\varepsilon}}_{xx}^{i,j,k}=\dot{\varepsilon}_{xx}(x_i, y_j, z_k)\,,\quad \dot{\bar{\varepsilon}}_{yy}^{i,j,k}=\dot{\varepsilon}_{yy}(x_i, y_j, z_k)\,,\quad \dot{\bar{\varepsilon}}_{zz}^{i,j,k}=\dot{\varepsilon}_{zz}(x_i, y_j, z_k)\,,\\
&\dot{\bar{\varepsilon}}_{xy}^{i,j,k}=\dot{\varepsilon}_{xy}(x_i+\Delta x/2, y_j+\Delta y/2, z_k)\,,\\
&\dot{\bar{\varepsilon}}_{xz}^{i,j,k}=\dot{\varepsilon}_{xz}(x_i+\Delta x/2, y_j, z_k+\Delta z/2)\,,\\
&\dot{\bar{\varepsilon}}_{yz}^{i,j,k}=\dot{\varepsilon}_{yz}(x_i, y_j+\Delta y/2, z_k+\Delta z/2)\,. \qquad (2.67)
\end{aligned}$$

Assuming an isotropic medium, the stress and strain rate tensors are connected via the constitutive relation

$$\sigma_{ij}=\sum_{k,l=1}^{3}\left(\lambda\,\delta_{ij}\delta_{kl}+\mu\,\delta_{ik}\delta_{jl}+\mu\,\delta_{il}\delta_{jk}\right)\varepsilon_{kl}\,, \qquad (2.68)$$

where λ and μ are the Lamé parameters. Using (2.68) we can compute the components of the discrete stress rate tensor:

$$\begin{aligned}
&\dot{\bar{\sigma}}_{xx}^{ij}=(\lambda+2\mu)\,\dot{\bar{\varepsilon}}_{xx}^{ij}+\lambda\,(\dot{\bar{\varepsilon}}_{yy}^{ij}+\dot{\bar{\varepsilon}}_{zz}^{ij})\,,\\
&\dot{\bar{\sigma}}_{yy}^{ij}=(\lambda+2\mu)\,\dot{\bar{\varepsilon}}_{yy}^{ij}+\lambda\,(\dot{\bar{\varepsilon}}_{xx}^{ij}+\dot{\bar{\varepsilon}}_{zz}^{ij})\,,\\
&\dot{\bar{\sigma}}_{zz}^{ij}=(\lambda+2\mu)\,\dot{\bar{\varepsilon}}_{zz}^{ij}+\lambda\,(\dot{\bar{\varepsilon}}_{xx}^{ij}+\dot{\bar{\varepsilon}}_{yy}^{ij})\,,\\
&\dot{\bar{\sigma}}_{xy}^{ij}=2\mu\,\dot{\bar{\varepsilon}}_{xy}^{ij}\,,\quad \dot{\bar{\sigma}}_{xz}^{ij}=2\mu\,\dot{\bar{\varepsilon}}_{xz}^{ij}\,,\quad \dot{\bar{\sigma}}_{yz}^{ij}=2\mu\,\dot{\bar{\varepsilon}}_{yz}^{ij}\,. \qquad (2.69)
\end{aligned}$$

Equations (2.69) reveal that the discrete stress rate components, $\dot{\bar{\sigma}}_{mn}^{ij}$, are located at the same grid positions as the discrete strain rate components, $\dot{\bar{\varepsilon}}_{mn}^{ij}$. The collocation of stress and strain components only holds in isotropic media, and interpolation becomes necessary in the anisotropic case. Note also that the elastic parameters, λ and μ, are needed at different grid positions. We did not translate this requirement into the notation in order to keep the treatment readable. Based on (2.69) we can advance the discrete stress field, $\bar{\sigma}$, in time. It remains to discretise the stress divergence. For the x-component of $\nabla\cdot\sigma$ we find the following 4th-order approximation:

$$
\begin{aligned}
&[\nabla \cdot \sigma(x_i+\Delta x/2, y_j, z_k)]_x = \rho\, \dot{v}_x(x_i+\Delta x/2, y_j, z_k) - f_x(x_i+\Delta x/2, y_j, z_k) \\
&\approx \frac{9}{8\Delta x}\left[\sigma_{xx}(x_i+\Delta x, y_j, z_k) - \sigma_{xx}(x_i, y_j, z_k)\right] - \frac{1}{24\Delta x}\left[\sigma_{xx}(x_i+2\Delta x, y_j, z_k) - \sigma_{xx}(x_i-\Delta x, y_j, z_k)\right] \\
&+ \frac{9}{8\Delta y}\left[\sigma_{xy}(x_i+\Delta x/2, y_j+\Delta y/2, z_k) - \sigma_{xy}(x_i+\Delta x/2, y_j-\Delta y/2, z_k)\right] \\
&- \frac{1}{24\Delta y}\left[\sigma_{xy}(x_i+\Delta x/2, y_j+3\Delta y/2, z_k) - \sigma_{xy}(x_i+\Delta x/2, y_j-3\Delta y/2, z_k)\right] \\
&+ \frac{9}{8\Delta z}\left[\sigma_{xz}(x_i+\Delta x/2, y_j, z_k+\Delta z/2) - \sigma_{xz}(x_i+\Delta x/2, y_j, z_k-\Delta z/2)\right] \\
&- \frac{1}{24\Delta z}\left[\sigma_{xz}(x_i+\Delta x/2, y_j, z_k+3\Delta z/2) - \sigma_{xz}(x_i+\Delta x/2, y_j, z_k-3\Delta z/2)\right]. \qquad (2.70)
\end{aligned}
$$

Written in terms of the discrete field variables, equation (2.70) becomes

$$
\begin{aligned}
\rho\, \dot{\bar{v}}_x^{i,j,k} - f_x^{i,j,k} &= \frac{9}{8\Delta x}\left[\bar{\sigma}_{xx}^{i+1,j,k} - \bar{\sigma}_{xx}^{i,j,k}\right] - \frac{1}{24\Delta x}\left[\bar{\sigma}_{xx}^{i+2,j,k} - \sigma_{xx}^{i-1,j,k}\right] \\
&+ \frac{9}{8\Delta y}\left[\bar{\sigma}_{xy}^{i,j,k} - \bar{\sigma}_{xy}^{i,j-1,k}\right] - \frac{1}{24\Delta y}\left[\bar{\sigma}_{xy}^{i,j+1,k} - \sigma_{xy}^{i,j-2,k}\right] \\
&+ \frac{9}{8\Delta z}\left[\bar{\sigma}_{xz}^{i,j,k} - \bar{\sigma}_{xz}^{i,j,k-1}\right] - \frac{1}{24\Delta z}\left[\bar{\sigma}_{xz}^{i,j,k+1} - \sigma_{xz}^{i,j,k-2}\right]. \qquad (2.71)
\end{aligned}
$$

Following this scheme, we find that the grid positions of $\bar{v}_x^{i,j,k}$, $\bar{v}_y^{i,j,k}$ and $\bar{v}_z^{i,j,k}$ coincide with those that we originally defined in equation (2.61) to (2.63). The staggered grid is therefore in itself consistent. Figure 2.8 summarises the positions of the different discrete quantities.

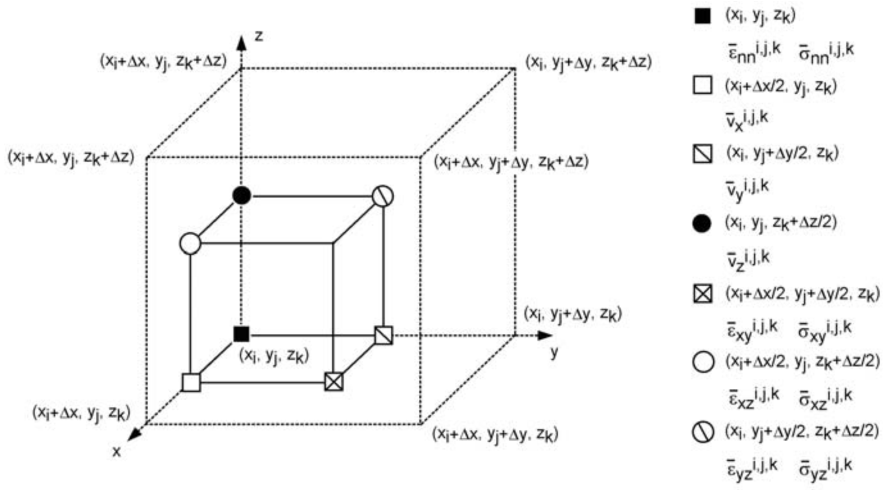

Fig. 2.8 Illustration of the staggered-grid concept. The diagonal elements of the discrete stress and strain tensors, $\bar{\varepsilon}_{nn}^{i,j,k}$ and $\bar{\sigma}_{nn}^{i,j,k}$ are located at the generic grid point $\mathbf{x}_{ijk} = (x_i, y_j, z_k)$. The remaining quantities are displaced from $\mathbf{x}_{ijk}$ by half a grid spacing.

The above development suggests a simple recipe for iteratively advancing the discrete wave field with the help of the leapfrog time discretisation: (see section 1.4.1): Starting from the zero initial conditions

1. compute the strain rate tensor, $\dot{\bar{\varepsilon}}(t)$, from the velocity, $\bar{\mathbf{v}}(t)$,

2. using the constitutive relation, compute the strain rate, $\dot{\bar{\sigma}}(t)$ from $\dot{\bar{\varepsilon}}(t)$,

3. using $\dot{\bar{\sigma}}(t)$, compute $\bar{\sigma}(t+\Delta t/2)$ via a second-order finite-difference approximation,
4. from $\bar{\sigma}(t+\Delta t/2)$ compute the discrete version of $\nabla \cdot \bar{\sigma}(t+\Delta t/2)$,
5. using $\nabla \cdot \bar{\sigma}(t+\Delta t/2)$ and the discrete momentum balance, compute $\bar{\mathbf{v}}(t+\Delta t)$ with the help of the second-order finite-difference approximation, and then

repeat as often as required. The beauty of the staggered grid is that it allows us to decrease the effective grid spacing without increasing the number of discrete field variables. In the conventional grid, used for instance in our 1D example (section 2.1), the effective grid spacing in x-direction is a multiple of $\pm\Delta x$. This is reduced to $\pm\Delta x/2$ and $\pm 2\Delta x/3$. The result is a much lower numerical dispersion that comes without any increase in computational costs.

2.3 Accuracy and efficiency

The efficiency of numerical modelling schemes is particularly important because their applicability is usually limited by the available computational resources. To be efficient, a numerical method should require as few grid points as possible, to achieve accurate solutions for the shortest spatial wavelength. The term *accurate* is naturally problem-specific, and so is the necessary number of grid points per minimum wavelength, n_{min}. Among other factors, n_{min} depends on the spatio-temporal discretisation, the types of waves to be modelled (body or surface waves), the presence of irregular topography, the distance travelled by the waves (see figure 2.5) and the data to which the synthetics are to be compared. The following paragraphs provide a collection of recommendations for n_{min} in a variety of scenarios.

For plane body waves propagating through the 4th-order staggered grid introduced in section 2.2.1, Robertsson et al. (1994) and Bohlen (2002) found that n_{min} should be around 4 to 8 in order to reduce the dispersion error below 2 to 5 %. Moczo et al. (2000) conclude that the group velocity error for S waves is 2.5 % for $n_{\text{min}} = 6$. Those values must be considered in the context of the heterogeneous Earth where velocity variations on the order of 1 % can be relevant. Thus, for applications in body wave delay time tomography, $n_{\text{min}} \gtrapprox 20$ is certainly recommendable.
In a systematic comparison of several finite-difference and finite-element schemes, Moczo et al. (2010) analysed the dependence of the numerical error on the P wave speed to S wave speed ratio, $v_{\text{P}}/v_{\text{S}}$. While staggered-grid approaches appear to be rather insensitive to $v_{\text{P}}/v_{\text{S}}$, the conventional grid produces excessive numerical errors when $v_{\text{P}}/v_{\text{S}}$ is larger than about 5, even when n_{min} is on the order of 10. This renders the conventional grid impractical for the simulation of wave propagation through sedimentary basins where $v_{\text{P}}/v_{\text{S}}$ can reach values of 5 and larger.
Numerous techniques aiming at the reduction of n_{min} can be found in the finite-difference literature. The most intuitive approach is to increase the length of the finite-difference operator. Based on a qualitative error analysis, Dablain (1986) suggested that 3 grid points per minimum wavelength are sufficient when a 10th-order finite-difference approximation is used for the 2D scalar wave equation. Holberg (1987) optimised the finite-difference coefficients such that the group velocity error is minimal. For a pre-defined maximum group velocity error, he found n_{min} to decrease roughly exponentially with increasing operator length. In the case of a 30-point operator, $n_{\text{min}} \approx 2.5$ is theoretically sufficient to reduce the group velocity error below 3 %. Furumura & Chen (2004) and Kennett & Furumura (2008) used a 16th-order finite-difference approximation to propagate waves in a 2D model across several thousand wavelengths.
In principle, the operator length can be increased indefinitely, so that the finite-difference method approaches the pseudo-spectral method where the discrete derivative operator is global, and where n_{min} is close to its theoretical minimum of 2 (see section 1.5).

The important conclusion to be drawn is that the numerical setup in general, and the number of grid points per wavelength in particular, is highly problem-dependent. The only rule is that the numerical grid should be chosen such that the resulting synthetics can be compared to the data in a meaningful way. This means that the numerical error should be much smaller than the typical differences that we expect between data and synthetics due to undiscovered Earth structure.

Chapter 3
Spectral-element methods

The spectral-element method is a high-order numerical method that allows us to solve the seismic wave equation in three-dimensional heterogeneous Earth models. The method enables adaptation of the mesh to the irregular surface topography and to the variable wavelengths inside the Earth. Moreover, the spectral-element method yields accurate solutions for surface waves without increasing the number of grid points per wavelength, therefore overcoming some of the most severe deficiencies of the finite-difference method.

Originally developed in fluid dynamics (Patera, 1984; Maday & Patera, 1989), the spectral-element method was first applied to the elastic wave equation and in a seismological context by Seriani et al. (1995), Faccioli et al. (1997) and Komatitsch (1997). Numerical solutions with high accuracy have been obtained in a large number of studies (e.g. Komatitsch & Vilotte, 1998; Seriani, 1998; Komatitsch et al., 2004; Fichtner et al. 2009). Applications to global wave propagation in the presence of self-gravitation, rotation and fluid regions can be found in Komatitsch & Tromp (2002), Chaljub et al. (2003) and Chaljub & Valette (2004). For excellent reviews of the spectral-element method the reader is referred to Komatitsch et al. (2005) and Chaljub et al. (2007).

We start our development with an illustration of the basic spectral-element concepts in one dimension. The necessary mathematical tools, including Lagrange interpolation and Gauss-Lobatto-Legendre quadrature can be found in Appendix A. The extension to the three-dimensional elastic case, treated in section 3.2, is straightforward but requires a few words on grid generation.

3.1 Basic concepts in one dimension

We follow the classical approach and introduce the basic concepts of the spectral-element method with an example in one dimension. For this we first introduce the *weak form* of the equations of motion and then continue with a description of the *Galerkin method*.

3.1.1 Weak solution of the wave equation

We consider the one-dimensional scalar wave equation

$$\rho(x)\,\ddot{u}(x,t) - \partial_x[\mu(x)\,\partial_x u(x,t)] = f(x,t)\,, \tag{3.1}$$

with the space variable $x \in G = [0,L]$ and time $t \in [0,T]$. The displacement field u is subject to the free-surface or Neumann boundary conditions

$$\partial_x u(x,t)|_{x=0} = \partial_x u(x,t)|_{x=L} = 0\,, \tag{3.2}$$

and the initial conditions

$$u|_{t=0} = \dot{u}|_{t=0} = 0\,. \tag{3.3}$$

Equation (3.1) together with the boundary and initial conditions (3.2) and (3.3) is referred to as the *strong form* of the wave equation. To derive the associated *weak* or *variational* form, we multiply (3.1) by an arbitrary, time-independent test function, $w : G \to \mathbb{R}$, and integrate over space:

$$\int_G \rho\, w\, \ddot{u}\, dx - \int_G w\, \partial_x (\mu\, \partial_x u)\, dx = \int_G w\, f\, dx\,. \tag{3.4}$$

Integrating the second term on the left-hand side by parts and inserting the boundary condition (3.2), gives

$$\int_G \rho\, w\, \ddot{u}\, dx + \int_G \mu\, \partial_x w\, \partial_x u\, dx = \int_G w\, f\, dx\,. \tag{3.5}$$

Solving the weak form of the wave equation now means, finding a wave field, u, such that it satisfies (3.5) *for any* suitable test function, w, and subject to the initial conditions

$$\int_G \rho\, w\, u|_{t=0}\, dx = \int_G \rho\, w\, \dot{u}|_{t=0}\, dx = 0\,. \tag{3.6}$$

The weak form of the wave equation has an important advantageous property from a numerical point of view: The free surface boundary condition (3.2) is implicitly satisfied, and it need not be treated explicitly as in finite-difference methods where the accurate implementation of the free surface can be a tedious task.

3.1.2 Spatial discretisation and the Galerkin method

Analytic solutions of both the strong and the weak forms of the wave equation often do not exist when the mass density ρ and the elastic parameter μ are spatially variable. In the *Galerkin method* we approximate the exact solution $u(x,t)$ by a finite superposition of n basis functions ψ_i $(i = 1,...,n)$ that depend only on space and not on time. We denote this approximation by $\bar{u}(x,t)$:

$$u(x,t) \approx \bar{u}(x,t) = \sum_{i=1}^{n} u_i(t)\, \psi_i(x)\,, \tag{3.7}$$

where $u_i(t)$ are the time-dependent expansion coefficients. The quality of this approximation depends on the choice of the basis functions, ψ_i, the source term f and on the medium properties ρ and μ. Instead of trying to solve the exact weak formulation, we limit ourselves to the requirement that $\bar{u}$ solves the approximate weak form

$$\int_G \rho\, \psi_i\, \ddot{\bar{u}}\, dx + \int_G \mu\, \partial_x \psi_i\, \partial_x \bar{u}\, dx = \int_G \psi_i\, f\, dx\,, \tag{3.8}$$

for all basis functions ψ_i, with $i = 1,...,n$, and subject to the initial conditions

$$\int_G \rho\, \psi_i\, \bar{u}|_{t=0}\, dx = \int_G \rho\, \psi_i\, \dot{\bar{u}}|_{t=0}\, dx = 0\,. \tag{3.9}$$

The basis functions, ψ_i, are thus used in the approximation of the wave field and as test functions in the weak formulation. Equation (3.7) together with the approximate weak formulation transforms the exact weak formulation into the following set of linear equations for the coefficients $u_i(t)$:

$$\sum_{i=1}^{n} \left[\ddot{u}_i(t) \int\limits_G \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx \right] + \sum_{i=1}^{n} \left[u_i(t) \int\limits_G \mu(x)\, \partial_x \psi_j(x)\, \partial_x \psi_i(x)\, dx \right] = \int\limits_G \psi_j(x)\, f(x,t)\, dx\,, \tag{3.10}$$

for all $j = 1,...,n$. Equation (3.10) is an algebro-differential equation that can conveniently be written in matrix notation:

$$\mathbf{M} \cdot \ddot{\mathbf{u}}(t) + \mathbf{K} \cdot \mathbf{u}(t) = \mathbf{f}(t)\,, \tag{3.11}$$

with the *mass matrix*

$$M_{ji} = \int\limits_G \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx\,, \tag{3.12}$$

the *stiffness matrix*

$$K_{ji} = \int\limits_G \mu(x)\, \partial_x \psi_j(x)\, \partial_x \psi_i(x)\, dx\,, \tag{3.13}$$

and the right-hand side

$$f_j(t) = \int\limits_G \psi_j(x)\, f(x,t)\, dx\,. \tag{3.14}$$

The vector **u** - not to be confused with the vectorial displacement field in the complete elastic wave equation - comprises the expansion coefficients u_i. The process of transforming the differential equation (3.1) plus the approximation (3.7) into the algebro-differential equation (3.11) is know as *Galerkin projection*. Once more we emphasise that the free surface boundary condition (3.2) is naturally contained in the stiffness matrix (equation 3.13) and does not require any additional work - as is needed in finite-difference methods. What distinguishes the spectral-element method among other numerical methods is the choice of the basis functions ψ_i and the integration scheme used to solve the integrals that appear in the mass and stiffness matrices.

In the next step we decompose the domain G into n_e disjoint subdomains G_e, called the *elements*. Equation (3.10) then transforms to

$$\sum_{i=1}^{n} \left[\ddot{u}_i(t) \sum_{e=1}^{n_e} \int\limits_{G_e} \rho(x)\, \psi_j(x)\, \psi_i(x)\, dx \right] + \sum_{i=1}^{n} \left[u_i(t) \sum_{e=1}^{n_e} \int\limits_{G_e} \mu(x)\, \partial_x \psi_j(x)\, \partial_x \psi_i(x)\, dx \right] = \sum_{e=1}^{n_e} \int\limits_{G_e} \psi_j(x)\, f(x,t)\, dx\,. \tag{3.15}$$

The disadvantage of equation (3.15) is that each expansion coefficient u_i depends on the integrals over all elements. We can circumvent this problem by choosing local basis functions, i.e., basis functions that are supported by one element only. The discrete equations can then be solved for each element individually. For this we define $N+1$ basis functions ψ_i^e $(i = 1, ..., N+1)$ on each of the n_e elements G_e. The displacement field within the element G_e is then approximated by

$$\bar{u}(x,t)|_{x \in G_e} = \sum_{i=1}^{N+1} u_i^e(t)\, \psi_i^e(x)\,, \tag{3.16}$$

so that equation (3.10) now holds for each element:

$$\sum_{i=1}^{N+1} \ddot{u}_i^e(t) \int\limits_{G_e} \rho(x)\, \psi_j^e(x)\, \psi_i^e(x)\, dx + \sum_{i=1}^{N+1} u_i^e(t) \int\limits_{G_e} \mu(x)\, \partial_x \psi_j^e(x)\, \partial_x \psi_i^e(x)\, dx = \int\limits_{G_e} \psi_j^e(x)\, f(x,t)\, dx\,, \tag{3.17}$$

Using matrix notation we write equation (3.17) in a more compact form:

$$\mathbf{M}^e \cdot \ddot{\mathbf{u}}^e(t) + \mathbf{K}^e \cdot \mathbf{u}^e(t) = \mathbf{f}^e(t)\,, \quad e = 1, ..., n_e\,, \tag{3.18}$$

where $\mathbf{u}^e$, $\mathbf{M}^e$ and $\mathbf{K}^e$ are the local coefficient vector, the local mass matrix and the local stiffness matrix. The total number of basis functions is now $n_e\,(N+1)$. Since the basis functions are locally supported by one element, the continuity of the discrete displacement $\bar{u}$ at the boundaries between the elements has to be imposed explicitly. To ensure that the wavelengths are sampled nearly uniformly, the size of the elements will usually be chosen proportional to the wave speed $\sqrt{\mu/\rho}$. The integrals in (3.17) can all be treated in the same way if we map each element G_e onto the *standard* or *reference interval* $[-1,1]$ via an element-specific transformation, F_e:

$$F_e : [-1,1] \rightarrow G_e\,, \quad x = F_e(\xi)\,, \quad \xi = \xi(x) = F_e^{-1}(x)\,, \quad e = 1, ..., n_e\,. \tag{3.19}$$

This transformation is illustrated in figure 3.1. Introducing the transformation into equation (3.17) gives

$$\sum_{i=1}^{N+1} \ddot{u}_i^e(t) \int_{-1}^{1} \rho[x(\xi)]\, \psi_j^e[x(\xi)]\, \psi_i^e[x(\xi)]\, \frac{dx}{d\xi}\, d\xi + \sum_{i=1}^{N+1} u_i^e(t) \int_{-1}^{1} \mu[x(\xi)]\, \frac{d\,\psi_j^e[x(\xi)]}{d\xi}\, \frac{d\,\psi_i^e[x(\xi)]}{d\xi} \left(\frac{d\xi}{dx}\right)^2 \frac{dx}{d\xi}\, d\xi$$

$$= \int_{-1}^{1} \psi_j^e[x(\xi)]\, f[(x(\xi)),t]\, \frac{dx}{d\xi}\, d\xi\,. \qquad (3.20)$$

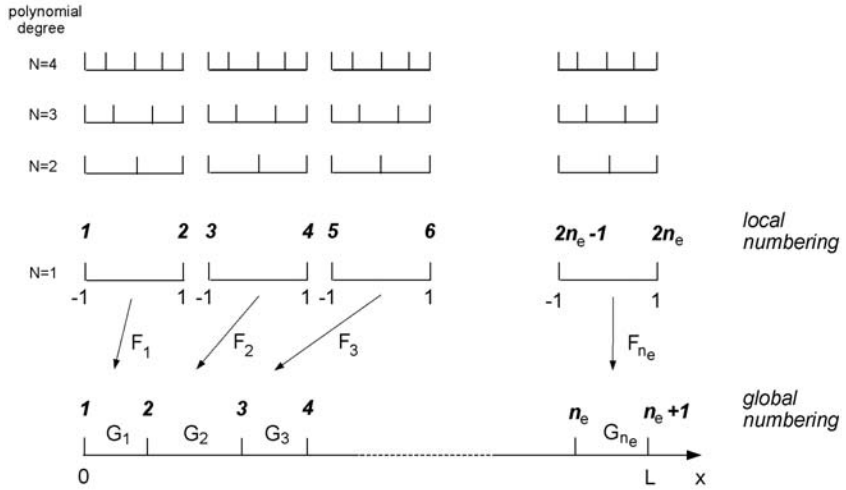

Fig. 3.1 Illustration of the element transformations, F_e, and the concepts of global and local numbering: The spatial domain $[0,L]$, shown in the lowermost part of the figure, is subdivided into n_e non-overlapping elements. Each element G_e is related to the reference interval $[-1,1]$ via an element-specific transformation F_e. The local numbering is used to address the node points of the individual elements, some of which are shared between two elements. The global numbering identifies the spatial locations of the node points. This illustration is for the simplest, though unrealistic, case where the elements have no internal node points (polynomial degree 1). Subdivisions of the reference interval for higher polynomial degrees are shown schematically above.

At this point of the development we specify the basis functions ψ_i. We choose the $N+1$ Lagrange polynomials of degree N that have the corresponding Gauss-Lobatto-Legendre points (GLL points) as collocation points:

$$\psi_i^e[x(\xi)] = \ell_i^{(N)}(\xi)\,, \quad \xi \in [-1,1]\,. \qquad (3.21)$$

This choice is motivated by a number of important results from numerical analysis. We summarise them here and give brief derivations in Appendix A:

1) Using the GLL points for polynomial interpolation ensures that the absolute value of the Lagrange polynomials $\ell_i^{(N)}$ is smaller than or equal to 1, for any polynomial order (see section A.2.3). This means that Runge's phenomenon can be suppressed.
2) The GLL points are Fekete points, i.e., they maximise the Vandermonde determinant (see Appendix A.2.1 and A.2.4). Thus, numerical errors right at the collocation points will have the smallest possible effect on the interpolated values between the collocation points.
3) The Lebesque constant associated with the GLL points grows slowly − in practice logarithmically − with increasing polynomial order (see Appendix A.2.5). This implies that the interpolation error decreases much more rapidly with increasing polynomial order than in the case of equidistant collocation points − at least when the interpolated function is well-behaved.
4) The GLL points are the collocation points of the GLL quadrature (section A.3.2). One can therefore apply the GLL quadrature formulas to obtain accurate approximations of the integrals in equation (3.20) and a *diagonal mass matrix*.

In the interest of a lighter notation we will henceforth omit the superscript $^{(N)}$ in $\ell_i^{(N)}$. Substituting $\ell_i(\xi)$ for $\psi_i[x(\xi)]$ in equation (3.20) gives

$$\sum_{i=1}^{N+1} \ddot{u}_i^e(t) \int_{-1}^{1} \rho'(\xi)\ell_j(\xi)\ell_i(\xi)\,\frac{dx}{d\xi}\,d\xi + \sum_{i=1}^{N+1} u_i^e(t) \int_{-1}^{1} \mu'(\xi)\,\dot{\ell}_j(\xi)\dot{\ell}_i(\xi)\left(\frac{d\xi}{dx}\right)^2 \frac{dx}{d\xi}\,d\xi = \int_{-1}^{1} \ell_j(\xi)\,f'(\xi,t)\,\frac{dx}{d\xi}\,d\xi\,, \tag{3.22}$$

where $\dot{\ell}$ denotes the derivative of ℓ with respect to ξ. The transformed density, ρ', elastic modulus, μ', and external force, f', are defined by

$$\rho'(\xi) := \rho[x(\xi)]\,, \quad \mu'(\xi) := \mu[x(\xi)]\,, \quad f'(\xi) := f[x(\xi)]\,. \tag{3.23}$$

With the GLL quadrature formula (A.74) we can approximate the integral in equation (3.22):

$$\begin{aligned}\sum_{i,k=1}^{N+1} \ddot{u}_i^e(t)\,w_k\rho'(\xi)\,\ell_j(\xi)\ell_i(\xi)\,\frac{dx}{d\xi}\bigg|_{\xi=\xi_k} &+ \sum_{i,k=1}^{N+1} w_k u_i^e(t)\mu'(\xi)\,\dot{\ell}_j(\xi)\dot{\ell}_i(\xi)\left(\frac{d\xi}{dx}\right)^2\frac{dx}{d\xi}\bigg|_{\xi=\xi_k} \\ &\approx \sum_{k=1}^{N+1} w_k\ell_j(\xi)\,f'(\xi,t)\,\frac{dx}{d\xi}\bigg|_{\xi=\xi_k}\,.\end{aligned} \tag{3.24}$$

The symbols ξ_k and w_k are the GLL points and their corresponding integration weights (see Appendix A). The numerical integration in (3.24) is not exact because the integrands are not polynomials of degree $2N-1$ or lower. In the following developments we will nevertheless replace $\approx$ by $=$, keeping in mind that this is an approximation. Recalling the cardinal interpolation property of the Lagrange polynomials, $\ell_i(\xi_k) = \delta_{ik}$, we can simplify (3.24):

$$\sum_{i=1}^{N+1} M_{ji}^e \ddot{u}_i^e(t) + \sum_{i=1}^{N+1} K_{ji}^e u_i^e(t) = f_j^e(t)\,, \quad e = 1,\ldots,n_e\,, \tag{3.25}$$

with

$$M_{ji}^e = w_j\rho'(\xi)\,\frac{dx}{d\xi}\,\delta_{ij}\bigg|_{\xi=\xi_j}\,, \tag{3.26}$$

$$K_{ji}^e = \sum_{k=1}^{N+1} w_k\mu'(\xi)\,\dot{\ell}_j(\xi)\dot{\ell}_i(\xi)\left(\frac{d\xi}{dx}\right)^2\frac{dx}{d\xi}\bigg|_{\xi=\xi_k}\,, \tag{3.27}$$

$$f_j^e(t) = w_j\,f'(\xi,t)\,\frac{dx}{d\xi}\bigg|_{\xi=\xi_j}\,. \tag{3.28}$$

Certainly the most advantageous property of the spectral-element discretisation is the diagonality of the local mass matrix $\mathbf{M}^e$. Since the inversion of M_{ji}^e is trivial and computationally inexpensive, we easily obtain an explicit formula for the second time derivative $\ddot{u}_i(t)$ that can then be discretised.

The numerical integration of equation (3.20) has so far been purely local, resulting in linear systems for each element (equation 3.25). Related to the individual treatment of the elements is a *local numbering* that allows us to address the GLL points of an element. Points shared between two neighbouring elements are counted twice because they appear in the discretised versions of two integrals. To ensure the continuity of the approximation $\bar{u}$ across the element boundaries we need to assemble a global system of equations. This is done by introducing a *global numbering* where each GLL point is counted once even when it is shared between elements. Associated to the global numbering is the global displacement vector $\mathbf{u}^{\text{global}}$. Figure 3.1 illustrates the concept of local and global numbering. The operation of assembling the local mass and stiffness matrices, $\mathbf{M}^e$ and $\mathbf{K}^e$, into their global versions, $\mathbf{M}^{\text{global}}$ and $\mathbf{K}^{\text{global}}$, then simply consists in summing the entries of the local matrices at coincident node points. This leads to a global system of equations,

$$\mathbf{M}^{\text{global}} \cdot \ddot{\mathbf{u}}^{\text{global}}(t) + \mathbf{K}^{\text{global}} \cdot \mathbf{u}^{\text{global}}(t) = \mathbf{f}^{\text{global}}(t)\,, \tag{3.29}$$

that we need to solve for $\mathbf{u}^{\mathrm{global}}$. We note that in practice the stiffness matrix is rarely computed explicitly because only its product with the vector of expansion coefficients is needed.

3.2 Extension to the three-dimensional case

3.2.1 Mesh generation

The spectral-element method in more than one dimension starts with the subdivision of the computational domain G into n_e non-overlapping elements G_e such that $G = \bigcup_{i=1}^{n_e} G_e$. This process is referred to as *mesh generation* or *meshing*. The design of the elements should be such that they follow major geologic features including sedimentary basins and faults. To ensure a spatially uniform sampling of the seismic waves in different parts of the model, the size of the elements should be proportional to the seismic velocity. Moreover, the faces of the elements should align with structural discontinuities because abrupt changes of the structure and the wave field can not be represented accurately by the smooth polynomials inside the elements. A mesh is said to *honour discontinuities* when the boundaries of the elements coincide with the discontinuities. A schematic meshing of a two-dimensional domain is shown in figure 3.2.

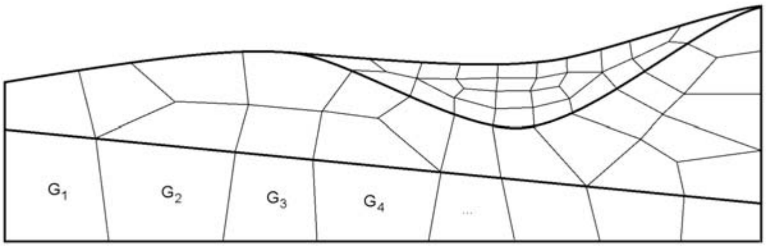

Fig. 3.2 Schematic representation of a mesh in a two-dimensional domain. Bold lines indicate structural discontinuities that are honoured by the elements. The seismic velocities in the top layer are small compared to the velocities in the bottom layer where the elements are largest.

The 3D mesh in figure (3.3) was used to analyse near-fault and strong-motion site effects in the Grenoble valley, France (Stupazzini, 2006; Stupazzini et al., 2009). The study region is characterised by strong topographic variations and seismic wave speeds that range between $v_{\mathrm{P}} \approx 1.5$ km/s and $v_{\mathrm{S}} \approx 0.3$ km/s at the surface of the alluvial basin (green) and $v_{\mathrm{P}} \approx 5.9$ km/s and $v_{\mathrm{S}} \approx 3.4$ km/s in the deep bedrock layers (yellow). This implies that the wavelengths within the model vary by nearly one order of magnitude. The size of the elements in the seismically slow alluvial basin is as small as 20 m and it reaches almost 900 m near the bottom of the model. The total number of elements is 216,972, and it allows for the propagation of frequencies up to 3 Hz when the polynomial degree is 4.
Today, powerful mesh generation tools, known as meshers, are available (e.g. CUBIT, developed by Sandia National Laboratories or GiD from the International Center for Numerical Methods in Engineering). They greatly simplify the mesh design for complex 3D models.

In order to apply the same numerical quadrature to all the elements G_e, they need to be mapped onto the unit cube $\Lambda = [-1,1]^3$. Mostly, this transformation will not be given analytically, especially in the case of complex geologic structures. It needs to be approximated instead, and this approximation is based on the concept of *shape functions* and *anchor nodes*: Each element G_e in the physical space is defined by a set of n_a anchor nodes $\mathbf{x}^a$ and their corresponding shape functions N^a. The 8 corners of an element are always used as anchor nodes. For elements with straight edges, 8 anchor nodes are sufficient to accurately represent the geometry. Additional anchor nodes on the edges or faces may be needed to represent elements with curved edges. A position vector $\mathbf{x}$ in G_e is related to a position vector ξ in the reference cube Λ via the transformation

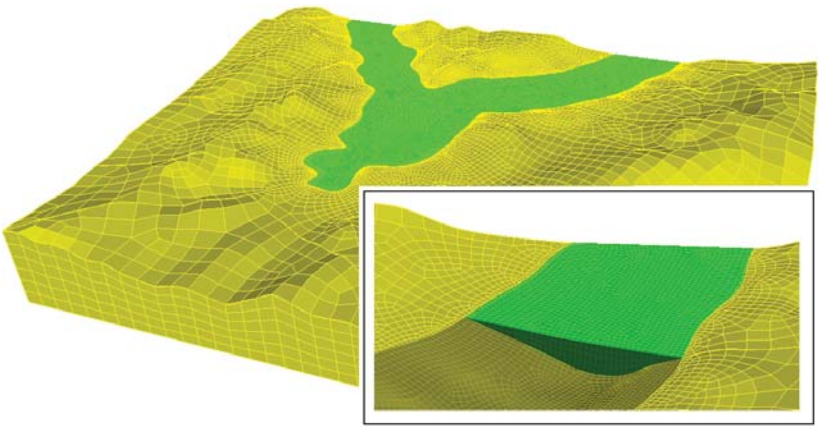

Fig. 3.3 Hexahedral mesh of the Grenoble valley used in a seismic ground motion study (Stupazzini et al., 2009). The wave speeds vary by almost one order of magnitude between the seismically slow alluvial basin (green) and the seismically fast bedrock (yellow). This variability translates to element sizes that range between 20 m at the surface of the basin to 900 m near the bottom of the model. Note that the seismic discontinuity between basin material and bedrock is honoured by the elements.

$$\mathbf{x}(\xi) = \mathbf{F}_e(\xi) = \sum_{a=1}^{n_a} N^a(\xi)\,\mathbf{x}^a\,, \tag{3.30}$$

where the shape functions satisfy the condition

$$N^a(\xi^b) = \delta_{ab}\,. \tag{3.31}$$

Equations (3.30) and (3.31) uniquely relate the anchor node $\mathbf{x}^a$ in the physical element G_e to an anchor node ξ^a in the reference cube Λ:

$$\mathbf{x}(\xi^a) = \mathbf{x}^a\,. \tag{3.32}$$

The shape functions N^a are most conveniently defined as products of 3 Lagrange polynomials, the collocation points of which are the coordinates ξ_i^a of the anchor nodes ξ^a in the reference cube:

$$N^a(\xi) = \ell_{1,a}(\xi_1)\ell_{2,a}(\xi_2)\ell_{3,a}(\xi_3)\,, \tag{3.33}$$

with

$$\ell_{i,a}(\xi_i) = \begin{cases} 1\,, & \text{for} \quad \xi_i = \xi_i^a\,, \\ 0\,, & \text{otherwise}\,, \end{cases} \qquad i = 1,2,3\,.$$

The degree of the Lagrange polynomials depends on the complexity of the element G_e. It is 1 for elements with straight edges and 2 for elements with curved edges. In most applications, the degree of the shape functions is lower than the degree of the polynomials used to interpolate the wave field inside the elements. The transformation F_e is therefore called *subparametric*.

Mesh generation is to some degree an art that can not be fully automated. The principal difficulty is to design a mesh that honours the complexities of the structural model while producing elements that are as large as possible. Very small elements can of course honour any complexity but they result in prohibitively short time steps because the CFL condition must always be satisfied in the numerical simulations. The choice of tensorised Lagrange polynomials as basis functions in 3D spectral elements furthermore imposes that only hexahedra can be used for the mesh even when tetrahedra may sometimes be better suited from a purely geometric point of view.

Generalisations of the spectral-element method in 2D allow for the combination of quadrangles and triangles within the same mesh (e.g. Komatitsch et al., 2001; Mercerat et al., 2006). In 3D this approach can be adapted to combine hexahedra and tetrahedra in order to allow for more geometrical flexibility.

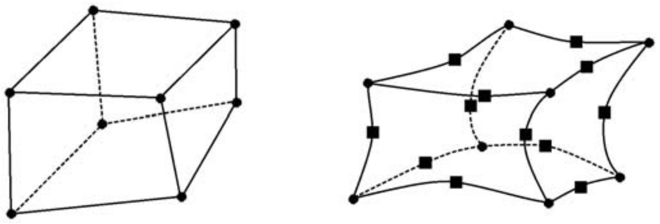

Fig. 3.4 **Left**: The element with straight edges is defined by 8 anchor nodes and linear shape functions. **Right**: The element with curved edges is defined by 8 anchor nodes in the corners plus 12 anchor nodes on the edges. The corresponding shape functions are of degree 2. Additional anchor nodes can be defined in the centre and on the faces, leading to an element with 27 anchor nodes.

Mesh generation has received much attention in recent years. In practice, however, it can be advantageous to take the proper design of a mesh not too seriously. It is a matter of fact that the precise location and sharpness of structural discontinuities inside the Earth are uncertain because our data have finite and often insufficient resolution. The implementation of a strict discontinuity in the spectral-element method can therefore lead to strong interface waves that are not observable in the real Earth where the discontinuity may not be as sharp as in our simplified models.

3.2.2 Weak solution of the elastic wave equation

As a preparatory step towards the spatial discretisation, we first derive the weak form of the elastic wave equation. We base our development on the strong displacement-stress variant of the equations of motion:

$$\rho(\mathbf{x})\,\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla\cdot\sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t) \tag{3.34}$$

$$\sigma(\mathbf{x},t) = \mathbf{C}(\mathbf{x}) : \nabla\mathbf{u}(\mathbf{x},t)\,, \tag{3.35}$$

subject to the boundary and initial conditions

$$\mathbf{n}\cdot\sigma|_{\mathbf{x}\in\partial G} = \mathbf{0}\,, \quad \mathbf{u}|_{t=0} = \dot{\mathbf{u}}|_{t=0} = \mathbf{0}\,. \tag{3.36}$$

For the moment we disregard dissipation, i.e. the time-dependence of the elastic tensor **C**. Multiplying equation (3.34) by an arbitrary, differentiable, time-independent test function, **w**, and integrating over G, gives

$$\int_G \rho\,\mathbf{w}\cdot\ddot{\mathbf{u}}\,d^3\mathbf{x} - \int_G \mathbf{w}\cdot(\nabla\cdot\sigma)\,d^3\mathbf{x} = \int_G \mathbf{w}\cdot\mathbf{f}\,d^3\mathbf{x}\,. \tag{3.37}$$

Invoking the identity

$$\mathbf{w}\cdot(\nabla\cdot\sigma) = \nabla\cdot(\mathbf{w}\cdot\sigma) - \nabla\mathbf{w} : \sigma\,, \tag{3.38}$$

together with Gauss' theorem, yields

$$\int_G \rho\,\mathbf{w}\cdot\ddot{\mathbf{u}}\,d^3\mathbf{x} - \int_{\partial G} \mathbf{w}\cdot\sigma\cdot\mathbf{n}\,d^2\mathbf{x} + \int_G \nabla\mathbf{w} : \sigma\,d^3\mathbf{x} = \int_G \mathbf{w}\cdot\mathbf{f}\,d^3\mathbf{x}\,. \tag{3.39}$$

Upon inserting the free surface boundary condition, equation (3.39) condenses to

$$\int_G \rho\,\mathbf{w}\cdot\ddot{\mathbf{u}}\,d^3\mathbf{x} + \int_G \nabla\mathbf{w} : \sigma\,d^3\mathbf{x} = \int_G \mathbf{w}\cdot\mathbf{f}\,d^3\mathbf{x}\,. \tag{3.40}$$

The same procedure is to be repeated for equation (3.35), but it does not require any additional transformations. Finding a weak solution to the equations of motion means to find a displacement field **u** that satisfies the integral relation (3.40) and

$$\int_G \mathbf{w} \cdot \sigma \, d^3\mathbf{x} = \int_G \mathbf{w} \cdot \mathbf{C} : \nabla \mathbf{u} \, d^3\mathbf{x}\,, \tag{3.41}$$

for any test function, **w**, and subject to the initial conditions

$$\int_G \rho \, \mathbf{w} \cdot \mathbf{u}|_{t=0} \, d^3\mathbf{x} = \int_G \rho \, \mathbf{w} \cdot \dot{\mathbf{u}} \, d^3\mathbf{x} = \mathbf{0}\,. \tag{3.42}$$

As in the one-dimensional case we note that the free surface condition is implicit in the weak formulation.

3.2.3 Discretisation of the equations of motion

In analogy to the Galerkin method for the one-dimensional case (see equation 3.7) we approximate the p-component u_p of the displacement field $\mathbf{u}$ by a superposition of basis functions

$$\psi_{ijk}(\mathbf{x}) = \psi_{ijk}(x_1, x_2, x_3)\,, \tag{3.43}$$

weighted by expansion coefficients u_p^{ijk}:

$$u_p(\mathbf{x},t) \approx \bar{u}_p(\mathbf{x},t) = \sum_{i,j,k=1}^{N+1} u_p^{ijk}(t)\, \psi_{ijk}(\mathbf{x})\,. \tag{3.44}$$

The corresponding approximation of the stress tensor components σ_{pq} is

$$\sigma_{pq}(\mathbf{x},t) \approx \bar{\sigma}_{pq}(\mathbf{x},t) = \sum_{i,j,k=1}^{N+1} \sigma_{pq}^{ijk}(t)\, \psi_{ijk}(\mathbf{x})\,. \tag{3.45}$$

Equations (3.44) and (3.45) already assume that u_p and σ_{pq} are considered inside an element $G_e \subset \mathbb{R}^3$, where they can be represented by $(N+1)^3$ basis functions. We therefore omit the superscript ^{e} that we used in chapter 3.1 to indicate local element-specific quantities.

To find a weak solution in the Galerkin sense, we replace the exact weak formulation from equations (3.40) to (3.42) by the requirement that the approximations $\bar{\mathbf{u}}$ and $\bar{\sigma}$ satisfy

$$\int_{G_e} \rho \, \psi_{ijk} \mathbf{e}_p \cdot \ddot{\bar{\mathbf{u}}} \, d^3\mathbf{x} + \int_{G_e} \nabla(\psi_{ijk} \mathbf{e}_p) : \bar{\sigma} \, d^3\mathbf{x} = \int_{G_e} \psi_{ijk} \mathbf{e}_p \cdot \mathbf{f} \, d^3\mathbf{x}\,, \tag{3.46}$$

and

$$\int_{G_e} \psi_{ijk} \mathbf{e}_p \cdot \bar{\sigma} \, d^3\mathbf{x} = \int_{G_e} \psi_{ijk} \mathbf{e}_p \cdot \mathbf{C} : \nabla \bar{\mathbf{u}} \, d^3\mathbf{x}\,, \tag{3.47}$$

for all basis functions, ψ_{ijk}, and for all unit vectors, $\mathbf{e}_p$, with $p = 1, 2, 3$. The weak initial conditions are

$$\int_{G_e} \rho \, \psi_{ijk} \mathbf{e}_p \cdot \bar{\mathbf{u}}|_{t=0} d^3\mathbf{x} = \int_{G_e} \rho \, \psi_{ijk} \mathbf{e}_p \, \dot{\bar{\mathbf{u}}} \, d^3\mathbf{x} = \mathbf{0}\,, \tag{3.48}$$

again for all basis functions and unit vectors. To keep the following formulas as readable as possible, we treat the summands in equations (3.46) and (3.47) individually. For the first term on the left-hand side of (3.46) we find

$$\mathbb{F}_{qrs}[\rho \ddot{u}_p] := \int_{G_e} \rho \, \psi_{qrs} \mathbf{e}_p \cdot \ddot{\bar{\mathbf{u}}} \, d^3\mathbf{x} = \sum_{i,j,k=1}^{N+1} \int_{G_e} \rho(\mathbf{x}) \ddot{u}_p^{ijk}(t)\, \psi_{ijk}(\mathbf{x})\, \psi_{qrs}(\mathbf{x}) \, d^3\mathbf{x}\,. \tag{3.49}$$

The symbol $\mathbb{F}$ is intended to express that $\mathbb{F}_{qrs}[\rho \ddot{u}_p]$ is a discrete local force, averaged over the element G_e. In the next step we relate the elements G_e ($e = 1, \ldots, n_e$) to the reference cube $\Lambda = [-1,\, 1]^3$ via invertible and element-specific transformations $\mathbf{F}_e$ that we already made explicit in equations (3.30) to (3.33):

$$\mathbf{F}_e : [-1,1]^3 = \Lambda \to G_e\,,\, \mathbf{x} = \mathbf{F}_e(\xi)\,,\, \xi = \xi(\mathbf{x}) = \mathbf{F}_e^{-1}(\mathbf{x})\,, \quad e = 1,...,n_e\,, \tag{3.50}$$

The action of the transformations $\mathbf{F}_e$ is illustrated in figure 3.5.

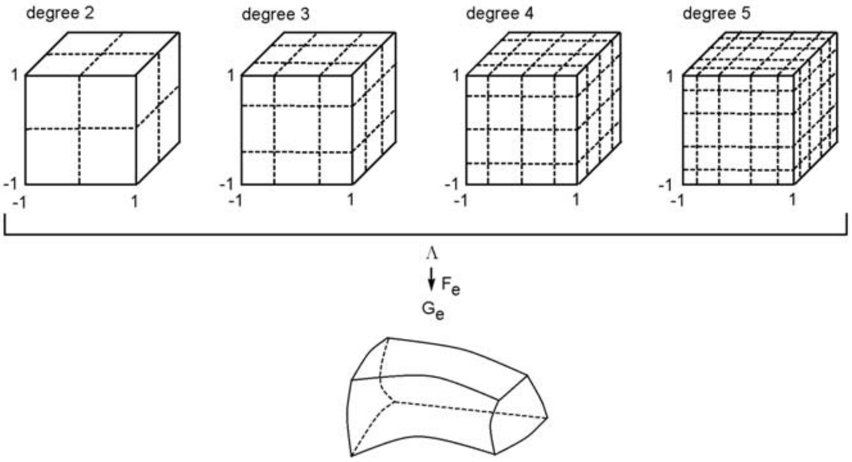

Fig. 3.5 Mapping of the deformed element G_e from the physical space to the reference cube $\Lambda = [-1,1]^3$. Dashed lines indicate the GLL nodal lines for different polynomial degrees.

Transforming the integral (3.49) according to the transformation (3.50) yields

$$\mathbb{F}_{qrs}[\rho \ddot{u}_p] = \sum_{i,j,k=1}^{N+1} \int_{\Lambda} \rho[\mathbf{x}(\xi)]\, \ddot{u}_p^{ijk}(t)\, \psi_{ijk}[\mathbf{x}(\xi)]\, \psi_{qrs}[\mathbf{x}(\xi)]\, J(\xi)\, d^3\xi\,, \tag{3.51}$$

where the symbol J in equation (3.51) denotes the Jacobian of $\mathbf{F}_e$. It is understood that the elements are chosen such that $J > 0$. For the basis functions $\psi_{ijk}[\mathbf{x}(\xi)]$ we choose the product of three Lagrange polynomials collocated at the GLL points:

$$\psi_{ijk}[\mathbf{x}(\xi)] = \ell_i(\xi_1)\ell_j(\xi_2)\ell_k(\xi_3)\,. \tag{3.52}$$

The resulting expression for $\mathbb{F}_{qrs}[\rho \ddot{u}_p]$ is then:

$$\mathbb{F}_{qrs}[\rho \ddot{u}_p] = \sum_{i,j,k=1}^{N+1} \int_{\Lambda} \rho'(\xi) \ddot{u}_p^{ijk}(t) \ell_i(\xi_1)\ell_j(\xi_2)\ell_k(\xi_3)\ell_q(\xi_1)\ell_r(\xi_2)\ell_s(\xi_3)\, J(\xi)\, d^3\xi\,, \tag{3.53}$$

with the transformed density defined by

$$\rho'(\xi) := \rho[\mathbf{x}(\xi)]\,. \tag{3.54}$$

Applying the GLL quadrature rule to equation (3.53) yields the following simple expression:

$$\mathbb{F}_{qrs}[\rho \ddot{u}_p] = w_q w_r w_s\, \rho'(\xi^{qrs}) \ddot{u}_p^{qrs}\, J(\xi^{qrs})\,, \tag{3.55}$$

where we need to evaluate the transformed density and the Jacobian at the GLL points

$$\xi^{qrs} := (\xi_1^q, \xi_2^r, \xi_3^s)\,. \tag{3.56}$$

Equation (3.56) is the three-dimensional equivalent of equation (3.26) that specifies the mass matrix for the one-dimensional case. The combination of GLL quadrature and Lagrange polynomials collocated at the GLL points again

ensures that the mass matrix in three dimensions is diagonal, meaning that the evaluation of $\mathbb{F}_{qrs}[\rho \ddot{u}_p]$ does not involve any expansion coefficients other than u_p^{qrs}.
We now consider the second term on the left-hand side of equation (3.46):

$$\mathbb{F}_{qrs}[(\nabla \cdot \sigma)_p] := \int_{G_e} \nabla(\psi_{qrs}\mathbf{e}_p) : \bar{\sigma}\, d^3\mathbf{x}. \tag{3.57}$$

Transforming to the reference cube Λ then gives

$$\mathbb{F}_{qrs}[(\nabla \cdot \sigma)_p] = \sum_{n,m=1}^{3} \int_{\Lambda} \frac{\partial \xi_m}{\partial x_n} \frac{\partial}{\partial \xi_m} [\ell_q(\xi_1)\ell_r(\xi_2)\ell_s(\xi_3)]\, \bar{\sigma}'_{np}(\xi)\, J(\xi)\, d^3\xi\,, \tag{3.58}$$

where we already substituted the Lagrange polynomials for the general basis functions, as proposed in equation (3.52). The transformed stress tensor, $\bar{\sigma}'$ is defined as

$$\bar{\sigma}'(\xi) := \bar{\sigma}[\mathbf{x}(\xi)]\,. \tag{3.59}$$

Approximating the integral in equation (3.58) via the GLL quadrature rule, results in a rather lengthy expression:

$$\begin{aligned}\mathbb{F}_{qrs}[(\nabla \cdot \sigma)_p] =& \sum_{n=1}^{3} \sum_{i=1}^{N+1} w_i w_r w_s \dot{\ell}_q(\xi_1^i)\, \sigma_{np}^{irs}\, J(\xi^{irs})\, \frac{\partial \xi_1}{\partial x_n}(\xi^{irs}) + \sum_{n=1}^{3} \sum_{i=1}^{N+1} w_q w_i w_s \dot{\ell}_r(\xi_2^i)\, \sigma_{np}^{qis}\, J(\xi^{qis})\, \frac{\partial \xi_2}{\partial x_n}(\xi^{qis}) \\ &+ \sum_{n=1}^{3} \sum_{i=1}^{N+1} w_q w_r w_i \dot{\ell}_s(\xi_3^i)\, \sigma_{np}^{qri}\, J(\xi^{qri})\, \frac{\partial \xi_3}{\partial x_n}(\xi^{qri})\,,\end{aligned} \tag{3.60}$$

Equation (3.60) involves a sum over $9(N+1)$ terms which makes it computationally much more expensive than (3.55). Repeating the above procedure for the source term $\mathbf{f}$ in equation (3.34) gives

$$\mathbb{F}_{qrs}(f_p) := \int_{G_e} \psi_{qrs}\mathbf{e}_p \cdot \mathbf{f}\, d^3\mathbf{x} = w_q w_r w_s\, f'_p(\xi^{qrs})\, J(\xi^{qrs})\,, \tag{3.61}$$

with the transformed force density

$$f'_p(\xi) := f_p[\mathbf{x}(\xi)]\,. \tag{3.62}$$

Equation (3.61) deserves special attention because its validity is not as obvious as it may initially appear. In fact, we must recall that $\int_{G_e} \psi_{qrs}\mathbf{e}_p \cdot \mathbf{f}\, d^3\mathbf{x} = w_q w_r w_s\, f'_p(\xi^{qrs})\, J(\xi^{qrs})$ is an approximation. Its accuracy is controlled by the quality of the GLL quadrature that we used to evaluate the projection integral.
It remains to consider the approximate weak form of the constitutive relation as specified by equation (3.47). For the Galerkin projection of the (mn)-component of the stress tensor σ we find

$$\mathbb{F}_{qrs}(\sigma_{mn}) := \int_{G_e} (\psi_{qrs}\mathbf{e}_m \cdot \bar{\sigma})_n\, d^3\mathbf{x} = w_q w_r w_s\, \sigma_{mn}^{qrs}\, J(\xi^{qrs})\,. \tag{3.63}$$

Projecting the (mn)-component of the right-hand side in the approximate weak constitutive relation (3.47) gives

$$\mathbb{F}_{qrs}[(\mathbf{C} : \nabla \mathbf{u})_{mn}] = \int_{G_e} (\psi_{qrs}\mathbf{e}_m \cdot \mathbf{C} : \nabla \bar{\mathbf{u}})_n\, d^3\mathbf{x} = \sum_{a,b=1}^{3} \sum_{i,j,k=1}^{N+1} \psi_{qrs}(\mathbf{x})\, C_{mnab} \frac{\partial}{\partial x_a}[u_b^{ijk}\, \psi_{ijk}(\mathbf{x})]\, d^3\mathbf{x}. \tag{3.64}$$

Transforming equation (3.64) and inserting the Lagrange polynomials as basis functions yields

$$\mathbb{F}_{qrs}[(\mathbf{C} : \nabla \mathbf{u})_{mn}] = \sum_{a,b,p=1}^{3} \sum_{i,j,k=1}^{N+1} \int_{\Lambda} u_b^{ijk} \ell_q(\xi_1)\ell_r(\xi_2)\ell_s(\xi_3) \cdot C'_{mnab}(\xi)\, \frac{\partial \xi_p}{\partial x_a} \frac{\partial}{\partial \xi_p} [\ell_i(\xi_1)\ell_j(\xi_2)\ell_k(\xi_3)]\, J(\xi)\, d^3\xi\,, \tag{3.65}$$

with

$$\mathbf{C}'(\xi) := \mathbf{C}[\mathbf{x}(\xi)] . \tag{3.66}$$

Then following the usual steps, we obtain the Galerkin projection

$$\mathbb{F}_{qrs}[(\mathbf{C}:\nabla\mathbf{u})_{mn}] = w_q w_r w_s \, J(\xi^{qrs}) \sum_{a,b=1}^{3} C'_{mnab}(\xi^{qrs}) \cdot \sum_{i=1}^{N+1} \left[u_n^{irs} \dot{\ell}_i(\xi_1^q) \frac{\partial \xi_1}{\partial x_m} + u_n^{qis} \dot{\ell}_i(\xi_2^r) \frac{\partial \xi_2}{\partial x_m} + u_n^{qri} \dot{\ell}_i(\xi_3^s) \frac{\partial \xi_3}{\partial x_m} \right] . \tag{3.67}$$

At this point we can assemble the discrete equations of motions:

$$[w_q w_r w_s \, \rho'(\xi^{qrs}) \, J(\xi^{qrs})] \, \ddot{u}_p^{qrs} = \mathbb{F}_{qrs}(f_p) - \mathbb{F}_{qrs}[(\nabla \cdot \sigma)_p] , \tag{3.68}$$

$$[w_q w_r w_s \, J(\xi^{qrs})] \, \sigma_{mn}^{qrs} = \mathbb{F}_{qrs}[(\mathbf{C} : \nabla\mathbf{u})_{mn}] . \tag{3.69}$$

The above equations are of general validity. They are useful when deformed elements are needed in order to mesh a structural discontinuity or topography. Equation (3.69) gives the expansion coefficients σ_{mn}^{qrs} of σ in terms of the expansion coefficients u_n^{qrs} of $\mathbf{u}$. The former can then be used in equation (3.68) to compute the acceleration $\ddot{u}_p^{qrs}$. As a result of using GLL node points together with GLL quadrature we obtain σ_{mn}^{qrs} and $\ddot{u}_p^{qrs}$ without inverting a large matrix. This is, as already mentioned in chapter 3.1, one of the principal advantages of the spectral-element method.
Since the explicit computation of the mass and stiffness matrices is unnecessary in practice, we work directly with equation (3.68) which is a local, i.e. element-wise, discrete force balance. The corresponding global force balance can be constructed as in the one-dimensional case.

3.3 Accuracy and efficiency

The accuracy of spectral-element solutions is controlled by both the size of the elements and the degree of the Lagrange polynomials. Decreasing the size of the elements and increasing the polynomial degree will mostly lead to higher accuracy. Exceptions are those cases where the numerical dispersion resulting from the low-order time-discretisation dominates over the numerical error of the spatial discretisation.
In general, the GLL quadrature is not exact even in the case of a homogeneous medium. This is because the integrands are products of two polynomials of degree N, and GLL quadrature is exact only for polynomials of degree-$(2N-1)$ and lower. This deficiency can become important when the material properties vary strongly inside an element.
To ensure that variable material properties are properly accounted for by the mesh, one may increase the polynomial degree. However, the maximum degree is very limited by the CFL stability condition and the available computational resources. This is because the distance between the first two GLL points decreases as $O(N^{-2})$, i.e. quadratically with increasing degree. Thus, choosing $N \gtrapprox 8$ usually results in unreasonably small step lengths. On the other hand, by choosing $N < 4$ one sacrifices much of the accuracy of the spectral-element method. A compromise based on experience, is to use polynomial degrees between 4 and 7.
Also based on experience is the rule of thumb that one should use at least 5 grid points per minimum wavelength in order to obtain accurate solutions when the propagation distance is on the order of 10 to 50 wavelengths. It is important keep in mind that numerical dispersion is cumulative, meaning that the numerical error grows steadily as the length of the simulation increases. Thus, a larger number of grid points per wavelength may become necessary when the propagation distance is significantly larger than 50 wavelengths.

Part II
The adjoint method and Fréchet kernels

Having solved the forward problem, i.e. the wave equation, it remains to exploit the wealth of waveform information contained in seismograms for the benefit of seismic tomography.

The first step towards this goal is the physically meaningful quantification of the differences between observed waveforms and waveforms that have been computed for an Earth model, $\mathbf{m}$. Mathematically, this waveform difference is expressed in terms of a misfit functional, χ, that generally depends non-linearly on $\mathbf{m}$. Solving the full waveform tomographic problem means to find an optimal Earth model, $\mathbf{m}_{\text{opt}}$, such that $\chi(\mathbf{m}_{\text{opt}})$ is the global minimum of χ. In this sense, full waveform inversion is a special non-linear minimisation problem. Owing to the large number of model parameters (thousands to tens of millions), the minimisation proceeds iteratively with the help of gradient-based algorithms such as the conjugate gradient or Newton methods.

All gradient-based minimisation algorithms critically rely on the derivative of the misfit functional with respect to the model parameters. The adjoint method (chapter 4) allows us to compute the derivative of χ with optimal efficiency: Starting from the wave equation we can derive the adjoint wave equation, the solution of which is the adjoint wave field. The properties of the adjoint wave field are determined by the adjoint source which is completely specified by the misfit, χ. Correlating the regular and adjoint wave fields, yields the derivative of χ with respect to $\mathbf{m}$. A generalisation of the adjoint method allows us to compute the Hessian of χ, which is the carrier of covariance information, and which plays a crucial role in the Newton method of non-linear minimisation.

Physical intuition is essential for the successful solution of any inverse problem. In the special case of full waveform inversion, our intuition is mostly based on the interpretation of Fréchet or sensitivity kernels, which are the volumetric densities of Fréchet derivatives. For isolated seismic phases such as P, S, sP, etc., Fréchet kernels are generally concentrated around the geometric ray path connecting source and receiver. The width of the kernels is proportional to the frequency band width of the observed waveforms (chapter 6).

Chapter 4
The time-domain continuous adjoint method

The adjoint method is a mathematical tool that allows us to compute the gradient of an objective functional with respect to the model parameters very efficiently. In this chapter we derive a general formulation of the adjoint method that is independent of a particular physical problem. We introduce important concepts such as time reversal, adjoint sources, adjoint fields and Fréchet or sensitivity kernels. We conclude this chapter with an application to the elastic wave equation including anisotropy and attenuation.

4.1 Introduction

The derivatives of an observable with respect to the parameters of the theory used for its prediction play a fundamental role in all physical sciences. Derivatives are indispensable in sensitivity analysis, numerous non-linear optimisation methods and in the development of the physical intuition that is crucial for the efficient and meaningful solution of inverse problems.

The adjoint method as a particularly efficient tool for the computation of the partial derivatives of an objective functional seems to have originated in the field of control theory. In control theory one considers an observable $\mathbf{u}$ that is the output of a dynamical system. The behaviour of the system depends on model parameters $\mathbf{m}$ that are linked to $\mathbf{u}$ via a physical theory symbolised by an operator $\mathbf{L}$,

$$\mathbf{L}(\mathbf{u},\mathbf{m}) = \mathbf{f}\,, \tag{4.1}$$

where $\mathbf{f}$ represents external forces. The goal is to choose the model parameters $\mathbf{m}$ as a function of the currently observed output, such that the system operates optimally. In mathematical terms optimality means that a problem-specific objective functional $\chi(\mathbf{m}) = \chi[\mathbf{u}(\mathbf{m})]$ is minimal. A change of χ in response to a change from $\mathbf{m}$ to $\mathbf{m} + \varepsilon\,\delta\mathbf{m}$ is approximated by the derivative of χ with respect to $\mathbf{m}$ in the direction $\delta\mathbf{m}$:

$$\nabla_m \chi(\mathbf{m})\,\delta\mathbf{m} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\left[\chi(\mathbf{m} + \varepsilon\,\delta\mathbf{m}) - \chi(\mathbf{m})\right]. \tag{4.2}$$

In many applications, including seismic tomography, the number of model parameters is large and the solution of the forward problem $\mathbf{L}(\mathbf{u},\mathbf{m}) = \mathbf{f}$ is computationally expensive. This renders a finite difference approximation of (4.2) for all possible directions $\delta\mathbf{m}$ practically impossible. This is where the adjoint method comes into play.

The adjoint state of a hyperbolic differential equation, such as the wave equation, can be found as early as 1968 in the book by J.-L. Lions on the optimal control of systems governed by partial differential equations. It was one of Lions students, G. Chavent, who may have been the first to use the adjoint method for the determination of distributed parameters. His thesis, entitled *Analyse Fonctionelle et Identification de Coefficients Répartis dans les Équations aux Dérivées Partielles* dates from 1971.

Probably one of the first geoscientific applications of control theory in conjunction with the adjoint method was presented by Chavent and his co-workers Dupuy and Lemonnier (1975). In the framework of petroleum engineering they "determined the permeability distribution by matching the history of pressure in a single-phase field, given flow production data." To highlight some of the basic concepts still found in modern applications, we briefly expand on their approach: Chavent, Dupuy and Lemonnier considered a reservoir Ω with a permeable boundary $\partial\Omega$ from which a fluid was extracted at *NW* production wells. The pressure p as a function of time and space was assumed to be governed by a diffusion equation and influenced by a transmissivity coefficient b, a known storage coefficient a and the production rate q_i of the wells. The goal was to find the transmissivity of the reservoir by matching the history of the computed pressure p at the wells to the truly observed pressure history, denoted p^0. As in optimal control theory, this was stated in the form of a minimisation problem involving an objective functional, $\chi(b, p^0, p)$, that quantifies the misfit between data, p^0, and the computed pressure estimate, p, as a function of b. The minimisation of χ with respect to b was achieved using a steepest descent method, and the direction of steepest descent was obtained via the adjoint method. Finite differencing would have been prohibitively expensive even though the discretised transmissivity model merely comprised 171 grid points.

Following a series of theoretical studies on the 1D seismic inverse problem (Bamberger et al., 1977, 1979), Bamberger, Chavent and Lailly (1982) presented what is likely to be the first application of the adjoint method for seismic imaging purposes. The *Inversion of normal incidence seismograms* was indeed set as an optimal control problem.

Since the mid 1980s, the adjoint method has been used in a variety of physical sciences, including meteorology (e.g. Talagrand & Courtier, 1987), ground water modelling (e.g. Sun, 1994) and geodynamics (e.g. Bunge et al., 2003). Examples of seismological applications of the adjoint method may be found in Tarantola (1988), Tromp et al. (2005), Fichtner et al. (2006a,b), Sieminski et al. (2007a,b), Liu & Tromp (2008), Stich et al. (2009), Tape et al. (2009,2010) and Fichtner et al. (2009,2010).

4.2 General formulation

We consider a physical observable $\mathbf{u}$ that depends on the position vector $\mathbf{x} \in G \subset \mathbb{R}^3$, time $t \in T = [t_0, t_1]$ and on model parameters $\mathbf{m} \in \mathfrak{M}$:

$$\mathbf{u} = \mathbf{u}(\mathbf{m}; \mathbf{x}, t)\,. \tag{4.3}$$

The model space $\mathfrak{M}$ contains all admissible parameters $\mathbf{m}$, and the semicolon in equation (4.3) indicates that $\mathbf{u}$ evolves in space and in time, whereas the model parameters are assumed to be fixed for a given realisation of $\mathbf{u}$. In seismology $\mathbf{u}$ represents an elastic wave field that is linked via the wave equation, symbolically written as $\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f}$, to external sources $\mathbf{f}$ and parameters $\mathbf{m}$ such as the mass density ρ and the elastic tensor $\mathbf{C}$ in the Earth (see section 1.2).

It is commonly not $\mathbf{u}$ itself, but a scalar objective functional $\chi(\mathbf{m}) = \chi[\mathbf{u}(\mathbf{m})]$ that we are interested in. Objective functionals serve two closely related purposes: First, they can represent a measurement process that translates a pure physical entity $\mathbf{u}$, such as a seismic wave, to a secondary observable, for instance the energy at a receiver position $\mathbf{x} = \mathbf{x}^r$:

$$\chi(\mathbf{m}) = \int_T \int_G \dot{\mathbf{u}}^2(\mathbf{m}; \mathbf{x}, t)\, \delta(\mathbf{x} - \mathbf{x}^r)\, dt\, d^3\mathbf{x}\,. \tag{4.4}$$

Second, when observed data $\mathbf{u}^0(\mathbf{x}, t)$ are available, χ can be used to quantify the discrepancy between the observation and the theoretical prediction $\mathbf{u}(\mathbf{m}; \mathbf{x}, t)$. The objective functional then plays the role of a misfit functional, that may, for instance, quantify the L_2 distance between observed and predicted seismograms at $\mathbf{x} = \mathbf{x}^r$:

$$\chi(\mathbf{m}) = \frac{1}{2} \int_T \int_G [\mathbf{u}(\mathbf{m}; \mathbf{x}, t) - \mathbf{u}^0(\mathbf{x}, t)]^2\, \delta(\mathbf{x} - \mathbf{x}^r)\, dt\, d^3\mathbf{x}\,. \tag{4.5}$$

While equations (4.4) and (4.5) are special cases of objective functionals, the appearance of an integral over space and time is of surprising generality not only in seismological applications. We therefore write χ in the form

$$\chi(\mathbf{m}) = \int_T \int_G \chi_1[\mathbf{u}(\mathbf{m}; \mathbf{x}, t)]\, dt\, d^3\mathbf{x} = \langle \chi_1(\mathbf{m}) \rangle\,, \tag{4.6}$$

where we introduced $\langle . \rangle$ as a short notation for the integral over $T \times G$. The derivative $\nabla_m \chi\, \delta\mathbf{m}$ of $\chi[\mathbf{u}(\mathbf{m})]$ with respect to $\mathbf{m}$ in a direction $\delta\mathbf{m}$ follows from the chain rule:

$$\nabla_m \chi \, \delta\mathbf{m} = \nabla_u \chi \, \delta\mathbf{u} = \langle \nabla_u \chi_1 \, \delta\mathbf{u} \rangle \tag{4.7}$$

where

$$\delta\mathbf{u} := \nabla_m \mathbf{u} \, \delta\mathbf{m} \tag{4.8}$$

denotes the derivative of **u** with respect to **m** in the direction $\delta\mathbf{m}$. The difficulty of equation (4.7) lies in the appearance of $\delta\mathbf{u}$ which is often hard to evaluate numerically. For a first-order finite-difference approximation of $\nabla_m \chi$ one needs to determine $\mathbf{u}(\mathbf{m} + \varepsilon \, \delta\mathbf{m})$ for each possible direction $\delta\mathbf{m}$. This, however, becomes infeasible in the case of numerically expensive forward problems and large model spaces. Consequently, we may not be able to compute $\nabla_m \chi$ unless we manage to eliminate $\delta\mathbf{u}$ from equation (4.7). For this purpose we differentiate the theoretical relationship $\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f}$ with respect to **m**. Again invoking the chain rule for differentiation gives:

$$\nabla_m \mathbf{L} \, \delta\mathbf{m} + \nabla_u \mathbf{L} \, \delta\mathbf{u} = \mathbf{0} \, . \tag{4.9}$$

The right-hand side of equation (4.9) vanishes because the external sources **f** do not depend on the model parameters **m**. We now multiply (4.9) by an arbitrary test function $\mathbf{u}^\dagger$ and then apply the integral $\langle \, . \, \rangle$:

$$\langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \rangle + \langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \, \delta\mathbf{u} \rangle = 0 \, . \tag{4.10}$$

Adding equations (4.7) and (4.10) gives

$$\nabla_m \chi \, \delta\mathbf{m} = \langle \nabla_u \chi_1 \, \delta\mathbf{u} \rangle + \langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \, \delta\mathbf{u} \rangle + \langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \rangle \, . \tag{4.11}$$

We can rewrite equation (4.11) using the adjoint operators $\nabla_u \chi_1^\dagger$ and $\nabla_u \mathbf{L}^\dagger$ which are defined by

$$\langle \nabla_u \chi_1 \, \delta\mathbf{u} \rangle = \langle \delta\mathbf{u} \cdot \nabla_u \chi_1^\dagger \rangle \tag{4.12}$$

and

$$\langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \, \delta\mathbf{u} \rangle = \langle \delta\mathbf{u} \cdot \nabla_u \mathbf{L}^\dagger \, \mathbf{u}^\dagger \rangle \, , \tag{4.13}$$

for any $\delta\mathbf{u}$ and $\mathbf{u}^\dagger$. We then obtain

$$\nabla_m \chi \, \delta\mathbf{m} = \langle \delta\mathbf{u} \cdot (\nabla_u \chi_1^\dagger + \nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger) \rangle + \langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \rangle \, . \tag{4.14}$$

We may now eliminate $\delta\mathbf{u}$ from equation (4.14) if we can determine a field $\mathbf{u}^\dagger$ to satisfy

$$\nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger = -\nabla_u \chi_1^\dagger \, . \tag{4.15}$$

Equation (4.15) is referred to as the *adjoint equation* of (4.1), and $\mathbf{u}^\dagger$ and $-\nabla_u \chi_1^\dagger$ are the *adjoint field* and the *adjoint source*, respectively. When the solution $\mathbf{u}^\dagger$ of the adjoint equation is found, then the derivative of the objective functional reduces to

$$\nabla_m \chi \, \delta\mathbf{m} = \langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \rangle \, . \tag{4.16}$$

By construction, $\nabla_m \chi \, \delta\mathbf{m}$ can now be computed for any differentiation direction $\delta\mathbf{m}$ without the explicit knowledge of $\delta\mathbf{u}$. This advantage comes at the price of having to find the adjoint operator $\nabla_u \mathbf{L}^\dagger$ and a solution of the adjoint problem (4.15).

Equation (4.15) can be simplified when the operator **L** is linear in **u**. It then follows that $\mathbf{L}^\dagger$ is also linear and therefore

$$\mathbf{L}^\dagger(\mathbf{u}^\dagger) = -\nabla_u \chi_1^\dagger \, . \tag{4.17}$$

Fréchet kernels

Much of our physical intuition is based on the interpretation of sensitivity or Fréchet kernels which are defined as the volumetric densities of the Fréchet derivative $\nabla_m \chi$:

$$K_m := \frac{d}{dV} \nabla_m \chi = \int_T \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, dt \,. \tag{4.18}$$

Using the notion of sensitivity kernels, we can recast equation (4.16) as follows:

$$\nabla_m \chi \, \delta\mathbf{m} = \int_G K_m(\mathbf{x}) \, \delta\mathbf{m}(\mathbf{x}) \, d^3\mathbf{x} \,. \tag{4.19}$$

The sensitivity kernels $K_m(\mathbf{x})$ reveal how the objective functional $\chi(\mathbf{m})$ is affected by model parameter changes at position $\mathbf{x}$ in the Earth. It is the study of $K_m(\mathbf{x})$ for different types of seismic waves and different objective functionals that allows us to design efficient inversion schemes and to interpret the results in a physically meaningful way. A collection of sensitivity kernels for different combinations of objective functionals and Earth model parameters is presented in chapter 6.

Translation to the discretised model space

In most applications, the model space, $\mathfrak{M}$, is discretised, meaning that the components, m_i, of the space-continuous model

$$\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}, m_2(\mathbf{x}), m_3(\mathbf{x}), \ldots] \in \mathfrak{M} \tag{4.20}$$

are expressed as a linear combination of $N < \infty$ basis functions, $b_j(\mathbf{x})$:

$$m_i(\mathbf{x}) = \sum_{j=1}^{N} \mu_{ij} \, b_j(\mathbf{x}) \,. \tag{4.21}$$

Commonly, the basis functions are spherical harmonics, blocks, wavelets or splines. With the representation (4.21), the model, $\mathbf{m}$, and the objective functional, χ, are fully determined by the coefficients or model parameters μ_{ij}. We are therefore interested in the partial derivatives $\partial\chi/\partial\mu_{ij}$. Using the definition of the classical derivative, we find

$$\begin{aligned} \frac{\partial \chi}{\partial \mu_{ij}} &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} [\chi(\ldots, \mu_{ij} + \varepsilon, \ldots) - \chi(\ldots, \mu_{ij}, \ldots)] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} [\chi(\ldots, m_i + \varepsilon b_j, \ldots) - \chi(\ldots, m_i, \ldots)] = \nabla_{m_i} \chi \, b_j \\ &= \int_G K_{m_i}(\mathbf{x}) \, b_j(\mathbf{x}) \, d^3\mathbf{x} \,. \end{aligned} \tag{4.22}$$

It follows from equation (4.22) that the gradient in the classical sense, $\partial\chi/\partial\mu_{ij}$, is given by the projection of the sensitivity kernel K_{m_i} onto the basis function, b_j.

Summary of the adjoint method

Owing to the outstanding importance of the adjoint method in the context of full waveform inversion, we briefly summarise the key equations:

Forward problem:

$$\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f} \,.$$

Objective functional:

$$\chi(\mathbf{m}) = \int_T \int_G \chi_1[\mathbf{u}(\mathbf{m}; \mathbf{x}, t)] \, dt \, d^3\mathbf{x} = \langle \chi_1 \rangle \,.$$

Adjoint equation:

$$\nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger = -\nabla_u \chi_1^\dagger \,.$$

Derivative of the objective functional:

$$\nabla_m \chi \, \delta\mathbf{m} = \langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \rangle = \int_G K_m \, \delta\mathbf{m} \, d^3\mathbf{x} \,.$$

Fréchet kernel:

$$K_m = \int_T \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L} \, \delta\mathbf{m} \, dt \,.$$

4.3 Application to the elastic wave equation

4.3.1 Derivation of the adjoint equations

The previous development of the adjoint method provides a simple and very general recipe that can be applied to a large class of objective functionals, χ, and physical theories, $\mathbf{L}(\mathbf{u},\mathbf{m}) = \mathbf{f}$. In the context of seismic tomography we are interested in the special case where $\mathbf{L}$ is the elastic wave equation and where $\mathbf{m}$ represents the structural properties of the Earth.

In our development we closely follow Tarantola (1988) and Fichtner et al. (2006): The elastic wave equation, as introduced in paragraph 1.2, is given by

$$\rho(\mathbf{x})\ddot{\mathbf{u}}(\mathbf{x},t) - \nabla \cdot \sigma(\mathbf{x},t) = \mathbf{f}(\mathbf{x},t) \,, \quad \mathbf{x} \in G \subset \mathbb{R}^3 \,, \quad t \in [t_0, \infty) \subset \mathbb{R} \,. \tag{4.23}$$

Equation (4.23) relates the elastic displacement field $\mathbf{u}$ in the Earth $G \subset \mathbb{R}^3$ to the mass density ρ, the stress tensor σ and an external force density $\mathbf{f}$. Under the assumption of a linear visco-elastic rheology, the stress tensor σ is related to the displacement gradient $\nabla\mathbf{u}$ via the constitutive relation

$$\sigma(\mathbf{x},t) = \int_{\tau=t_0}^{\infty} \dot{\mathbf{C}}(\mathbf{x}, t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau) \, d\tau. \tag{4.24}$$

Inserting equation (4.24) into equation (4.23) allows us to express the wave equation in terms of an operator $\mathbf{L}$:

$$\mathbf{L}(\mathbf{u}, \rho, \mathbf{C}) = \mathbf{f} \,, \tag{4.25}$$

$$\mathbf{L}(\mathbf{u}, \rho, \mathbf{C}) = \rho(\mathbf{x}) \, \ddot{\mathbf{u}}(\mathbf{x},t) - \nabla \cdot \int_{\tau=t_0}^{t} \dot{\mathbf{C}}(\mathbf{x}, t-\tau) : \nabla\mathbf{u}(\mathbf{x},\tau) \, d\tau \,. \tag{4.26}$$

The elastic wave operator, $\mathbf{L}$, is accompanied by the initial and boundary conditions

$$\mathbf{u}|_{t \le t_0} = \dot{\mathbf{u}}|_{t \le t_0} = \mathbf{0} \,, \quad \mathbf{n} \cdot \sigma|_{\mathbf{x} \in \partial G} = \mathbf{0} \,. \tag{4.27}$$

To compute the derivative of an objective functional, χ, with respect to ρ and $\mathbf{C}$ we need to find the adjoint operator $\nabla_u \mathbf{L}^\dagger$ of $\nabla_u \mathbf{L}$, as defined in equation (4.13):

$$\langle \mathbf{u} \cdot \nabla_u \mathbf{L}^\dagger \, \mathbf{u}^\dagger \rangle = \langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \, \mathbf{u} \rangle \,. \tag{4.28}$$

The explicit version of $\langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \, \mathbf{u} \rangle$ is

$$\langle \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \mathbf{u} \rangle = \int_T \int_G \mathbf{u}^\dagger \cdot \nabla_u \mathbf{L} \mathbf{u} \, dt \, d^3\mathbf{x} = \int_T \int_G \rho \, \mathbf{u}^\dagger \cdot \ddot{\mathbf{u}} \, dt \, d^3\mathbf{x} - \int_T \int_G \mathbf{u}^\dagger \cdot \left[\nabla \cdot \int_{\tau=t_0}^{t} \dot{\mathbf{C}}(t-\tau) : \nabla\mathbf{u}(\tau) \, d\tau \right] dt \, d^3\mathbf{x} \,, \tag{4.29}$$

where we have omitted all dependencies on **x** in the interest of clarity. Our goal is to isolate **u** in equation (4.29) so that it is not differentiated. We start with the first term on the right-hand side, i.e. with

$$\langle \mathbf{u}^\dagger \cdot \rho \ddot{\mathbf{u}} \rangle = \int_T \int_G \rho\, \mathbf{u}^\dagger \cdot \ddot{\mathbf{u}}\, d^3\mathbf{x}\, dt \,. \tag{4.30}$$

Repeated integration by parts yields

$$\int_T \int_G \rho\, \mathbf{u}^\dagger \cdot \ddot{\mathbf{u}}\, d^3\mathbf{x}\, dt = \int_T \int_G \rho\, \mathbf{u} \cdot \ddot{\mathbf{u}}^\dagger\, d^3\mathbf{x}\, dt + \int_G \rho\, \dot{\mathbf{u}} \cdot \mathbf{u}^\dagger\, d^3\mathbf{x} \bigg|_{t=t_1} - \int_G \rho\, \mathbf{u} \cdot \dot{\mathbf{u}}^\dagger\, d^3\mathbf{x} \bigg|_{t=t_1} , \tag{4.31}$$

where we already used the initial conditions $\mathbf{u}|_{t\le t_0} = \dot{\mathbf{u}}|_{t\le t_0} = \mathbf{0}$. By imposing the *terminal conditions*

$$\mathbf{u}^\dagger|_{t\ge t_1} = \dot{\mathbf{u}}^\dagger|_{t\ge t_1} = \mathbf{0} \,, \tag{4.32}$$

we can obtain the first ingredient of the adjoint operator:

$$\langle \mathbf{u}^\dagger \cdot \rho \ddot{\mathbf{u}} \rangle = \langle \mathbf{u} \cdot \rho \ddot{\mathbf{u}}^\dagger \rangle \,. \tag{4.33}$$

We now turn to the spatial differentiation of **u**. For this, we consider the expression

$$\Upsilon := \langle \mathbf{u}^\dagger \cdot (\nabla \cdot \sigma) \rangle = \int_T \int_G \mathbf{u}^\dagger \cdot \left[\nabla \cdot \int_{\tau=t_0}^{t} \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}(\tau)\, d\tau \right] d^3\mathbf{x}\, dt \,. \tag{4.34}$$

Upon invoking the symmetries of the elastic tensor, **C**, we find the relation

$$\mathbf{u}^\dagger \cdot [\nabla \cdot (\dot{\mathbf{C}} : \nabla \mathbf{u})] = \nabla \cdot (\mathbf{u}^\dagger \cdot \dot{\mathbf{C}} : \nabla \mathbf{u}) - \nabla \cdot (\mathbf{u} \cdot \dot{\mathbf{C}} : \nabla \mathbf{u}^\dagger) + \mathbf{u} \cdot [\nabla \cdot (\dot{\mathbf{C}} : \nabla \mathbf{u}^\dagger)] \,, \tag{4.35}$$

which we can use to transform equation (4.34):

$$\begin{aligned} \Upsilon = &\int_G \int_{t=t_0}^{t_1} \int_{\tau=t_0}^{t} \nabla \cdot [\mathbf{u}^\dagger(t) \cdot \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}(\tau)]\, d\tau\, dt\, d^3\mathbf{x} \\ &- \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_0}^{t} \nabla \cdot [\mathbf{u}(\tau) \cdot \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}^\dagger(t)]\, d\tau\, dt\, d^3\mathbf{x} + \int_G \int_{t=t_0}^{t_1} \int_{\tau=t_0}^{t} \mathbf{u}(\tau) \cdot [\nabla \cdot (\dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}^\dagger(\tau))]\, d\tau\, dt\, d^3\mathbf{x} \,. \end{aligned} \tag{4.36}$$

With the help of Gauss' theorem and the identity

$$\int_{\tau=t_0}^{t} \int_{t=t_0}^{t_1} dt\, d\tau = \int_{t=\tau}^{t_1} \int_{\tau=t_0}^{t_1} dt\, d\tau \tag{4.37}$$

we write Υ in a form that allows us to eliminate two of the contributing integrands:

$$\begin{aligned} \Upsilon = &\int_{\partial G} \int_{t=t_0}^{t_1} \mathbf{u}^\dagger(t) \cdot \left[\int_{\tau=t_0}^{t} \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}(\tau)\, d\tau \right] \cdot \mathbf{n}\, dt\, d^2\mathbf{x} \\ &- \int_{\partial G} \int_{\tau=t_0}^{t_1} \mathbf{u}(\tau) \cdot \left[\int_{t=\tau}^{t_1} \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}^\dagger(t)\, dt \right] \cdot \mathbf{n}\, d\tau\, d^2\mathbf{x} + \int_G \int_{\tau=t_0}^{t_1} \mathbf{u}(\tau) \cdot \left[\nabla \cdot \int_{t=\tau}^{t_1} \dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}^\dagger(t)\, dt \right] d\tau\, d^3\mathbf{x} \,. \end{aligned} \tag{4.38}$$

The first term in equation (4.38) is equal to zero because the expression in square brackets is the stress tensor σ, and $\sigma \cdot \mathbf{n} = \mathbf{0}$ on ∂G as a consequence of the free surface boundary condition. Since $\mathbf{u}^\dagger$ is so far only constrained by the terminal conditions (4.32) we are free to impose a boundary condition that forces the second term to zero:

$$\mathbf{n} \cdot \sigma^\dagger|_{\mathbf{x}\in\partial G} = \mathbf{0} \,, \tag{4.39}$$

with the *adjoint stress tensor* $\boldsymbol{\sigma}^\dagger$ defined by

$$\boldsymbol{\sigma}^\dagger(t) = \int_{\tau=t}^{t_1} \dot{\mathbf{C}}(\tau - t) : \nabla \mathbf{u}^\dagger(\tau)\, d\tau\,. \tag{4.40}$$

The third integrand in equation (4.38) is already of the form that we require. It thus remains to assemble the complete adjoint $\nabla_u \mathbf{L}^\dagger$: Combining equations (4.29), (4.33) and equations (4.38) to (4.40) yields

$$\langle \mathbf{u} \cdot \nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger \rangle = \langle \mathbf{u} \cdot \rho \ddot{\mathbf{u}}^\dagger \rangle - \Upsilon = \langle \mathbf{u} \cdot (\rho\, \ddot{\mathbf{u}}^\dagger - \nabla \cdot \boldsymbol{\sigma}^\dagger) \rangle\,. \tag{4.41}$$

It follows that the adjoint operator $\nabla_u \mathbf{L}^\dagger$ is given by

$$\nabla_u \mathbf{L}^\dagger \mathbf{u}^\dagger = \rho\, \ddot{\mathbf{u}}^\dagger - \nabla \cdot \boldsymbol{\sigma}^\dagger\,. \tag{4.42}$$

Thus, to compute the derivative of the objective functional, $\nabla_m \chi$, one needs to solve the adjoint equation

$$\rho \ddot{\mathbf{u}}^\dagger - \nabla \cdot \boldsymbol{\sigma}^\dagger = -\nabla_u \chi_1^\dagger\,, \tag{4.43}$$

subject to the terminal and boundary conditions

$$\mathbf{u}^\dagger|_{t \geq t_1} = \dot{\mathbf{u}}^\dagger|_{t \geq t_1} = \mathbf{0}\,, \quad \mathbf{n} \cdot \boldsymbol{\sigma}^\dagger|_{\mathbf{x} \in \partial G} = \mathbf{0}\,. \tag{4.44}$$

In non-dissipative media the elastic wave operator $\mathbf{L}$ is *self-adjoint*, meaning that $\mathbf{L} = \mathbf{L}^\dagger$. Since the adjoint equation (4.43) is still of the wave equation type, it can be solved numerically using the same methods as for the solution of the regular wave equation.

The obvious numerical difficulty in solving the adjoint equation is the occurrence of the terminal conditions (4.44) that require that the adjoint field be zero at time $t = t_1$ when the observation ends. In practice, this condition can only be met by solving the adjoint equation backwards in time, that is by reversing the time axis from $t_0 \to t_1$ to $t_1 \to t_0$. The terminal conditions then act as zero initial conditions, at least in the numerical simulation.

Time reversal appears in numerous applications including reverse time migration (e.g. Baysal et al., 1983) and the time reversal imaging of seismic sources (e.g. Larmat et al., 2006). Most of these are closely related to the adjoint method. The adjoint sources are fully specified by the objective functional, that may be used, for instance, to measure the misfit between observed and synthetic seismograms. In this context the adjoint method is often classified as a *back projection technique* where the waveform residuals are propagated backwards in time towards the location from where they originated. The adjoint operator is sometimes referred to as *back projection operator*.

Equation (4.40) reveals that the adjoint stress tensor $\boldsymbol{\sigma}^\dagger$ at time t depends on future strain from t to t_1. This results in a growth of elastic energy when the wave field propagates in the regular time direction from t_0 to t_1. In reversed time, however, the elastic energy decays, so that numerical instabilities do not occur (e.g. Tarantola, 1988).

In our analysis we did not consider perturbations in the topography of the Earth's internal discontinuities. For a derivation of Fréchet kernels with respect to topographical variations the reader is referred to Liu & Tromp (2008).

4.4 First derivatives with respect to selected structural parameters

The most general expression for the derivative of an objective functional $\chi(\mathbf{m})$ in the direction $\delta\mathbf{m}$ is given by equation (4.16) that we repeat here for convenience:

$$\nabla_m \chi\, \delta\mathbf{m} = \langle \mathbf{u}^\dagger \cdot \nabla_m \mathbf{L}\, \delta\mathbf{m} \rangle\,. \tag{4.45}$$

Substituting the wave operator defined in equation (4.26) yields the explicit formula

$$\nabla_m \chi\, \delta\mathbf{m} = \int_T \int_G \mathbf{u}^\dagger(t) \cdot \left[\delta\rho\, \ddot{\mathbf{u}} - \nabla \cdot \int_{\tau=t_0}^{t} \delta\dot{\mathbf{C}}(t-\tau) : \nabla \mathbf{u}(\tau)\, d\tau \right] dt\, d^3\mathbf{x}\,, \tag{4.46}$$

with $\delta\mathbf{m} = (\delta\rho, \delta\mathbf{C})$. Again, we omitted spatial dependencies in the interest of readability. Integrating by parts provides a more symmetric and more useful version of equation (4.46):

$$\nabla_m \chi\, \delta\mathbf{m} = -\int_T \int_G \delta\rho\, \dot{\mathbf{u}}^\dagger(t) \cdot \dot{\mathbf{u}}(t)\, dt\, d^3\mathbf{x} + \int_T \int_G \left[\int_{\tau=t_0}^{t} \nabla\mathbf{u}^\dagger(t) : \delta\dot{\mathbf{C}}(t-\tau) : \nabla\mathbf{u}(\tau)\, d\tau \right] dt\, d^3\mathbf{x}. \tag{4.47}$$

The Fréchet kernels associated with (4.47) are

$$K_\rho = -\int_T \dot{\mathbf{u}}^\dagger(t) \cdot \dot{\mathbf{u}}(t)\, dt\,, \tag{4.48}$$

and

$$\mathbf{K}_C(\tau) = \int_T \nabla\mathbf{u}^\dagger(t) \otimes \nabla\mathbf{u}(t+\tau)\, dt\,, \tag{4.49}$$

where $\otimes$ denotes the tensor or dyadic product. Both kernels are non-zero only within the *primary influence zone* where the regular and adjoint wave fields interact at some time between t_0 and t_1. The primary influence zone, illustrated in figure 4.1, is the region where a model perturbation, $\delta\mathbf{m}$, causes the regular wave field, $\mathbf{u}$, to generate a first-order or single scattered wave that affects the measurement at the receiver. A perturbation located outside the primary influence zone has no first-order effect on the measurement.

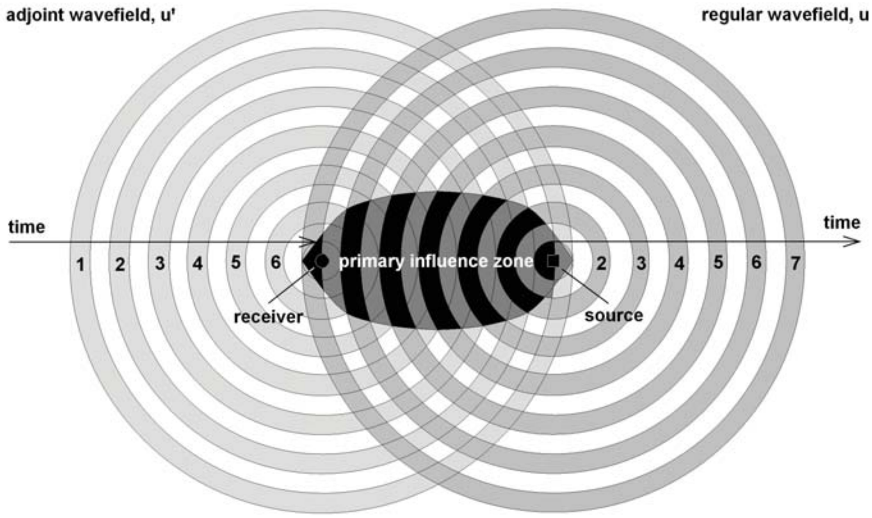

Fig. 4.1 Illustration of the primary influence zone where the regular wave field, $\mathbf{u}$, interacts with the adjoint wave field, $\mathbf{u}^\dagger$. Numbers are used to mark the regular and adjoint wavefronts at successive points in time. As time goes on, the regular wave field propagates away from the source while the adjoint wave field collapses into the receiver. In numerical simulations the adjoint equations are solved backwards in time in order to satisfy the terminal conditions. On the reverse time axis, the adjoint field propagates away from the receiver, starting at the final observation time. The primary influence zone marks the region where a model perturbation, $\delta\mathbf{m}$, generates a first-order scattered wave field that affects the measurement at the receiver. Perturbations located outside the primary influence zone have no first-order effect on the measurement.

For most seismic phases, the primary influence zone is a roughly cigar-shaped region connecting the source and the receiver. Its precise geometry depends on many factors including the frequency content, the length of the considered time window, the type of measurement and the reference Earth model, $\mathbf{m}$. Specific examples for common seismic phases and measurements can be found in chapter 6.
In what follows we derive the Fréchet kernels for several widely used rheologies, i.e. for special choices of $\mathbf{C}(t)$.

4.4.1 Perfectly elastic and isotropic medium

The defining property of a perfectly elastic medium is that present stresses do not depend on past deformation. This is expressed mathematically by a time-dependence of $\mathbf{C}$ and $\delta\mathbf{C}$ that has the form of a unit-step function:

$$\mathbf{C}(\mathbf{x},t) = \mathbf{C}(\mathbf{x})\,H(t)\,, \quad \delta\mathbf{C}(\mathbf{x},t) = \delta\mathbf{C}(\mathbf{x})\,H(t)\,. \tag{4.50}$$

Upon inserting (4.50) into equation (4.47) we obtain a simplified expression for $\nabla_m\chi$:

$$\nabla_m\chi\,\delta\mathbf{m} = -\int_T\int_G \delta\rho\,\dot{\mathbf{u}}^\dagger(t)\cdot\dot{\mathbf{u}}(t)\,dt\,d^3\mathbf{x} + \int_T\int_G \nabla\mathbf{u}^\dagger(t) : \delta\mathbf{C} : \nabla\mathbf{u}(t)\,dt\,d^3\mathbf{x}\,. \tag{4.51}$$

In an isotropic medium the components of $\mathbf{C}$ are given by (e.g. Aki & Richards, 2002)

$$C_{ijkl} = \lambda\,\delta_{ij}\delta_{kl} + \mu\,\delta_{ik}\delta_{jl} + \mu\,\delta_{il}\delta_{jk}\,. \tag{4.52}$$

The symbols λ and μ denote the Lamé parameters. It follows that the complete derivative of χ is composed of three terms:

$$\nabla_m\chi\,\delta\mathbf{m} = \nabla_\rho\chi\,\delta\rho + \nabla_\lambda\chi\,\delta\lambda + \nabla_\mu\chi\,\delta\mu\,, \tag{4.53}$$

with

$$\nabla_\rho\chi\,\delta\rho = -\int_T\int_G \delta\rho\,\dot{\mathbf{u}}^\dagger\cdot\dot{\mathbf{u}}\,dt\,d^3\mathbf{x}\,, \tag{4.54a}$$

$$\nabla_\lambda\chi\,\delta\lambda = \int_T\int_G \delta\lambda\,(\nabla\cdot\mathbf{u})(\nabla\cdot\mathbf{u}^\dagger)\,dt\,d^3\mathbf{x}\,, \tag{4.54b}$$

$$\nabla_\mu\chi\delta\mu = \int_T\int_G \delta\mu\,[(\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u}) + (\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u})^T]\,dt\,d^3\mathbf{x}\,. \tag{4.54c}$$

We obtain the corresponding sensitivity kernels by simply dropping the spatial integration:

$$K_\rho^0 = -\int_T \dot{\mathbf{u}}^\dagger\cdot\dot{\mathbf{u}}\,dt\,, \tag{4.55a}$$

$$K_\lambda^0 = \int_T (\nabla\cdot\mathbf{u})(\nabla\cdot\mathbf{u}^\dagger)\,dt\,, \tag{4.55b}$$

$$K_\mu^0 = \int_T [(\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u}) + (\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u})^T]\,dt\,. \tag{4.55c}$$

The superscript ^{0} symbolises that the Fréchet kernels correspond to the fundamental parameterisation $\mathbf{m} = (\rho,\lambda,\mu)$. Equations (4.55a) to (4.55c) confirm the intuitive expectation that K_λ^0 vanishes for pure shear waves, while both shear and compressional waves contribute to K_μ^0. Depending on personal preference and numerical convenience, one may rewrite some of the above equations in terms of the regular strain tensor $\varepsilon = \frac{1}{2}[(\nabla\mathbf{u}) + (\nabla\mathbf{u})^T]$ and the adjoint strain tensor $\varepsilon^\dagger = \frac{1}{2}[(\nabla\mathbf{u}^\dagger) + (\nabla\mathbf{u}^\dagger)^T]$. For this we can use the relations

$$(\nabla\cdot\mathbf{u})(\nabla\cdot\mathbf{u}^\dagger) = (\mathrm{tr}\,\varepsilon)\,(\mathrm{tr}\,\varepsilon^\dagger)\,, \tag{4.56}$$

and

$$(\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u}) + (\nabla\mathbf{u}^\dagger) : (\nabla\mathbf{u})^T = 2\varepsilon^\dagger : \varepsilon\,. \tag{4.57}$$

Based on equations (4.55a) to (4.55c) we easily derive formulas for the Fréchet kernels with respect to other parameters, such as the bulk modulus, $\kappa = \lambda + 2\mu/3$, the S wave speed, $v_\mathrm{S} = \sqrt{\mu/\rho}$ or the P wave speed, $v_\mathrm{P} = \sqrt{(\lambda+2\mu)/\rho}$. Defining ρ, κ and μ as independent parameters, we find the following set of kernels:

$$K_\rho = K_\rho^0\,, \quad K_\kappa = K_\lambda^0\,, \quad K_\mu = K_\mu^0 - \frac{2}{3}K_\lambda^0\,. \tag{4.58}$$

With ρ, v_P and v_S as independent parameters we have

$$K_\rho = K_\rho^0 + (v_\mathrm{P}^2 - 2v_\mathrm{S}^2)\,K_\lambda^0 + v_\mathrm{S}^2\,K_\mu^0\,, \quad K_{v_\mathrm{S}} = 2\rho\,v_\mathrm{S}\,K_\mu^0 - 4\rho\,v_\mathrm{S}\,K_\lambda^0\,, \quad K_{v_\mathrm{P}} = 2\rho\,v_\mathrm{P}\,K_\lambda^0\,. \tag{4.59}$$

4.4.2 Perfectly elastic medium with radial anisotropy

Anisotropy is the dependence of the elastic tensor on the orientation of the coordinate system. Its most direct seismological expressions are the splitting of shear waves and the dependence of seismic velocities on the propagation and polarisation directions. Polarisation anisotropy is particularly pronounced in the Earth's uppermost mantle, and it leads to the readily observable *Love-Rayleigh discrepancy*: The propagation speeds of Love and Rayleigh waves travelling in the same direction can rarely be explained with an isotropic model. Pure polarisation anisotropy occurs in elastic media with radial symmetry axis that are fully described by 5 independent elastic tensor components. They can be summarised in two 3×3 matrices (e.g. Babuska & Cara, 1991):

$$\begin{pmatrix} C_{rrrr} & C_{rr\phi\phi} & C_{rr\theta\theta} \\ C_{\phi\phi rr} & C_{\phi\phi\phi\phi} & C_{\phi\phi\theta\theta} \\ C_{\theta\theta rr} & C_{\theta\theta\phi\phi} & C_{\theta\theta\theta\theta} \end{pmatrix} = \begin{pmatrix} \lambda+2\mu & \lambda+c & \lambda+c \\ \lambda+c & \lambda+2\mu+a & \lambda+a \\ \lambda+c & \lambda+a & \lambda+2\mu+a \end{pmatrix} \tag{4.60a}$$

$$\begin{pmatrix} C_{\phi\theta\phi\theta} & C_{\phi\theta r\theta} & C_{\phi\theta r\phi} \\ C_{r\theta\phi\theta} & C_{r\theta r\theta} & C_{r\theta r\phi} \\ C_{r\phi\phi\theta} & C_{r\phi r\theta} & C_{r\phi r\phi} \end{pmatrix} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu+b & 0 \\ 0 & 0 & \mu+b \end{pmatrix} \tag{4.60b}$$

All components of $\mathbf{C}$ that do not appear in equation (4.60) are equal to zero. Many of the examples given in the following chapters will be based on this type of anisotropy. For $a=b=c=0$ we retrieve the elastic tensor of an isotropic medium. In homogeneous media we can give intuitive interpretations of λ, μ, a, b and c: The velocity of a horizontally propagating plane S wave is

$$v_\mathrm{SV} = \sqrt{\frac{\mu+b}{\rho}} \quad \text{or} \quad v_\mathrm{SH} = \sqrt{\frac{\mu}{\rho}} \tag{4.61}$$

depending on whether it is vertically (v_SV) or horizontally (v_SH) polarised. For vertically and horizontally propagating P waves we find the propagation speeds v_PV and v_PH, given by

$$v_\mathrm{PV} = \sqrt{\frac{\lambda+2\mu}{\rho}} \quad \text{and} \quad v_\mathrm{PH} = \sqrt{\frac{\lambda+2\mu+a}{\rho}}\,, \tag{4.62}$$

respectively. The parameter c can only be determined from P waves that do not travel in exactly radial or horizontal directions. It is commonly absorbed in a new parameter, η, defined by (e.g. Dziewonski & Anderson, 1981)

$$\eta := \frac{\lambda+c}{\lambda+a}\,. \tag{4.63}$$

After some tedious but straightforward algebraic manipulations we find the following set of Fréchet kernels with respect to ρ, λ, μ, a, b and c:

$$K_\rho^0 = -\int_T \dot{\mathbf{u}}^\dagger \cdot \dot{\mathbf{u}}\, dt\,, \tag{4.64a}$$

$$K_\lambda^0 = \int_T (\mathrm{tr}\,\varepsilon^\dagger)(\mathrm{tr}\,\varepsilon)\, dt\,, \tag{4.64b}$$

$$K_\mu^0 = 2\int_T \varepsilon^\dagger : \varepsilon\, dt\,, \tag{4.64c}$$

$$K_a^0 = \int_T (\varepsilon_{\phi\phi}^\dagger + \varepsilon_{\theta\theta}^\dagger)(\varepsilon_{\phi\phi} + \varepsilon_{\theta\theta})\, dt\,, \tag{4.64d}$$

$$K_b^0 = 4\int_T (\varepsilon_{r\theta}^\dagger \varepsilon_{r\theta} + \varepsilon_{r\phi}^\dagger \varepsilon_{r\phi})\, dt\,, \tag{4.64e}$$

$$K_c^0 = \int_T [\varepsilon_{rr}^\dagger(\varepsilon_{\phi\phi} + \varepsilon_{\theta\theta}) + \varepsilon_{rr}(\varepsilon_{\phi\phi}^\dagger + \varepsilon_{\theta\theta}^\dagger)]\, dt\,. \tag{4.64f}$$

Fréchet kernels for the seismologically more relevant parameters $v_{\rm SV}, v_{\rm SH}, v_{\rm PV}, v_{\rm PH}$ and η can now be expressed as linear combinations of the kernels in equations (4.64a) to (4.64f):

$$\begin{aligned} K_\rho &= K_\rho^0 + v_{\rm SH}^2 K_\mu^0 + (v_{\rm PV}^2 - 2v_{\rm SH}^2)K_\lambda^0 + (v_{\rm PH}^2 - v_{\rm PV}^2)K_a^0 \\ &\quad + (v_{\rm SV}^2 - v_{\rm SH}^2)K_b^0 + [2(1-\eta)v_{\rm SH}^2 + \eta v_{\rm PH}^2 - v_{\rm PV}^2]K_c^0 \\ &= K_\rho^0 + \rho^{-1}(\mu K_\mu^0 + \lambda K_\lambda^0 + aK_a^0 + bK_b^0 + cK_c^0)\,, \end{aligned} \tag{4.65a}$$

$$K_{v_{\rm SH}} = 2\rho v_{\rm SH}[K_\mu^0 - 2K_\lambda^0 - K_b^0 + 2(1-\eta)K_c^0]\,, \tag{4.65b}$$

$$K_{v_{\rm SV}} = 2\rho v_{\rm SV} K_b^0\,, \tag{4.65c}$$

$$K_{v_{\rm PH}} = 2\rho v_{\rm PH}(K_a^0 + \eta K_c^0)\,, \tag{4.65d}$$

$$K_{v_{\rm PV}} = 2\rho v_{\rm PV}(K_\lambda^0 - K_a^0 - K_c^0)\,, \tag{4.65e}$$

$$K_\eta = \rho\,(v_{\rm PH}^2 - 2v_{\rm SH}^2)\,K_c^0 = (\lambda + a)\,K_c^0\,. \tag{4.65f}$$

Equations (4.65a) to (4.65f) can be simplified in the case of an isotropic reference medium where $v_{\rm SH} = v_{\rm SV} = v_{\rm S}$, $v_{\rm PH} = v_{\rm PV} = v_{\rm P}$ and $\eta = 1$:

$$K_\rho = K_\rho^0 + v_{\rm S}^2 K_\mu^0 + (v_{\rm P}^2 - 2v_{\rm S}^2)K_\lambda^0\,, \tag{4.66a}$$

$$K_{v_{\rm SH}} = 2\rho v_{\rm S}(K_\mu^0 - 2K_\lambda^0 - K_b^0)\,, \tag{4.66b}$$

$$K_{v_{\rm SV}} = 2\rho v_{\rm S} K_b^0\,, \tag{4.66c}$$

$$K_{v_{\rm PH}} = 2\rho v_{\rm P}(K_a^0 + K_c^0)\,, \tag{4.66d}$$

$$K_{v_{\rm PV}} = 2\rho v_{\rm P}(K_\lambda^0 - K_a^0 - K_c^0)\,, \tag{4.66e}$$

$$K_\eta = \rho\,(v_{\rm P}^2 - 2v_{\rm S}^2)\,K_c^0\,. \tag{4.66f}$$

The comparison of equations (4.59) and (4.66a) to (4.66f) yields two interesting relations between the Fréchet kernels for $v_{\rm PV}, v_{\rm PH}, v_{\rm SV}, v_{\rm SH}, v_{\rm P}$ and $v_{\rm S}$ in the case of an isotropic background medium:

$$K_{v_{\rm P}} = K_{v_{\rm PH}} + K_{v_{\rm PV}}\,, \tag{4.67}$$

$$K_{v_{\rm S}} = K_{v_{\rm SH}} + K_{v_{\rm SV}}\,. \tag{4.68}$$

Depending on the data used, the kernels $K_{v_{\rm P}}$ and $K_{v_{\rm S}}$ can be small, in the sense that the data are hardly affected by changes in $v_{\rm P}$ or $v_{\rm S}$. The above relations, however, imply that nearly vanishing Fréchet kernels with respect to $v_{\rm P}$ and $v_{\rm S}$ may in fact be composed of significantly non-zero Fréchet kernels with respect to $v_{\rm PV}, v_{\rm PH}, v_{\rm SV}$ and $v_{\rm SH}$. Certain P and S wave sensitivities can become alive when the model is allowed to be anisotropic. An example of this phenomenon was found by Dziewonski & Anderson (1981) who noted that the P wave sensitivity of fundamental mode Rayleigh waves with periods around 120 s is restricted to the uppermost 100 km. Sensitivities with respect to the PH and PV wave speeds, however, are large down to 400 km depth but opposite in sign.

Chapter 5
Misfit functionals and adjoint sources

The early developments of full waveform inversion for two-dimensional acoustic problems (e.g. Tarantola, 1984; Gauthier et al., 1986) were almost immediately followed by the recognition that the choice of a suitable misfit functional is crucial for a successful application to real data. In one of the first large-scale full waveform inversions, Crase et al. (1990) proposed a series of robust misfit measures that are comparatively insensitive to seismic noise. Their study was extended by Brossier et al. (2010). The L_2 distance between observed and synthetic seismograms is efficient for the detection of sharp material contrasts, but the recovery of long-wavelength Earth structure requires misfit functionals that explicitly extract phase information. In their pioneering work, Luo & Schuster (1991) therefore proposed to measure the cross-correlation time shift between observed and synthetic waveforms (section 5.3). This idea was extended by van Leeuwen & Mulder (2010). An alternative approach based on the measurement of time-frequency misfits was proposed by Fichtner et al. (2008). Closely related is the quantification of waveform differences based on the instantaneous phase and envelope (Bozdağ & Trampert, 2010).
In the following sections we present a selection of misfit functionals and derive the corresponding adjoint sources. This is intended to offer both physical insight and concrete solutions. Colourful examples for the resulting Fréchet kernels are shown in section 6. To avoid clutter, we consider measurements made at a single station. The generalisation to measurements at multiple stations is straightforward.

As a preparatory step, we introduce the concept of the *adjoint Greens function*, that will play a key role in our subsequent developments. The adjoint Greens function, $\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)$, is defined as the solution of the adjoint equation (4.17) with a source that acts in i-direction at the position $\mathbf{x}=\xi$ and at time $t=\tau$:

$$\mathbf{L}^\dagger[\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)] := \mathbf{e}_i\,\delta(\mathbf{x}-\xi)\,\delta(t-\tau)\,. \tag{5.1}$$

The symbol $\mathbf{e}_i$ denotes the unit vector in i-direction. The importance of the definition (5.1) is related to the fact that we can express the adjoint field, $\mathbf{u}^\dagger$, corresponding to an arbitrary adjoint source, $\mathbf{f}^\dagger(\mathbf{x},t)=\sum_{j=1}^3 \mathbf{e}_j f_j^\dagger(\mathbf{x},t)$, in terms of $\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)$. To see this, we multiply equation (5.1) by $f_i^\dagger(\xi,\tau)$ and sum over i from 1 to 3:

$$\sum_{i=1}^3 f_i^\dagger(\xi,\tau)\,\mathbf{L}^\dagger[\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)] = \mathbf{f}^\dagger(\xi,\tau)\,\delta(\mathbf{x}-\xi)\,\delta(t-\tau)\,. \tag{5.2}$$

The adjoint wave operator, $\mathbf{L}^\dagger$, is linear and it does not involve derivatives with respect to ξ and τ. We can therefore rewrite (5.2) in the form

$$\mathbf{L}^\dagger\left[\sum_{i=1}^3 f_i^\dagger(\xi,\tau)\,\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)\right] = \mathbf{f}^\dagger(\xi,\tau)\,\delta(\mathbf{x}-\xi)\,\delta(t-\tau)\,. \tag{5.3}$$

Integrating equation (5.4) over time, τ, and space, ξ, yields

$$\mathbf{L}^\dagger\left[\sum_{i=1}^3 \int_T \int_G f_i^\dagger(\xi,\tau)\,\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)\,d\tau\,d^3\xi\right] = \mathbf{f}^\dagger(\mathbf{x},t)\,. \tag{5.4}$$

This implies that the adjoint field

$$\mathbf{u}^\dagger(\mathbf{x},t) = \sum_{i=1}^{3} \int_T \int_G f_i^\dagger(\xi,\tau)\,\mathbf{g}_i^\dagger(\xi,\tau;\mathbf{x},t)\,d\tau\,d^3\xi \tag{5.5}$$

is the solution of the adjoint equation (4.17) with the adjoint source $\mathbf{f}^\dagger(\mathbf{x},t)$. Equation (5.5) is the *representation theorem* for adjoint fields. We are now set for the study of specific misfit functionals and their adjoint sources.

5.1 Derivative of the pure wave field and the adjoint Greens function

The adjoint method can be used as a tool to linearise the forward problem. For this we let χ be the i-component of the elastic displacement field, $\mathbf{u}$, measured at the receiver position $\mathbf{x} = \mathbf{x}^r$ and at time $t = t^r$:

$$\chi(\mathbf{m}) = u_i(\mathbf{m};\mathbf{x}^r,t^r)\,. \tag{5.6}$$

To derive the adjoint source corresponding to the objective functional defined in equation (5.6), we write χ in integral form, as proposed in equation (4.6):

$$\chi(\mathbf{m}) = \langle\chi_1(\mathbf{m})\rangle = \int\limits_T\int\limits_G \chi_1(\mathbf{m})\,dt\,d^3\mathbf{x} = \int\limits_T\int\limits_G \mathbf{e}_i\cdot\mathbf{u}(\mathbf{m};x,t)\,\delta(\mathbf{x}-\mathbf{x}^r)\delta(t-t^r)\,dt\,d^3\mathbf{x}\,. \tag{5.7}$$

The symbol $\mathbf{e}_i$ denotes the unit vector in i-direction. It follows from equation (5.7) that the integrand, χ_1, is given by

$$\chi_1(\mathbf{m}) = \mathbf{e}_i\cdot\mathbf{u}(\mathbf{m};x,t)\,\delta(\mathbf{x}-\mathbf{x}^r)\delta(t-t^r)\,, \tag{5.8}$$

Applying the recipe from equation (4.15) to the above expression for χ_1, yields the adjoint source, $\mathbf{f}^\dagger$, that corresponds to the objective functional $\chi(\mathbf{m}) = u_i(\mathbf{m};x^r,t^r)$:

$$\mathbf{f}^\dagger(\mathbf{x},t) = -\nabla_u\chi_1 = -\mathbf{e}_i\,\delta(\mathbf{x}-\mathbf{x}^r)\,\delta(t-t^r)\,. \tag{5.9}$$

The adjoint source, $\mathbf{f}^\dagger$, is point-localised in both space and time. It acts at the receiver position, $\mathbf{x}^r$, and at the observation time, t^r. The direction of the adjoint source is opposite to the direction in which the observation was made. Substituting (5.9) into (4.17) yields the adjoint equation

$$\mathbf{L}^\dagger(\mathbf{u}^\dagger) = -\mathbf{e}_i\,\delta(\mathbf{x}-\mathbf{x}^r)\,\delta(t-t^r)\,. \tag{5.10}$$

The comparison of (5.10) with the definition (5.1) implies that the adjoint field, $\mathbf{u}^\dagger$, is equal to the negative adjoint Greens function:

$$\mathbf{u}^\dagger(\mathbf{x},t) = -\mathbf{g}_i^\dagger(\mathbf{x}^r,t^r;\mathbf{x},t)\,. \tag{5.11}$$

For the derivative of χ we then find

$$\nabla_m\chi\,\delta\mathbf{m} = \nabla_m u_i(\mathbf{x}^r,t^r)\,\delta\mathbf{m} = -\langle\mathbf{g}_i^\dagger(\mathbf{x}^r,t^r)\cdot\nabla_m\mathbf{L}\,\delta\mathbf{m}\rangle\,. \tag{5.12}$$

In the special case of a homogeneous and unbounded 3D medium, the adjoint Greens function consists of spherical waves with infinitesimally short wavelength (e.g. Aki & Richards, 2002). This scenario is shown schematically in figure (5.1). The sensitivity kernels are products of the regular wave field, $\mathbf{u}$, and the adjoint Greens function $\mathbf{g}_i^\dagger$. They are non-zero only within the primary influence zone where the adjoint field and the regular field overlap at a given time between t_0 and t^r. The width of the primary influence zone is proportional to the wavelength of the regular wave $\mathbf{u}$, and it tends to zero as the wavelength of $\mathbf{u}$ decreases.

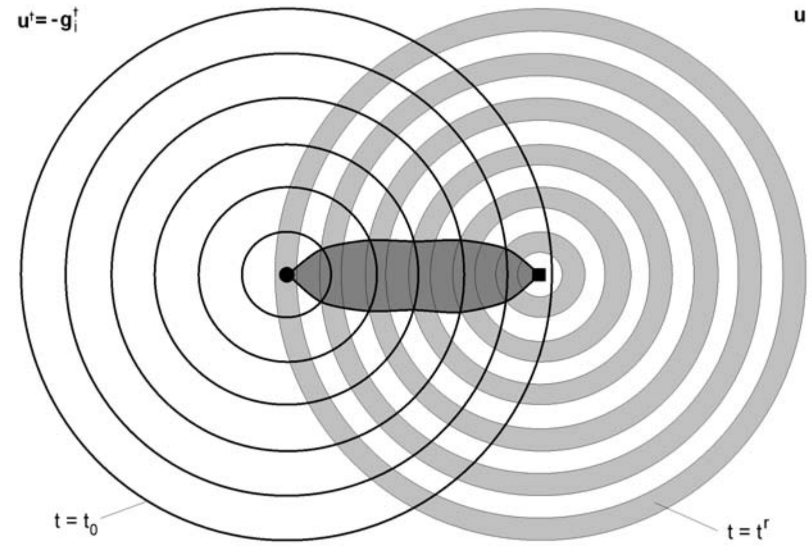

Fig. 5.1 Illustration of a sensitivity kernel that arises from the interaction of the regular wave field $\mathbf{u}$ emanating from the source (■) and the negative adjoint Greens function, $\mathbf{u}^\dagger - \mathbf{g}_i^\dagger$, that propagates from the receiver (•) towards the source. The kernel is non-zero only in the dark shaded region where the regular and adjoint fields overlap.

5.2 L_2 waveform difference

The misfit functional classically used in full waveform inversion (e.g. Bamberger et al., 1982; Tarantola, 1984; Igel et al., 1996) is the L_2 distance between the observed seismogram, $\mathbf{u}^0$, and the synthetic seismogram, $\mathbf{u}$, at the receiver position $\mathbf{x}^r$:

$$\chi(\mathbf{m}) = \frac{1}{2} \int\limits_T [\mathbf{u}^0(\mathbf{x}^r,t) - \mathbf{u}(\mathbf{m};\mathbf{x}^r,t)]^2 \, dt \,. \tag{5.13}$$

In the hypothetical case where χ is equal to zero, the data are explained perfectly by the Earth model. The integrand χ_1 corresponding to χ is

$$\chi_1(\mathbf{m}) = \frac{1}{2} [\mathbf{u}^0(\mathbf{x}^r,t) - \mathbf{u}(\mathbf{m};\mathbf{x}^r,t)]^2 \, \delta(\mathbf{x} - \mathbf{x}^r) \,. \tag{5.14}$$

Invoking the recipe from equation (4.15) yields the adjoint source that corresponds to the misfit functional defined in (5.13):

$$\mathbf{f}^\dagger(\mathbf{x},t) = -\nabla_u \chi_1 = [\mathbf{u}(\mathbf{m};\mathbf{x},t) - \mathbf{u}^0(\mathbf{x}^r,t)] \, \delta(\mathbf{x} - \mathbf{x}^r) \,. \tag{5.15}$$

The adjoint source is again point-localised at the receiver position, and its time evolution is determined by the residual time series $\mathbf{u}(t) - \mathbf{u}^0(t)$. The common expression that the adjoint method consists in *propagating the residuals back in time* largely results from equation (5.15).

While intuitively plausible, the L_2 distance can be problematic in practice. It is, first of all, not robust, meaning that outliers in the data can become dominant. Second, the numerical value of χ as defined in (5.13) is controlled by the large-amplitude waveforms. This means that the invaluable information contained, for instance, in the time shifts of low-amplitude P waves, is almost entirely lost. The L_2 distance furthermore emphasises the non-linearity that is already inherent in the forward problem. This results in multiple local minima of the misfit functional, which is a very disadvantageous property in the context of gradient-based misfit minimisation algorithms. The existence of local minima is closely related to the incapability of the L_2 distance to detect the long-wavelength structure of the Earth. To retrieve long-wavelength features, one must adopt a multi-scale approach, where the inversion starts with the longest possible periods that are then successively decreased (e.g. Dessa et al., 2004; Bleibinhaus et al., 2007). A possible alternative and complement is the combination of full waveform inversion based on the L_2 distance with classical traveltime ray tomography (e.g. Pratt & Goulty, 1991; Zhou et al., 1995; Korenaga et al., 1997; Dessa & Pascal. 2003). Following the recognition that the L_2 distance depends very non-linearly on long-wavelength structure, various objective functionals have been designed that explicitly extract traveltime information. These are the subjects of the following paragraphs.

5.3 Cross-correlation time shifts

Despite the inherent problems of the L_2 waveform distance, it remains desirable to extract as much information as possible from the difference between observed and synthetic seismograms. A milestone towards this goal was Luo & Schuster's (1991) realisation that phase information needs to be included explicitly in the objective functional. Separating the phases of seismic waveforms from their amplitudes is required to overcome the excessive non-linearity introduced by the L_2 waveform difference as defined in (5.13). Luo & Schuster's method is based on the estimation of delay times by cross-correlating data and numerically computed synthetics – a technique reminiscent of ideas expressed earlier by Dziewonski et al. (1972), Lerner-Lam & Jordan (1983) or Cara & Lévêque (1987) in the context of surface wave analysis. The cross-correlation approach was further formalised by Gee & Jordan (1992) and then directly applied to data by Zhou et al. (1995) and Chen et al. (2007). It was also used for the computation of finite-frequency delay time kernels (Dahlen et al., 2000; Tromp et al., 2005; Liu & Tromp, 2008; Sieminski et al., 2007a,b). In our development we closely follow the concept introduced by Luo & Schuster (1991). We consider the *i*-component of an observed waveform, $u_i^0(\mathbf{x}^r,t)$, and the corresponding synthetic, $u_i(\mathbf{m};\mathbf{x}^r,t)$, at a receiver position $\mathbf{x}^r$. It is implicitly assumed that a specific waveform, such as the direct P or S wave, has been isolated from both the observed and synthetic seismograms. The *cross-correlation time shift* $\mathcal{T}$ is defined as the time where the cross-correlation function

$$\mathcal{C}(u_i^0,u_i)(\tau) := \int\limits_T u_i^0(\mathbf{x}^r,t)u_i(\mathbf{m};\mathbf{x}^r,t+\tau)\,dt \tag{5.16}$$

attains its global maximum. We therefore have $\mathcal{T} > 0$ when the synthetic waveform arrives later than the observed waveform, and $\mathcal{T} < 0$ when the synthetic waveform arrives earlier than the observed waveform. It is assumed that both synthetic and observed waveforms have been properly filtered and isolated. The misfit functional that we now wish to minimise, is

$$\chi(\mathbf{m}) = \frac{1}{2}\mathcal{T}^2(\mathbf{m})\,. \tag{5.17}$$

This is the misfit commonly used in traveltime tomography. The definition (5.16) does not provide an explicit expression for $\mathcal{T}$. We therefore need to derive a *connective function* that links u_i and u_i^0 to the secondary observable $\mathcal{T}$. For this, we note that $\mathcal{C}(u_i^0,u_i)(\tau)$ attains a maximum for $\tau = \mathcal{T}$, and it therefore satisfies the necessary condition

$$0 = \frac{d}{d\tau}\mathcal{C}(u_i^0,u_i)(\tau)|_{\tau=\mathcal{T}} = \int\limits_T u_i^0(\mathbf{x}^r,t)\dot{u}_i(\mathbf{m};\mathbf{x}^r,t+\mathcal{T})\,dt = -\int\limits_T \dot{u}_i^0(\mathbf{x}^r,t-\mathcal{T})u_i(\mathbf{m};\mathbf{x}^r,t)\,dt\,. \tag{5.18}$$

Equation (5.18) defines $\mathcal{T}$ implicitly, at least when there is only one maximum. That $\mathcal{T}$ corresponds indeed to the global maximum of the correlation function and not to one of several local extrema needs to be ensured by the careful observer. Invoking the implicit function differentiation yields the derivative of χ:

$$\nabla_m\chi\,\delta\mathbf{m} = \mathcal{T}\,\nabla_m\mathcal{T}(\mathbf{m})\,\delta\mathbf{m} = \frac{\mathcal{T}\int_T \dot{u}_i^0(\mathbf{x}^r,t-\mathcal{T})\,\nabla_m u_i(\mathbf{m};\mathbf{x}^r,t)\,\delta\mathbf{m}\,dt}{\int_T \ddot{u}_i^0(\mathbf{x}^r,t-\mathcal{T})\,u_i(\mathbf{m},\mathbf{x}^r,t)\,dt}\,. \tag{5.19}$$

Under the assumption that u_i^0 and u_i are purely time-shifted and not otherwise distorted with respect to each other, we have $u_i^0(\mathbf{x}^r,t-\mathcal{T}) = u_i(\mathbf{m};\mathbf{x}^r,t)$, and equation (5.19) reduces to

$$\nabla_m\chi\,\delta\mathbf{m} = -\frac{\mathcal{T}}{||\dot{u}_i||_2^2}\int\limits_T \dot{u}_i(\mathbf{m};\mathbf{x}^r,t)\,\nabla_m u_i(\mathbf{m};\mathbf{x}^r,t)\,\delta\mathbf{m}\,dt\,, \tag{5.20}$$

with

$$||\dot{u}_i||_2^2 = \int_T \dot{u}_i^2(\mathbf{m};\mathbf{x}^r,t)\,dt\,. \tag{5.21}$$

For the term $\nabla_m u_i(\mathbf{m};\mathbf{x}^r,t)\,\delta\mathbf{m}$ in equation (5.20) we substitute the result from equation (5.12):

$$\begin{aligned}\nabla_m\chi\,\delta\mathbf{m} &= \frac{\mathscr{T}}{||\dot{u}_i||_2^2}\int_T \dot{u}_i(\mathbf{m};\mathbf{x}^r,t)\,\langle\mathbf{g}_i^\dagger(\mathbf{x}^r,t)\cdot\nabla_m\mathbf{L}\,\delta\mathbf{m}\rangle\,dt\\ &= \frac{\mathscr{T}}{||\dot{u}_i||_2^2}\int_T\int_T\int_G \dot{u}_i(\mathbf{m};\mathbf{x}^r,t)\,\mathbf{g}_i^\dagger(\mathbf{x}^r,t;\mathbf{x},\tau)\cdot\nabla_m\mathbf{L}\,\delta\mathbf{m}\,d\tau\,dt\,d^3\mathbf{x}\,.\end{aligned} \tag{5.22}$$

Defining the adjoint field

$$\mathbf{u}^\dagger(\mathbf{x},\tau) := \frac{\mathscr{T}}{||\dot{u}_i||_2^2}\int_T \dot{u}_i(\mathbf{m};\mathbf{x}^r,t)\,\mathbf{g}_i^\dagger(\mathbf{x}^r,t;\mathbf{x},\tau)\,dt\,, \tag{5.23}$$

allows us to write the derivative of χ in its canonical form (4.16):

$$\nabla_m\chi\,\delta\mathbf{m} = \int_T\int_G \mathbf{u}^\dagger(\mathbf{x},t)\cdot\nabla_m\mathbf{L}\,\delta\mathbf{m}\,dt\,d^3\mathbf{x}\,. \tag{5.24}$$

Equation (5.23) has the form of the representation theorem (5.5), where the adjoint wave field, $\mathbf{u}^\dagger$, is expressed in terms of an integral over the adjoint source times the adjoint Greens function. It therefore follows that the adjoint source corresponding to χ is given by

$$\mathbf{f}^\dagger(\mathbf{x},t) = \frac{\mathscr{T}\,\mathbf{e}_i}{||\dot{u}_i||_2^2}\,\dot{u}_i(\mathbf{m};\mathbf{x}^r,t)\,\delta(\mathbf{x}-\mathbf{x}^r)\,, \tag{5.25}$$

with $\mathbf{e}_i$ denoting the unit vector in the observation direction. As a result of the point-wise measurement, the adjoint source is space-localised at the receiver position. Its time evolution is determined by the synthetic displacement velocity. The factor $||\dot{u}_i||_2^{-2}$ effectively normalises the sensitivity kernels. This ensures that their amplitude does not depend on the amplitude of the regular wave field.

The geometry of the kernels is quasi-independent of the data because the adjoint source does not contain the data, except for the factor $\mathscr{T}$. This quasi-independence rests, of course, on the assumption that the observed and synthetic waveforms are sufficiently similar to allow for the replacement of $u_i^0(\mathbf{x}^r,t-\mathscr{T})$ by $u_i(\mathbf{m};\mathbf{x}^r,t)$, which then led to equation (5.20). Teleseismic body waveforms often reveal the necessary similarity between observation and synthetic, so that measurements of cross-correlation time shifts are physically meaningful. Depending on the magnitude of the events, it may become necessary to account for the source time function, using, for instance, a matched-filter approach (e.g. Sigloch & Nolet, 2006). In the case of surface waves, a sufficient waveform similarity is usually not present. It can, however, be enforced by rigorous band-pass filtering (e.g. Gee & Jordan, 1992).

Several recent applications of cross-correlation time shifts (e.g. Sigloch et al., 2008; Tape et al., 2010) were based on measurements in multiple frequency bands. This strategy increases the amount of useful information substantially, thus leading to higher spatial resolution of tomographic images.

Chapter 6
Fréchet kernel gallery

Following several chapters replete with purely theoretical developments, we now delve into the description of concrete Fréchet kernels for several seismologically relevant combinations of measurements and Earth model parameters. This is, first of all, intended to advance the intuition necessary for the meaningful solution of any tomographic problem.
Throughout this chapter we will be concerned with measurements of cross-correlation time shifts, $\mathscr{T}$. This choice is motivated by the simplicity of the corresponding adjoint sources and the similarity with other time- and amplitude-like measurements that are based, for instance, on time-frequency representations (Fichtner et al., 2008) or generalised seismological data functionals (Gee & Jordan, 1992). Furthermore, we restrict our attention to single-station measurements. The Fréchet kernels for multi-station measurements are simply the superposition of the individual kernels.
For the exemplary kernel calculations presented in the following paragraphs, we use a spectral-element method (Fichtner et al., 2009, section that solves the elastic wave equation in a spherical section. The Earth model is the isotropic version of PREM (Dziewonski & Anderson, 1981). For simplicity we will refer to Fréchet kernels from cross-correlation measurements as traveltime kernels, and to Fréchet kernels from L_2 amplitude measurements as amplitude kernels.

Following a brief description of the computational setup, we study the anatomy of traveltime and amplitude kernels with respect to v_P and v_S for a variety of body wave phases. Section 6.2 is devoted to the sensitivity of surface waves to both isotropic and anisotropic perturbations.

6.1 Body waves

The 3D sensitivity distributions of body waves with a finite frequency content have been analysed extensively during the past two decades. In a series of pioneering studies, Yomogida (1992), Dahlen et al. (2000) and Dahlen & Baig (2002) used the ray approximation for the computation of body wave Fréchet kernels in laterally homogeneous media. For an iterative waveform inversion, kernels need to be computed also in 3D heterogeneous media, which requires the use of fully numerical methods. Numerous examples of body wave sensitivity kernels computed with the help of finite-difference and spectral-element modelling can be found in Zhao et al. (2005), Liu & Tromp (2006), Liu & Tromp (2008) or Sieminski et al. (2007b).
For our study of body wave sensitivities we consider a geometric setup where several body wave phases are easily observable. The synthetic source is located at 400 km depth beneath western Turkey. It radiates elastic waves recorded in western Spain at an epicentral distance of 25.23°. The surface projection of the ray path for this source-receiver pair and the P wave radiation pattern are shown in figure 6.1. The corresponding three-component synthetic velocity seismograms in figure 6.2 exhibit various prominent body wave phases, including the direct P and S waves, as well as the surface-reflected phases sP and sS. The dominant period is 15 s.

Cross-correlation time shifts or variants thereof, are one of the most widely used measurement of finite-frequency waveform misfit (e.g. Luo & Schuster, 1991; Sigloch & Nolet, 2006; Nolet, 2008). This popularity results from the robustness of the measurement and its quasi-linear relation to Earth structure that facilitates the solution of the tomographic inverse problem.

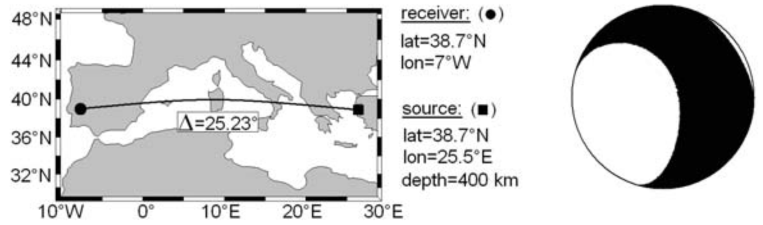

Fig. 6.1 Source-receiver geometry for a deep source beneath western Turkey and a receiver located in western Spain. The epicentral distance is 25.23°. The P wave radiation pattern is plotted to the right.

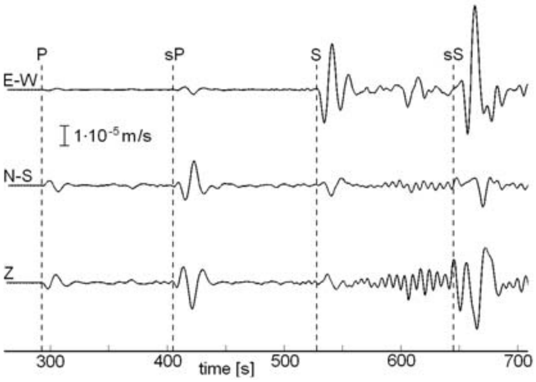

Fig. 6.2 Synthetic velocity seismograms for the source-receiver geometry shown in figure 6.1. The most prominent phases, P, S, sP and sS, are indicated. The dominant period is 15 s.

The critical component of cross-correlations is the isolation of a clearly identifiable waveform that has its own identity. Examples include direct P and S waves or multiply-reflected body waves that do not interfere with other phases. So far, the isolation of one specific waveform has been implicit in most of our developments. Figure 6.3 therefore illustrates in detail the construction of the adjoint source corresponding to the cross-correlation time shift measurement on the vertical component of the direct P wave displacement. The complete vertical-component displacement seismogram, $u_z(t)$, is shown in figure 6.3a. It contains, besides the direct P wave, a clearly distinguishable sP phase. To construct the adjoint source time function, the P wave is isolated from the velocity seismogram, $\dot{u}_z(t)$, using a standard cosine taper, $W(t)$. The windowing produces the waveform, $W(t)\dot{u}_z(t)$, shown in figure 6.3c, which is then scaled by the inverse squared norm of the tapered velocity seismogram, $||W\dot{u}_z||^{-2}$. In the final step, $W(t)\dot{u}_z(t)\,||W\dot{u}_z||^{-2}$ is reversed in time to produce the adjoint source time function (figure 6.3e) that can be used in numerical calculations of Fréchet kernels, where the adjoint equation is solved backwards in time.

Cross-sections through the resulting P wave speed Fréchet kernel, K_{v_P}, are presented in figure 6.4. The large spatial extension of the kernel is in contrast to kernels from infinite-frequency ray theory where all sensitivity is concentrated along the ray. To emphasise this difference, Fréchet kernels corresponding to waveforms with a finite frequency content are commonly referred to as *finite-frequency kernels*. The kernel from figure 6.4 is the prime example of a sensitivity distribution that contradicts our intuition that is largely founded on the visualisation of seismic waves by rays. Along the geometric ray path, indicated by a dashed curve connecting source and receiver, the sensitivity is exactly equal to zero, as noted already by Yomogida (1992). It is strongest in the outer parts of the first Fresnel zone. The negative sign is expected because increasing v_P should lead to earlier-arriving P waves and a reduction in the cross-correlation time shift, as defined in equation (5.16). Marquering et al. (1999) jokingly described the shape of the sensitivity kernel as resembling a hollow banana parallel to the ray and a doughnut perpendicular to the ray. The expression *banana-doughnut kernel* has since then diffused into the seismological literature very efficiently. As a result of the doughnut hole, an anomaly concentrated in the outer part of the first Fresnel zone can give rise to a larger time shift than one located directly on the ray. This result was verified by Hung et al. (2000) with the help of numerical wave propagation. Hung et al. (2000) also noted that a zero time shift for perturbations along the ray path would not be present if it were measured by hand-picking the onset of the arrivals. In a cross-correlation measurement, however, the complete waveform contributes to the time shift $\mathscr{T}$. The cross-correlation time shift can thus be identically zero while the difference

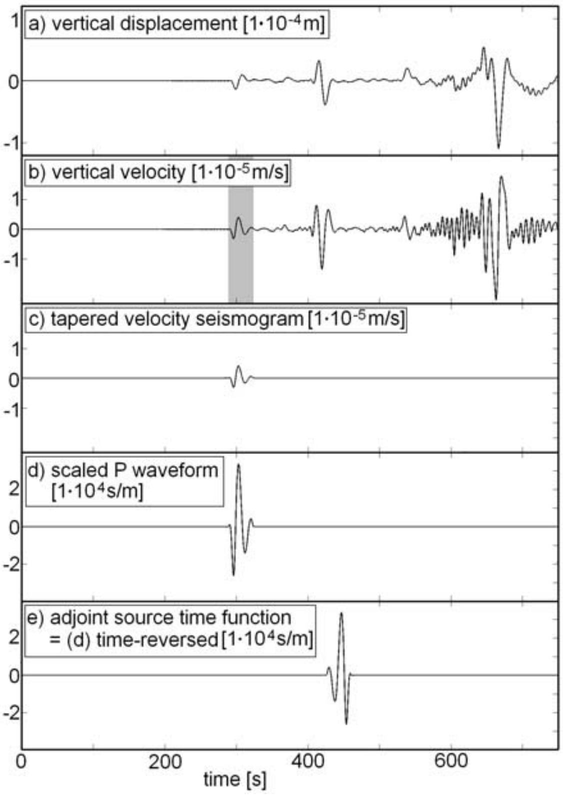

Fig. 6.3 Illustration of the different steps that lead to the construction of the adjoint source time function for a cross-correlation time shift measurement on the 15 s P waveform. **a)** The complete vertical-component displacement seismogram, $u_z(t)$, contains clear P, sP and surface waves. **b)** The velocity seismogram, $\dot{u}_z(t)$, is required for the computation of the adjoint source time function. The location of the window function, $W(t)$, used to isolate the P waveform in the next step, is marked by the gray shaded area. **c)** The isolated P waveform, $W(t)\dot{u}_z(t)$, results from the application of a window function, $W(t)$, to the velocity seismogram. **d)** Isolated P waveform scaled by the inverse squared norm of itself, $||W\dot{u}_z||^{-2}$. **e)** Adjoint source time function, i.e. the time-reversed version of **d)**.

between the P wave onset times is not. A particularly perplexing feature are the weak positive sensitivities within the second Fresnel zone. They can lead to an advance of the synthetic waveform despite the presence of a negative P wave speed anomaly.

Another notable feature in figure 6.4 is the asymmetry of the kernel relative to the geometric ray path. It is most pronounced in the slice at 20°E perpendicular to the ray plane. The asymmetry results from the asymmetric radiation pattern of the forward P wave. Regions of weak sensitivity correspond to nodal planes in the radiation pattern.

A similar asymmetry is only very weakly developed in the vicinity of the receiver where the shape of the kernel is dominated by the characteristics of the adjoint wave field. The adjoint source is a vertical single point force that generates a more isotropic radiation pattern that causes the kernel to be nearly symmetric with respect to the geometric ray.

In our next example we consider the prominent S wave arriving on the E-W component around 530 s. For the construction of the adjoint source time function we proceed as in the case of the direct P wave, the only difference being the location of the time window, $W(t)$, around the E-W-component S waveform. The resulting Fréchet kernel with respect to the S wave speed, K_{v_S}, is shown in figure 6.5. Its general anatomy is similar to the P wave kernel. Only the width of the kernel is reduced. This is due to both the radiation pattern and the shorter wavelength of S waves compared to the wavelength of P waves at the same period.

The variations of this theme are numerous and replete of interesting physics. Particularly noteworthy cases are those where an observed waveform results from P-to-S or S-to-P conversions at material interfaces. The vertical-component sP phase from figure 6.2, for instance, correponds to an S wave that propagates from the source to the surface where it partly converts to a P wave that travels towards the receiver.

The two branches of sP are clearly visible in the Fréchet kernels with respect to the P wave speed, K_{v_P}, and with respect to the S wave speed, K_{v_S}. P wave speed sensitivity concentrates along the ray path that connects the surface reflection point to the receiver. Along the path segment where sP propagates in the form of an S wave, i.e. between the source and the surface reflection point, P wave speed sensitivity is small. S wave speed sensitivity, K_{v_S}, is complimentary to

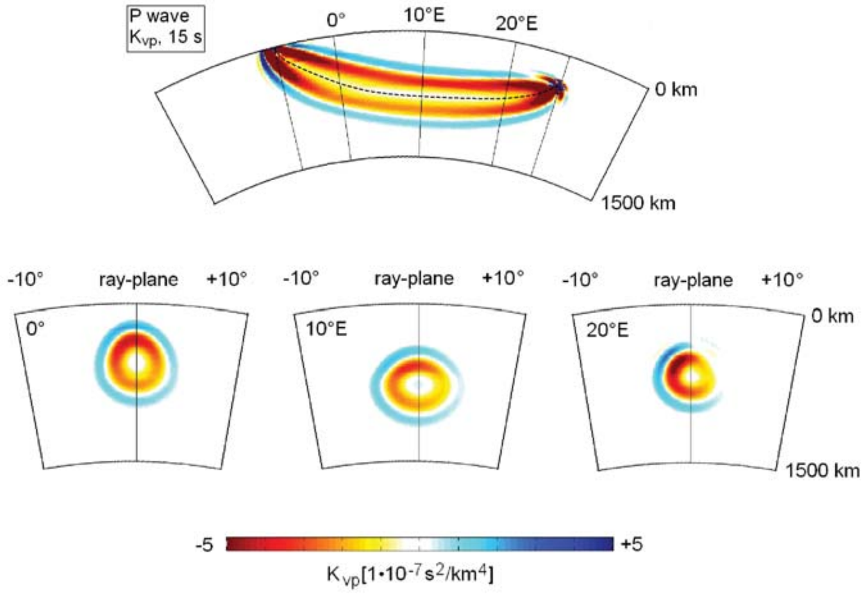

Fig. 6.4 Traveltime kernel, K_{v_P}, corresponding to a 15 s P wave from figure 6.2. The dashed curve marks the geometric ray path. **Top**: Vertical slice through the ray plane connecting source and receiver. Sensitivity is exactly zero along the ray path. **Bottom**: Vertical slices perpendicular to the ray plane at 0°, 10°E and 20°E. Owing to its characteristic shape, the traveltime kernel is often referred to as banana-doughnut kernel.

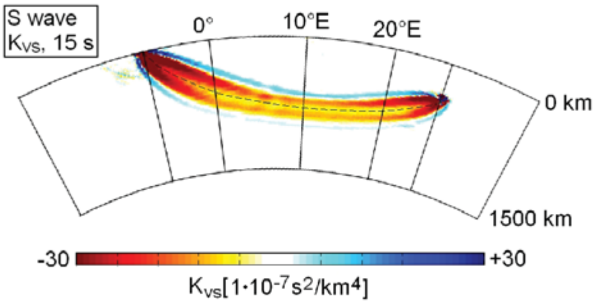

Fig. 6.5 Cross-section through the traveltime Fréchet kernel, K_{v_S}, that corresponds to the E-W-component S wave with a dominant period of 15 s, as shown in figure 6.2

P wave speed sensitivity, K_{v_P}. It is large between the source and the surface reflection point but nearly zero along the P wave segment of the path.

Highly oscillatory sensitivity near the source and the receiver is clearly visible especially in the P wave speed kernel, K_{v_P}. This is because the finite-frequency sP wave is spatially not as isolated as an infinite-frequency sP wave in the ray-theoretical framework. S-to-P scattered waves from the near-source region, for instance, can arrive within the sP time window, and thus affect the cross-correlation measurement. From a practical point of view, highly oscillatory sensitivity is hardly relevant in the solution of a tomographic problem. The projection of the Fréchet kernels onto the basis functions (see chapter 4 and equation 4.22) usually eliminates most of the strong oscillations and emphasises the longer wavelength structure of the kernels.

In our next example for cross-correlation measurements on body waves, we consider the prominent sS arrival on the E-W component. The corresponding Fréchet kernel, shown in figure 6.7, reveals a complex pattern of positive and negative sensitivity that is not as obviously associated with the geometric ray path as the kernels for P, S and sP. Yet, the large sensitivity near the reflection point indicates the implication of the surface in the generation of the sS phase.

All of the examples studied so far, illustrate the delicate relationship between infinite-frequency ray theory and finite-frequency kernels. Despite its limited range of validity, ray theory is an indispensable aid in the interpretation of Fréchet kernels for measurements on isolated seismic phases. Regions of non-zero sensitivity mostly follow

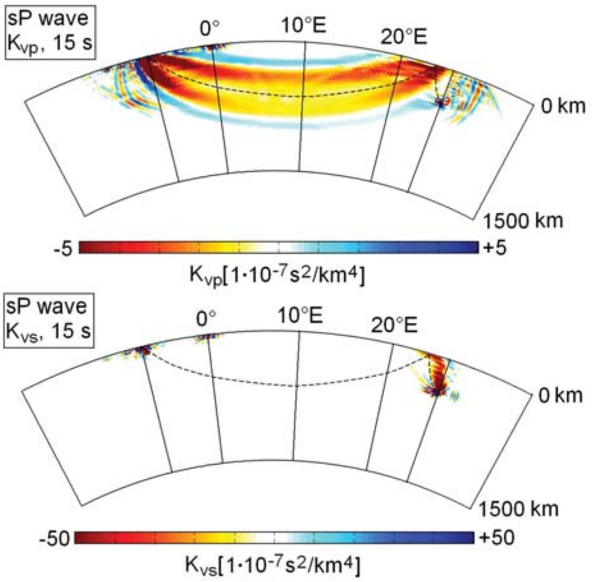

Fig. 6.6 Cross-sections through traveltime Fréchet kernels for the vertical-component sP waveform with a dominant period of 15 s, as shown in figure 6.2. **Top**: Sensitivity with respect to the P wave speed, v_P, (K_{v_P}) is concentrated along the P branch of the ray path. **Bottom**: S wave speed sensitivity (K_{v_S}) is restricted to a region around the s branch of the ray path that represents the S wave propagating from the source to the reflection point at the surface. Highly oscillatory sensitivity appears near the receiver, but is hardly relevant from a practical point of view.

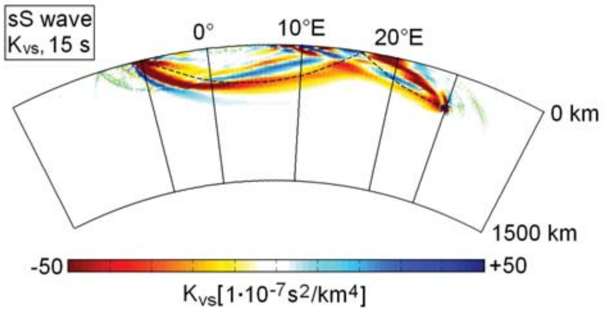

Fig. 6.7 Cross-section through the traveltime Fréchet kernel, K_{v_S}, for the 15 s sS wave observed on the E-W-component in figure 6.2.

the infinite-frequency ray path. However, the fine structure of the kernels is strongly affected by the frequency content, the source radiation pattern, interference effect and the shape of the window function used to isolate a waveform.
As demonstrated by Dahlen et al. (2000), finite-frequency traveltime kernels collapse into infinitely thin rays as the dominant period, T_d, tends to zero. Figure 6.8 illustrates this trend. It shows traveltime kernels for the direct P wave with dominant periods of $T_d \sim 7$ s (top) and $T_d \sim 25$ s (bottom). The width of the first Fresnel zone is proportional to $\sqrt{T_d}$ which explains why the 25 s kernel is nearly twice as wide as the 7 s kernel.

6.2 Surface waves

Following the analysis of body wave Fréchet kernels, we focus on the surface wave part of longer-period seismograms. The finite-frequency sensitivity of surface waves has been studied extensively in recent years. Friederich (1999) calculated 3D Fréchet kernels in radially symmetric Earth models for both body and surface waves within a normal mode framework, and applied his method to the imaging of S velocity structure in the East Asian upper mantle (Friederich, 2003). Also based on normal mode theory, Zhou et al. (2004) derived Fréchet kernels for multi-taper measurements on fundamental-mode surface waves. The kernels were then used to compute a global surface wave tomographic model (Zhou et al., 2005). An extension to higher-mode surface waves can be found in Zhou (2009). Working with the potential representation of surface waves, Yoshizawa & Kennett (2005) were able to derive surface wave kernels for laterally variable media, thus highlighting the importance of using the proper kernels in order to account for the heterogeneity

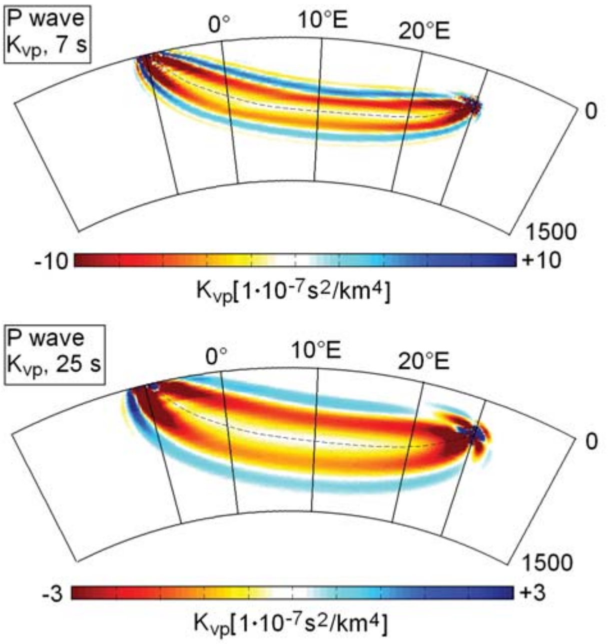

Fig. 6.8 Illustration of the frequency-dependence of P wave speed Fréchet kernels, K_{v_P}, for cross-correlation measurements. **Top**: The dominant period is 7 s, which results in a comparatively slim kernel. **Bottom**: For a dominant period of 25 s the kernel extends far from the geometric ray path and the doughnut hole is particularly pronounced.

in the real Earth. A further improvement was made by Sieminski et al. (2007b) who computed surface wave sensitivity with respect to anisotropic parameters using global spectral-element simulations.

The setup of our numerical modelling is the same as in the previous section (see figure 6.1), the only difference being that the source is shallow (50 km depth) so that strong fundamental-mode surface waves are excited. Figure 6.9 displays the three-component displacement velocity recorded at an epicentral distance of 25.23°. The dominant period is 50 s. Rayleigh waveforms are clearly visible on both the E-W and vertical components. The Love wave is restricted to the N-S component because the propagation is strictly in E-W direction.

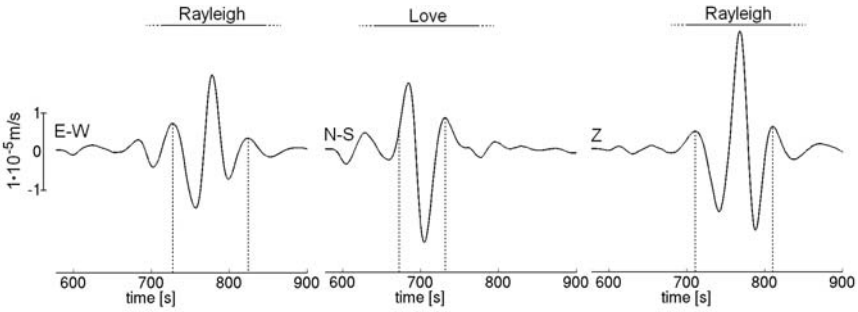

Fig. 6.9 Synthetic surface wave trains with a dominant period of 50 s. The source-receiver geometry is the same as in figure 6.1, but the source depth is only 50 km, so that strong fundamental-mode surface waves are excited. The dashed vertical lines indicate the boundaries of the tapers used to isolate waveforms for the computation of Fréchet kernels.

6.2.1 Isotropic Earth models

To compute the sensitivity of the surface wave trains with respect to perturbations in isotropic Earth structure, we isolate the Love and Rayleigh waves using a cosine taper, the boundaries of which are displayed in the form of vertical dashed lines in figure 6.9. The tapered waveforms mostly consist of fundamental-mode surface waves, but we do not make any special effort to eliminate higher modes using, for instance, multi-taper techniques (e.g. Zhou et al., 2004).

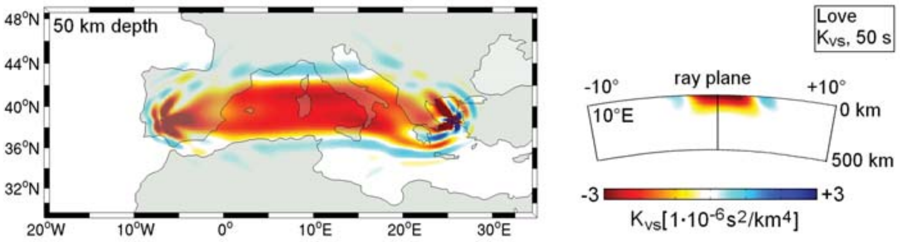

Fig. 6.10 Traveltime sensitivity kernel with respect to the S wave speed, v_S, for the 50 s Love wave from figure 6.9. **Left**: Horizontal slice at 50 km depth. **Right**: Vertical slice at 10°E longitude.

Figure 6.10 shows horizontal and vertical cross sections through the S wave speed traveltime kernel, K_{v_S}, corresponding to the isolated Love waveform in figure 6.9. Sensitivity extends in a bent cigar-shaped region (or banana-shaped, depending on personal preference) from the source to the receiver, and it is restricted to the uppermost 200 km of the Earth model. The kernel exhibits the typical alternating positive and negative sensitivity bands that are separated by zero-sensitivity surfaces where first-order scattering has no effect on the cross-correlation measurement. The characteristic doughnut hole found in the body wave traveltime kernels (e.g. figures 6.4 and 6.5) is not present in the surface wave kernels. This is due to the 2D propagation nature of surface waves (e.g. Zhou et al., 2004). The side band structure of the kernels is most affected by the details of the measurement process, such as the width and slope of the taper used to isolate the waveform. However, within the first Fresnel zone the kernels are rather independent of the measurement details.

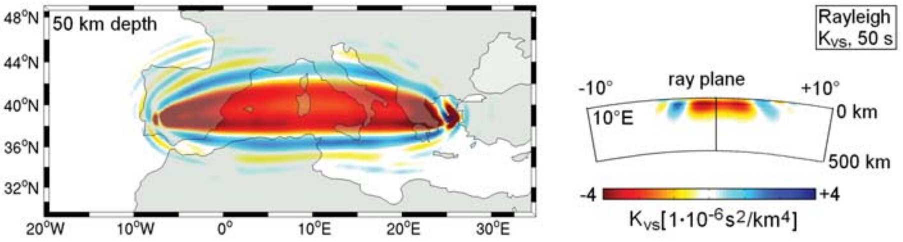

Fig. 6.11 Traveltime sensitivity kernel with respect to the S wave speed, v_S, for the 50 s vertical-component Rayleigh wave from figure 6.9. **Left**: Horizontal slice at 50 km depth. **Right**: Vertical slice at 10°E longitude.

The S wave speed sensitivity of the vertical-component Rayleigh wave is shown in figure 6.11. The anatomy of the kernel parallel to the propagation direction is similar to K_{v_S} for the Love wave (figure 6.10). However, the sensitivity drops to exactly zero near the surface (see also figure 6.13). This is in accordance with analytically derived one-dimensional sensitivity distributions for plane waves (e.g. Takeuchi & Saito, 1972).
As can be seen in figure 6.12, the vertical-component Rayleigh wave is weakly sensitive to the P wave speed, v_P, within the crustal part of the model, that is, above 50 km depth. Love waves, in contrast, exhibit no P wave sensitivity whatsoever. It is therefore reasonable to restrict a surface wave tomography to the S wave structure of the Earth.

The sensitivity distributions of surface waves exhibit a characteristic frequency dependence: The depth extent of the kernels increases with increasing dominant period. This effect is visualised in figure 6.13, which shows vertical profiles through the v_S kernels for Rayleigh waves at periods of 50 s, 100 s and 200 s. A 50 s Rayleigh wave is most sensitive

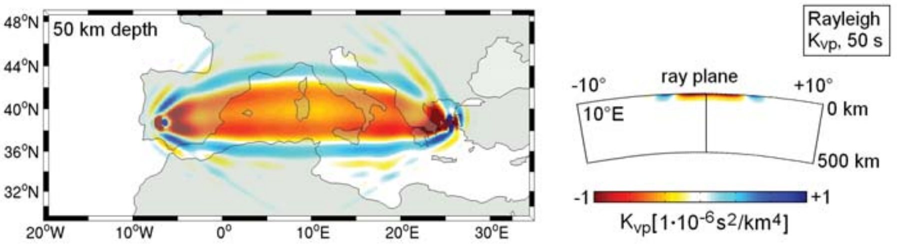

Fig. 6.12 Traveltime sensitivity kernel with respect to the P wave speed, v_{P}, for the 50 s vertical-component Rayleigh wave from figure 6.9. **Left**: Horizontal slice at 50 km depth. **Right**: Vertical slice at 10°E longitude.

to structure around 40 km depth, but is practically unaffected by v_{S} perturbations below 200 km. The 200 s kernel, in contrast, extends to more than 500 km depth, with maximum sensitivity near 200 km. It is important to note that the sharp drop in sensitivity with increasing depth is due to the dominance of the fundamental-mode surface waves within the analysis time window. The depth dependence of Fréchet kernels for higher-mode surface waves is generally more complex (e.g. Takeuchi & Saito, 1972; Zhou, 2009).

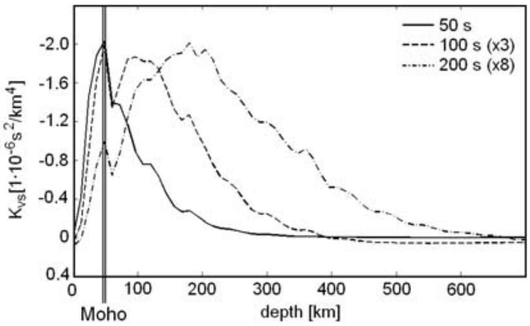

Fig. 6.13 Vertical profiles through the v_{S} kernels for Rayleigh waves at periods of 50 s (solid), 100 s (dashed) and 200 s (dash-dotted). The profiles are located in the centre of the first Fresnel zone at 10°E longitude. To allow for easier comparison, the profiles for 100 s and 200 s are amplified by factors of 3 and 8, respectively. The vertical double line indicates the location of the Moho at around 40 km depth.

6.2.2 *Radial anisotropy*

With the help of equations (4.66) we can calculate the sensitivity of the surface wave trains with respect to anisotropic perturbations, for instance in v_{SH} and v_{SV}.
As expected, the vertical-component Rayleigh wave is primarily sensitive to v_{SV}, which is the propagation speed of a vertically polarised plane shear wave (figure 6.11). Visually, $K_{v_{\mathrm{SV}}}$ and $K_{v_{\mathrm{S}}}$ from figure 6.11 are hardly distinguishable. From equation (4.68) we know that the difference $K_{v_{\mathrm{S}}} - K_{v_{\mathrm{SV}}}$ is equal to $K_{v_{\mathrm{SH}}}$; and indeed, $K_{v_{\mathrm{SH}}}$ as displayed in the lower part of figure 6.11 is barely visible. The small non-zero contributions to $K_{v_{\mathrm{SH}}}$ are mostly the result of Love-Rayleigh coupling (Sieminski et al., 2007a). A Love wave leaving the source is scattered off a δv_{SH} perturbation and partly converted into a vertically polarised shear wave that arrives within the Rayleigh wave window.

In the case of Love waves we observe a similar phenomenon (figure 6.15): The sensitivity with respect to v_{SH} is largest because Love waves are horizontally polarised. From a visual comparison of $K_{v_{\mathrm{SH}}}$ and $K_{v_{\mathrm{S}}}$ from figure 6.10 we find that $K_{v_{\mathrm{SV}}} = K_{v_{\mathrm{S}}} - K_{v_{\mathrm{SH}}}$ should be small, which is confirmed by the slices shown in the upper part of figure 6.15. The v_{SV} kernel suggests that Rayleigh-Love coupling occurs in the vicinity of the source and the receiver, but it is highly inefficient along the ray path.

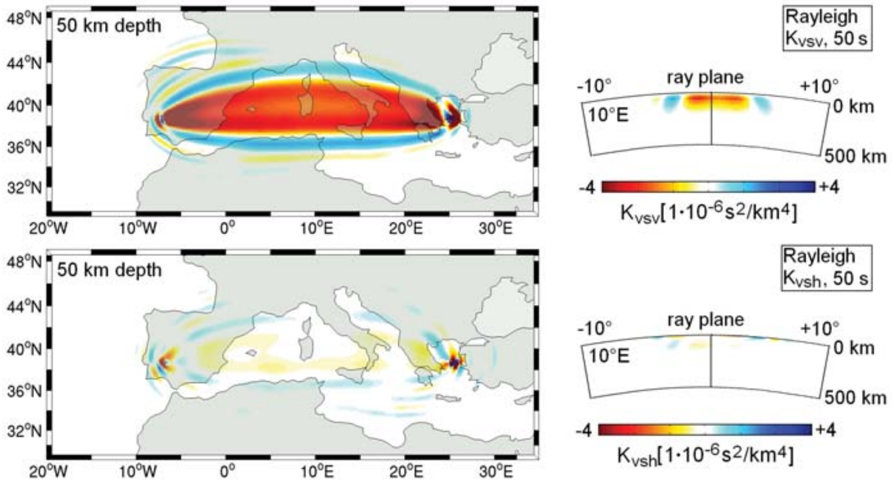

Fig. 6.14 Traveltime sensitivity kernels for the vertical-component Rayleigh wave from figure 6.9. **Top**: Sensitivity with respect to the *SV* wave speed, v_{SV}. **Bottom**: Sensitivity with respect to the *SH* wave speed, v_{SH}.

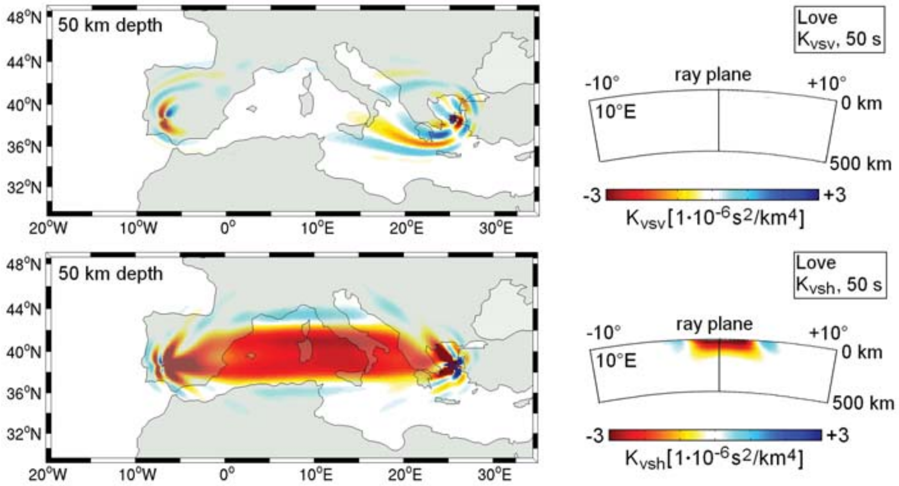

Fig. 6.15 Traveltime sensitivity kernels for the Love wave from figure 6.9. **Top**: Sensitivity with respect to the *SV* wave speed, v_{SV}. **Bottom**: Sensitivity with respect to the *SH* wave speed, v_{SH}.

Appendix A
Mathematical background for the spectral-element method

This appendix gives a brief introduction to the mathematical foundations of the spectral-element method. It is far from being exhaustive but sufficient for most practical purposes. For more complete treatments the reader is referred to Quarteroni et al. (2000), Karniadakis & Sherwin (2005), Pozrikidis (2005) or Allaire (2007).

A.1 Orthogonal polynomials

We consider a family of polynomials, p_n, defined on the interval $[a,b] \subset \mathbb{R}$, and where p_n is of degree $n \geq 0$. The polynomials are said to be *orthogonal* when any of their mutual projections satisfies the condition

$$\int_a^b w(x)\, p_n(x)\, p_m(x)\, dx = A_n \delta_{nm}\,. \tag{A.1}$$

The symbol $w(x)$ denotes a positive weighting function, and A_n is a normalisation constant. Each integration weight together with particular integration limits uniquely determines a family of orthogonal polynomials. The two families that are most relevant in the context of the spectral-element method are the *Legendre polynomials* and the *Lobatto polynomials*.

Legendre polynomials, denoted by $L_n(x)$, are orthogonal with respect to the flat integration weight $w(x) = 1$ and the integration interval $[-1,1]$. In symbols:

$$\int_{-1}^{1} L_n(x)\, L_m(x)\, dx = A_n \delta_{nm}\,. \tag{A.2}$$

The Legendre polynomials are explicitly given by

$$L_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2-1)^n\,, \tag{A.3}$$

and they can be shown to satisfy the *Legendre differential equation*

$$\frac{d}{dx}[(x^2-1)\, \frac{d}{dx} L_n(x)] = n(n+1)\, L_n(x)\,. \tag{A.4}$$

Lobatto polynomials, $Lo_n(x)$, are defined in terms of the Legendre polynomials:

$$Lo_n(x) := \frac{d}{dx} L_{n+1}(x)\,. \tag{A.5}$$

Thus, the Lobatto polynomials satisfy the differential equation

$$\frac{d}{dx}[(x^2-1)Lo_{n-1}(x)] = n(n+1)\,L_n(x)\,. \tag{A.6}$$

The Lobatto polynomials are the family that is orthogonal with respect to the integration weight $w(x) = (1-x^2)$ and the integration interval $[-1,1]$:

$$\int_{-1}^{1} (1-x^2)\,Lo_n(x)\,Lo_m(x)\,dx = A_n\delta_{nm}\,. \tag{A.7}$$

In the following sections we will see that the roots of the Lobatto polynomials play an important role in polynomial interpolation and numerical quadrature.

A.2 Function interpolation

A.2.1 Interpolating polynomial

Finite-element methods in general and the spectral-element method in particular use interpolating functions for the representation of continuous functions that are known exactly only on a finite set of collocation points or grid points. The properties of a finite-element method depend strongly on the interpolation scheme.

Let $f(x)$ be a function that is known at $N+1$ data points x_i, where $i = 1,2,\ldots,N+1$. We want to interpolate the function at an arbitrary point $x \in [x_1, x_{N+1}]$. For this, we replace $f(x)$ by an interpolating function $g(x)$ that satisfies the $N+1$ *interpolation or matching conditions*

$$g(x_i) = f(x_i)\,, \tag{A.8}$$

for $i = 1,2,\ldots,N+1$. The properties of the interpolating function are chosen in accordance with the requirements of a particular application. The most straightforward choice of an interpolating function $g(x)$ is the polynomial of degree N,

$$P_N(x) = a_1x^N + a_2x^{N-1} + \ldots + a_Nx + a_{N+1}\,. \tag{A.9}$$

Applying the interpolation condition (A.8) to the representation of the Nth degree polynomial (A.9) gives a set of $N+1$ linear equations for the polynomial coefficients $a_1, a_2, \ldots, a_{N+1}$:

$$\begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ x_1 & x_2 & \ldots & x_N & x_{N+1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_1^{N-1} & x_2^{N-1} & \ldots & x_N^{N-1} & x_{N+1}^{N-1} \\ x_1^N & x_2^N & \ldots & x_N^N & x_{N+1}^N \end{pmatrix}^T \cdot \begin{pmatrix} a_{N+1} \\ a_N \\ \ldots \\ a_2 \\ a_1 \end{pmatrix} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \ldots \\ f(x_N) \\ f(x_{N+1})\,, \end{pmatrix} \tag{A.10}$$

where the symbol ^{T} indicates matrix transposition. The $(N+1)\times(N+1)$ matrix on the left-hand side of equation (A.10) is the *Vandermonde matrix*, denoted by V. Computing the solution of the system (A.10) via Cramer's rule we find

$$a_i = \frac{\det V_i}{\det V}\,. \tag{A.11}$$

The symbol V_i denotes the Vandermonde matrix where the i^{th} row has been replaced by the right-hand side of equation (A.10). By induction one can show that the *Vandermonde determinant* is explicitly given by

$$\det V = \prod_{i=1}^{N}\prod_{j=i+1}^{N}(x_i - x_j)\,. \tag{A.12}$$

Thus, when the collocation points x_i are mutually distinct, the Vandermonde matrix is nonsingular, and the interpolation problem is well posed.

When the Vandermonde determinant is small, small variations in the right-hand side of equation (A.10) will result in large changes of the polynomial coefficients a_i and in large changes of the interpolated values between the collocation points. This implies that numerical errors, e.g. discretisation errors, will have a smaller effect when the Vandermonde determinant is large. The collocation points that maximise the Vandermonde matrix are called *Fekete points* (see section A.2.4).

A.2.2 Lagrange interpolation

Lagrange interpolation allows us to find an interpolating polynomial without explicitly computing the coefficients of the monomials in equation (A.9). For this we introduce the family of N^{th} degree *Lagrange polynomials*

$$\ell_i^{(N)}(x) := \prod_{k=1,k\neq i}^{N+1} \frac{x-x_k}{x_i-x_k}\,, \qquad i = 1,2,...,N+1\,. \tag{A.13}$$

The Lagrange polynomials satisfy the *cardinal interpolation property*

$$\ell_i^{(N)}(x_j) = \delta_{ij}\,. \tag{A.14}$$

With the definition (A.13), the interpolating polynomial, $P_N(x)$, takes the form

$$P_N(x) = \sum_{i=1}^{N+1} f(x_i)\ell_i^{(N)}(x)\,. \tag{A.15}$$

The Lagrange interpolation is, by construction, exact when $f(x)$ is a polynomial of degree $\leq N$. Choosing, as a special case, $f(x) = (x-a)^m$, where a is a constant and $m = 0,1,...,N$, gives

$$(x-a)^m = \sum_{i=1}^{N+1} (x_i-a)^m \ell_i^{(N)}(x)\,. \tag{A.16}$$

For $m = 0$ we obtain the *first Cauchy relation*

$$1 = \sum_{i=1}^{N+1} \ell_i^{(N)}(x)\,, \tag{A.17}$$

and for $a = x$ the *second Cauchy relation*

$$0 = \sum_{i=1}^{N+1} (x_i-x)^m \ell_i^{(N)}(x)\,. \tag{A.18}$$

The Cauchy relations (A.17) and (A.18) will play an important role in the following paragraphs. As an alternative to the definition (A.13), we can represent the Lagrange polynomials in term of the *generating polynomial*

$$\Phi_{N+1}(x) := \prod_{i=1}^{N+1} (x-x_i) = (x-x_1)(x-x_2)\cdot\ldots\cdot(x-x_{N+1})\,. \tag{A.19}$$

Differentiating (A.19) and substituting the result into (A.13), gives the identity

$$\ell_i^{(N)}(x) = \frac{1}{(x-x_i)}\frac{\Phi_{N+1}(x)}{\dot{\Phi}_{N+1}(x_i)}\,. \tag{A.20}$$

The dot in equation (A.19) denotes a differentiation with respect to the independent variable x. Equation (A.20) can be used to derive expressions for the derivatives of the Lagrange polynomials, $\dot{\ell}_i^{(N)}$, at the collocation points x_j:

$$\dot{\ell}_i^{(N)}(x_j) = \begin{cases} \frac{1}{(x_j - x_i)} \frac{\dot{\Phi}_{N+1}(x_j)}{\dot{\Phi}_{N+1}(x_i)} & \text{if} \quad x_i \neq x_j \\ \frac{\ddot{\Phi}_{N+1}(x_i)}{2\dot{\Phi}_{N+1}(x_i)} & \text{if} \quad x_i = x_j \, . \end{cases} \tag{A.21}$$

For our discussion of interpolation errors and Fekete points we will need a representation of the Lagrange polynomials in terms of the Vandermonde matrix. For this, we set $a = 0$ in equation (A.16) and obtain a set of $N+1$ equations $-$ one for each $m = 0, 1, ..., N$:

$$\begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ x_1 & x_2 & ... & x_N & x_{N+1} \\ ... & ... & ... & ... & ... \\ x_1^{N-1} & x_2^{N-1} & ... & x_N^{N-1} & x_{N+1}^{N-1} \\ x_1^N & x_2^N & ... & x_N^N & x_{N+1}^N \end{pmatrix} \cdot \begin{pmatrix} \ell_1^{(N)}(x) \\ \ell_2^{(N)}(x) \\ ... \\ \ell_N^{(N)}(x) \\ \ell_{N+1}^{(N)}(x) \end{pmatrix} = \begin{pmatrix} 1 \\ x \\ ... \\ x^{N-1} \\ x^N \, . \end{pmatrix} \tag{A.22}$$

Applying Cramer's rule to the linear system (A.22) gives explicit expressions for the Lagrange polynomials:

$$\ell_i^{(N)}(x) = \frac{\det V(x_i = x)}{\det V} \, . \tag{A.23}$$

Having introduced the basic concepts of polynomial interpolation, we will now consider suitable choices of the collocation points x_i.

A.2.3 Lobatto interpolation

In section (A.2.1) we already mentioned that it is desirable to use Fekete points as collocation points for polynomial interpolation. To systematically derive Fekete points, we first have to introduce Lobatto interpolation as a special case of Lagrange interpolation.

Without loss of generality, we restrict our attention to the interval $[-1, 1]$. For a given set of collocation points x_i, with $i = 1, 2, ..., N+1$, the Lagrange polynomials satisfy the cardinal interpolation property $\ell_i^{(N)}(x_j) = \delta_{ij}$. In addition to this mandatory condition we shall now require that the collocation point distribution is such that the $N-1$ optional conditions

$$\dot{\ell}_i^{(N)}(x_i) = 0 \, , \tag{A.24}$$

are satisfied for $i = 2, 3, ..., N$, i.e. for the internal nodes. Property (A.24) ensures that the Lagrange polynomial $\ell_i^{(N)}(x)$ reaches a local maximum value of 1 at the internal collocation point x_i. Interestingly, this additional requirement uniquely specifies the internal node points as the roots of the Lobatto polynomial $Lo_{N-1}(x)$.

To prove this statement, we first note that condition (A.24) together with equation (A.21) implies that the second derivative of the generating polynomial, $\Phi_{N+1}(x)$ vanishes at the node points x_i, with $i = 2, 3, ..., N$:

$$\ddot{\Phi}_{N+1}(x_i) = 0 \, , \tag{A.25}$$

Since $\ddot{\Phi}(x)$ is a polynomial of degree $N-1$ that vanishes at the internal node points, just as $\Phi_{N+1}(x)$ itself, we must have

$$\ddot{\Phi}_{N+1}(x) = c \, \frac{\Phi_{N+1}(x)}{(x-1)(x+1)} = -c \, \frac{\Phi_{N+1}(x)}{(1-x^2)} \, , \tag{A.26}$$

where c is a constant. Equation (A.26) holds because a polynomial is uniquely determined by its zeros. The coefficient of the highest-power monomial x^{N+1} of Φ_{N+1} is equal to 1. This implies $c = N(N+1)$, and therefore

$$(1-x^2)\ddot{\Phi}_{N+1}(x)+N(N+1)\Phi_{N+1}(x)=0\,. \tag{A.27}$$

The generating polynomial satisfies the differential equation (A.27). From equations (A.4) and (A.5) we infer that the solution of (A.27) is

$$\Phi_{N+1}(x)=a\,(x^2-1)\,\dot{L}_N(x)=a\,(x^2-1)\,Lo_{N-1}(x)\,, \tag{A.28}$$

where a is a normalisation constant. Thus, the collocation points that satisfy the optional condition (A.24) are the roots of the *completed Lobatto polynomial*

$$Lo^c_{N-1}(x):=(1-x^2)\,Lo_{N-1}(x)\,. \tag{A.29}$$

They are commonly referred to as *Gauss-Lobatto-Legendre points* (GLL points). Figure A.1 shows the Lagrange polynomials of degree 3 to 6 with the GLL points as collocation points.

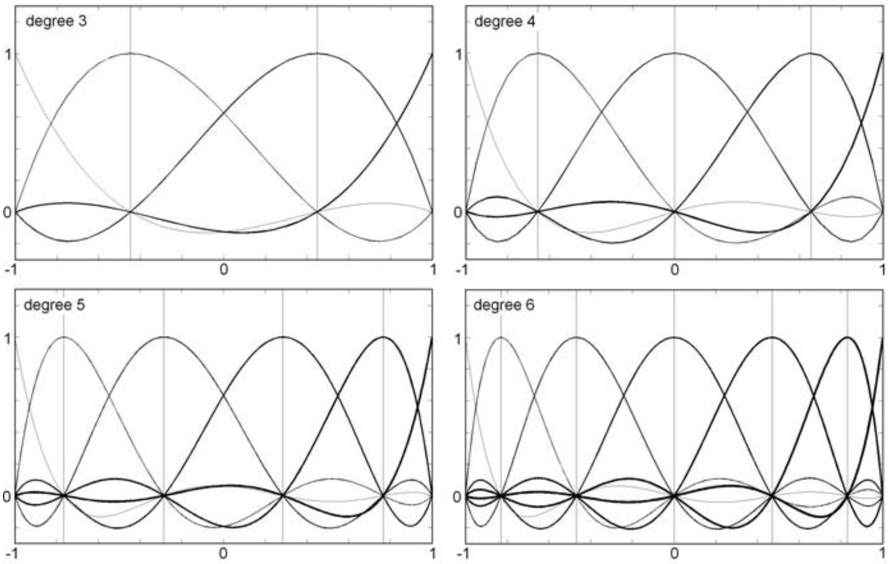

Fig. A.1 The Lagrange polynomials of degree 3 to 6. The collocation points, indicated by vertical lines, are the Gauss-Lobatto-Legendre (GLL) points.

Using the GLL points for polynomial interpolation has an important consequence: The absolute values of the corresponding Lagrange polynomials are smaller or equal to 1; in symbols:

$$|\ell_i^{(N)}(x)|\leq 1\,,\qquad x\in[-1,1]\,. \tag{A.30}$$

This property is interesting in the context of Runge's phenomenon. Runge's phenomenon, illustrated in figure A.2, consist in the over-shooting of high-order interpolants near the edges of the interpolation interval when equidistant collocation points are used. This undesirable effect can be avoided with Lobatto interpolation. The possibility to suppress Runge's phenomenon is essential for the use of high-order interpolation in the spectral-element method.

To prove relation (A.30) we consider the $2N$-degree polynomial

$$Q_{2N}(x):=[\ell_1^{(N)}(x)]^2+[\ell_2^{(N)}(x)]^2+...+[\ell_{N+1}^{(N)}(x)]^2-1\,. \tag{A.31}$$

The cardinal interpolation property ensures that

$$Q_{2N}(x_i)=0\,,\qquad i=1,2,...,N+1\,. \tag{A.32}$$

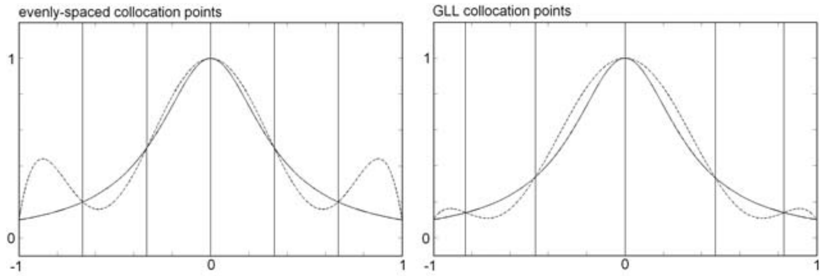

Fig. A.2 Illustration of Runge's phenomenon. Runge's function, $f(x) = (1+ax)^{-1}$ with $a = 3$ is plotted as solid curve and the interpolants as dashed curves. The interpolation of Runge's function with equidistant collocation points (left) leads to a strong over-shooting of the interpolant near the edges of the interval $[-1,1]$. This undesirable effect can be suppressed by choosing the GLL points as collocation points (right).

Evaluating the derivative of $Q_{2N}(x)$ at the internal collocation points, gives

$$\dot{Q}_{2N}(x_j) = 2\sum_{i=1}^{N+1} \dot{\ell}_i^{(N)}(x_j)\,\ell_i^{(N)}(x_j)\,, \tag{A.33}$$

and in the light of the optional condition (A.24)

$$\dot{Q}_{2N}(x_j) = 0\,, \qquad j = 2,3,\ldots,N\,. \tag{A.34}$$

Equations (A.32) and (A.34) imply that the internal collocation points are double roots of $Q_{2N}(x)$. We can therefore express $Q_{2N}(x)$ in terms of the generating polynomial, $\Phi_{N+1}(x)$ as follows:

$$Q_{2N}(x) = d\,\frac{\Phi_{N+1}^2(x)}{(x+1)(x-1)} = -d\,\frac{\Phi_{N+1}^2(x)}{(1-x^2)}\,, \tag{A.35}$$

where d is a constant. Substituting (A.28) into (A.35) yields

$$Q_{2N}(x) = -b\,(1-x^2)\,Lo_{N-1}^2(x)\,, \tag{A.36}$$

with a new constant b. To determine b, we evaluate $\dot{Q}_{2N}$ at the collocation point $x = 1$ with the help of equation A.21:

$$\dot{Q}_{2N}(1) = 2\dot{\ell}_{N+1}^{(N)}(1) = \frac{\ddot{\Phi}_{N+1}(1)}{\dot{\Phi}_N(1)}\,. \tag{A.37}$$

Again substituting $a\,(x^2-1)\,Lo_{N-1}(x)$ for $\Phi_{N+1}(x)$, gives

$$\dot{Q}_{2N}(1) = \left.\frac{\partial_x^2[(x^2-1)\,Lo_{N-1}(x)]}{\partial_x[(x^2-1)\,Lo_{N-1}(x)]}\right|_{x=1}\,. \tag{A.38}$$

The differential equation (A.6) satisfied by the Legendre and Lobatto polynomials, allows us to simplify (A.38):

$$\dot{Q}_{2N}(1) = \frac{Lo_{N-1}(1)}{L_N(1)}\,. \tag{A.39}$$

The Legendre polynomials are normalised such that $L_N(1) = 1$, and for the Lobatto polynomials evaluated at $x = 1$ we infer from (A.6)

$$N(N+1)\,L_N(1) = N(N+1) = 2x\,Lo_{N-1}(x)|_{x=1} + (x^2-1)\,\partial_x Lo_{N-1}(x)|_{x=1} = 2Lo_{N-1}(1)\,. \tag{A.40}$$

Thus, we have

$$\dot{Q}_{2N}(1) = \frac{1}{2}N(N+1)\,. \tag{A.41}$$

Directly differentiating (A.36) yields

$$\dot{Q}_{2N}(1) = 2b\,Lo_{N-1}^2(1) = \frac{1}{2}bN^2(N+1)^2\,, \tag{A.42}$$

and therefore

$$b = \frac{1}{N(N+1)}\,. \tag{A.43}$$

Combining (A.36) and (A.43) gives the final expression of $Q_{2N}(x)$ in terms of the generating polynomial:

$$Q_{2N}(x) = -\frac{1-x^2}{N(N+1)}\,Lo_{N-1}^2(x)\,. \tag{A.44}$$

Equation (A.44) implies $Q_{2N}(x) \leq 0$ and therefore

$$\sum_{i=1}^{N+1}[\ell_i^{(N)}(x)]^2 \leq 1\,. \tag{A.45}$$

Thus, the relation $|\ell_i^{(N)}(x)| \leq 1$ holds for each individual $\ell_i^{(N)}(x)$.

A.2.4 Fekete points

Fekete points are the collocation points that maximise the Vandermonde determinant (see section A.2.1). Taking the Fekete points as collocation points ensures that small variations of the $f(x_i)$ − due for example to numerical inaccuracies or measurement errors − result in the smallest possible variations of the interpolated values between the grid points. In this paragraph we demonstrate that the GLL points are the Fekete points of the Vandermonde determinant. The grid points $x_1 = -1$ and $x_{N+1} = 1$ are assumed fixed.

The argument is very simple: Inside the interval $[-1,1]$ the Lagrange polynomials are smaller than or equal to 1 (relation A.30), provided that the internal collocation points are the zeros of $Lo_{N-1}(x)$. The cardinal interpolation property ensures that $\ell_i^{(N)}(x_i) = 1$. Therefore, the Lagrange polynomial $\ell_i^{(N)}(x)$, with i between 2 and $N-1$, reaches a local maximum at $x = x_i$. Thus, we have $\dot{\ell}_i^{(N)}(x_i) = 0$. (This is just the optional condition (A.24).) Using the expression of the Lagrange polynomials in terms of the Vandermonde determinant (equation A.23), yields

$$\frac{d}{dx}\ell_i^{(N)}(x)|_{x=x_i} = \frac{d}{dx}\frac{\det V(x_i = x)}{\det V}\bigg|_{x=x_i} = \frac{1}{\det V}\frac{d}{dx_i}\det V = 0\,, \tag{A.46}$$

for $i = 2,3,...,N-1$, and therefore

$$\frac{d}{dx_i}V = 0\,, \qquad i = 2,3,...,N-1\,. \tag{A.47}$$

Since the internal collocation points are local maxima, we also have

$$0 > \frac{d^2}{dx^2}\ell_i^{(N)}(x)|_{x=x_i} = \frac{d^2}{dx^2}\frac{\det V(x_i = x)}{\det V}\bigg|_{x=x_i} = \frac{1}{\det V}\frac{d^2}{dx_i^2}\det V\,, \tag{A.48}$$

for $i = 2,3,...,N-1$, and

$$\frac{d^2}{dx_i^2}\det V < 0\,, \qquad i = 2,3,...,N-1\,. \tag{A.49}$$

Relations (A.47) and (A.49) imply that the zeros of $Lo_{N-1}(x)$ are indeed the Fekete points. The maximum of the Vandermonde matrix is global, but a proof of this statement is clearly beyond the scope of this brief overview (see Szegö, 1975).

A.2.5 Interpolation error

We consider the max norm of the interpolation error $e(x)$:

$$||e(x)||_\infty := \max_{x\in[a,b]} |f(x) - P_N(f,x)|\,. \tag{A.50}$$

In equation (A.50) the dependence of the interpolating polynomial P_N on the function $f(x)$ is made explicit through the notation $P_N = P_N(f,x)$. Of all N^{th}-degree polynomials approximating the function $f(x)$, there is an optimal polynomial $P_N^{opt}(f,x)$ such that $||e(x)||_\infty$ is minimal. This minimal value of $||e(x)||_\infty$ is the *minimax error*, denoted by ε_N. The optimal polynomial $P_N^{opt}(f,x)$ is not necessarily an interpolating polynomial. The max norm $||e(x)||_\infty$ is related to the minimax error ε_N through the relation

$$\begin{aligned}||e(x)||_\infty &= ||P_N(f,x) - f(x)||_\infty = ||P_N(f,x) - P_N^{opt}(f,x) + P_N^{opt}(f,x) - f(x)||_\infty \\ &\le ||P_N(f,x) - P_N^{opt}(f,x)||_\infty + ||P_N^{opt}(f,x) - f(x)||_\infty = ||P_N(f,x) - P_N^{opt}(f,x)||_\infty + \varepsilon_N\,.\end{aligned} \tag{A.51}$$

The first summand can be transformed as follows:

$$||P_N(f,x) - P_N^{opt}(f,x)||_\infty = ||P_N(f,x) - P_N(P_N^{opt},x)||_\infty \le ||P_N||_\infty ||f(x) - P_N^{opt}(f,x)||_\infty = ||P_N||_\infty \varepsilon_N\,. \tag{A.52}$$

In equation (A.52) P_N is interpreted as an operator that acts on the function $f(x)$, and $||P_N||_\infty$ is the operator norm of P_N. For a fixed set of collocation points, P_N is linear in $f(x)$ because the polynomial coefficients are linear functions of the data values $f(x_i)$. Combining (A.51) and (A.52) gives

$$||e(x)||_\infty \le (1 + ||P_N||_\infty)\,\varepsilon_N\,. \tag{A.53}$$

The max norm of the interpolation operator, $||P_N||_\infty$, can be expressed in terms of the Lagrange polynomials:

$$\begin{aligned}||P_N||_\infty &= \sup_f \frac{||P_N(f,x)||_\infty}{||f(x)||_\infty} = \sup_f \frac{||\sum_{i=1}^{N+1} f(x_i)\ell_i^{(N)}(x)||_\infty}{||f(x)||_\infty} \\ &= \sup_f \frac{\max_{x\in[a,b]} |\sum_{i=1}^{N+1} f(x_i)\ell_i^{(N)}(x)|}{\max_{x\in[a,b]} |f(x)|} \le \max_{x\in[a,b]} \left|\sum_{i=1}^{N+1} \ell_i^{(N)}(x)\right| \le \max_{x\in[a,b]} \sum_{i=1}^{N+1} |\ell_i^{(N)}(x)|\,.\end{aligned} \tag{A.54}$$

The function

$$L_N(x) := \sum_{i=1}^{N+1} |\ell_i^{(N)}(x)|\,, \tag{A.55}$$

is the *Lebesque function* and its max norm, denoted by Λ_N, is the *Lebesque constant*. Using this terminology, equation (A.53) can be transformed to

$$||e(x)||_\infty \le (1 + \Lambda_N)\,\varepsilon_N\,. \tag{A.56}$$

Relation (A.56) reveals that there are two separate contributions to the interpolation error: (1) The minimax error ε_N depends only on the smoothness of the function $f(x)$. Smooth functions produce smaller minimax errors than rough functions. We will not further consider the influence of ε_N because we have no freedom of choice concerning the function that we need to interpolate. (2) The Lebesque constant, Λ_N, depends only on the locations of the collocation points. It is desirable to choose the collocation points such that Λ_N is minimised.

The dependence of the Lebesque constant on the polynomial order has received much attention in the literature. (See

Hesthaven (1998) for a summary of important results.) When the collocation points are equidistant, the associated Lebesque constant, Λ_N^{equi}, can be shown to exhibit the asymptotic behaviour

$$\Lambda_N^{equi} \sim \frac{2^N}{N \log N}\,. \tag{A.57}$$

This renders high-order polynomial interpolation with equidistant nodes practically useless. That the Lebesque constant can not be made arbitrarily small is the content of Erdös' theorem. It states that

$$\Lambda_N > \frac{2}{\pi} \log(N+1) - c\,, \tag{A.58}$$

where c is a positive constant. This holds for any set of collocation points. A nearly logarithmic growth of Λ_N can be achieved when the roots of the Chebyshev polynomials are used as collocation points. For the GLL points we infer from equation (A.30) that

$$\left(\Lambda_N^{GLL}\right)^2 = \left(\sum_{i=1}^{N+1} |\ell_i^{(N)}(x)|\right)^2 \leq \sum_{i=1}^{N+1} |\ell_i^{(N)}(x)|^2 \leq N+1\,, \tag{A.59}$$

and therefore

$$\Lambda_N^{GLL} \leq \sqrt{N+1}\,. \tag{A.60}$$

The estimate (A.60) seems to be too pessimistic in practice. Based on numerical experiments, Hesthaven (1998) conjectures that Λ_N^{GLL} is bounded as

$$\Lambda_N^{GLL} \leq \frac{2}{\pi} \log(N+1) + 0.685\,. \tag{A.61}$$

Thus, the Lebesque constant associated with the GLL points is nearly optimal.

A.3 Numerical integration

The following paragraphs are concerned with the derivation of numerical quadrature formulas that are used in the context of the spectral-element method. The general strategy is to replace the function that we wish to integrate, $f(x)$, by an interpolating polynomial and to solve the resulting integral analytically.

A.3.1 Exact numerical integration and the Gauss quadrature

Consider the weighted integral of a degree-$(2N+1)$ polynomial Q_{2N+1}

$$\int_a^b w(x) Q_{2N+1}(x)\, dx\,, \tag{A.62}$$

where $w(x)$ is a positive weighting function. We can write Q_{2N+1} as the sum of a degree-N polynomial $P_N(x)$ that interpolates Q_{2N+1} at the collocation points $x_1, x_2, ..., x_{N+1}$ and another polynomial of degree $2N+1$:

$$Q_{2N+1}(x) = P_N(x) + R_N(x)\,(x-x_1)(x-x_2)\cdot\ldots\cdot(x-x_{N+1})\,. \tag{A.63}$$

The polynomial $R_N(x)$ is of degree N. We are interested in the integration error, ε, incurred by replacing the degree-$(2N+1)$ polynomial Q_{2N+1} by the degree-N polynomial $P_N(x)$:

$$\varepsilon = \int_a^b w(x) Q_{2N+1}(x)\, dx - \int_a^b w(x) P_N(x)\, dx = \int_a^b w(x) R_N(t)\, (x - x_1)(x - x_2) \cdot \ldots \cdot (x - x_{N+1})\, dx\,. \tag{A.64}$$

The integration error is equal to zero when the $N+1$ collocation points x_i are the roots of the degree-$(N+1)$ orthogonal polynomial $p_{N+1}(x)$ that corresponds to the integration weight $w(x)$. To prove this assertion, we note that $p_{N+1}(x)$ is proportional to $(x-x_1)(x-x_2)\cdot\ldots\cdot(x-x_{N+1})$ when each x_i is a root of $p_{N+1}(x)$. Therefore, we have

$$(x - x_1)(x - x_2) \cdot \ldots \cdot (x - x_{N+1}) = c\, p_{N+1}(x)\,, \tag{A.65}$$

where c is a constant. The degree-N polynomial $R_N(x)$ can be expressed through the orthogonal polynomials up to degree N:

$$R_N(x) = \sum_{i=1}^{N} c_i p_i(x)\,. \tag{A.66}$$

The numbers c_i, with $i = 1, 2, ..., N$, are the expansion coefficients. Combining equations (A.64) to (A.66) and using the orthogonality of the polynomials $p_i(x)$, yields

$$\varepsilon = \sum_{i=1}^{N} c \int_a^b c_i p_i(x) p_{N+1}(x)\, dx = 0\,. \tag{A.67}$$

Thus, we can integrate a degree-$(2N+1)$ polynomial exactly with only $N+1$ collocation points, given that the collocation points are the roots of the degree-$(N+1)$ orthogonal polynomial that corresponds to the integration weight $w(x)$.
In the case where $f(x)$ is any function, not necessarily a polynomial, we can construct working formulas that approximate the integral. For this, we replace $f(x)$ by its interpolating polynomial

$$P_N(x) = \sum_{i=1}^{N+1} f(x_i)\, \ell_i^{(N)}(x)\,, \tag{A.68}$$

and introduce this approximation into the weighted integral,

$$\int_a^b w(x) f(x)\, dx \approx \int_a^b w(x) P_N(x)\, dx = \sum_{i=1}^{N+1} w_i f(x_i)\,. \tag{A.69}$$

The *integration weights*, w_i, are independent of $f(x)$:

$$w_i = \int_a^b w(x)\, \ell_i^{(N)}(x)\, dx\,. \tag{A.70}$$

Even when $f(x)$ is not a polynomial, the collocation points should be the roots of an orthogonal polynomial, simply because $f(x)$ is closer to a degree-$(2N+1)$ polynomial than to a degree-N polynomial. Equation (A.69) is known as *Gauss quadrature rule*.

A.3.2 Gauss-Legendre-Lobatto quadrature

A disadvantage of the Gauss quadrature – especially in the context of the spectral-element method – is that the roots of orthogonal polynomials are generally located inside the integration interval $[a, b]$ but never directly on its boundaries. Explicitly imposing the requirement that two collocation points coincide with the boundaries leads to

Gauss-Lobatto-Legendre (GLL) quadrature formulas.

In the interest of simplicity we consider the integration interval $[-1,1]$ and the flat weighting function $w(x)=1$. The integral over a degree-$(2N-1)$ polynomial Q_{2N-1} is then

$$\int_{-1}^{1} Q_{2N-1}(x)\,dx\,. \tag{A.71}$$

Following the developments of the previous section, we decompose Q_{2N-1} into an interpolating polynomial of degree N and a polynomial of degree $2N-1$:

$$Q_{2N-1}(x) = P_N(x) + R_{N-2}(x)\,(x+1)(x-x_2)\cdot\ldots\cdot(x-x_N)(x-1)\,. \tag{A.72}$$

The collocation points -1 and 1 are now imposed explicitly. For the integration error we find

$$\varepsilon = -\int_{-1}^{1}(1-x^2)\,R_{N-2}(x)\,(x-x_2)(x-x_3)\cdot\ldots\cdot(x-x_N)\,dx\,. \tag{A.73}$$

To make ε vanish, we need to choose the $N-1$ internal collocation points $x_2,x_3,\ldots,x_N$ such that they are the roots of the degree-$(N-1)$ orthogonal polynomial that corresponds to the integration weight $(1-x^2)$, i.e. to the Lobatto polynomial Lo_{N-1}. Thus, a polynomial of degree $2N-1$ can be integrated exactly when replaced by the degree-N interpolating polynomial $P_N(x)$. The collocation points are -1, 1 and the $N-1$ roots of the Lobatto polynomial Lo_{N-1}.

In the case of an arbitrary function $f(x)$, we approximate the integral over $f(x)$ with the integral over the interpolating polynomial $P_N(x)$:

$$\int_{-1}^{1} f(x)\,dx \approx \int_{-1}^{1} P_N(x)\,dx = \sum_{i=1}^{N+1} w_i f(x_i)\,, \tag{A.74}$$

The integration weights, w_i, are

$$w_i = \int_{-1}^{1} \ell_i^{(N)}(x)\,dx\,. \tag{A.75}$$

Equation (A.74) is referred to as *Gauss-Lobatto-Legendre quadrature rule*.

Cited literature and further reading material

1. Abarbanel, S., Gottlieb,D.: A mathematical analysis of the PML method. J. Comp. Phys. **134**, 357-363 (1997)
2. Abdulah, A.: Seismic body wave attenuation tomography beneath the Australasian region. PhD Thesis, The Australian National University (2007)
3. Aki, K., Christoffersson, A., Husebye, E. S.: Determination of the three-dimensional seismic structure of the lithosphere. J. Geophys. Res. **82**, 277-296 (1977)
4. Aki, K., Richards, P. G.: Quantitative Seismology, second edition. University Science Books (2002)
5. Alford, R. M., Kelly, K. R., Boore, D. M.: Accuracy of finite difference modeling of the acoustic wave equation. Geophysics **39**, 834-842 (1974)
6. Allaire, G.: Numerical analysis and optimization. Oxford University Press (2007)
7. Alterman, Z., Karal, F. C.: Propagation of elastic waves in layered media by finite-difference methods. Bull. Seis. Soc. Am. **58**, 367-398 (1968)
8. Anderson, D. L., Archambeau, C.: Anelasticity of Earth. J. geophys. Res. **69**, 2071-2084 (1964)
9. Askan, A., Akcelik, V., Bielak, J., Ghattas, O.: Full waveform inversion for seismic velocity and anelastic losses in heterogeneous media. Bull. Seis. Soc. Am. **97**, 1990-2008 (2007)
10. Askan, A., Bielak, J.: Full anelastic waveform tomography including model uncertainty. Bull. Seis. Soc. Am. **98**, 2975-2989 (2008)
11. Babuska, V., Cara, M.: Seismic Anisotropy in the Earth. Kluwer Academic Publishers, Dordrecht, The Netherlands (1991)
12. Backus, G. E.: Long-wave elastic anisotropy produced by horizontal layering. J. Geophys. Res. **67**(11), 4427-4440 (1962)
13. Baig, A. M., Dahlen, F. A., Hung, S.-H.: Traveltimes of waves in three-dimensional random media. Geophys. J. Int. **153**, 467-482 (2003)
14. Bamberger, A., Chavent, G., Lailly, P.: Une application de la théorie du contrôle à un problème inverse sismique. Ann. Geophys. **33**, 183-200 (1977)
15. Bamberger, A., Chavent, G., Lailly, P.: About the stability of the inverse problem in 1-D wave equations - application to the interpretation of seismic profiles. Appl. Math. Opt. **5**, 1-47 (1979)
16. Bamberger, A., Chavent, G., Hemons, Ch., Lailly, P.: Inversion of normal incidence seismograms. Geophysics **47**(5), 757-770 (1982)
17. Bao, H., Bielak, J., Ghattas, O., Kallivokas, L. F., O'Hallaron, D. R., Shewchuk, J. R., Xu, J.: Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. Comp. Meth. Appl. Mech. Eng. **152**, 85-102 (1998)
18. Bassin, C., Laske, G. and Masters, G.: The Current Limits of Resolution for Surface Wave Tomography in North America. EOS Trans. AGU **81**, F897 (2000)
19. Båth, M.: Introduction to seismology, 2nd edition. Birkhäuser, Basel, Boston, Stuttgart (1979)
20. Bécache, E., Fauqueux, S., Joly, P.: Stability of perfectly matched layers, group velocities and anisotropic waves. J. Comp. Phys. **188**, 399-433 (2003)
21. Ben Hadj Ali, H., Operto, S., Virieux, J., Sourbier, F.: Efficient 3D frequency-domain full waveform inversion with phase encodings. In: 71. Conference and Technical Exhibition, EAGE, Extended Abstracts, 5812 (2009a)
22. Ben Hadj Ali, H., Operto, S., Virieux, J., Sourbier, F.: Three-dimensional frequency-domain full waveform inversion with phase encoding. In 79. SEG Meeting, Extended Abstracts, 2288-2292 (2009b)
23. Bérenger, J.-P.: A perfectly matched layer for the absorption of electromagnetic waves. J. Comp. Phys. **114**, 185-200 (1994)
24. Blanch, J. O., Robertsson, J. O. A., Symes, W. W.: Modelling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. Geophysics **60**(1), 176-184 (1995)
25. Bleibinhaus, F., Hole, J. A., Ryberg, T., Fuis, G. S.: Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging. J. Geophys. Res. **112**, doi:10.1029/2006JB004611 (2007)
26. Bleibinhaus, F., Lester, R. W., Hole, J. A.: Applying waveform inversion to wide-angle seismic surveys. Tectonophysics **472**, 238-248 (2009)
27. Bleibinhaus, F., Rondenay, S.: Effects of surface scattering in full waveform inversion. Geophysics **74**, WCC87-WCC95 (2009)
28. Bohlen, T.: Parallel 3-D viscoelastic finite difference modelling. Computers & Geosciences, **28** 887-899 (2002)
29. Bolt, B. A.: Velocity of the seismic waves Lg and Rg across Australia. Nature **180**, 495 (1958)
30. Bolt, B. A., Doyle, H. A., Sutton, D. J.: Seismic observations from the 1956 atomic explosions in Australia, Geophys. J. R. Astr. Soc. **1** (1958)
31. Bolt, B. A., Niazi, M.: Dispersion of Rayleigh waves across Australia. Geophys. J. R. Astr. Soc. **9**, 21-35 (1964)
32. Boore, D. M.: Love waves in nonuniform waveguides: finite difference calculations. J. Geophys. Res. **75**, 1512-1527 (1970)
33. Boore, D. M.: Finite-difference methods for seismic wave propagation in heterogeneous materials. In: Methods in Computational Physics **11**, Academic Press, New York (1972)
34. Boschi, L.: Global multiresolution models of surface wave propagation: comparing equivalently regularized Born and ray theoretical solutions. Geophys. J. Int. **167**, 238-252 (2006)
35. Bozdağ, E., Trampert, J.: On crustal corrections in surface wave tomography. Geophys. J. Int. **172**, 1066-1082 (2008)
36. Bozdağ, E., Trampert, J.: Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. Geophys. J. Int., in press (2010)
37. Brenders, A. J., Pratt, R. G.: Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. Geophys. J. Int. **168**, 133-151 (2007)
38. Brenders, A. J., Pratt, R. G.: Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data. Geophys. J. Int. **168**, 152-170 (2007)

39. Brossier, R., Operto, S., Virieux, J.: Seismic imaging of complex onshore structures by 2D elastic frequency-domain full waveform inversion. Geophysics **74**, WCC105-WCC118 (2009)
40. Brossier, R., Operto, S., Virieux, J.: Robust elastic frequency-domain full waveform inversion using the l-1 norm. Geophys. Res. Lett. **36**, L20310 (2009)
41. Brossier, R., Operto, S., Virieux, J.: Which data residual norm for robust elastic frequency-domain full waveform inversion? Geophysics **75**, R37-R46 (2010)
42. Bullen, K. E., Bolt, B.: An introduction to the theory of seismology, 4th edition. Cambridge University Press (1985)
43. Bunge, H.-P., Davies, J. H.: Tomographic images of a mantle circulation model. Geophys. Res. Lett. **28**, 77-80 (2001)
44. Bunks, C., Saleck, F. M., Zaleski, S., Chavent, G.: Multiscale seismic waveform inversion. Geophysics **60**, 1457-1473 (1995)
45. Burstedde, C., Ghattas, O.: Algorithmic strategies for full waveform inversion: 1D experiments. Geophysics **74**, WCC37-WCC46 (2009)
46. Capdeville, Y., Chaljub, E., Vilotte, J. P., Montagner, J.-P.: Coupling the spectral-element method with a modal solution for elastic wave propagation in global earth models. Geophys. J. Int. **152**, 34-67 (2003)
47. Capdeville, Y., Romanowicz, B., To, A.: Coupling spectral elements and modes in a spherical Earth: An extension to the ”sandwich” case. Geophys. J. Int. **154**, 44-57 (2003b)
48. Capdeville, Y., Gung, Y., Romanowicz, B.: Towards global earth tomography using the spectral element method: a technique based on source stacking. Geophys. J. Int. **162**, 541-554 (2005)
49. Capdeville, Y., Marigo, J.-J.: Second-order homogenisation of the elastic wave equation for non-periodic layered media. Geophys. J. Int. **170**, 823-838 (2007)
50. Capdeville, Y., Marigo, J. J.: Shallow layer correction for spectral element like methods. Geophys. J. Int. **172**, 1135-1150 (2008)
51. Capdeville, Y.: Contributions aux problès direct et inverse en sismologie. Thesis, Université Paris 7 (2010)
52. Capdeville, Y., Guillot, L., Marigo, J.-J.: 1-D non-periodic homogenisation for the seismic wave equation. Geophys. J. Int. **181**, 897-910 (2010a)
53. Capdeville, Y., Guillot, L., Marigo, J.-J.: 2-D non-periodic homogenisation to upscale elastic media for P-SV waves. Geophys. J. Int. **182**, 903-922 (2010b)
54. Cara, M.: Filtering of dispersed wave trains. Geophys. J. R. astr. Soc. **33**, 65-80 (1973)
55. Cara, M., Lévêque, J. J.: Waveform inversion using secondary observables. Geophys. Res. Lett. **14**(10), 1046-1049 (1987)
56. Cerjan, C., Kosloff, D., Kosloff, R., Reshef, M.: A nonreflecting boundary condition for discrete acoustic and elastic wave equations. Geophysics **50**, 705-708 (1985)
57. Chaljub, E., Capdeville, Y., Vilotte, J. P.: Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral-element approximation on non-conforming grids. J. Comp. Phys. **187**, 457-491 (2003)
58. Chaljub, E., Valette, B.: Spectral element modelling of three-dimensional wave propagation in a self-gravitating Earth with an arbitrarily stratified outer core. Geophys. J. Int. **158**, 131-141 (2004)
59. Chaljub, E., Komatitsch, D., Capdeville, Y., Vilotte, J. P., Valette, B., Festa, G.: Spectral-element analysis in seismology. In: Advances in Geophysics **48**, 365-419. Editors: Wu, R.-S., Maupin, V. (2007)
60. Charney, J., Frjørtoft, R., von Neumann, J.: Numerical integration of the barotropic vorticity equation. Tellus **2**, 237-254 (1950)
61. Charpentier, I.: Checkpointing schemes for adjoint codes: Application to the meteorological model Meso-NH. SIAM J. Sci. Comp. **22**(6), 2135-2151 (2001)
62. Chen, P., Zhao, L., Jordan, T. H.: Full 3D tomography for the crustal structure of the Los Angeles region. Bull. Seis. Soc. Am. **97**(4), 1094-1120 (2007)
63. Cheng, H.-X., Kennett, B. L. N.: Frequency dependence of seismic wave attentuation in the upper mantle beneath the Australian region. Geophys. J. Int. **150**, 45-57 (2002)
64. Choi, Y., Min, D.-J., Shin, C.: Two-dimensional waveform inversion of multi-component data in acoustic coupled media. Geophys. Prosp. **56**, 863-881 (2008)
65. Červený, V.: Seismic ray theory. Cambridge University Press (2001)
66. Chevrot, S., van der Hilst, R. D.: The Poisson ratio of the Australian crust: geological and geophysical implications. Earth Planet. Sci. Lett. **183**, 121-132 (2000)
67. Claerbout, J. F., Johnson, A. G.: Extrapolation of time dependent waveforms along their path of propagation. Geophys. J. **26**, 285-295 (1971)
68. Claerbout, J. F.: Fundamentals of geophysical data processing. McGraw-Hill, New York (1976)
69. Clayton, R. W., Engquist, B.: Absorbing boundary conditions for acoustic and elastic wave equations. Bull. Seis. Soc. Am. **67**, 1529-1540 (1977)
70. Clayton, R. W., Engquist, B.: Absorbing boundary conditions for wave-equation migration. Geophysics **45**, 895-904 (1980)
71. Cleary, J. R.: P times to Australian stations from nuclear explosions. Bull. Seis. Soc. Am. **57**, 773-781 (1967)
72. Cleary, J. R., Simpson, D. W., Muirhead, K. J.: Variations in the Australian upper mantle structure, from observations of the Cannikin explosion. Nature **236**, 111-112 (1972)
73. Clitheroe, G., Gudmundsson, O., Kennett, B. L. N.: The crustal thickness of Australia. J. Geophys. Res. **105**, 13,697-13,713 (2000)
74. Cole, J. B.: A nearly exact second-order finite-difference time-domain wave propagation algorithm on a coarse grid. Comp. in Phys. **8**, 730-734 (1994)
75. Cole, J. B.: High accuracy solution of Maxwell’s equations using nonstandard finite differences. Comp. in Phys. **11**, 287-292 (1997)
76. Collino, F.: High-order absorbing boundary conditions for wave propagation models. Straight line boundary and corner cases. In: Proceedings of the Second International Conference on Mathematical and Numerical Aspects of Wave Propagation, SIAM, Delaware, 161-171 (1993)

77. Collino, F. & Tsogka, C.: Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. Geophysics **66**(1), 294-307 (2001)
78. Cormack, A. M.: Representation of a function by its line integrals, with some radiological applications. J. Appl. Phys. **34**(9), 2722-2727 (1963)
79. Courant, R., Friedrichs, K., Lewy, H.: Über die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen **100**, 32-74 (1928)
80. Crank, J., Nicolson, P.: A practical method for numerical evaluation of solutions of partial differential equations of heat conduction type. Proc. Camb. Phil. Soc. **43**, 50-67 (1947)
81. Crase, E., Pica, A., Noble, M., McDonald, J., Tarantola, A.: Robust elastic nonlinear waveform inversion - Application to real data. Geophysics **55**(5), 527-538 (1990)
82. Cummins, P., Geller, R., Hatori, T., Takeuchi, N.: DSM synthetic seismograms: SH, spherically symmetric case. Geophys. Res. Lett. **21**, 533-536 (1994a)
83. Cummins, P., Geller, R., Takeuchi, N.: DSM complete synthetic seismograms: P-SV, spherically symmetric case. Geophys. Res. Lett. **21**, 1663-1666 (1994b)
84. Dablain, M. A.: The application of high-order differencing to the scalar wave equation. Geophysics **51**, 54-66 (1986)
85. Dahlen, F. A., Hung, S.-H., Nolet, G.: Fréchet kernels for finite-frequency traveltimes - I. Theory. Geophys. J. Int. **141**, 157-174 (2000)
86. Dahlen, F. A., Tromp, J.: Theoretical global seismology. Princeton University Press (2001)
87. Dahlen, F. A., Baig, A. M.: Fréchet kernels for body-wave amplitudes. Geophys. J. Int. **150**, 440-466 (2002)
88. Dai, N. X., Vafidis, A., Kanasewich, E. R.: Seismic migration and absorbing boundaries with a one-way wave system for heterogeneous media. Geophys. Prosp. **44**, 719-739 (1996)
89. De Basabe, J. D., Sen, M. K.: Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations. Geophysics **72**, T81-T95 (2007)
90. Debayle, E., Kennett, B. L. N.: The Australian continental upper mantle: Structure and deformation inferred from surface waves. J. Geophys. Res. **105**, 25423-25450 (2000)
91. Debayle, E., Sambridge, M.: Inversion of massive surface wave data sets: Model construction and resolution assessment. J. Geophys. Res. **109**, B02316 (2004)
92. de Hoop, A. T.: Representation theorems for the displacement in an elastic solid and their applications to elastodynamic diffraction theory. D.Sc Thesis, Technische Hogeschool Delft (1958)
93. de Jersey, N. J.: Seismological evidence bearing on the crustal thickness in the South-West Pacific. University of Queensland, Brisbane, Paper v. 3, no. 2 (1946)
94. de la Puente, J., Käser, M., Dumbser, M., Igel, H.: An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - IV. Anisotropy. Geophys. J. Int. **169**, 1210-1228 (2007)
95. de la Puente, J., Dumbser, M., Käser, M., Igel, H.: Discontinuous Galerkin methods for wave propagation in poroelastic media. Geophysics **73**, T77-T97 (2008)
96. de la Puente, J., Ampuero, J.-P., Käser, M.: Dynamic rupture modeling on unstructured meshes using a discontinuous Galerkin method. J. Geophys. Res. **114**, doi:10.1029/2008JB006271 (2009)
97. Deng, H. L.: A complexity analysis of generic optimization problems: Characterizing the topography of high-dimensional functions. PhD Thesis, Center for Wave Phenomena, Colorado School of Mines (1997)
98. Dessa, J. X., Pascal, G.: Combined traveltime and frequency-domain seismic waveform inversions: a case of study on multioffset ultrasonic data. Geophys. J. Int. **154**, 117-133 (2003)
99. Dessa, J. X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J., Kaneda, Y.: Multiscale seismic imaging of the eastern Nankai trough by full waveform inversion. Geophys. Res. Lett. **31**, doi:10.1029/2004GL020453 (2004)
100. Devaney, A. J.: Geophysical diffraction tomography. IEEE Transactions on Geoscience and Remote Sensing **22**(1), 3-13 (1984)
101. Djikpéssé, H., Tarantola, A.: Multiparameter ℓ_1-norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms. Geophysics **64**, 1023-1035 (1999)
102. Dumbser, M., Käser, M.: An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - II. The three-dimensional isotropic case. Geophys. J. Int. **167**, 319-336 (2006)
103. Dumbser, M., Käser, M., de la Puente, J.: Arbitrary high-order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D. Geophys. J. Int. **171**(2), 665-694 (2007)
104. Dziewonski, A. M., Mills, J, Bloch, S.: Residual dispersion measurement - A new method of surface wave analysis. Bull. Seis. Soc. Am. **62**, 129-139 (1972)
105. Dziewonski, A. M., Hales, A. L., Lapwood, E. R.: Parametrically simple Earth models consistent with geophysical data. Phys. Earth Planet. Int. **10**, 12-48 (1975)
106. Dziewonski, A. M., Hager, B. H., O'Connell, R. J.: Large-scale heterogeneities in the lower mantle. J. Geophys. Res. **82**, 239-255 (1977)
107. Dziewonski, A. M., Anderson, D. L.: Preliminary reference Earth model. Phys. Earth Planet. Int. **25**, 297-356 (1981)
108. Emerman, S. H., Stephen, R. A.: Comment on 'Absorbing boundary conditions for acoustic and elastic wave equations' by R. Clayton and B. Engquist. Bull. Seis. Soc. Am. **73**, 661-665 (1983)
109. Engquist, B., Majda, A.: Absorbing boundary conditions for the numerical simulation of waves. Math. Comp. **31**, 629-651 (1977)
110. Epanomeritakis, I., Akcelik, V., Ghattas, O., Bielak, J.: A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. Inverse Problems **24**, doi:10.1088/0266-5611/24/3/034015 (2008)
111. Faccioli, E., Maggio, F., Paolucci, R., Quarteroni, A.: 2D and 3D elastic wave propagation by a spseudo-spectral domain decomposition method. Journal of Seismology **1**, 237-251 (1997)

112. Festa, G., Nielsen, S.: PML absorbing boundaries. Bull. Seis. Soc. Am. **93**, 891-903 (2003)
113. Festa, G. & Vilotte, J.-P.: The Newmark scheme as velocity-stress time-staggering: an efficient PML implementation for spectral element simulations for elastodynamics. Geophys. J. Int. **161**, 789-812 (2005)
114. Festa, G., Delavaud, E., Vilotte, J.-P.: Interaction between surface waves and absorbing boundaries for wave propagation in geological basins: 2D numerical simulations. Geophys. Res. Lett. **32**, doi:10.1029/2005GL024091 (2005)
115. Flanagan, M. P., Wiens, D. A.: Attenuation of broadband P and S waves in Tonga: Observation of frequency-dependent Q. Pure Appl. Geophys. **153**, 345-375 (1998)
116. Fichtner, A., Bunge, H.-P., Igel, H.: The adjoint method in seismology - I. Theory. Phys. Earth Planet. Int. **157**, 86-104 (2006a)
117. Fichtner, A., Bunge, H.-P., Igel, H.: The adjoint method in seismology - II. Applications: traveltimes and sensitivity functionals. Phys. Earth Planet. Int. **157**, 105-123 (2006b)
118. Fichtner, A., Igel, H.: Efficient numerical surface wave propagation through the optimization of discrete crustal models - a technique based on non-linear dispersion curve matching (DCM). Geophys. J. Int. **173**, 519-533 (2008)
119. Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P.: Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. Geophys. J. Int. **175**, 665-685 (2008)
120. Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P.: Spectral-element simulation and inversion of seismic waves in a spherical section of the Earth. Journal of Numerical Analysis Industrial and Applied Mathematics **4**, 11-22 (2009)
121. Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P.: Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. Geophys. J. Int. **179**, 1703-1725 (2009)
122. Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P.: Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle. Earth Planet. Sci. Lett. **290**, 270-280 (2010)
123. Fink, M.: Time-reversed acoustics. Scientific American **281**, 91-97 (1999)
124. Fishwick, S., Kennett, B. L. N., Reading, A. M.: Contrasts in lithospheric structure within the Australian craton - insights from surface wave tomography. Earth Planet. Sci. Lett. **231**, 163-176 (2005)
125. Fishwick, S., Reading, A. M.: Anomalous lithosphere beneath the Proterozoic of western and central Australia: A record of continental collision and intraplate deformation? Precambrian Research **166**, 111-121 (2008)
126. Fishwick, S., Heintz M., Kennett, B.L.N., Reading, A.M., Yoshizawa, K.: Steps in lithospheric thickness within eastern Australia, evidence from surface wave tomography. Tectonics **27**, TC0049, doi:10.129/2007TC002116 (2008)
127. Flanagan, M. P., Wiens, D. A.: Attenuation of broadband P and S waves in Tonga: Observation of frequency-dependent Q. Pure Appl. Geophys. **153**, 345-375 (1998)
128. Fletcher, R., Reeves, C. M.: Function minimization by conjugate gradients. The Computational Journal **7**, 149-154 (1964)
129. Fletcher, R.: Practical methods of optimization. John Wiley & Sons, New York (1987)
130. Frankel, A., Clayton, R. W.: A finite-difference simulation of wave propagation in two-dimensional random media. Bull. Seis. Soc. Am. **74**, 2167-2186 (1984)
131. Frankel, A., Clayton, R. W.: Finite-difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity. J. Geophys. Res. **91**, 6465-6489 (1986)
132. Frankel, A., Vidale, J.: A 3-dimensional simulation of seismic waves in the Santa Clara valley, California, from a Loma Prieta aftershock. Bull. Seis. Soc. Am. **82**, 2045-2074 (1992)
133. Friederich, W., Dalkolmo, J.: Complete synthetic seismograms for a spherically symmetric earth by a numerical computation of the Green's function in the frequency domain. Geophys. J. Int. **122**, 537-550 (1995)
134. Friederich, W.: Propagation of seismic shear and surface waves in a laterally heterogeneous mantle by multiple forward scattering. Geophys. J. Int. **136**, 180-204 (1999)
135. Friederich, W.: The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms. Geophys. J. Int. **153**, 88-102 (2003)
136. Furumura, T., Kennett, B. L. N., Furumura, M.: Seismic wavefield calculation for laterally heterogeneous whole earth models using the pseudospectral method. Geophys. J. Int. **135**, 845-860 (1998)
137. Furumura, T., Chen, L.: Large-scale parallel simulation and visualisation of 3D seismic wave field using the Earth Simulator. Comp. Mod. Eng. Sci. **6**, 153-168 (2004)
138. Furumura, T., Kennett, B. L. N.: Subduction zone guided waves and the heterogeneity structure of the subducted plate: Intensity anomalies in northern Japan. J. Geophys. Res. **110**, doi:10.1029/2004JB003486 (2005)
139. Gao, F., Levander, A., Pratt, R. G., Zelt, C. A., Fradelizio, G. L.: Waveform tomography at a groundwater contamination site: Surface reflection data. Geophysics **72**, G45-G55 (2007)
140. Gauthier, O., Virieux, J., Tarantola, A.: Two-dimensional nonlinear inversion of seismic waveforms: numerical results. Geophysics **51**(7), 1387-1403 (1986)
141. Gazdag, J.: Wave equation migration with the phase shift method. Geophysics **43**, 1342-1351.
142. Gee, L. S., Jordan, T. H.: Generalized seismological data functionals. Geophys. J. Int. **111**, 363-390 (1992)
143. Geller, R. J., Ohminato, T.: Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the direct solution method. Geophys. J. Int. **116**, 421-446 (1994)
144. Geller, R. J., Takeuchi, N.: A new method for computing highly accurate DSM synthetic seismograms. Geophys. J. Int. **123**, 449-470 (1995)
145. Geller, R. J., Takeuchi, N.: Optimally accurate second-order time-domain finite difference schemes for the elastic equation of motion: one-dimensional case. Geophys. J. Int. **135**, 48-62 (1998)
146. Gill, P. E., Murray, W., Wright, M. H.: Practical optimization. Academic Press, London (1981)
147. Givoli, D.: High-order local non-reflecting boundary conditions: a review. Wave Motion **39**, 319-326 (2003)

148. Goncz, J. H., Cleary, J. R.: Variations in the structure of the upper mantle beneath Australia from Rayleigh wave observations. Geophys. J. R. Astr. Soc. **44**, 507-516 (1976)
149. Gottschämmer, E., Olsen, K. B.: Accuracy of the explicit planar free-surface boundary condition implemented in a fourth-order staggered-grid velocity-stress finite-difference scheme. Bull. Seis. Soc. Am. **91**, 617-623 (2001)
150. Graves, R. W.: Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. Bull. Seis. Soc. Am. **86**, 1091-1106 (1996)
151. Griewank, A., Walther, A.: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. Trans. Math. Software, **26**, 19-45 (2000)
152. Gudmundsson, O., Kennett, B. L. N., Goody, A.: Broadband observations of upper-mantle seismic phases in northern Australia and the attenuation structure in the upper mantle. Phys. Earth Planet. Int. **84**, 207-226 (1994)
153. Guillot, L., Capdeville, Y., Marigo, J.-J.: 2-D non-periodic homogenisation of the elastic wave equation - SH case. Geophys. J. Int., in press (2010)
154. Gutenberg, B.: Über die Konstitution des Erdinnern, erschlossen aus Erdbebenbeobachtungen. Physikalische Zeitschrift **14**, 1217-1218 (1913)
155. Gutenberg, B.: Untersuchungen zur Frage bis zu welcher Tiefe die Erde kristallin ist. Zeitschrift für Geophysik, **2**, 24-29 (1926)
156. Gutenberg, B.: Seismological evidence for roots of mountains. Bull. Geol. Soc. Am. **54**, 473-498 (1943)
157. Hestenes, M. R., Stiefer, E. L.: Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Stand. **49**, 409-436 (1952)
158. Hesthaven, J. S.: From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. SIAM J. Numer. Anal. **35**(2), 655-676 (1998)
159. Hestholm, S., Ketcham, S., Greenfield, R., Moran, M., McMechan, G.: Quick and accurate parameterization in viscoelastic wave modeling. Geophysics **71**, 147-150 (2006)
160. Hicks, G. J., Pratt, R. G.: Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit. Geophysics **66**, 598-612.
161. Hicks, G. J.: Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions. Geophysics **67**, 156-166 (2002)
162. Higdon, R. L.: Radiation boundary conditions for elastic wave propagation. SIAM J. Num. An. **27**, 831-870 (1990)
163. Higdon, R. L.: Absorbing boundary conditions for elastic waves. Geophysics **56**, 231-241 (1991)
164. Holberg, O.: Computational aspects of the choice of operator and sampling interval for numerical differentiation in large-scale simulation of wave phenomena. Geophys. Prosp. **35**, 629-655 (1987)
165. Hole, J. A., Ryberg, T., Fuis, G. S., Bleibinhaus, F., Sharma, A. K.: Structure of the San Andreas fault zone at SAFOD from a seismic refraction survey. Geophys. Res. Lett. **33** (2006)
166. Hu, F. Q.: On absorbing boundary conditions for linearized Euler equations by a perfectly matched layer. J. Comp. Phys. **129**, 201-219 (1996)
167. Hung, S. H., Dahlen, F. A., Nolet, G.: Fréchet kernels for finite-frequency traveltimes - II. Examples. Geophys. J. Int. **141**, 175-203 (2000)
168. Igel, H., Mora, P., Riollet, B.: Anisotropic wave propagation through finite-difference grids. Geophysics **60**, 1203-1216 (1995)
169. Igel, H., Djikpesse, H., Tarantola, A.: Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio. Geophys. J. Int. **124**(2), 363-371 (1996)
170. Igel, H.: Wave propagation in three-dimensional spherical sections by the Chebyshev spectral method. Geophys. J. Int. **136**, 559-566 (1999)
171. Igel, H., Nissen-Meyer, T., Jahnke, G.: Wave propagation in 3D spherical sections. Effects of subduction zones. Phys. Earth Planet. Int. **132**, 219-234 (2002)
172. Ikelle, L., Diet, J. P., Tarantola, A.: Linearized inversion of multi-offset seismic reflection data in the $\omega - k$ domain: Depth-dependent reference medium. Geophysics **53**, 50-64 (1988)
173. Jackson, D. D.: Attenuation of seismic waves by grain boundary relaxation. Proc. Nat. Acad. Sci. **68**, 1577-1579 (1971)
174. Jackson, I.: Laboratory measurements of seismic wave dispersion and attenuation: Recent progress. In: Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale, Geophysical Monograph **117**, 265-289 (2000)
175. Jaiswal, P., Zelt, C. A., Dasgupta, R., Nath, K. K.: Seismic imaging of the Naga thrust using multiscale waveform inversion. Geophysics **74**, WCC129-WCC140 (2009)
176. Jeffreys, H.: The rigidity of the Earth's central core. Monthly Notices of the Royal Astronomical Society, Geophysical Supplement **1**, 371-383 (1926)
177. Jeffreys, H., Bullen, K.: Seismological Tables. British Association for the Advancement of Science, London (1940)
178. Jo, C.-H., Shin, C., Suh, J. H.: An optimal 9-point, finite-difference, frequency-space 2-D scalar wave extrapolator. Geophysics **61**, 529-537 (1996)
179. Jordan, T. H.: Composition and development of the continental tectosphere. Nature **274**, 544-548 (1978)
180. Kaiho, Y., Kennett, B. L. N.: Three-dimensional structure beneath the Australasian region from refracted wave observations. Geophys. J. Int. **142**, 651-668 (2000)
181. Kane, C., Marsden, J. Ortiz, M., West, M.: Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. Int. J. Numer. Meth. Eng. **49**, 1295-1325 (2003)
182. Karniadakis, G. M., Sherwin, S.: Spectral/hp element methods for computational fluid dynamics, 2nd edition. Oxford University Press (2005)
183. Käser, M., Dumbser, M., de la Puente, J., Igel, H.: An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - III. Viscoelastic attenuation. Geophys. J. Int. **168**, 224-242 (2007)

184. Kawai, K., Takeuchi, N., Geller, R. J.: Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media. Geophys. J. Int. **164**, 411-424 (2006)
185. Kawai, K., Geller, R. J.: Waveform inversion for localised seismic structure and an application to D" structure beneath the Pacific. J. Geophys. Res. **115**, doi:10.1029/2009JB006503 (2010)
186. Kawakatsu, H., Montagner, J.-P.: Time-reversal seismic-source imaging and moment tensor inversion. Geophys. J. Int. **175**, 686-688 (2008)
187. Keilis-Borok, V. I., Yanovskaya, T. B.: Inverse problems in seismology. Geophys. J. **13**, 223-234 (1967)
188. Kelley, C. T.: Iterative methods for optimization. SIAM, Philadelphia (1999)
189. Kelly, K. R., Ward, R. W., Treitel, S., Alford, R. M.: Synthetic seismograms: a finite-difference approach. Geophysics **41**, 2-27 (1976)
190. Kennett, B. L. N.: Seismic waves in a stratified half space. Geophys. J. R. astr. Soc. **57**, 557-583 (1979)
191. Kennett, B. L. N.: Seismic waves in a stratified half space II. - Theoretical seismograms. Geophys. J. R. astr. Soc. **61**, 1-10 (1980)
192. Kennett, B. L. N.: Elastic wave propagation is stratified media. In: Advances in Mechanics **21**, 79-167, Academic Press, New York (1981)
193. Kennett, B. L. N., Engdahl, E. R., Buland, R.: Constraints on seismic velocities in the Earth from traveltimes. Geophys. J. Int. **122**, 108-124 (1995)
194. Kennett, B. L. N.: On the density distribution within the Earth. Geophys. J. Int. **132**(2), 374-382 (1998)
195. Kennett, B. L. N.: The seismic wavefield I. - Introduction and theoretical development. Cambride University Press (2001)
196. Kennett, B. L. N., Furumura, T.: Stochastic waveguide in the lithosphere: Indonesian subduction zone to Australian craton. Geophys. J. Int. **172**, 363-382 (2008)
197. Keys, R. G.: Absorbing boundary conditions for acoustic media: Geophysics **50**, 892-902.
198. Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P.: Optimization by Simulated Annealing. Science **220**, 671-680 (1983)
199. Klingelhoefer, F., Lafoy, Y., Collot, J., Cosquer, E., Géli, L., Nouzé, H., Vially, R.: Crustal structure of the basin and ridge system west of New Caledonia (southwest Pacific) from wide-angle and reflection seismic data. J. Geophys. Res. **112**, doi:10.1029/2007JB005093 (2007)
200. Komatitsch, D.: Méthodes spectrales et éléments spectraux pour l'équation de l'élastodynamique 2D et 3D en milieu hétérogène. PhD thesis, Institut de Physique du Globe de Paris (1997)
201. Komatitsch, D., Vilotte, J. P.: The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. Bull. Seis. Soc. Am. **88**, 368-392 (1998)
202. Komatitsch, D., Barnes, C., Tromp, J.: Wave propagation near a fluid-solid interface: A spectral-element approach. Geophysics **65**, 623-631 (2000)
203. Komatitsch, D., Martin, R., Tromp, J., Taylor, M. A., Wingate, B. A.: Wave propagation in 2D elastic media using a spectral element method with triangles and quadrangles. J. Comp. Acoust. **9**, 703-718 (2001)
204. Komatisch, D., Tromp, J.: Spectral-element simulations of global seismic wave propagation - I. Validation. Geophys. J. Int. **149**(2), 390-412 (2002)
205. Komatitsch, D., Tromp, J.: Spectral-element simulations of global seismic wave propagation-II. 3-D models, oceans, rotation and self-gravitation. Geophys. J. Int. **150**, 303-318 (2002)
206. Komatitsch, D., Tromp, J.: A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation. Geophys. J. Int. **154**, 146-153 (2003)
207. Komatitsch, D., Liu, Q., Tromp, J., Süss, P., Stidham, C., Shaw, J. H.: Simulation of ground motion in the Los Angeles basin based upon the spectral-element method. Bull. Seis. Soc. Am. **94**, 187-206 (2004)
208. Komatitsch, D., Tsuboi, S., Tromp, J.: The spectral-element method in seismology. In: Seismic Earth: Array Analysis of Broadband Seismograms, Eds. A. Levander, G. Nolet, AGU Monograph **157**, 205-228 (2005)
209. Konishi, K., Kawai, K., Geller, R. J., Fuji, N.: MORB in the lowermost mantle beneath the western Pacific: Evidence from waveform inversion. Earth Planet. Sci. Lett. **278**, 219-225 (2009)
210. Korenaga, J., Holbrook, W., Singh, S., Minshull, T.: Natural gas hydrates on the Southeast U.S. margin: Constraints from full waveform and travel time inversions of wide-angle seismic data. J. Geophys. Res. **102**, 15345-15365 (1997)
211. Kosloff, D. D., Baysal, E.: Forward modelling by a Fourier method. Geophysics **47**, 1402-1412 (1982)
212. Kosloff, D. D., Kessler, D., Filho, A. Q., Tessmer, E., Behle, A., Strahilevitz, R.: Solution of the equations of dynamic elasticity by a Chebyshev spectral method. Geophysics **55**, 734-748 (1990)
213. Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Baumstein, A., Lacasse, M. D., Lee, S.: Fast full-wavefield seismic inversion using encoded sources. Geophysics **74**, WCC177, doi:10.1190/1.3230502 (2009)
214. Kristek, J., Moczo, P., Archuleta, R. J.: Efficient methods to simulate planar free surface im the 3D 4th-order staggered-grid finite-difference schemes. Stud. geophys. geod. **46**, 355-381 (2002)
215. Kristek, J. Moczo, P.: On the accuracy of the finite-difference schemes: The 1D elastic problem. Bull. Seis. Soc. Am. **96**, 2398-2414 (2006)
216. Kristek, J., Moczo, P., Galis, M.: A brief summary of some PML formulations and discretizations for the velocity-stress equation of seismic motion. Stud. Geophys. Geod. **53**, 459-474 (2009)
217. Kristekova, M., Kristek, J., Moczo, P., Day, S. M.: Misfit criteria for quantitative comparison of seismograms. Bull. Seis. Soc. Am. **96**, 1836-1850 (2006)
218. Kristekova, M., Kristek, J., Moczo, P.: Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals. Geophys. J. Int. **178**, 813-825 (2009)
219. Lamb, H.: On the propagation of tremors over the surface of an elastic solid. Phil. Trans. Roy. Soc. London **A203**, 1-42 (1904)

220. Lambeck, K., Burgess, G., Shaw, R. D.: Teleseismic travel-time anomalies and deep crustal structure in central Australia. Geophys. J. Int. **94**, 105-124 (1988)
221. Larmat, C., Montagner, J.-P., Fink, M., Capdeville, Y., Tourin, A., Clévédé, E.: Time-reversal imaging of seismic sources and application to the great Sumatra earthquake. Geophys. Res. Lett. **33**, doi:10.1029/2006GL026336 (2006)
222. Larmat, C., Guyer, R. A., Johnson, P. A.: Tremor source location using time reversal: Selecting the appropriate imaging field. Geophys. Res. Lett. **36**, doi:10.1029/2009GL040099 (2009)
223. Larner, K., Beasley, C.: Cascaded migration: Improving the accuracy of finite-difference migration. Geophysics **52**, 618-643 (1987)
224. Lay, T., Wallace, T. C.: Modern global seismology. Academic Press, San Diego (1995)
225. Lerner-Lam, A. L., Jordan, T. H.: Earth structure from fundamental and higher-mode waveform analysis. Geophys. J. R. astr. Soc. **75**, 759-797 (1983)
226. Levander, A. R.: Fourth-order finite-difference P-SV seismograms. Geophysics **53**, 1425-1436 (1988)
227. Levenberg, K.: A method for the solution of certain non-linear problems in least-squares. Quart. Appl. Math. **2**, 164-168 (1944)
228. Lévêque, J. J., Rivera, L., Wittlinger, G.: On the use of the checkerboard test to assess the resolution of tomographic inversions. Geophys. J. Int. **115**, 313-318 (1993)
229. Li, X., Romanowicz, B.: Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. J. Geophys. Res. **101**, 11,245-11,271 (1996)
230. Li, Y. G., Vidale, J. E., Cochran, E. S.: Low-velocity structure of the San Andreas fault at Parkfield from fault zone trapped waves. Geophys. Res. Lett. **31**, L12S06 (2004)
231. Liu, H.-P., Anderson, D. L., Kanamori, H.: Velocity dispersion due to anelasticity; implications for seismology and mantle composition. Geophys. J. R. astr. Soc., **47** 41-58 (1976)
232. Liu, Q., Tromp.: Finite-frequency kernels based on adjoint methods. Bull. Seis. Soc. Am. **96**, 2383-2397 (2006)
233. Liu, Q., Tromp, J.: Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, Geophys. J. Int. **174**, 265-286 (2008)
234. Lokmer, I., O'Brien, G. S., Stich, D., Bean, C. J.: Time reversal imaging of synthetic volcanic tremor sources. Geophys. Res. Lett. **36**, doi:10.1029/2009GL038178 (2009)
235. Luo, Y. & Schuster, G. T.: Wave-equation traveltime inversion. Geophysics **56**, 645-653 (1991)
236. Lysmer, J., Drake, L. A.: A finite-element method for seismology. In: Methods in Computational Physics **11**, Academic Press, New York (1972)
237. Madariaga, R.: Dynamics of an expanding circular fault. Bull. Seis. Soc. Am. **67**, 163-182 (1976)
238. Maday, Y., Patera, A. T.: Spectral element methods for the incompressible Navier-Stokes equations. In: State of the art survey in computational mechanics, 71-143, editors: Noor, A. K., Oden, J. T. (1989)
239. Mahrer, K. D.: An empirical study of instability and improvement of absorbing boundary conditions for the elastic wave equation. Geophysics **51**, 1499-1501 (1986)
240. Mahrer, K. D.: Numerical time step instability and Stacey's and Clayton-Engquist's absorbing boundary conditions. Bull. Seis. Soc. Am. **80**, 213-217 (1990)
241. Malin, P., Shalev, E., Balven, H., Lewis-Kenedi, C.: Structure of the San Andreas fault at SAFOD from P wave tomography and fault-guided wave mapping. Geophys. Res. Lett. **33**, L13314 (2006)
242. Malinowski, M., Operto, S.: Quantitative imaging of the permo-mesozoic complex and its basement by frequency-domain waveform tomography of wide-aperture seismic data from the Polish basin. Geophys. Prosp. **56**, 805-825 (2008)
243. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**, 431-441 (1963)
244. Marquering, H., Nolet, G., Dahlen, F. A.: Three-dimensional waveform sensitivity kernels. Geophys. J. Int. **132**, 521-534 (1998)
245. Marquering, H., Dahlen, F. A., Nolet, G.: Three-dimensional sensitivity kernels for finite-frequency traveltime: the banana-doughnut paradox. Geophys. J. Int. **137**, 805-815 (1999)
246. Maupin, V.: A multiple-scattering scheme for modelling surface wave propagation in isotropic and anisotropic three-dimensional structures. Geophys. J. Int. **146**, 332-348 (2001)
247. McDougall, I., Duncan, R. A.: Age progressive volcanism in the Tasmantid Seamounts. Earth Planet. Sci. Lett. **89**, 207-220 (1988)
248. Mégnin, C., Romanowicz, B.: The 3D shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms. Geophys. J. Int. **143**, 709-728 (2000)
249. Meier, U., Curtis, A., Trampert,J.: 2007a. Global crustal thickness from neural network inversion of surface wave data. Geophys. J. Int. **169**, 706-722 (2007a)
250. Meier, U., Curtis, A., Trampert,J.: Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. Geophys. Res. Lett. **34**, doi:10.1029/2007GL030989 (2007b)
251. Meju, M. A.: Regularised extremal bounds analysis (REBA): An approach to quantifying uncertainty in nonlinear geophysical inverse problems. Geophys. Res. Lett. **36**, doi:10:1029/2008GL036407 (2009)
252. Mercerat, E. D., Vilotte, J. P., Sánchez-Sesma, F. J.: Triangular spectral-element simulation of two-dimensional elastic wave propagation using unstructured triangular grids. Geophys. J. Int. **166**, 679-698 (2006)
253. Meza-Fajardo, K. C., Papageorgiou, A. S.: A nonconvolutional, split-field, perfectly matched layer for wave propagation in isotropic and anisotropic elastic media: stability analysis. Bull. Seis. Soc. Am. **98**(4), 1811-1836 (2008)
254. Michell, J.: Conjectures concerning the cause, and observations upon the phaenomena of earthquakes; particularly of that great earthquake of the first of November, 1755, which proved so fatal to the city of Lisbon, and whose effects were felt as far as Africa, and more or less throughout almost all Europe. Phil. Trans. Roy. Soc. London **51**, 566-634 (1760)
255. Mizutani, H. Geller, R. J., Takeuchi, N.: Comparison of accuracy and efficiency of time-domain schemes for calculating synthetic seismograms. Phys. Earth Planet. Int. **119**, 75-97 (2000)

256. Moczo, P., Bystrický, Kristek, J., Carcione, J. M., Bouchon, M.: Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures. Bull. Seis. Soc. Am. **87**, 1305-1323 (1997)
257. Moczo, P., Kristek, J., Halada, L.: 3D fourth-order staggered-grid finite-difference schemes: stability and grid dispersion. Bull. Seis. Soc. Am. **90**, 587-603 (2000)
258. Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R. J., Halada, J.: 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli. Bull. Seis. Soc. Am. **92**, 3042-3066 (2002)
259. Moczo, P., Kristek, J., Gális, M.: Simulation of the planar free surface with near-surface lateral discontinuities in the finite-difference modeling of seismic motion. Bull. Seis. Soc. Am. **94**, 760-768 (2004)
260. Moczo, P., Kristek, J.: On the rheological methods used for the time-domain methods of seismic wave propagation. Geophys. Res. Lett. **32**(1), Art. No. L01306 (2005)
261. Moczo, P., Kristek, J., Galis, M., Pazak, P., Balazovjech, M.: The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. Acta Physica Slovaca **57**(2), 177-406 (2007)
262. Moczo, P., Kristek, J., Galis, M., Pazak, P.: On accuracy of the finite-difference and finite-element schemes with respect to P-wave to S-wave speed ratio. Geophys. J. Int. **182**, 493-510 (2010)
263. Mohorovičić, A.: Das Beben vom 8. Oktober 1909. Jahrbuch des meteorologischen Observatoriums in Zagreb für 1909, **4**(1), 1-67 (1910) (translated title)
264. Montagner, J.-P.: Seismic anisotropy of the Pacific Ocean inferred from long-period surface waves dispersion. Phys. Earth Planet. Int. **38**, 28-50 (1985)
265. Montagner, J.-P.: Upper mantle low anisotropy channels below the Pacific Plate. Earth Planet. Sci. Lett. **202**, 263-274 (2002)
266. Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., Hung, S. H.: Finite-frequency tomography reveals a variety of plumes in the mantle. Science **303**, 338-343 (2004)
267. Mora, P.: Elastic finite differences with convolutional operators. Stanford Expl. Proj. Rep. **48**, 277-289 (1986)
268. Mora, P.: Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics **52**, 1211-1228 (1987)
269. Mora, P.: Elastic wave-field inversion of reflection and transmission data. Geophysics **53**, 750-759 (1988)
270. Mora, P.: Inversion=migration+tomography. Geophysics, **54**, 1575-1586 (1989)
271. Morelli, A., Dziewonski, A. M.: Body wave traveltimes and a spherically symmetric P- and S-wave velocity model. Geophys. J. Int. **122**, 178-194 (1993)
272. Muir, F., Dellinger, J., Etgen, J., Nichols, D.: Modeling elastic fields across irregular boundaries. Geophysics **57**, 1189-1193 (1992)
273. Myers, J. S., Shaw, R. D., Tyler, I. M.: Tectonic evolution of Proterozoic Australia. Tectonics **15**, 1431-1446 (1996)
274. Nataf, H. C., Ricard, Y.: 3SMAC: an a priori tomographic model of the upper mantle based on geophysical modeling. Phys. Earth Planet. Int. **95**, 101-122 (1995)
275. Nehrbass, J. W., Lee, J. F., Lee, R., 1996. *Stability analysis for perfectly matched layered absorbers*, Electromagnetics, **16**(4), 385-397.
276. Nettles, M., Dziewonski, A. M.: Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America. J. Geophys. Res. **113**, B02303, doi:10.1029/2006JB004819 (2008)
277. Newmark, N. M.: A method of computation for structural dynamics. ASCE J. of the Engineering Mechanics Division **85**, 67-94 (1959)
278. Nissen-Meyer, T.: Numerical simulation of 3D seismic wave propagation through subduction zones. Diploma thesis, Ludwig-Maximilians-Universität München (2001)
279. Nissen-Meyer, T., Fournier, A., Dahlen, F. A.: A two-dimensional spectral-element method for computing spherical-earth seismograms - I. Moment-tensor source. Geophys. J. Int. **168**, 1067-1092 (2007a)
280. Nissen-Meyer, T., Dahlen, F. A., Fournier, A.: Spherical Earth Fréchet sensitivity kernels. Geophys. J. Int. **168**, 1051-1066 (2007b)
281. Nissen-Meyer, T., Fournier, A., Dahlen, F. A.: A two-dimensional spectral-element method for computing spherical-earth seismograms - II. Waves in solid-fluid media. Geophys. J. Int. **174**, 873-888 (2008)
282. Nolet, G., Montelli, R.: Optimal parametrization of tomographic models. Geophys. J. Int. **161**, 1-8 (2005)
283. Nolet, G.: A breviary of seismic tomography. Cambridge Univ. Press (2008)
284. Nyman, D. C., Landisman, M.: The display-equalized filter for frequency-time analysis. Bull. Seis. Soc. Am. **67**, 393-404 (1977)
285. Ohminato, T., Chouet, B.: A free-surface boundary condition for including 3D topography in the finite-difference method. Bull. Seis. Soc. Am. **87**, 494-515 (1997)
286. Oldham, R. D.: On the propagation of earthquake motion to great distances. Phil. Trans. Roy. Soc. London **194**, 135-174 (1900)
287. Oldham, R. D.: The constitution of the interior of the Earth as revealed by earthquakes. Quart. J. Geol. Soc. London **62**, 456-475 (1906)
288. Operto, S. and Virieux, J. and Dessa, J. X. and Pascal, G.: Crustal Seismic Imaging from Multifold Ocean Bottom Seismometer Data by Frequency Domain Full Waveform Tomography: Application to the Eastern Nankai Trough. J. Geophys. Res. **111**, doi:10.1029/2005JB003835 (2006)
289. Pan, G. S., Phinney, R. A., Odom, R. I.: Full-waveform inversion of plane-wave seismograms in stratified acoustic media: Theory and feasibility. Geophysics **53**, 21-31 (1988)
290. Patera, A. T.: A spectral element method for fluid dynamics: laminar flow in a channel expansion. J. Comp. Phys. **54**, 468-488 (1984)
291. Pica, A., Diet, J. P., Tarantola, A.: Non-linear inversion of seismic reflection data in a laterally invariant medium. Geophysics **55**, 284-292 (1990)
292. Polak, E., Ribière, G.: Note sur la convergence de la méthode de directions conjuguées. Revue Fr. Inf. Rech. Oper. **16-R1**, 35-43 (1969)
293. Polak, E.: Optimization. Springer Verlag, New York (1997)

294. Pozrikidis, C.: Finite and spectral element methods using Matlab. Chapman & Hall/CRC (2005)
295. Pratt, R. G., Worthington, M. H.: Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave equation method. Geophys. Prosp. **38**, 287-310 (1990)
296. Pratt, R. G., Song, Z. M., Williamson, P., Warner, M.: Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. Geophys. J. Int. **124**, 323-340 (1996)
297. Pratt, R. G., Shin, C., Hicks, G. J.: Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. Geophys. J. Int. **133**, 341-362 (1998)
298. Pratt, R. G.: Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Geophysics **64**, 888-901 (1999)
299. Pratt, R. G., Shipp, R. M.: Seismic waveform inversion in the frequency domain, Part 2: Fault delineation in sediments using crosshole data. Geophysics **64**(3), 902-914 (1999)
300. Press, F.: Earth models obtained by Monte Carlo Inversion. J. Geophys. Res. **73**, 5223-5234 (1968)
301. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P.: Numerical recipies, 3rd edition. Cambridge University Press (2007)
302. Priolo, E., Carcione, J. M., Seriani, G.: Numerical simulation of interface waves by high-order spectral modeling techniques. J. Acoust. Soc. Am. **95**, 681-693 (1994)
303. Quarteroni, A., Sacco, R., Saleri, F.: Numerical mathematics. Springer Verlag, New York (2000)
304. Ravaut, C. and Operto, S. and Improta, L. and Virieux, J. and Herrero, A. and Dell'Aversana, P.: Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. Geophys. J. Int. **159**, 1032-1056 (2004)
305. Ritsema, J., van Heijst, H. J., Woodhouse, J. H.: Complex shear velocity structure imaged beneath Africa and Iceland. Science **286**, 1925-1928 (1999)
306. Ritsema, J., van Heijst, H. J.: Seismic imaging of structural heterogeneity in Earth's mantle: Evidence for large-scale mantle flow. Sci. Progr. **83**, 243-259 (2000)
307. Robertsson, J. O. A., Blanch, J. O., Symes, W. W.: Viscoelastic finite-difference modeling. Geophysics **59**(9), 1444-1456 (1994)
308. Robertsson, J. O. A.: A numerical free-surface condition for elastic/visco-elastic finite-difference modeling in the presence of topography. Geophysics **61**, 1921-1934 (1996)
309. Romanowicz, B.: The upper mantle degree 2: Constraints and inferences from global mantle wave attenuation measurements. J. Geophys. Res. **95**, 11051-11071 (1990)
310. Ronchi, C., Ianoco, R., Paolucci, P. S.: The 'Cubed Sphere': a new method for the solution of partial differential equations in spherical geometry. J. Comp. Phys. **124**, 93-114 (1996)
311. Sadourny, R.: Conservative finite-difference approximations of the primitive equation on quasi-uniform spherical grids. Mon. Weather Rev. **100**, 136-144 (1972)
312. Sambridge, M., Mosegaard, K.: Monte Carlo methods in geophysical inverse problems. Rev. Geophys. **40**, doi:10.1029/2000RG000089 (2002)
313. Santosa, F., Symes, W. W.: Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology. Inverse Problems **4**, 211-233.
314. Schoenberg, M., Muir, F.: A calculus for finely layered anisotropic media. Geophysics **54**(5), 581-589 (1989)
315. Schuberth, B. S. A., Bunge, H.-P., Steinle-Neumann, G., Moder, C., Oeser, J.: Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: High plume excess temperatures in the lowermost mantle. Geochem. Geophys. Geosys. **10**, Q01W01, doi:10.1029/2008GC002235 (2009a)
316. Schuberth, B. S. A., Bunge, H.-P., Ritsema, J.: Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? Geochem. Geophys. Geosys. **10**, Q05W03, doi:10.1029/2009GC002401 (2009b)
317. Seriani, G., Priolo, E., Pregarz, A.: Modelling waves in anisotropic media by a spectral element method. Proceedings of the 3rd International Conference on Mathematical and Numerical Aspects of Wave Propagation, 289-298 (1995).
318. Seriani, G.: 3-D large-scale wave propagation modeling by a spectral element method on a Cray T3E multiprocessor. Comp. Meth. Appl. Mech. Eng. **164**, 235-247 (1998)
319. Seriani, G., Oliveira, S. P.: Dispersion analysis of spectral element methods for elastic wave propagation. Wave Motion **45**, 729-744 (2008)
320. Sheng, J. M., Leeds, A., Buddensiek, M., Schuster, G. T.: Early arrival waveform tomography on near-surface refraction data. Geophysics **71**, U47-U57 (2006)
321. Shibutani, T., Sambridge, M., Kennett, B. L. N.: Genetic algorithm inversion for receiver functions with application to crust and uppermost mantle structure beneath Eastern Australia: Geophys. Res. Lett. **23**, 1829-1832 (1996)
322. Shin, C., Min, D. J.: Waveform inversion using a logarithmic wave field. Geophysics **71**, R31-R42 (2006)
323. Shin, C., Cha, Y. H.: Waveform inversion in the Laplace domain. Geophys. H. Int. **173**, 922-931 (2008)
324. Shito, A., Karato, S.-I., Park, J.: Frequency dependence of Q in Earth's upper mantle inferred from continuous spectra of body waves. Geophys. Res. Lett. **31**, doi:10.1029/2004GL019582 (2004)
325. Sieminski, A., Lévêque, J.-J., Debayle, E.: Can finite-frequency effects be accounted for in ray theory surface wave tomography? Geophys. Res. Lett. **31**(24), L24614 (2004)
326. Sieminski, A., Liu, Q., Trampert, J., Tromp, J.: Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods. Geophys. J. Int. **168**(3), 1153-1174 (2007a)
327. Sieminski, A., Liu, Q., Trampert, J., Tromp, J.: Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods. Geophys. J. Int. **171**(1), 368-389 (2007b)
328. Sigloch, K., Nolet, G.: Measuring finite-frequency body-wave amplitudes and traveltimes. Geophys. J. Int. **167**, 271-287 (2006)

329. Sigloch, K., McQuarrie, N., Nolet, G.: Two-stage subduction history under North America inferred from multiple-frequency tomography. Nature Geoscience **1**, 458-462 (2008)
330. Simone, A., Hestholm, S.: Instabilities in applying absorbing boundary conditions to high-order seismic modelling algorithms. Geophysics **63**, 1017-1023 (1998)
331. Simons, F. J., Zielhuis, A., van der Hilst, R. D.: The deep structure of the Australian continent from surface wave tomography. Lithos **48**, 17-43 (1999)
332. Simons, F. J., van der Hilst, R. D., Montagner, J.-P., Zielhuis, A.: Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle. Geophys. J. Int. **151**, 738-754 (2002)
333. Song, Z. M., Williamson, P. R., Pratt, R. G.: Frequency-domain acoustic wave modeling and inversion of crosshole data: Part 2 - inversion method, synthetic experiments and real-data results. Geophysics **60**, 796-809 (1995)
334. Spetzler, J., Trampert, J., Snieder, R.: Are we exceeding the limits of the great circle approximation in global surface wave tomography? Geophys. Res. Lett. **28**(12), 2341-2344 (2001)
335. Šteckl, I., Pratt, R. G.: Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators. Geophysics **63**, 1779-1794 (1998)
336. Stich, D., Danecek, P., Morelli, A., Tromp, J.: Imaging lateral heterogeneity in the northern Apennines from time reversal of reflected surface waves. Geophys. J. Int. **177**, 543-554 (2009)
337. Stolt, R. H.: Migration by Fourier transform. Geophysics **43**, 23-48 (1978)
338. Strang, G., Nguyen, T.: Wavelets and Filter Banks, Wellesley-Cambridge Press (1996)
339. Stuart, A., Pavliotis, G.: Multiscale Methods: Averaging and Homogenization. Springer (2007)
340. Stupazzini, M.: 3D ground motion simulation of the Grenoble valley by GeoELSE. Proceedings of the 3rd International Symposium on the Effects of Surface Geology on Seismic Motion (ESG), Grenoble, France.
341. Stupazzini, M., Paolucci, R., Igel, H.: Near-fault earthquake ground-motion simulation in the Grenoble valley by a high-performance spectral-element code. Bull. Seis. Soc. Am. **99**, 286-301.
342. Szegö, G.: Orthogonal Polynomials. American Mathematical Society, 5th edition, Providence, RI (1975)
343. Takeuchi, H., Saito, M.: Seismic surface waves. in: Methods in Computational Physics **11**, 217-295, Academic Press (1972)
344. Takeuchi, N., Geller, R. J.: Optimally accurate second-order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media. Phys. Earth Planet. Int. **119**, 99-131 (2000)
345. Takeuchi, N., Geller, R. J., Cummins, P. R.: Complete synthetic seismograms for 3-D heterogeneous Earth models computed using modified DSM operators and their applicability to inversion for Earth structure. Phys. Earth Planet. Int. **119**, 25-36 (2000)
346. Tape, C., Liu, Q. Y., Maggi, A., Tromp, J.: Adjoint tomography of the southern California crust. Science **325**, 988-992 (2009)
347. Tape, C., Liu, Q. Y., Maggi, A., Tromp, J.: Seismic tomography of the southern California crust based upon spectral-element and adjoint methods. Geophys. J. Int. **180**, 433-462 (2010)
348. Tarantola, A., Valette, B.: Generalised nonlinear inverse problems solved using the least squares criterion. Rev. Geophys. **20**, 219-232 (1982)
349. Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. Geophysics **49**, 1259-1266 (1984)
350. Tarantola, A.: Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. Pure Appl. Geophys. **128**, 365-399 (1988)
351. Tarantola, A.: Inverse problem theory and methods for model parameter estimation, 2nd edition. Society for Industrial and Applied Mathematics, Philadephia (2005)
352. Teixeira, F. L. & Chew, W. C.: Systematic derivation of anisotropic pml absorbing media in cylindrical and spherical coordinates. IEEE Microwave and Guided Wave Letters **7**, 371-373 (1997)
353. Tessmer, E., Kosloff, D. D.: 3D elastic modelling with surface topography by a Chebyshev spectral method. Geophysics **59**, 464-473 (1994)
354. Thomson, D. J.: Spectrum estimation and harmonic analysis. Proc. IEEE **70**, 1055-1096 (1982)
355. Thurber, C., Roecker, S., Roberts, K., Gold, M., Powell, L., Rittger, K.: Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: Preparing for SAFOD. Geophys. Res. Lett. **30**, doi:10.1029/2002GL016004 (2003)
356. Tibuleac, I. M., Nolet, G., Michaelson, C., Koulakov, I.: P wave amplitudes in a 3-D earth. Geophys. J. Int. **155**, 1-10 (2003)
357. Tikhonov, A. N.: Solution of incorrectly formulated problems and the regularisation method. Dokl. Akad. Nauk. SSSR **151**, 501-504 (1963)
358. Toshinawa, T., Ohmachi, T.: Love-wave propagation in a three-dimensional sedimentary basin. Bull. Seis. Soc. Am. **82**, 1661-1677 (1992)
359. Trampert, J., Spetzler, J.: Surface wave tomography: finite-frequency effects lost in the null space. Geophys. J. Int. **164**(2), 394-400 (2006)
360. Tromp, J., Tape, C., Lui, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophys. J. Int. **160**, 195-216 (2005)
361. Udías, A.: Principles of seismology. Cambridge University Press (1999)
362. van der Hilst, R. D., Widiyantoro, S., Engdahl, E. R.: Evidence for deep mantle circulation from global tomography. Nature **386**, 578-584 (1997)
363. van der Hilst, R. D., Krason, H.: Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model. Science **283**, 1885-1888 (1999)
364. van der Hilst, R. D., de Hoop, M. V.: Banana-doughnut kernels and mantle tomography. Geophys. J. Int. **163**(3), 956-961 (2005)

365. van Leeuwen, T., Mulder, W. A.: A correlation-based misfit criterion for wave-equation traveltime tomography. Geophys. J. Int. **182**, 1383-1394 (2010)
366. Virieux, J.: SH-wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics **49**, 1933-1957 (1984)
367. Virieux, J.: P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics **51**, 889-901 (1986)
368. Virieux, J., Operto, S.: An overview of full waveform inversion in exploration geophysics. Geophysics **74**, WCC1-WCC26 (2009)
369. Vogel, C. R., Oman, M. E.: Iterative methods for total variation denoising. SIAM J. Sci. Comp. **17**, 227-238 (1996)
370. von Rebeur-Paschwitz, E. L. A.: The earthquake of Tokio 18. April 1889. Nature **40**, 32-32 (1889)
371. Wang, Y. H., Rao, Y.: Crosshole seismic waveform tomography I.: strategy for real data application. Geophys. J. Int. **166**, 1224-1236 (2006)
372. Wang, H. J., Igel, H., Gallovic, F., Cochard, A., Ewald, M.: Source-related variations of ground motions in 3D media: application to the Newport-Inglewood fault, Los Angeles basin. Geophys. J. Int. **175**, 202-214 (2008)
373. Williamson, P. R.: A guide to the limits of resolution imposed by scattering in ray tomography. Geophysics **56**(2), 202-207 (1991)
374. Williamson, P. R., Worthington, M. H.: Resolution limits in ray tomography due to wave behavior: Numerical experiments. Geophysics **58**(5), 727-735 (1993)
375. Woodhouse, J. H., Dahlen, F. A.: Effect of a general aspherical perturbation on free oscillations of the Earth. Geophys. J. R. astr. Soc. **53**, 335-354 (1978)
376. Woodhouse, J.: The joint inversion of seismic waveforms for lateral variations in Earth structure and earthquake source parameters. In: Physics of the Earth's Interior **85**, 366-397; North Holland (1983)
377. Woodhouse, J. H.: The calculation of eigenfrequencies and eigenfunctions of the free oscillations of the earth and the sun. In: Seismological Algorithms, 321-370, Academic Press, London (1988)
378. Woodward, M. J.: Wave-equation tomography. Geophysics **57**, 15-26 (1992)
379. Wu, R.-S., Toksöz, M. N.: Diffraction tomography and multisource holography applied to seismic imaging. Geophysics **52**(1), 11-25 (1987)
380. Yomogida, K.: Fresnel zone inversion for lateral heterogeneities in the Earth. Pure Appl. Geophys. **138**, 391-406 (1992)
381. Yoshizawa, K., Kennett, B. L. N.: Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects. J. Geophys. Res. **109**, doi:10.1029/2002JB002254 (2004)
382. Yoshizawa, K., Kennett, B. L. N.: Sensitivity kernels for finite-frequency surface waves. Geophys. J. Int. **162**, 910-926 (2005)
383. Zahradník, J., Urban, L.: Effect of a simple mountain range on underground seismic motion. Geophys. J. R. Astr. Soc. **79**, 167-183 (1984)
384. Zahradník, J., Moczo, P., Hron, F.: Testing four elastic finite difference schemes for behaviour at discontinuities. Bull. Seis. Soc. Am. **83**, 107-129 (1993)
385. Zhang, G. Q.: High-order approximation of one-way wave equations. J. Comp. Math. **3**, 90-97 (1985)
386. Zhao, L., Jordan, T. H., Chapman, C. H.: Three-dimensional Fréchet differential kernels for seismic delay times. Geophys. J. Int. **141**, 558-576 (2000)
387. Zhao, L., Jordan, T. H., Olsen, K. B., Chen, P.: Fréchet kernels for imaging regional Earth structure based on three-dimensional reference models. Bull. Seis. Soc. Am. **95**, 2066-2080 (2005)
388. Zheng, Y. & Huang, X.: Anisotropic perfectly matched layers for elastic waves in cartesian and curvilinear coordinates. MIT Earth Resources Laboratory, Consortium Report (2002)
389. Zhou, C., Cai, W., Luo, Y., Schuster, G., Hassanzadeh, S.: Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data. Geophysics **60**, 765-773 (1995)
390. Zhou, Y., Dahlen, F. A., Nolet, G.: Three-dimensional sensitivity kernels for surface wave observables. Geophys. J. Int. **158**, 142-168 (2004)
391. Zhou, Y., Dahlen, F. A., Nolet, G., Laske, G.: Finite-frequency effects in global surface wave tomography. Geophys. J. Int. **163**, 1087-1111 (2005)
392. Zhou, Y.: Multimode surface wave sensitivity kernels in radially anisotropic earth media. Geophys. J. Int. **176**, 865-888 (2009)
393. Zielhuis, A., Nolet, G.: Shear-wave velocity variations in the upper mantle beneath central Europe. Geophys. J. Int. **117**, 695-715 (1994)
394. Zielhuis, A., van der Hilst, R. D.: Upper-mantle shear velocity beneath eastern Australia from inversion of waveforms from SKIPPY portable arrays. Geophys. J. Int. **127**, 1-16 (1996)
395. Zoeppritz, K. B.: Über Erdbebenwellen II. Laufzeitkurven. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 529-549 (1907)
396. Zoeppritz, K. B., Geiger, L.: Über Erdbebenwellen III. Berechnung von Weg und Geschwindigkeit der Vorläufer. Die Poissonsche Konstante im Erdinnern. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 400-428 (1909)
397. Zoeppritz, K. B., Geiger, L., Gutenberg, B.: Über Erdbebenwellen V. Konstitution des Erdinnern, erschlossen aus dem Bodenverrückungsverhalten der einmal reflektierten zu den direkten longitudinalen Erdbebenwellen, und einige andere Beobachtungen über Erdbebenwellen. Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 121-206 (1912)