

Spectral-Element Simulations of 3-D Seismic Wave Propagation and Applications to Source and Structural Inversions

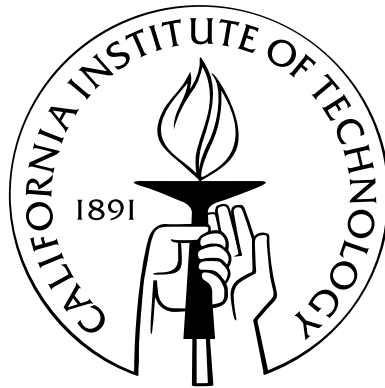
Thesis by

Qinya Liu

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2006

(Defended May 9, 2006)

© 2006

Qinya Liu

All Rights Reserved

To My Parents

Acknowledgements

My six years of graduate study at the Seismo Lab has finally come to an end. These six years of valuable experience has shaped and will continue to shape the course of my life. I would not have been able to come to this point and finish this thesis without the extensive help and care I received from many people, including my parents, as well as my colleagues and friends in the Seismo Lab. I am grateful to all of you.

I could not ask for a better adviser than Dr. Jeroen Tromp, who not only opened in front of me a vast world of 3-D numerical simulations and innovative methods to apply these great tools to improve both the earthquake source and the earth structure, but also has been a great support for me through very hard times with his constant encouragement, persistent understanding, and great confidence in me. I am always deeply affected by his enthusiasm and optimism towards our research work. Being more like a friend to all the students and post-docs in our group, he provides us a very harmonic working environment to pursue our research interest with easy access to help.

I would also like to thank my other thesis committee members. Joann Stock, my academic adviser has kindly assisted me in choosing the courses I needed to take for the first few years; I have had several fruitful discussions with Rob Clayton, who was also the first faculty member I met when I first arrived at Caltech six years ago. I would also like to thank Hiroo Kanamori for his great teaching and his kindness to review one of my papers. Don Helmberger, who is also the advisor of my second oral project, always amazes me with his persistence in exploring the earth through wiggle-to-wiggle understanding of seismic waveforms.

My officemates have been a constant source of support for me in every aspect. The nice small talks in the office have always lifted up my spirit and created a very delightful working environment. Carl Tape, Yoshi Kaneko, and Lydia Taylor have extended their

hands to me and pulled me through difficulties. Anita Ai and Shengnian Luo also assisted me in many aspects.

I also feel lucky that we have a very amazing research group, not only because of all the fruitful research collaborations and interesting group discussions, but also because of the very good personal friendships built on mutual sincerity. I feel grateful to have been able to work with Vala Hjorleifsdottir, Carl Tape, Min Chen, Ying Zhou, Carene Larmat, Swami Krishnan, Alessia Maggi and many others, and thank them for their constant support and encouragement, and I look forward to continue collaborating with them in the future. I would also like to express my sincere thanks to Ying Zhou, Vala Hjorleifsdottir, and Brian Savage for their help, understanding, and encouragement.

I would like to thank all my other labmates: Eun-seo Choi and Ozgun Konca have always been enjoyable to talk to; Ying Tan, Zhimei Yan, Eh Tan, and Daoyuan Sun have been very kind to me; Chen Ji and Yaru Hsu have helped me a lot in my personal life. And I have also learned a lot from all of them.

I would like to thank Viola Carter, Elisa Loeffen, Evelina Cui, Sue Yamada, and other Seismo Lab staff for their support and great work that has made Seismo Lab a very enjoyable place for both study and research. I want to give my special thanks to Viola Carter, who not only does a wonderful job assisting our research group, but also provided me with a lot of insightful advice and valuable information in many circumstances.

I have also benefited a lot from collaborating with other researchers, including Dimitri Komatitsch, who helped me a lot in understanding the specfem code; Jasha Polet, who collaborated with me on the spectral-element moment-tensor inversions; Peter Lovely, who provided us with a detailed Salton Trough model; Paul Friberg, who showed great patience in trying to understand my moment-tensor inversion code and make a neat database out of it, and many others.

Finally I would like to thank my parents, who have been caring for and worrying about me since the day I was born. Their unselfish love has always been a great comfort and support throughout my life, particularly during difficult times. I should especially thank my mother, who sacrificed a lot to take care of me in the past few months. This thesis and some of the research work would not have been possible without her.

Abstract

This thesis presents a concise introduction to the spectral-element method and its applications to the simulation of seismic wave propagation in 3-D earth models. The spectral-element method is implemented in the regional scale for a 3-D integrated southern California velocity model. Significantly better waveform fits are achieved for the 3-D synthetics calculated compare to the 1-D synthetics generated from the 1-D standard southern California model, especially for many basin stations where strong amplifications are observed due to the very low wave-speed sediments. A hypothetical earthquake rupturing from northeast to southwest at the southern end of the San Andreas fault is simulated to improve our understanding of the seismic hazards in the Salton Trough region.

With the improved 3-D Green's function, we perform source inversions for both the source mechanisms and event depths of $M_w \geq 3.5$ earthquakes in southern California. The inversion results generally agree well with the results obtained by other traditional methods, but with significantly more stations used in the inversions. Time shifts are generally required to align the data and the synthetics, which provides a great dataset for the improvement of the 3-D velocity models in southern California.

We use the adjoint method to formulate the tomographic inverse problem based upon a 3-D initial model. We calculate the sensitivity kernels, a key component of the tomographic inversion, that relate the perturbations of observations to the perturbations of the model parameters. These kernels are constructed by the interaction of the regular forward wavefield and the adjoint wavefield generated by putting the time-reversed signals at the receivers as simultaneous adjoint sources. We compute the travel-time sensitivity kernels for typical phases in both regional and global problems for educational purposes, and outline the procedures of applying the conjugate-gradient method to solve both source and structural inversion problems.

Contents

Acknowledgements	iv
Abstract	vi
1 Introduction	1
2 Introduction to the Spectral Element Method	3
2.1 Abstract	3
2.2 Review of Existing Numerical Techniques	3
2.3 The Theory of the Spectral-Element Method	5
2.3.1 Strong Formulation – the Equation of Motion	5
2.3.2 Weak Formulation	6
2.3.3 Earthquake Source Representation	7
2.3.4 Attenuation	8
2.3.5 Absorbing Boundary Conditions	9
2.3.6 Discretization	9
2.3.7 Representation of Functions	11
2.3.8 Integration Quadrature	12
2.3.9 Global Test Functions	12
2.3.10 Applications to the Weak Form of the Wave Equation	13
2.3.10.1 Mass Matrix	14
2.3.10.2 Strain and Stress	14
2.3.10.3 Stiffness Matrix	15
2.3.10.4 Force Terms	16

2.3.10.5	Boundary terms	17
2.3.11	Assembling and Time Marching	17
2.3.12	Accuracy Analysis	19
3	Spectral-Element Simulation for Earthquakes in Southern California	20
3.1	Abstract	20
3.2	Model Description	20
3.3	Numerical Implementations	23
3.3.1	Mesh Design for Realistic 3-D Velocity Models	23
3.3.1.1	Adaption to Velocity Discontinuities	23
3.3.1.2	Doubling Scheme	23
3.3.2	Parallel Computing	24
3.4	Yorba Linda Earthquake	24
3.5	Hypothetical Major Rupture on the Southern San Andreas Fault	26
3.6	Discussion	29
4	Spectral-Element Moment-Tensor Inversions for Earthquakes in Southern California	32
4.1	Abstract	32
4.2	Introduction	33
4.3	Moment-Tensor Inversion Theory	34
4.4	Time Shifts	36
4.5	Numerical Tests	38
4.5.1	Synthetic Test: Latitude	39
4.5.2	Synthetic Test: Depth	39
4.5.3	Synthetics Test: Envelope Misfit Versus Waveform Misfit Without Time Shifts	40
4.5.4	Data Test: Depth	41
4.6	Inversion Procedures and Technical Details	41
4.6.1	Calculation and Processing of Synthetics	41
4.6.2	Data Processing	43

4.6.3	Inversion	43
4.7	Hollywood Earthquake	44
4.8	Big Bear Main Shock	49
4.9	Diamond Bar Earthquake	50
4.10	Discussion	54
4.11	Conclusion	57
5	Finite-frequency Kernels Based upon Adjoint Method for Regional Wave Sim- ulations	58
5.1	Abstract	58
5.2	Introduction	59
5.3	Lagrange Multiplier Method	60
5.4	Spectral Element Method	66
5.5	Numerical Implementation of the Adjoint Method	67
5.6	Application to a 3-D Homogeneous Model	68
5.6.1	S Kernel	71
5.6.2	SS Kernel	71
5.6.3	P Kernel	72
5.6.4	PS+SP Kernel	73
5.7	Application to a 3-D Southern California Model	75
5.7.1	P Kernels	75
5.7.2	S Kernels	76
5.7.3	Surface-wave Kernels	77
5.8	Tomographic Inversions	78
5.9	Conclusions	80
6	Finite-frequency Kernels Based upon Adjoint Method for Global Wave Simu- lations	82
6.1	Abstract	82
6.2	Introduction	83
6.3	Equation of Motion	84

6.4	Adjoint Equations	87
6.5	Anelasticity	97
6.6	Global Seismic Wave Propagation Based upon the Spectral Element Method	99
6.7	Numerical Implementation of the Adjoint Method	100
6.8	Applications to Spherically Symmetric Models	101
6.8.1	P Kernel	103
6.8.2	P_{diff} Kernel	105
6.8.3	PKP Kernel	106
6.8.4	S and ScS Kernels	109
6.8.5	SKS Kernel	109
6.8.6	Depth Phases	111
6.8.7	Kernels for Surface-reflected Phases	112
6.8.8	P, P, P Kernel	115
6.9	Conjugate-gradient Method for Tomography Inversions	115
6.10	Discussion	118
A	Attenuation by Standard Linear Solids	119
A.1	Complex Wave Number	119
A.2	Theory of viscoelasticity	119
A.3	Standard Linear Solids	120
A.4	Multiple Standard Linear Solid	121
A.5	Constant Q	122
A.6	Implementations of Anelasticity in SEM	124
B	More on Spectral-Element Moment-Tensor Inversion Theory	126
B.1	Linear inversions	126
B.2	Non-linear inversions	127
C	More on the Adjoint Method	131
C.1	Absorbing Boundaries	131
C.2	Finite Source	133

List of Figures

3.1	Topographic map and cross-sections of the 3D integrated southern California velocity model	22
3.2	Mesh Doubling Scheme used in the mesh design for the 3D integrated southern California velocity	25
3.3	Comparison of waveform fits for 3D vs. 1D synthetics for basin stations . . .	27
3.4	Waveform fits between the data and the 3D synthetics for representative stations in southern California	28
3.5	Movie frames of a hypothetical earthquake on the San Andreas fault	30
3.6	Shakemaps of the hypothetical rupture event on the San Andreas fault	31
4.1	Numerical tests of waveform misfit functions	42
4.2	Correlation between data and synthetics for the initial source mechanisms of the Hollywood event	48
4.3	Variance reduction for transverse component synthetics calculated based upon the inverted source parameters for the 2003 Big Bear main shock	51
4.4	Waveform fits for transverse component data and synthetics for the inverted source parameters of the Big Bear main shock	52
4.5	Stations used in the preliminary source inversion and the standard source inversions for the Diamond Bar event	55
4.6	Misfit reduction for transverse component seismograms by the source parameters for the Diamond Bar event	56
5.1	Simultaneous access to the forward and adjoint wave field	69
5.2	Experiment setup for the computation of kernels for homogeneous models . .	70

5.3	Synthetic seismograms calculated for a homogeneous velocity model	70
5.4	S sensitivity kernels for a homogeneous velocity model	71
5.5	SS sensitivity kernels for a homogeneous velocity model	72
5.6	P sensitivity kernels for a homogeneous velocity model	73
5.7	PS+SP sensitivity kernels for a homogeneous velocity model	74
5.8	P Sensitivity kernel for the integrated southern California model	76
5.9	S sensitivity kernel for the integrated southern California model	77
5.10	Surface wave kernels for the integrated southern California model	78
6.1	Ray geometry for some of the arrivals considered in this study	104
6.2	K_α kernels for P phases at $\Delta = 60^\circ$	105
6.3	Synthetic seismograms for P_{diff} and PKP phases	106
6.4	K_α kernels for P_{diff} phases at $\Delta = 103^\circ$	107
6.5	Travel-time curves for PKP phase	108
6.6	K_α kernels for PKP phases at $\Delta = 170^\circ$	108
6.7	Synthetic seismograms for S, ScS, sS, and SS phases	109
6.8	K_β kernels for S and ScS phases at $\Delta = 60^\circ$	110
6.9	Synthetic seismograms for SKS and sP phases	111
6.10	Sensitivity kernels for SKS phase at $\Delta = 115^\circ$	112
6.11	Sensitivity kernels for depth phases	113
6.12	Synthetic seismograms for SS, SP, and PP phases	113
6.13	Sensitivity kernels for surface-reflected phases	114
6.14	Synthetic seismograms for $P, P,$ phase	115
6.15	K_α kernel for the $P, P,$ phase at $\Delta = 70^\circ$	116
A.1	Standard linear solid	121
A.2	Multiple standard linear solids	123

List of Tables

4.1	Table of the source parameters obtained by various inversion methods for the Hollywood earthquake	47
4.2	Table of the source parameters obtained by various inversion methods for the Big Bear mainshock	50
4.3	Table of the source parameters obtained by various inversion methods for the Diamond Bar earthquake	54

Chapter 1

Introduction

Accurate simulation of seismic wave propagation in a 3-D heterogeneous medium has always been a crucial yet difficult task in seismology. By precisely computing the synthetic seismograms given any 3-D earth model for an earthquake, and comparing them to the actual seismograms recorded at the stations, we can evaluate either the quality of the velocity models or the accuracy of the earthquake source parameters, making it possible to improve the velocity models or the source parameters by further matching the data and the synthetics. Recently, the spectral-element method was introduced to simulate both regional and global wave propagation in seismology [e.g., *Komatitsch and Tromp, 2002a,b; Chaljub et al., 2003; Komatitsch et al., 2004*]. This method combines the geometrical flexibility of the finite-element method with an accurate representation of the wave field in terms of high-degree Lagrange polynomials. Its successful application to model global seismic wave propagation in 3-D heterogeneous tomography models has demonstrated significant improvement in both travel-time prediction and waveform fits compared to the results obtained from semi-analytical solutions to the wave equation for a 1-D laterally homogeneous earth model [*Komatitsch and Tromp, 2002a,b; Ritsema et al., 2002; Komatitsch et al., 2003; Ji et al., 2005*].

Chapter 2 presents a summary of the regional application of the spectral-element method to an integrated 3-D velocity model of southern California. This complicated 3-D heterogeneous model is generated by combining high-resolution structural-geology basin models with a regional tomographic model. The spectral-element mesh is carefully designed to incorporate different velocity discontinuities, e.g., the surface topography as well as the

Moho topography. A typical regional earthquake is simulated, and the synthetics at various seismic stations are compared to the ground motion recorded, producing generally good waveform fits throughout the southern California region at periods of 6 seconds and longer. With a high-resolution sedimentary basin model implemented for the Salton Trough region, a hypothetical scenario earthquake on the southern segment of the San Andreas fault is simulated to assess the possible intensity of shaking, especially in the Salton Trough region.

This great forward modeling tool presents us with the opportunity to further solve the most important problems in seismology, the source and structural inversions. In Chapter 4, we take advantage of the accuracy of the 3-D wave simulations for the integrated 3-D velocity model of southern California, and compute the 3D Green's functions required by moment-tensor inversions for the Southern California Seismic Network (SCSN) stations. With careful selection and weighting of the three component seismograms, we are able to invert the location and source mechanisms for some typical southern California earthquakes, using significantly more data compared to those used in the traditional body-wave and surface-wave source estimates.

Because of the limitations and accuracy of the forward wave modeling methods, traditional tomographic inversions are limited by the types and quantities of the data that can be collected for the inversion, and the approximation that has to be made in the inversion procedures. Using a full 3-D wave simulation approach completely eliminates these limitations. However, the computational cost of the spectral-element method as a forward modeling tool in the conventional framework of tomographic inversions may be prohibitive at this stage, given that a vast number of forward simulations is needed to perform the inversions. This leads to the application of the adjoint method to compute sensitivity kernels, which relate the perturbations of the misfit measurements between observations and the synthetics to the perturbations of the structural parameters. Chapter 5 and Chapter 6 present these sensitivity kernels for regional and global phases respectively, in both 1-D and 3-D velocity models. These kernels are the basic components in the construction of the gradient of the misfit function; therefore, they are the key ingredients for the adjoint tomographic inversions.

Chapter 2

Introduction to the Spectral Element Method

2.1 Abstract

This chapter serves as an introduction to the spectral element method and its applications to simulate seismic wave propagation. It starts by reviewing existing numerical schemes that calculate synthetic seismograms for 3-D velocity models. Then the theory and numerical implementations of the spectral element methods are introduced. The spectral element method describes the geometry of the mesh elements by the product of low-degree Lagrange polynomials, while it represents the function field by the product of high-degree Lagrange polynomials defined on the Gauss-Lobatto-Legendre (GLL) points, which renders higher accuracy compared to the classical finite element methods. Because of the choices of these GLL points, the mass matrix of the assembled discrete system is exactly diagonal, which significantly reduces the computational cost associated with solving a large linear system and makes it relatively easy to implement the numerical scheme in parallel.

2.2 Review of Existing Numerical Techniques

A large collection of numerical techniques is available for the simulation of wave propagation in 3-D elastic media. The most widely used approach is the finite difference method (FD), in which the strong form of the wave equation is discretized both spatially and tem-

porally [Kelly *et al.*, 1976; Virieux, 1986]. It has been successfully applied to solve both regional and global problems [e.g., Olsen and Archuleta, 1996; Graves, 1996; Igel and Weber, 1996], but it also suffers from issues like "grid dispersion" near large gradients of the wave field [e.g., Frankel, 1993; Olsen and Archuleta, 1996], "numerical anisotropy", and at least 15 points per wavelength is required at the upper half-power frequency to maintain numerical stability [e.g., Alford *et al.*, 1974; Kelly *et al.*, 1976]. The use of the strong formulation of the wave equation makes the method geometrically inflexible; therefore, it is difficult to incorporate 3-D surface topography in the wave simulation [e.g., Robertsson, 1996].

Pseudospectral methods are also popular in calculating the synthetic seismograms for 3-D earth models. Originally proposed in fluid dynamics, and later applied to elastodynamics and seismology [e.g., Gazdag, 1981; Tessmer and Kosloff, 1994; Furumura *et al.*, 1998], these methods, although often very accurate, are restricted to models with smooth variations, and suffer from the same geometrical inflexibility as the finite difference method.

The classical finite element method (FEM) has also been attempted by seismologists to calculate wave propagation in 3-D [e.g., Bao *et al.*, 1998]. It is based upon the variational formulation of the wave equation, and therefore can easily handle complicated geometries associated with the 3-D velocity model, e.g., the topographies of the free surface and any internal interface. However, the finite element methods, when used in low order, suffer from numerical dispersions, and when used in high order, produce spurious waves, and in both cases affect the accuracy of the simulation. The direct solution method proposed by Geller and Ohminato [1994] is another numerical approach for the calculation of synthetic seismograms for 3-D velocity models, similar to the finite element method. Both methods require the ability to solve large linear systems, and therefore are computationally costly.

2.3 The Theory of the Spectral-Element Method

2.3.1 Strong Formulation – the Equation of Motion

The most fundamental equation solved in seismology is the elastodynamics equation, given in its most general form by

$$\rho \partial_t^2 \mathbf{s} = \nabla \cdot \mathbf{T} + \mathbf{F} \quad \text{in } \Omega, \quad (2.1)$$

where Ω is the domain of interest, and $\mathbf{s}(\mathbf{x}, t)$ is the displacement field of the wave motion that needs to be solved; $\rho(\mathbf{x})$ denotes the density distribution throughout the domain, and $\mathbf{F}(\mathbf{x}, t)$ denotes the external body force that acts on the medium and excites the wave motion. In pure elastic simulations, the stress tensor \mathbf{T} can be related to the strain field by the following constitutive relation:

$$\mathbf{T} = \mathbf{c} : \boldsymbol{\epsilon} = \mathbf{c} : \nabla \mathbf{s}, \quad (2.2)$$

where \mathbf{c} denotes the fourth-order elastic tensor. The last equality of the above equation holds because of the symmetry of the elastic tensor. The wave equation is generally solved subject to the the initial conditions stating that the medium is at rest before the start of the simulation

$$\begin{aligned} \mathbf{s}(\mathbf{x}, t = 0) &= \mathbf{0} \\ \dot{\mathbf{s}}(\mathbf{x}, t = 0) &= \mathbf{0}. \end{aligned} \quad (2.3)$$

If the wave equation (2.1) is solved for a medium volume with boundaries (including internal boundaries), it is also subject to the boundary conditions

$$\begin{aligned} \hat{\mathbf{n}} \cdot \mathbf{T} &= \mathbf{0} \quad \text{on } \Sigma^0 \\ [\hat{\mathbf{n}} \cdot \mathbf{T}]_{-}^{+} &= \mathbf{0} \quad \text{on } \Sigma, \end{aligned} \quad (2.4)$$

where Σ^0 is the external boundary of the model volume Ω , Σ denotes the collection of all discontinuities inside Ω , and $\hat{\mathbf{n}}$ denotes the unit outward normal to the boundaries. Equation (2.1) subject to conditions (2.3) and (2.4) is also called the *strong formulation* or *differential formulation* of the wave equation.

2.3.2 Weak Formulation

In the finite element method, rather than using the equation of motion and the associated boundary conditions listed in section 2.3.1, one discretizes and solves the wave equation in its weak form. Dotting both sides of equation (2.1) with a test vector function $\mathbf{w}(\mathbf{x})$ and integrating by parts over the model volume Ω , we obtain the following expression:

$$\int_{\Omega} \rho \mathbf{w} \cdot \ddot{\mathbf{s}} dV = - \int_{\Omega} \nabla \mathbf{w} : \mathbf{T} d\Omega + \int_{\Omega} \mathbf{w} \cdot \mathbf{F} d\Omega + \int_{\Gamma} \mathbf{w} \cdot (\hat{\mathbf{n}} \cdot \mathbf{T}) d\Gamma, \quad (2.5)$$

where Γ is the absorbing boundaries, which will be discussed in section 2.3.5. Notice that the boundary condition (2.4) has already been used in the process of integration by parts and is naturally incorporated in the above expression. When combined with the initial condition in equation (2.3), equation (2.5) is also called the *weak formulation* or the *integral formulation* of the wave equation. And this is the common starting point of a collection of numerical techniques that use different test functions and/or discretize the wave field on different basis functions. Sections 2.3.6 through 2.3.12 illustrate the selections of test functions and basis functions by the spectral element method and the associated numerical implementations.

The purpose of sections 2.3.3 through 2.3.5 is to describe different aspects of the weak form (2.5), including how to express the body force $\mathbf{F}(\mathbf{x})$ in the context of seismology, how to incorporate attenuation to accommodate viscoelastic effects on the wave propagation, and how to deal with artificial boundary conditions when bounded domains are used to simulate wave propagation in an unbounded medium.

2.3.3 Earthquake Source Representation

Earthquake generally occurs as a result of dislocations on a fault. In the far field, the region where the faulting occurs can be considered as a point source. For example, a single force at location x_0 produces a force field given by

$$F_j(\mathbf{x}, t, \mathbf{f}) = f_j g(t) \delta(\mathbf{x} - \mathbf{x}_0), \quad (2.6)$$

where f_j is the amplitude of the force, and $g(t)$ is the normalized source time function that satisfies the requirement that $\int g(t) dt = 1$. A more realistic description of a seismic point source is given by the moment tensor \mathbf{M} , and the force field arising from a moment-tensor point force at location x_0 can be expressed as

$$F_j(\mathbf{x}, t, \mathbf{m}) = -M_{jk} g(t) \partial_k \delta(\mathbf{x} - \mathbf{x}_0). \quad (2.7)$$

When the length scale of the faulting caused by the earthquake rupture is substantially large, the earthquake source has to be represented by a finite fault. We mesh the finite fault into rectangular blocks, and define basis functions $P_{IJ}(\mathbf{x})$ associated with each rectangular block. We select a time window of the rupture history of the finite fault, and define the associated basis function with each time block $B_\sigma(t)$. Now the moment density tensor function can be expanded in terms of the basis functions:

$$m_{jk}(\mathbf{x}, t) = \sum_{I, J, \sigma} m_{jk}^{IJ\sigma} P_{IJ}(\mathbf{x}) B_\sigma(t). \quad (2.8)$$

Once we obtain the Green's functions associated with the earthquake location and a particular station, the response at the station for this earthquake can be calculated by the convolution of the Green's functions with the seismic sources represented in this section.

2.3.4 Attenuation

In a viscoelastic medium, the constitutive relation (2.2) needs to be modified such that the stress is determined not by the instantaneous strain, but by the entire strain history:

$$\mathbf{T}(t) = \int_{-\infty}^t \mathbf{c}(t-t') : \nabla \mathbf{s}(t') dt'. \quad (2.9)$$

Since the quality factor Q is observed to be constant over the wide range of frequencies of interest in seismology, the viscoelastic effects can be mimicked by a series of standard linear solids [Liu *et al.*, 1976], and the elastic tensor c_{ijkl} can be rewritten as

$$c_{ijkl}(t) = c_{ijkl}^R \left[1 - \sum_{l=1}^L (1 - \tau_{ijkl}^{el}/\tau^{\sigma l}) e^{-t/\tau^{\sigma l}} \right] H(t), \quad (2.10)$$

where c_{ijkl}^R denotes the relaxed modulus and $H(t)$ is the Heaviside function. In practice, only three linear solids ($L = 3$) are used to obtain a constant Q effect. The constitutive relation then becomes

$$\mathbf{T} = \mathbf{c}^U : \nabla \mathbf{s} - \sum_{l=1}^L \mathbf{R}^l, \quad (2.11)$$

where the unrelaxed modulus c_{ijkl}^U is given by

$$c_{ijkl}^U = c_{ijkl}^R \left[1 - \sum_{l=1}^L (1 - \tau_{ijkl}^{el}/\tau^{\sigma l}) \right], \quad (2.12)$$

and the memory variables \mathbf{R}^l satisfy the following first-order differential equations:

$$\partial_t \mathbf{R}^l = -\mathbf{R}^l/\tau^{\sigma l} + \delta \mathbf{c} : \nabla \mathbf{s}/\tau^{\sigma l}, \quad (2.13)$$

where the modulus defect δc_{ijkl}^l is defined by

$$\delta c_{ijkl}^l = -c_{ijkl}^R (1 - \tau_{ijkl}^{el}/\tau^{\sigma l}). \quad (2.14)$$

For detailed derivation of equations (2.10) through (2.14), please refer to Appendix A.

2.3.5 Absorbing Boundary Conditions

In order to simulate wave propagation in an unbounded medium, absorbing boundary conditions are introduced for the edges of the bounded model volume. Waves can propagate outward transparently through these artificial boundaries, avoiding reflections from these edges. The most popular approximated absorbing boundary conditions were introduced by *Clayton and Engquist* [1977] and *Stacey* [1988], who used the paraxial approximations of the elastic wave equations. The approximations we implement are similar to those of *Clayton and Engquist* [1977] with some small modifications. Under these approximations, traction on these artificial boundaries Γ , can be expressed as

$$\mathbf{t} = \rho [v_n(\hat{\mathbf{n}} \cdot \dot{\mathbf{s}})\hat{\mathbf{n}} + v_1(\hat{\mathbf{t}}_1 \cdot \dot{\mathbf{s}})\hat{\mathbf{t}}_1 + v_2(\hat{\mathbf{t}}_2 \cdot \dot{\mathbf{s}})\hat{\mathbf{t}}_2], \quad (2.15)$$

where v_n is the quasi-P wave speed of waves traveling in the $\hat{\mathbf{n}}$ direction, v_1 is the quasi-S wave speed of waves polarized in the $\hat{\mathbf{t}}_1$ direction, and v_2 is the quasi-S wave speed of waves polarized in the $\hat{\mathbf{t}}_2$ direction. These approximations are exact for the waves incident vertically on the boundary and become less reliable with increasing incident angle. Other more efficient treatments of the absorbing boundary conditions are also available, such as the perfect matched layer absorbing boundary condition introduced by [*Bérenger*, 1994; *Komatitsch and Tromp*, 2003]. However, they are more expensive to implement compared to equation (2.15).

2.3.6 Discretization

We design the SEM mesh similarly to the FEM mesh. The model volume Ω is subdivided into a series of hexahedral non-overlapping elements Ω_e , $e = 1, \dots, n_e$, such that $\Omega = \bigcup_{e=1}^{n_e} \Omega_e$. Consequently the absorbing boundary Γ is also subdivided into a series of quadrilateral surface elements Γ_b . We map the boundary elements into the standard 2-D biunit square and define the 2-D shape functions $N_a(\xi, \eta)$, $a = 1, \dots, n_b$ on the square as

the double product of Lagrange polynomials given by

$$N_1(\xi) = \frac{1}{2}(1 + \xi), \quad N_2(\xi) = \frac{1}{2}(1 - \xi) \quad (2.16)$$

for $n_b = 4$ and

$$N_1(\xi) = \frac{1}{2}\xi(\xi - 1), \quad N_2(\xi) = 1 - \xi^2, \quad N_3(\xi) = \frac{1}{2}\xi(\xi + 1) \quad (2.17)$$

for $n_b = 9$. Notice that $-1 \leq \xi \leq 1$. Now the shape of the boundary elements can be expressed as

$$\mathbf{x}(\xi, \eta) = \sum_{a=1}^{n_b} N_a(\xi, \eta) \mathbf{x}_a. \quad (2.18)$$

The Jacobian of this transform is given by

$$\mathbf{J}_b = \left(\frac{\partial \mathbf{x}}{\partial \xi} \times \frac{\partial \mathbf{x}}{\partial \eta} \right), \quad (2.19)$$

and the scalar Jacobian of this transform is just the determinant of the Jacobian matrix

$$J_b = \det[\mathbf{J}_b]. \quad (2.20)$$

The outward normal of the boundary elements can be expressed as

$$\mathbf{n}(\xi, \eta) = \frac{\frac{\partial \mathbf{x}}{\partial \xi} \times \frac{\partial \mathbf{x}}{\partial \eta}}{\left\| \frac{\partial \mathbf{x}}{\partial \xi} \times \frac{\partial \mathbf{x}}{\partial \eta} \right\|}. \quad (2.21)$$

Similarly, we map the volume elements into the standard 3-D biunit cubic square and define the 3-D shape functions $N_a(\xi, \eta, \zeta)$, $a = 1, \dots, n_v$ as the triple product of the Lagrange polynomial functions given by (2.16) for $n_v = 8$ or (2.17) for $n_v = 27$. The shape of the volume elements can be represented by a transform from a biunit cubic to the volume elements

$$\mathbf{x}(\xi, \eta, \zeta) = \sum_{a=1}^{n_v} N_a(\xi, \eta, \zeta) \mathbf{x}_a, \quad (2.22)$$

and the Jacobian matrix and the scalar Jacobian of this transform can be expressed as

$$\mathbf{J}_v = \left[\frac{\partial \mathbf{x}(\xi, \eta, \zeta)}{\partial (\xi, \eta, \zeta)} \right], \quad J_v = \det[\mathbf{J}_v]. \quad (2.23)$$

One critical requirement on the mesh design is that the boundary and volume elements should never degenerate, i.e., $J_b \neq 0$ and $J_v \neq 0$, which ensures a unique and invertible mapping between the elements and standard biunit domains.

2.3.7 Representation of Functions

Although the discretization schemes described in section 2.3.6 are the same as in the finite element methods, the SEM differs from the FEM in that it represents the function field by higher-order Lagrange interpolants, therefore rendering higher accuracy of the simulation compared to the classical FEM. We first choose a set of control points $\xi_\alpha, \alpha = 1, \dots, n_l$ on the $[-1, 1]$ interval to be the n_l Gauss-Lobatto-Legendre (GLL) points that satisfy

$$(1 - \xi^2)\dot{P}_N(\xi) = 0, \quad (2.24)$$

then the high-order Lagrange polynomial functions $l_\alpha(\xi), l_\beta(\eta), l_\gamma(\zeta)$ constructed from these control points can be used to expand any function defined on the boundary element Σ_e or on the volume element Ω_e :

$$f(\mathbf{x}(\xi, \eta))|_{\Sigma_e} = \sum_{\alpha\beta} f^{\alpha\beta} l_\alpha(\xi) l_\beta(\eta) \quad (2.25)$$

$$f(\mathbf{x}(\xi, \eta, \zeta))|_{\Omega_e} = \sum_{\alpha\beta\gamma} f^{\alpha\beta\gamma} l_\alpha(\xi) l_\beta(\eta) l_\gamma(\zeta), \quad (2.26)$$

and the derivatives of the volumetric function field can also be approximated by these Lagrange interpolants:

$$\frac{\partial f(\mathbf{x}(\xi, \eta, \zeta))}{\partial x_i} = \sum_{\alpha\beta\gamma} f^{\alpha\beta\gamma} \left[\dot{l}_\alpha(\xi) l_\beta(\eta) l_\gamma(\zeta) \frac{\partial \xi}{\partial x_i} + l_\alpha(\xi) \dot{l}_\beta(\eta) l_\gamma(\zeta) \frac{\partial \eta}{\partial x_i} + l_\alpha(\xi) l_\beta(\eta) \dot{l}_\gamma(\zeta) \frac{\partial \zeta}{\partial x_i} \right], \quad (2.27)$$

where $\dot{l}_\alpha(\xi)$ denotes the differentiation of function $l_\alpha(\xi)$. Notice the $\frac{\partial \xi}{\partial x_i}$, $\frac{\partial \eta}{\partial x_i}$, $\frac{\partial \zeta}{\partial x_i}$ are just the matrix elements of \mathbf{J}_v^{-1} .

2.3.8 Integration Quadrature

We use the Gauss-Lobatto-Legendre integration rule to approximate integrations over surface elements

$$\int_{\Sigma_e} f(\mathbf{x}) d\mathbf{x} = \sum_{\alpha\beta} \omega_\alpha \omega_\beta f^{\alpha\beta} J_b^{\alpha\beta}, \quad (2.28)$$

where ω_α , $\alpha = 0, \dots, n_l$ denote the weights associated with the GLL points of integration, and $J_b^{\alpha\beta} = J_b(\xi_\alpha, \eta_\beta)$. Similarly integrations over volume element Ω_e can be approximated by

$$\int_{\Omega_e} f(\mathbf{x}) d\mathbf{x} = \sum_{\alpha\beta\gamma} \omega_\alpha \omega_\beta \omega_\gamma f^{\alpha\beta\gamma} J_v^{\alpha\beta\gamma}, \quad (2.29)$$

where $J_v^{\alpha\beta\gamma} = J_v(\xi_\alpha, \eta_\beta, \zeta_\gamma)$.

2.3.9 Global Test Functions

Although section 2.3.7 demonstrated how to represent a function field by the local Lagrangian polynomials at the elemental level, the test function used in the weak formulation (2.5) is defined globally throughout the volume. Therefore we define the I th global scalar test function by its restrictions on the individual elements:

$$w^I(\mathbf{x})|_{\Omega_e} = \begin{cases} l_\alpha(\xi)l_\beta(\eta)l_\gamma(\zeta) & \text{if } I \in \Omega_e, \text{ and } I|_{\Omega_e} = (\alpha, \beta, \gamma) \\ 0 & \text{if } I \notin \Omega_e \end{cases}, \quad (2.30)$$

where I is the global index of a particular grid point, $I = 1, \dots, N_g$. If the I th global grid point resides in an element and is not shared by elements, i.e., it has zero valence, then the global test function $w^I(\mathbf{x})$ is simply the 3-D local Lagrange function $l_\alpha(\xi)l_\beta(\eta)l_\gamma(\zeta)$ extended to the whole space. However, if the grid point is shared by neighboring elements and has valence n_{val} , then the global test function $w^I(\mathbf{x})$ will be composed of $n_{val} + 1$ pieces of local Lagrange function connecting on this shared point at the edge of the elements.

To compute integrations, we establish the mapping between local grid points and the global grid points,

$$(\alpha_i, \beta_j, \gamma_k; \Omega_e) \longrightarrow (I). \quad (2.31)$$

Notice this projection is one-to-one for grid points inside any element ($n_{val} = 0$), and multiple-to-one for the grid points on the element boundaries that are shared by neighboring elements ($n_{val} \geq 1$). In order to assemble the contributions from all the elements for the integrations involving the I th global test function, one needs first to loop over all spectral elements $\Omega_e, e = 1, \dots, n_e$, and then all the GLL points ($n_l \times n_l \times n_l$) to count in all the contributions from local grid points that correspond to this I th global point. Also notice that $W^I(\mathbf{x})$ when evaluated at local grid points of Ω_e will have

$$w^I|_{\Omega_e}(\alpha', \beta', \gamma') = \delta_{\alpha\alpha'} \delta_{\beta\beta'} \delta_{\gamma\gamma'}, \quad (2.32)$$

where $(\alpha', \beta', \gamma')$ is another local grid point of element Ω_e .

2.3.10 Applications to the Weak Form of the Wave Equation

In this section, we will numerically perform the integrations on the left and right hand sides of the weak formulation (2.5) using the discretization and integrations schemes listed in previous sections. We first expand the displacement wave field as

$$\mathbf{s}(\mathbf{x}(\xi, \eta, \zeta), t) = \sum_{j=1}^3 \hat{\mathbf{x}}_j \sum_{\sigma, \tau, \mu=0}^{n_l} s_j^{\sigma\tau\mu}(t) l_\sigma(\xi) l_\tau(\eta) l_\mu(\zeta). \quad (2.33)$$

The goal of our simulation is to solve the discretized displacement field $s_j^{\sigma\tau\mu}(t)$. We choose the global test vector in equation (2.5), corresponding to the k th component of the I th global mesh point, to be the function defined by (2.32):

$$\mathbf{w}(\mathbf{x}(\xi, \eta, \zeta)) = w \hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k \sum_{\substack{\Omega_e \\ (\alpha, \beta, \gamma; \Omega_e) \rightarrow I}} l_\alpha(\xi) l_\beta(\eta) l_\gamma(\zeta). \quad (2.34)$$

2.3.10.1 Mass Matrix

The mass matrix term on the left hand side of equation (2.5) can be numerically integrated as follows:

$$\begin{aligned}
\int_{\Omega} \rho \mathbf{w} \cdot \ddot{\mathbf{s}} \, d\mathbf{x} &= \int_{\Omega} \rho w \ddot{s}_k(t) \, d\mathbf{x} = \sum_p \int_{\Omega_p} \rho w \ddot{s}_k(t) \, d\mathbf{x} \\
&= \sum_p \sum_{\alpha' \beta' \gamma'} \omega_{\alpha'} \omega_{\beta'} \omega_{\gamma'} J^{\alpha' \beta' \gamma'} \rho^{\alpha' \beta' \gamma'} w^{\alpha' \beta' \gamma'} \ddot{s}_k^{\alpha' \beta' \gamma'} \\
&= \left[\sum_{(\alpha, \beta, \gamma; \Omega_e) \rightarrow I} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} J^{\alpha \beta \gamma} \rho^{\alpha \beta \gamma} \right] \ddot{s}_k^I(t), \tag{2.35}
\end{aligned}$$

where the property of the Lagrange polynomials $l_{\alpha}(\xi'_{\alpha}) = \delta_{\alpha\alpha'}$ has been used in the last step. In other words, because the GLL points associated with the Lagrange polynomials are also the GLL quadrature points, the mass matrix that pre-multiplies the acceleration vector $[\ddot{s}_k^I(t)]$ will be diagonal. More discussion about this property will be presented in section 2.3.11.

2.3.10.2 Strain and Stress

The computation of strain and stress is necessary for both the boundary and volume integrations on the right hand side of the weak form (2.5). We first calculate the gradient of the displacement field according to the expansion given by (2.33),

$$\begin{aligned}
\partial_i s_j(\mathbf{x})|_{\Omega_e}^{\alpha\beta\gamma} &= \partial_i \left[\sum_{\alpha' \beta' \gamma'} s_j^{\alpha' \beta' \gamma'} l_{\alpha'}(\xi) l_{\beta'}(\eta) l_{\gamma'}(\zeta) \right] |_{\Omega_e}^{\alpha\beta\gamma} \\
&= \left[\sum_{\sigma} s_j^{\sigma\beta\gamma} \dot{l}_{\sigma}(\xi_{\alpha}) \right] \partial_i \xi(\xi_{\alpha}, \eta_{\beta}, \zeta_{\gamma}) \\
&+ \left[\sum_{\sigma} s_j^{\alpha\sigma\gamma} \dot{l}_{\sigma}(\eta_{\beta}) \right] \partial_i \eta(\xi_{\alpha}, \eta_{\beta}, \zeta_{\gamma}) \\
&+ \left[\sum_{\sigma} s_j^{\alpha\beta\sigma} \dot{l}_{\sigma}(\zeta_{\gamma}) \right] \partial_i \zeta(\xi_{\alpha}, \eta_{\beta}, \zeta_{\gamma}), \tag{2.36}
\end{aligned}$$

then the strain tensor ϵ is given by

$$\epsilon_{ij} = \frac{1}{2}(\partial_i s_j + \partial_j s_i), \quad (2.37)$$

and for the elastic wave propagation problem, the stress tensor can be expressed as

$$T_{kl} = C_{klij}\epsilon_{ij}. \quad (2.38)$$

For the isotropic and no pre-stress case, it can be reduced to

$$T_{kl} = (\kappa - \frac{2}{3}\mu)\delta_{kl}\epsilon_{ii} + 2\mu\epsilon_{kl}. \quad (2.39)$$

And when attenuation is introduced to the wave equation by standard linear solids, equation (2.38) is replaced by (2.11).

2.3.10.3 Stiffness Matrix

Now we compute the volume integrals associated with the stress on the right hand side of equation 2.5.

$$\begin{aligned} \int_{\Omega} \nabla \mathbf{w} \cdot \mathbf{T} \, d\Omega &= \int_{\Omega} \partial_j w T_{jk} \, d\Omega \\ &= \sum_p \int_{\Omega_p} \partial_j w T_{jk} \, d\Omega_e \\ &= \sum_{(\alpha, \beta, \gamma; \Omega_e) \rightarrow I} \int_{\Omega_e} [\dot{l}_\alpha \partial_j \xi l_\beta l_\gamma + l_\alpha \dot{l}_\beta \partial_j \eta l_\gamma + l_\alpha l_\beta \dot{l}_\gamma \partial_j \zeta] T_{jk} \, d\mathbf{x} \\ &= \sum_{(\alpha, \beta, \gamma; \Omega_e) \rightarrow I} \left\{ \left[\sum_{\sigma} \omega_{\sigma} \dot{l}_\alpha(\xi_{\sigma}) \partial_j \xi(\xi_{\sigma}, \eta_{\beta}, \zeta_{\gamma}) T_{jk}^{\sigma\beta\gamma} J^{\sigma\beta\gamma} \right] \omega_{\beta} \omega_{\gamma} \right. \\ &\quad + \left[\sum_{\sigma} \omega_{\sigma} \dot{l}_\beta(\eta_{\sigma}) \partial_j \eta(\xi_{\alpha}, \eta_{\sigma}, \zeta_{\gamma}) T_{jk}^{\alpha\sigma\gamma} J^{\alpha\sigma\gamma} \right] \omega_{\alpha} \omega_{\gamma} \\ &\quad \left. + \left[\sum_{\sigma} \omega_{\sigma} \dot{l}_\gamma(\zeta_{\sigma}) \partial_j \zeta(\xi_{\alpha}, \eta_{\beta}, \zeta_{\sigma}) T_{jk}^{\alpha\beta\sigma} J^{\alpha\beta\sigma} \right] \omega_{\alpha} \omega_{\beta} \right\}. \quad (2.40) \end{aligned}$$

2.3.10.4 Force Terms

The integration associated with the product of the global test function and the force terms can be performed similarly to the previous sections. For a point force $\mathbf{F}(\mathbf{x}, t) = \mathbf{f}(t)\delta(\mathbf{x} - \mathbf{x}_0)$,

$$\begin{aligned} \int_{\Omega} \mathbf{w} \cdot \mathbf{F} \, d\Omega &= w(\mathbf{x}_0) f_k(t) \\ &= \begin{cases} l_{\alpha}(\xi(\mathbf{x}_0)) l_{\beta}(\xi(\mathbf{x}_0)) l_{\gamma}(\xi(\mathbf{x}_0)) f_i(t) & \mathbf{x}_0 \in \Omega_e \\ 0 & \mathbf{x}_0 \notin \Omega_e \end{cases}, \end{aligned} \quad (2.41)$$

where \mathbf{x}_0 , if resides in an element Ω_e , can be mapped to the $(\xi(\mathbf{x}_0), \eta(\mathbf{x}_0), \zeta(\mathbf{x}_0))$ point on the standard biunit domain.

Similarly for moment tensor point source $F_k(\mathbf{x}, t) = -M_{kj} \partial_j \delta(\mathbf{x} - \mathbf{x}_0) g(t)$,

$$\int_{\Omega} \mathbf{w} \cdot \mathbf{F} \, d\Omega = \int_{\Omega} w F_k \, d\Omega = -M_{kj} \partial_j w(\mathbf{x}_0) g(t), \quad (2.42)$$

where we need to further evaluate $\partial_j w$. Realizing that $\partial_j w = \frac{\partial w}{\partial \xi_l} \frac{\partial \xi_l}{\partial x_j}$ and defining $G_{kl}(\xi_0, \eta_0, \zeta_0) = M_{kj} \frac{\partial \xi_l}{\partial x_j}(\mathbf{x}_0)$, where (ξ_0, η_0, ζ_0) are the local coordinates of point \mathbf{x}_0 in the element Ω_p , we obtain

$$\begin{aligned} M_{kj} \partial_j W^I(\mathbf{x}_0) &= G_{kl} \frac{\partial W}{\partial \xi_l}(\xi_0, \eta_0, \zeta_0) \\ &= \sum_{r,t,v} l_r(\xi_0) l_t(\eta_0) l_v(\zeta_0) G_{kl}(\xi_r, \eta_t, \zeta_v) \partial_{\xi_l} (l_{\alpha}(\xi_r) l_{\beta}(\eta_t) l_{\gamma}(\zeta_v)) \\ &= \sum_{r,t,v} l_r(\xi_0) l_t(\eta_0) l_v(\zeta_0) [G_{k1}(\xi_r, \eta_t, \zeta_v) \dot{l}_{\alpha}(\xi_r) l_{\beta}(\eta_t) l_{\gamma}(\zeta_v) \\ &\quad + G_{k2}(\xi_r, \eta_t, \zeta_v) l_{\alpha}(\xi_r) \dot{l}_{\beta}(\eta_t) l_{\gamma}(\zeta_v) \\ &\quad + G_{k3}(\xi_r, \eta_t, \zeta_v) l_{\alpha}(\xi_r) l_{\beta}(\eta_t) \dot{l}_{\gamma}(\zeta_v)]. \end{aligned} \quad (2.43)$$

For a general force field $F_i(\mathbf{x}, t)$, the integration becomes

$$\int_{\Omega} \mathbf{w} \cdot \mathbf{F} \, d\Omega = \int_{\Omega} w F_k \, d\Omega = \sum_p \int_{\Omega_p} w F_k \, d\Omega_e \quad (2.44)$$

$$= \sum_{(\alpha, \beta, \gamma; \Omega_e) \rightarrow I} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} J^{\alpha\beta\gamma} F_k^{\alpha\beta\gamma}(t). \quad (2.45)$$

2.3.10.5 Boundary terms

Now only the absorbing boundary surface integration is left on the right hand side of equation (2.5). Notice that only when the I th global grid point is on a boundary element Γ_e will this integration not vanish, in which case it can be performed as follows,

$$\begin{aligned} \int_{\Gamma} \mathbf{w} \cdot (\hat{\mathbf{n}} \cdot \mathbf{T}) \, d\Gamma &= \int_{\Gamma} w \hat{n}_j T_{jk} \\ &= \sum_p \int_{\Gamma_p} w t_k \, d\Gamma_e \\ &= \sum_{(\alpha, \beta, \gamma; \Gamma_e) \rightarrow I} \omega_{\alpha} \omega_{\beta} J^{\alpha\beta} t_k^{\alpha\beta}. \end{aligned} \quad (2.46)$$

Normal stress t_i is zero for the free surface, and continuous for any internal boundaries. For global problems, the fluid outer core couples with the solid inner core, and we may solve the fluid and solid sides independently, while the interchange of the two fields will occur through this boundary condition (please refer to *Komatitsch and Tromp* [1999] for detailed implementations). As for regional problems, all the artificial boundaries are treated as "absorbing" boundaries, on which the traction is given by equation (2.15).

2.3.11 Assembling and Time Marching

With all the parts of the weak formulation (2.5) resolved, we assemble both sides of the equations for the global test vector $\mathbf{w}(\mathbf{x})$, corresponding to the global grid point I ,

$$\left[\sum_{(\alpha, \beta, \gamma; \Omega_e) \rightarrow I} \omega_{\alpha} \omega_{\beta} \omega_{\gamma} J^{\alpha\beta\gamma} \rho^{\alpha\beta\gamma} \right] \ddot{s}_k^I(t) = \mathcal{T}_k^I + \mathcal{F}_k^I(t) + \mathcal{B}_k^I, \quad (2.47)$$

where the expressions for \mathcal{T}_k^I , $\mathcal{F}(t)_k^I$, and \mathcal{B}_k^I are given by equations (2.40), (2.42) and (2.46).

Now we define the displacement vector $\underline{s}(t) = [s_k^I(t)]$, which is a collection of the displacement as a function of time at all the global grid points, and assemble both sides of equation (2.47) into a matrix form given by

$$\mathcal{M} \ddot{\underline{s}}(t) = \mathcal{RHS}, \quad (2.48)$$

where \mathcal{RHS} denotes the vector version of the right hand side of equation (2.47), $\mathcal{RHS} = \mathcal{T} + \mathcal{F}(t) + \mathcal{B}$. Therefore, in practice, spectral-element simulations involve assembling the mass matrix \mathcal{M} , computing \mathcal{RHS} , and solving the linear system given by (2.48) at every time step. As mentioned earlier, because of the choices of GLL points and the GLL quadrature rules, the mass matrix is exactly diagonal. Therefore this linear system can be solved by simple division:

$$\ddot{\underline{s}}^I(t) = \mathcal{M}_I^{-1} \mathcal{RHS}^I, \quad (2.49)$$

where \mathcal{M}_I is the diagonal element of the mass matrix \mathcal{M} corresponding to the I th global grid point.

Equation (2.49) gives the formula to compute acceleration at any given time. We can then march the wavefield by finite differencing in the time domain, and use the Newmark scheme to update the velocity and displacement at $(n+1)$ th time step, given by the predictor phase at the beginning of each time step:

$$\begin{aligned} d^{n+1} &= d^n + v^n \Delta t + \frac{1}{2} a^n (\Delta t)^2 \\ v^{n+1} &= v^n + \frac{1}{2} a^n \Delta t \\ a^{n+1} &= 0, \end{aligned} \quad (2.50)$$

and the corrector phase at the end of each time step:

$$\begin{aligned}
 a^{n+1} &= \mathcal{M}^{-1}(\mathcal{T} + \mathcal{F}(t) + \mathcal{B}) \\
 v^{n+1} &= v^{n+1} + \frac{1}{2}a^{n+1}\Delta t \\
 d^{n+1} &= d^{n+1}.
 \end{aligned} \tag{2.51}$$

2.3.12 Accuracy Analysis

Two key parameters to evaluate the quality of the mesh design and the accuracy of the simulation are the *number of grid points per wavelength* and the *Courant number*. The number of grid points per wavelength, which determines the resolution of the mesh in terms of how it samples the wavefield, is defined by

$$N = \tau \left(\frac{v}{\Delta h} \right)_{min}, \tag{2.52}$$

where τ is the shortest period present in the source time function, v is the wave speed, and Δh is the spatial between neighboring grids. A general rule of thumb for the SEM method is that a spacing sampling of approximately 5 points per minimum wavelength, when Lagrange polynomials of degree order $n_l = 8$ is used [Komatitsch and Vilotte, 1998], are required for the stability and accuracy of the simulation. The Courant number, which gives the number of grid points advanced in each time step, is defined by:

$$C = \Delta t \left(\frac{v}{\Delta h} \right)_{max}. \tag{2.53}$$

Generally, Δt should be chosen such that $C \leq 0.3$ to maintain a stable simulation. A common practice to evaluate the quality of the mesh design is to plot these two quantities for the mesh, and by examining how these two above requirements are satisfied one can estimate the accuracy and stability of the simulation.

Chapter 3

Spectral-Element Simulation for Earthquakes in Southern California

3.1 Abstract

In this section, we apply the spectral element method outlined in Chapter 2 to a 3-D southern California velocity model. We simulate seismic wave propagation in this integrated model for two recent southern California earthquakes and demonstrate the accuracy of our numerical method and the significant improvement of waveform fits achieved by the introduction of the detailed 3-D model. Most of the results from this chapter are published in *Komatitsch et al.* [2004] and *Lovely et al.* [2006].

3.2 Model Description

We use an integrated crustal velocity model of southern California. We adopt the *Hauksson* [2000] tomographic model as the regional background model and, within this model, we embed a recent Los Angeles basin P-velocity model developed by *Süss and Shaw* [2003], which extends from 119.3°W to 117.3°W longitude and from 33.0°N to 34.8°N latitude. This model further contains a higher-resolution block within the Los Angeles basin, extending from 118.4°W to 117.9°W in longitude and from 33.7°N to 34.1°N in latitude. This model was constructed in GOCAD (a 3D structural modeling tool [*Mallet*, 1992]) and contains a detailed description of the sedimentary basement shape determined from hun-

dreds of petroleum industry well logs and more than twenty-thousand kilometers of seismic reflection profiles. We also include the velocity model in the Salton Trough region developed by *Lovely et al.* [2006]. We scale the P velocity of the model through an empirical relation [*Komatitsch et al.*, 2004] to obtain the shear velocity. The mesh for this integrated model honors the shape of the Moho determined by *Zhu and Kanamori* [2000], and the topography and bathymetry obtained from the USGS [*Komatitsch et al.*, 2004]. Figure 3.1(a) shows the integrated model in map view, and Figure 3.1(b) shows P-velocity variations in two cross-sections, A A' and B B', through the model. Cross-section A A' starts from the Coast Ranges, goes through the Ventura and Los Angeles basins, and ends in the Peninsular Ranges. As seen clearly in Figure 3.1(b), both basins have very slow P velocities (warm colors) due to the thick sediments. Cross-section B B' starts north of the East California Shear Zone, goes through the Mojave Desert, the San Gabriel Mountains, the Los Angeles Basin, and ends at the Continental Borderland. In the corresponding P-velocity cross-section, the Los Angeles basin again appears with distinctly slow P velocities at the surface. Notice the significant variations in surface topography, the difference in elevation between the San Gabriel Mountains and the Continental Borderland being about 3 km. The Moho depth in southern California also varies significantly, from about 20 km underneath the Continental Borderland to roughly 40 km underneath the San Bernardino Mountains [*Zhu and Kanamori*, 2000].

This integrated model involves substantial 3D variations, precluding the use of simple quasi-analytic solutions or perturbation theory. The accurate simulation of seismic wave propagation in this complicated 3D model is crucial to minimize the effects of numerical imperfection on the inversion for both source and structural parameters. Therefore, we use the spectral-element method to simulate wave propagation in this integrated 3D velocity model.

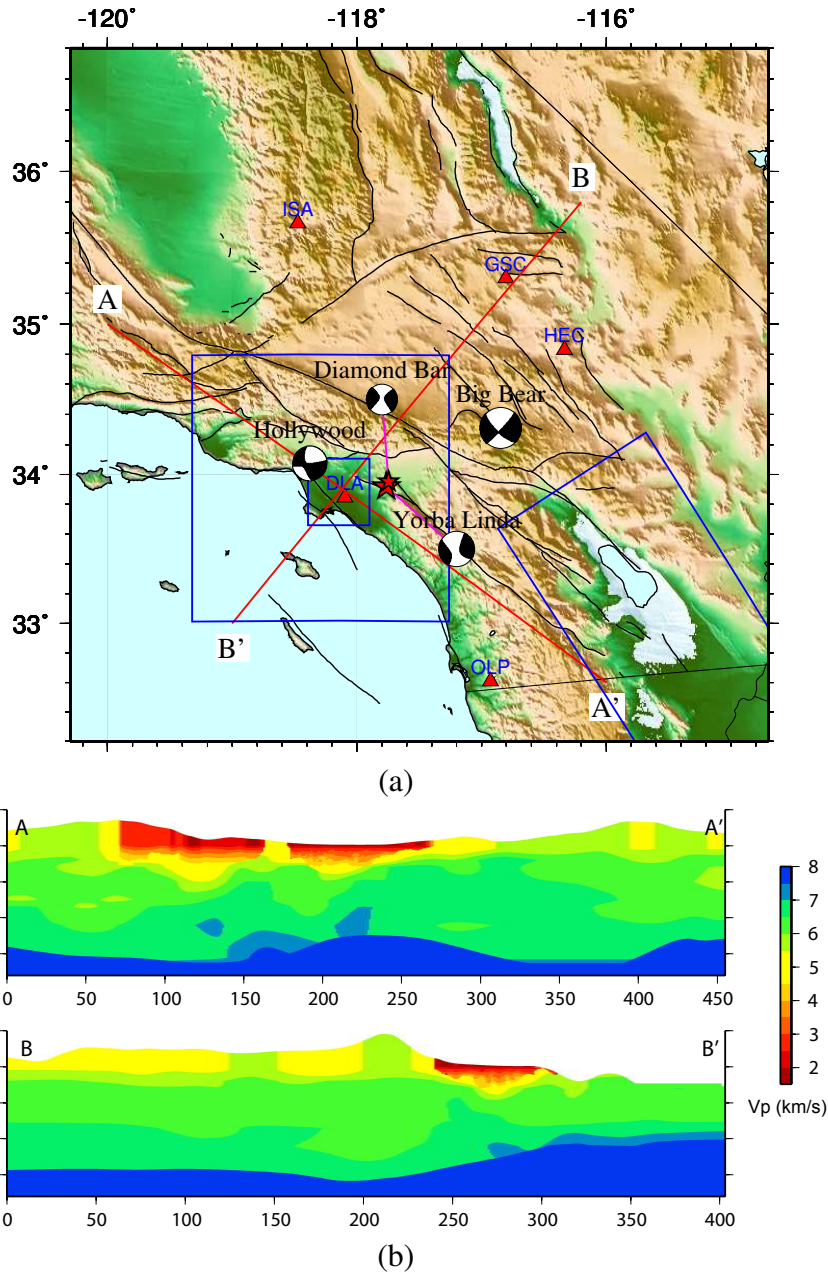


Figure 3.1: (a) Topographic map with shaded relief of southern California showing the extent of our 3D integrated velocity model. The largest blue box indicates the area covered by the medium-resolution GOCAD model developed by *Süss and Shaw* [2003]. The smaller blue box indicates the high-resolution part of the GOCAD model centered on the Los Angeles basin. The tilted blue box in the Salton Trough region indicates the range of a detailed GOCAD model provided by *Lovely et al.* [2006]. The epicentral locations and source mechanisms of the four earthquakes discussed in this chapter and the next chapter are denoted by beachballs in the plot. The major late Quaternary faults [*Jennings, 1975*] are also displayed. (b) Compressional-wave speed variations along two cross-sections A A' and B B' as described in section 3.2. Notice the significant 3D wave-speed variations, and in particular the very low wave-speed sedimentary basins in cross-section A A'. Also notice the shallow Moho under the continental borderland and the substantial topography and bathymetry (exaggerated five times) variations in the cross-section B B'.

3.3 Numerical Implementations

3.3.1 Mesh Design for Realistic 3-D Velocity Models

3.3.1.1 Adaption to Velocity Discontinuities

We compile the surface shape files for topography, basin basement, and the Moho depth, and interpolate evenly-spaced horizontal surfaces in between these given surfaces according to their absolute depth. Figure 3.2(a) shows two slices of the topography, basin basement, and Moho surfaces in the Los Angeles basin region (for the definition of slices, please refer to section 3.3.2). Then the mesh is designed to adapt to the given three surfaces as well as the horizontal surfaces interpolated in between them.

3.3.1.2 Doubling Scheme

For realistic velocity models, the seismic velocity increases with depth in most cases; therefore, for a mesh designed to have relatively homogeneous grid spacing throughout the model, the number of grid points per wavelength will inevitably increase with depth, meaning the top of the model mesh will have fewer grid points per wavelength compared to the bottom. Hence the accuracy of the simulation is determined by the N number (equation (2.52)) at the top of the mesh. However, since an excessive number of grid points is available for the mesh at depth, much more computation cost is needed to propagate waves through these regions of the model without much gain in the accuracy of the simulation. The ideal situation to achieve computational efficiency is to have an almost uniform number of grid points per wavelength throughout the entire model, which means that the grid spacing will generally have to increase with depth. One simple way to achieve this is to use the mesh doubling scheme. Figure 3.2(b) shows how the doubling is achieved in two layers for a 3-D Cartesian mesh, with doubling in every horizontal direction achieved in one layer.

3.3.2 Parallel Computing

Because of the substantial computational power required by the spectral-element simulations for realistic 3-D seismic velocity models, we implement the spectral-element method in parallel computers. The model region is divided into $N_\xi \times N_\eta$ number of slices from the surface, and each slice is assigned to one computation unit, and the communication between slices is accomplished by the message passing interface (MPI).

3.4 Yorba Linda Earthquake

As an example, we simulate seismic wave propagation in the 3-D integrated southern California model for the Sep. 3, 2002, $M_w = 4.2$ Yorba Linda earthquake. This event occurred at a depth of 7 km, and was well-recorded by the Southern California Seismic Network (SCSN). The source mechanism we used is inverted using the spectral-element moment-tensor inversion method outlined in Chapter 4. Figure 3.3(a) shows the transverse displacement waveform fits between data and synthetics for the 3-D velocity model, in comparison with figure 3.3(b) showing the waveform fits between data and synthetics generated for a standard 1-D southern California velocity model [Hadley and Kanamori, 1977; Dreger and Helmberger, 1990] for the zoomed-in Los Angeles region. Both data and synthetics have been filtered between 6 seconds and 35 seconds. For the 3-D synthetics, time shifts have been applied to align them with the corresponding data to take into account the inadequacy of the 3-D velocity model. Notice that strong basin resonance is clearly seen in the 3-D synthetics for most of the basin stations, which is not present for the 1-D synthetics generated for the 1-D standard velocity model mostly compiled with data from rock sites. Significant improvement of the waveform fits has been achieved for basin stations such as PDR, KAF, STS, LGB, BRE, and LLS, where the observed amplitude can be up to 20 times larger than the 1D predictions. This suggests that the 3-D velocity model and the spectral-element simulation provide very accurate 3-D Green's functions for these basin stations, which helps to facilitate the inversion of source mechanisms in Chapter 4. Figure 3.4 shows the overall very nice waveform fits throughout the SCSN stations for this event for the stations out-

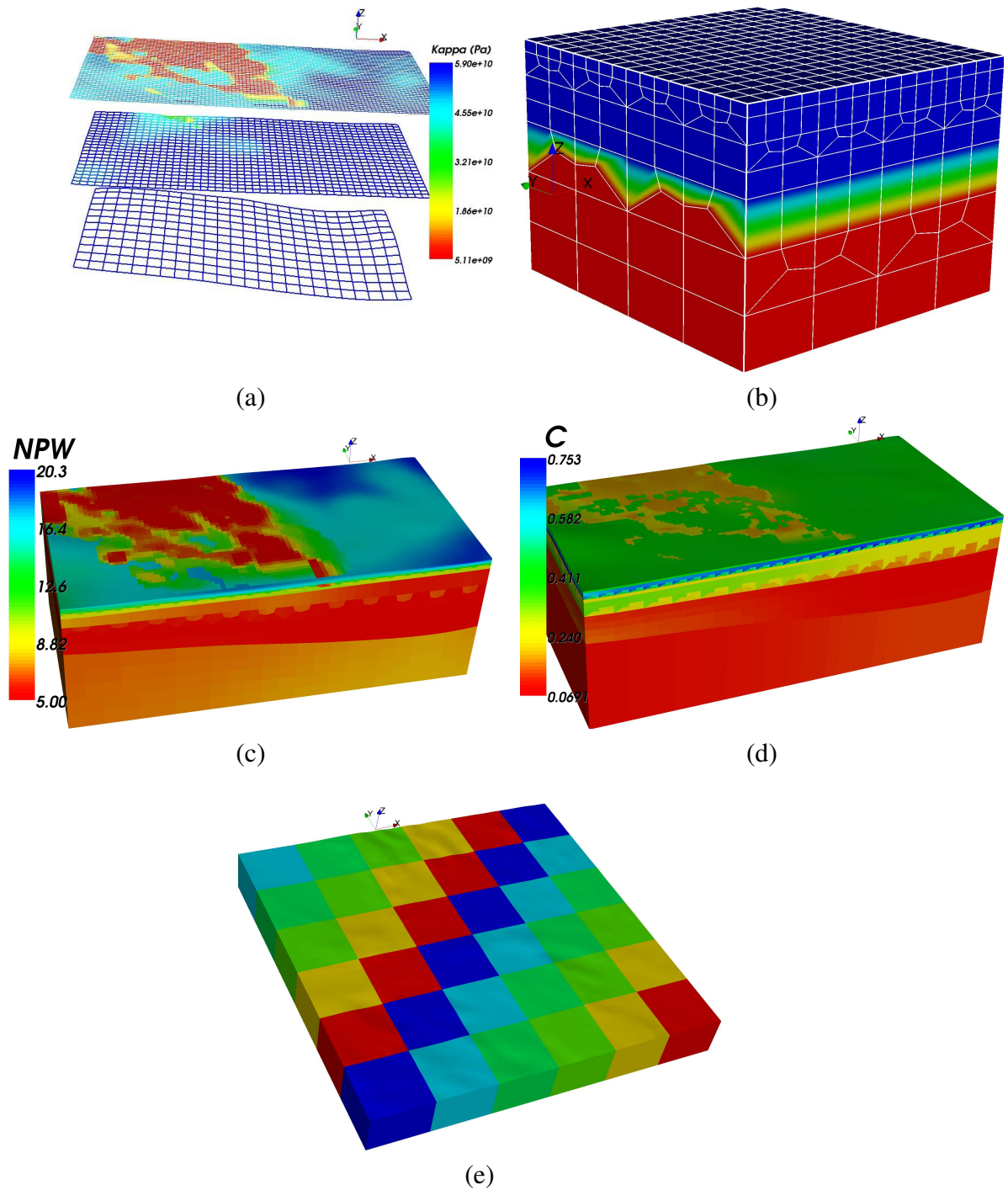


Figure 3.2: (a) Two slices of the 3D spectral-element mesh designed for the southern California region going through the Los Angeles basin. The color of the mesh shows the values of the bulk modulus at the grid points. (b) A slice of the spectral-element mesh showing the doubling scheme. The size of the elements is doubled in the x direction in one layer and then doubled in the y direction in the next layer. (c) The number of points per wavelength (N) for two slices of the mesh. (d) The Courant number (C) describing the number of grid points advanced in one time step for two slices of the mesh. (e) We divide our model into $N_\xi \times N_\eta$ (6×6) slices on the surface. The simulation for each slice is run on one processor, and communications between processors are accomplished by the Message Passing Interface (MPI) on a Beowulf cluster.

side the Los Angeles basin. Another noticeable feature is that the simulation also captures the strong resonance visible in some of the stations in the Imperial Valley such as THX, SAL, ERR, and BTC, which is due to the high resolution of the sedimentary basin model (constructed by *Lovely et al.* [2006]) present in our integrated 3-D model. This area will be further studied in the next section.

3.5 Hypothetical Major Rupture on the Southern San Andreas Fault

The implementations of detailed structural geology models of the basin velocity structure for both the Los Angeles basin and the Salton Trough have made it possible to assess realistic seismic hazard generated by scenarios for large plausible earthquake on the San Andreas fault. For demonstration purposes, we simulate a $M_w = 7.9$ earthquake along the southern San Andreas fault by mapping the rupture history and slip distribution of the 2002, $M_w = 7.9$ Denali earthquake [*Tsuboi et al.*, 2003] onto the San Andreas fault. The magnitude of this hypothetical event is comparable to the 1857 $M_w = 7.9$ Fort Tejon earthquake which ruptured an adjacent northern segment of the fault. The rupture initiates in the north-west and propagates towards the southeast, which focuses more energy into the Salton Trough area because of the rupture directivity. Figure 3.5 shows a series of snapshots of the velocity field generated by this scenario earthquake as simulated by our spectral-element method. Apparently, most of the energy rolls along the rupture segment of the San Andreas fault as indicated by the red line, while expanding outward at the same time. Strong directivity is indicated by the large amplitude of the motion at the wavefront towards the rupture direction, while very little motion at the wavefront opposite to the rupture direction. Also, a strong basin amplification phenomenon is observed for both the Los Angeles basin and the Salton Trough, where strong resonance of the wave field persists over a long time, even after the wave front has passed these regions. The shakemaps (Figure 3.6) illustrates the peak ground velocity and acceleration generated by this scenario earthquake. The motion in the Salton Trough region is significantly amplified compared

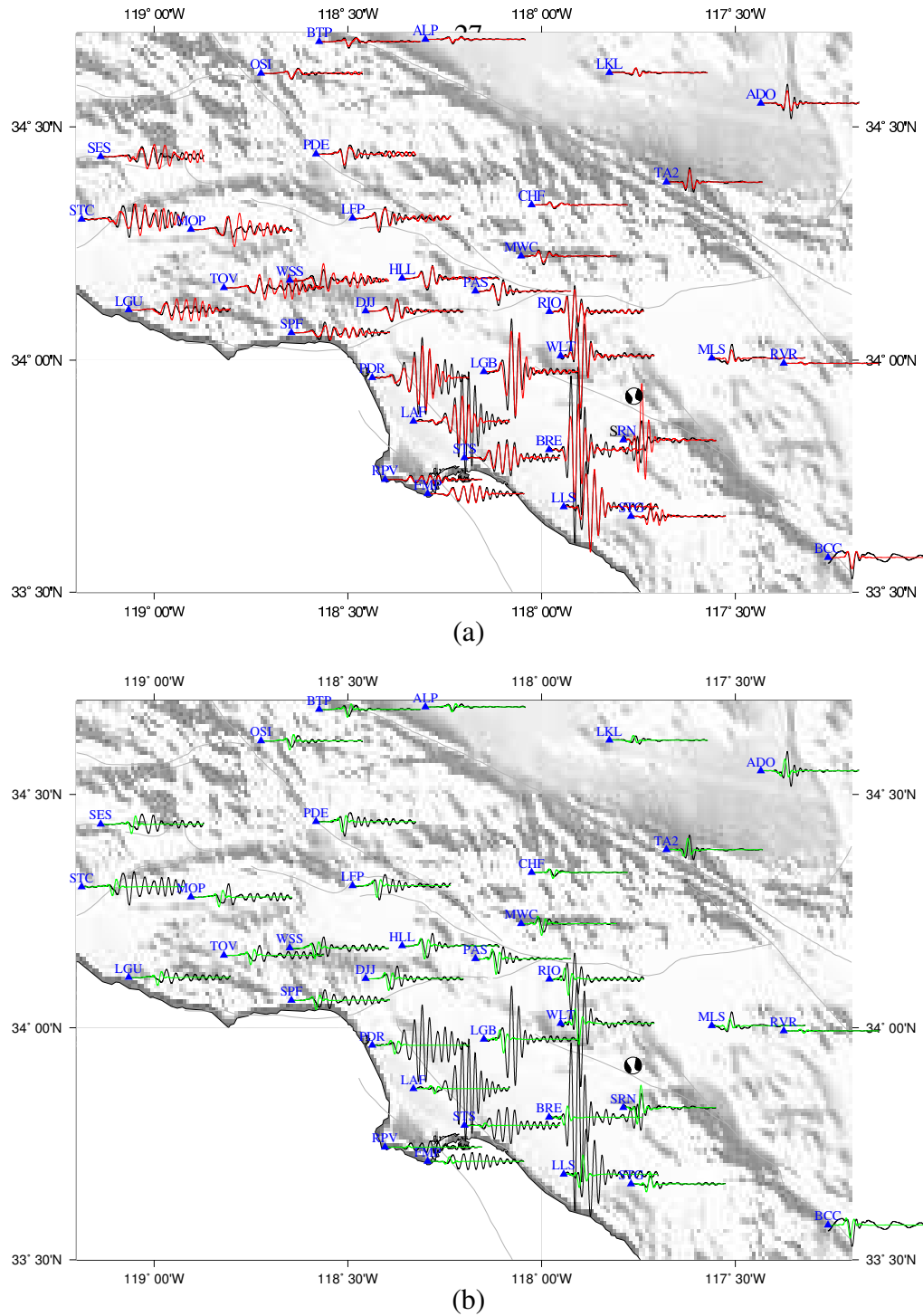


Figure 3.3: (a) Waveform fits of displacement seismograms between the data (in black) and the 3-D synthetics (in red) computed for the 3-D integrated southern California model by the spectral-element method for the Yorba Linda earthquake at some of the stations in the vicinity of the Los Angeles basin. The 3-D synthetics have been shifted to achieve best correlation with the data. The event location and source mechanism is denoted by the small beachball at the edge of the Los Angeles basin. (b) Waveform fits of displacement seismograms between the data (in black) and the 1-D synthetics (in green) computed for the 1-D standard southern California model.

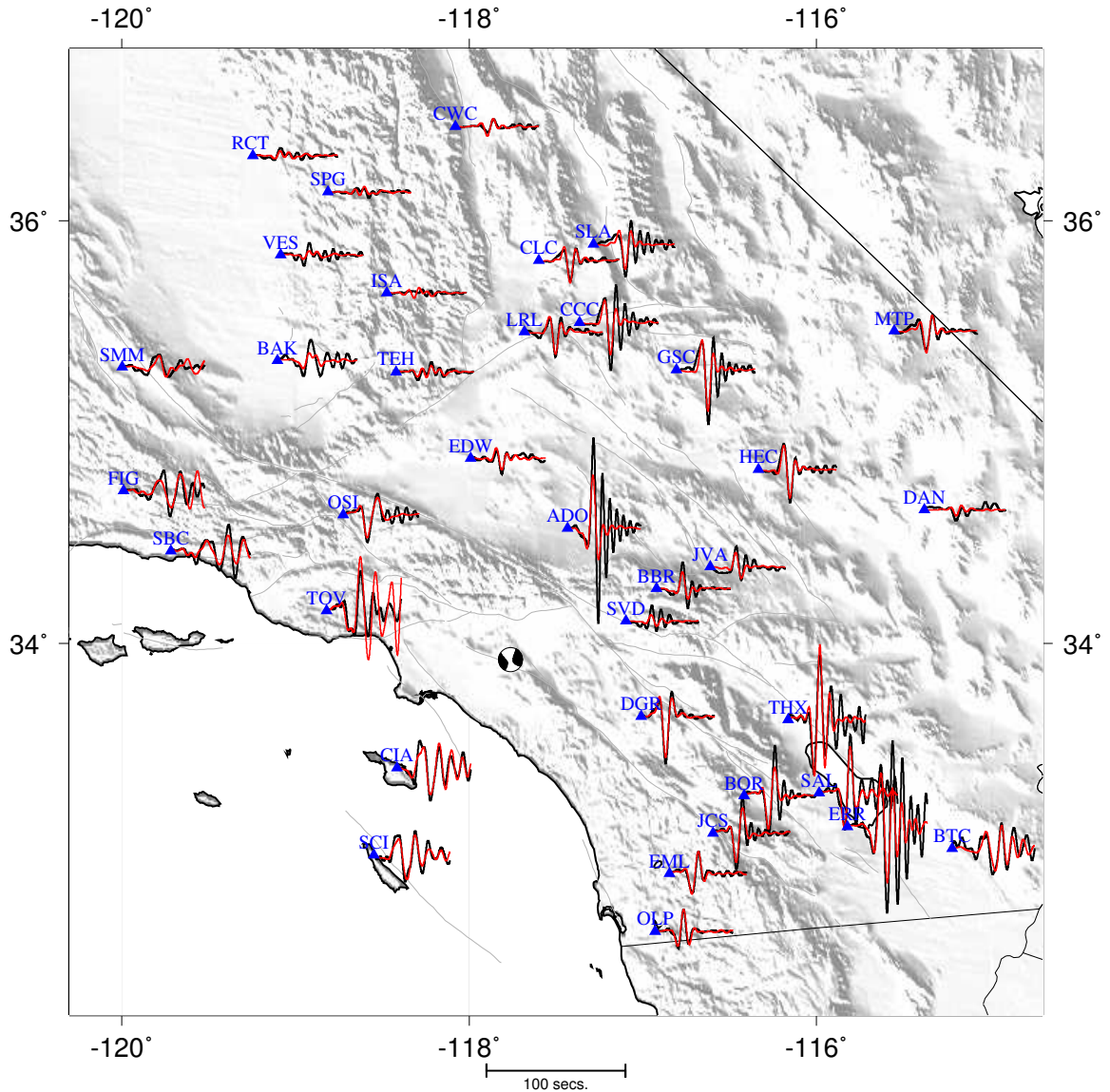


Figure 3.4: The waveform fits of displacement seismograms between the data and the 3-D synthetics calculated for some representative stations throughout southern California for the Yorba Linda earthquake (data in black and synthetics in red). Both the data and the synthetics have been filtered between 6 seconds and 35 seconds. The event location and mechanism is denoted by the beachball at the edge of the Los Angeles basin.

to the surrounding bedrock regions, caused by both the amplification effect of the Salton Trough basin which is at about 6 km depth at the deepest point [Lovely *et al.*, 2006], and also the directivity effect that focuses the energy from the northwest into this sedimentary basin. For cases like this, the spectral-element method, combined with the detailed velocity models, can provide more accurate assessment of the potential damage from possible catastrophic events.

3.6 Discussion

We simulate wave propagation for both the Yorba Linda earthquake and a hypothetical earthquake through an integrated 3-D velocity model. Comparisons of the 3-D synthetics generated by these simulations with the data show generally great improvements of waveform fits, while the synthetics generated from a 1-D model developed to fit waveforms at rock sites show poor waveform fits for most of the basin stations, both in the Los Angeles basin and the Salton Trough region. Although the accuracy of our simulation up to 1 second can be easily achieved with the parallel implementation of the spectral-element method, the highest frequency at which we can achieve decent fits to the observed waveforms is limited by the accuracy of the 3-D velocity model itself. Apparently, by introducing detailed 3-D sedimentary basin models for both the Los Angeles basin and the Salton Trough area, we have predicted much better the amplification and resonance effects at the stations in these regions, making it possible to perform realistic seismic hazard assessments. On the other hand, these improved 3-D Green's functions can also be used in the source mechanism inversions for earthquakes throughout southern California, especially small magnitude events in the basins.

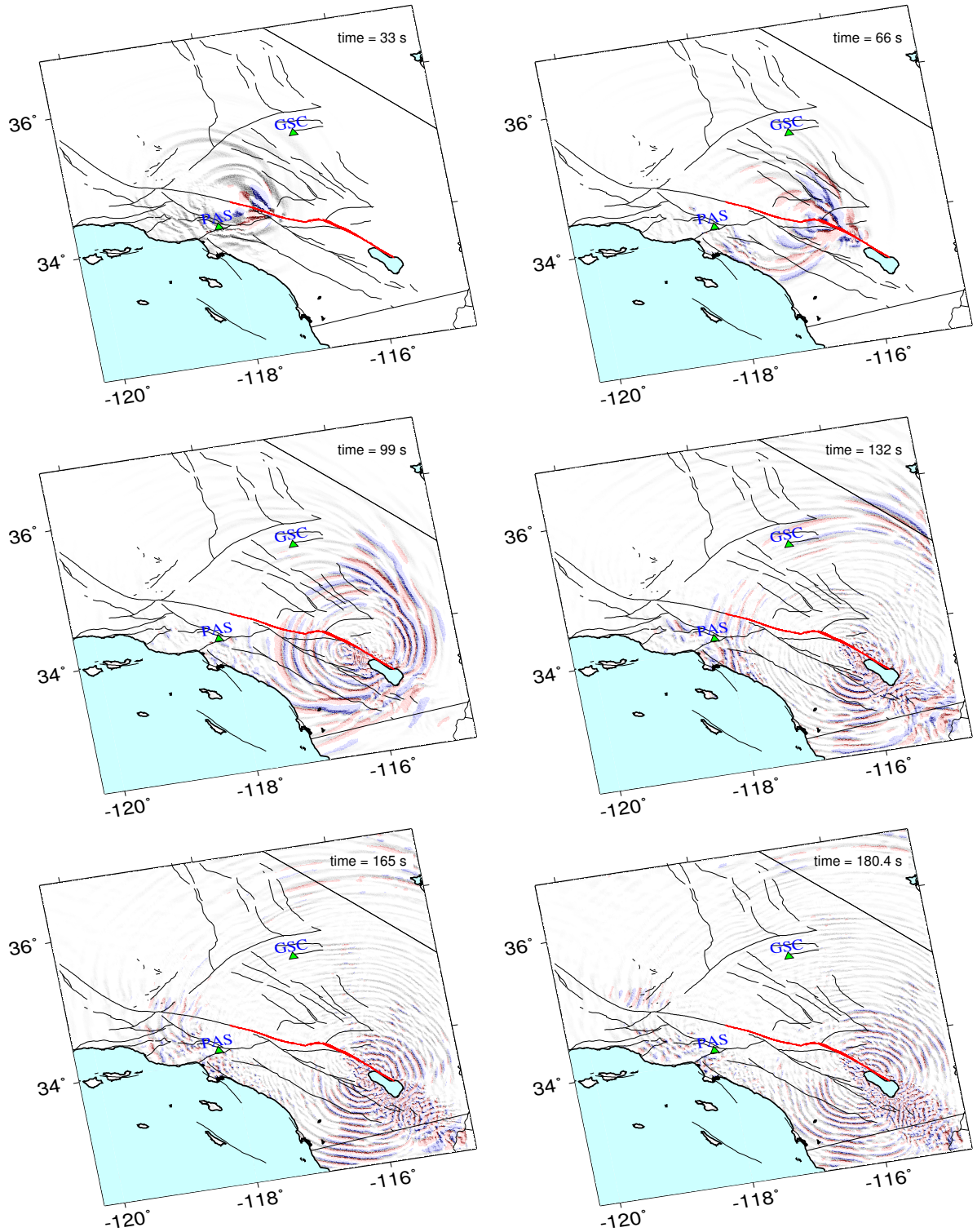


Figure 3.5: Snapshots of the hypothetical earthquake rupture from northwest toward southeast on the southern end of the San Andreas fault. All frames are plotted on the same color scale. Notice the large amplification that occurs in both the Salton Trough region and the Los Angeles basin.

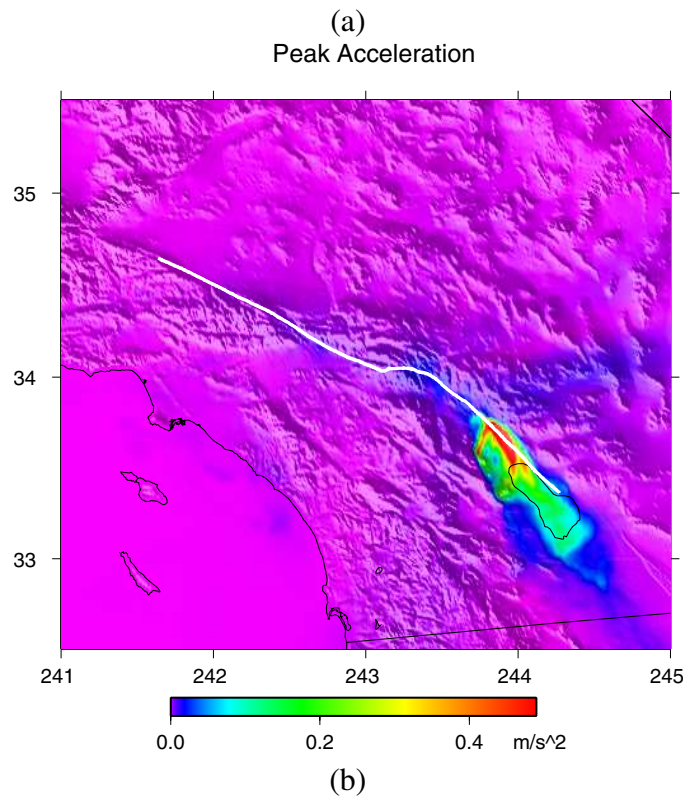
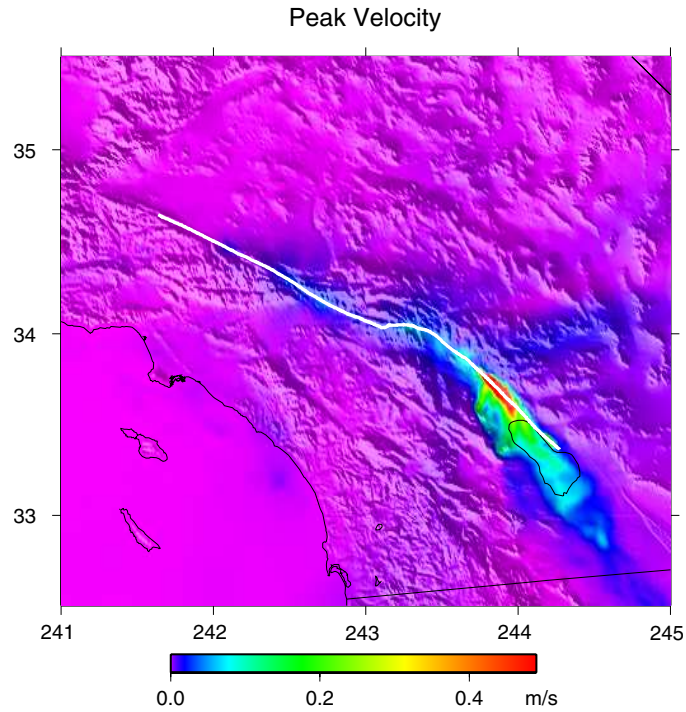


Figure 3.6: (a) Peak velocity shakemap generated for this hypothetical rupture on the San Andreas fault from the northeast to the southwest. The ruptured section of the San Andreas fault is denoted by the white line. (b) Peak acceleration shakemap for this hypothetical event.

Chapter 4

Spectral-Element Moment-Tensor Inversions for Earthquakes in Southern California

4.1 Abstract

We have developed and implemented an automated moment-tensor inversion procedure to determine source parameters for southern California earthquakes. The method is based upon spectral-element simulations of regional seismic wave propagation in an integrated three-dimensional (3-D) southern California velocity model. Sensitivity to source parameters is determined by numerically calculating the Fréchet derivatives required for the moment-tensor inversion. We minimize a waveform misfit function, and allow limited time shifts between data and corresponding synthetics to accommodate additional 3-D heterogeneity not included in our model. The technique is applied to three recent southern California earthquakes: the September 9, 2001, $M_l = 4.2$ Hollywood event, the February 22, 2003, $M_l = 5.4$ Big Bear event, and the December 14, 2001, $M_l = 4.0$ Diamond Bar event. Using about half of the available three-component data at periods of 6 seconds and longer, we obtain focal mechanisms, depths, and moment magnitudes that are generally in good agreement with estimates based upon traditional body-wave and surface-wave inversions. Most of the results from this chapter are published in *Liu et al.* [2004].

4.2 Introduction

Southern California is characterized by geologic and tectonic complexity. Every year about 40 earthquakes with local magnitude greater than 3.5 occur on the fault systems in the region, and the determination of the source parameters of these earthquakes is important for understanding the associated tectonic processes. Southern California Seismic Network (SCSN) first-motion data can be used to determine focal mechanisms of earthquakes [Hauksson *et al.*, 2002]. Regional long-period (10 – 50 seconds) surface waves have been used to invert for source parameters by correcting for the effects of 3-D heterogeneity with a regional phase-velocity map [Thio and Kanamori, 1995]. However, in general this method is only suitable for earthquakes with $M_w \geq 3.7$, because the signal-to-noise ratio becomes too low for smaller earthquakes at such long periods. Regional broadband waveform data have also been used to determine source parameters. The early part of the P wave — i.e., P_{nl} , the combination of the P_n and PL phases [Helmberger and Engen, 1980] — on the vertical and radial components of the seismograms is relatively insensitive to crustal heterogeneity. For this reason, this phase has been used to determine the strike, dip, and rake of moderate-size earthquakes [Wallace and Helmberger, 1981]. Grid searching over strike, dip, rake, and depth was introduced to deal with the non-linearity associated with these source parameters for southern California earthquakes [Dreger and Helmberger, 1991] based upon the one-dimensional standard southern California velocity model [Hadley and Kanamori, 1977; Dreger and Helmberger, 1990]. In order to utilize the whole waveform and accommodate the imperfect 1-D Green's function, the vertical and radial components of the seismograms at local and regional distances are cut into P_{nl} and surface-wave segments. By allowing for time shifts between these two segments, source mechanisms can be resolved by fitting them simultaneously [Zhao and Helmberger, 1994; Zhu and Helmberger, 1996].

Many more high-quality stations have been deployed by the SCSN since the early '90s. A large portion of these is located in or near the Los Angeles basin, where complicated 3-D velocity structure produces complicated waveforms. Simple time shifts do not adequately accommodate the imperfect 1-D Green's function in the source inversions, and 3-D Green's

functions are required for a significant portion of the SCSN stations. For this purpose, this paper introduces an adaptation of the centroid-moment tensor (CMT) formalism [Gilbert and Dziewonski, 1975; Dziewonski *et al.*, 1981; Ritsema and Lay, 1993] to determine the source mechanisms of small- to moderate-size earthquakes in southern California. The calculation of synthetic seismograms and Fréchet derivatives (i.e., derivatives of the synthetics with respect to source parameters) is accomplished based upon the spectral-element method (SEM) [e.g., Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999] and recently developed 3-D southern California crustal velocity models [Hauksson, 2000; Süss and Shaw, 2003] as described in Chapter 3.

4.3 Moment-Tensor Inversion Theory

Suppose we have N observed time series $\{d_i(t); i = 1, \dots, N\}$ and N corresponding synthetic seismograms $\{s_i(t, \mathbf{m}); i = 1, \dots, N\}$ to constrain a point source model $\mathbf{m} = \{m_j; j = 1, \dots, n\}$. In the case of an inversion for the six elements of the moment tensor $\mathbf{M} = \{M_{rr}, M_{\theta\theta}, M_{\phi\phi}, M_{r\theta}, M_{r\phi}, M_{\theta\phi}\}$, we have $n = 6$ and $\mathbf{m} = \{\mathbf{M}\}$. An inversion for the moment tensor \mathbf{M} and event depth d_s involves $n = 7$ and $\mathbf{m} = \{\mathbf{M}, d_s\}$. An inversion for the moment tensor \mathbf{M} and the event location \mathbf{r}_s results in $n = 9$ and $\mathbf{m} = \{\mathbf{M}, \mathbf{r}_s\}$. Although theoretically and numerically feasible, the period range used in our simulations does not enable us to invert for the source half-duration of small- to moderate-size earthquakes.

Given a set of source model parameters \mathbf{m} , the misfit between the data and the synthetics may be assessed based upon the least-square waveform misfit function

$$E_1(\mathbf{m}) = \frac{1}{2A_1} \sum_{i=1}^N w_i \int [d_i(t) - s_i(t, \mathbf{m})]^2 dt. \quad (4.1)$$

A_1 is a normalization factor, which can take the value of $\frac{1}{2} \sum_{i=1}^N w_i \int d_i(t)^2 dt$ to ensure that $E_1(\mathbf{m})$ generally takes values between zero and one. Weights $w_i = w_i^a w_i^d w_i^c$, $i = 1, \dots, N$, may be assigned based upon azimuth-, w_i^a , epicentral distance-, w_i^d , and component-related criteria, w_i^c . The disadvantage of the waveform misfit function (4.1) is that it is susceptible to cycle skips (i.e., the synthetic is advanced or delayed by one wave cycle or more relative

to the data).

An alternative misfit function may be defined based upon the envelope of the waveforms. For example, the data envelope function is defined by

$$e(d_i(t)) = \left[d_i^2(t) + \hat{d}_i^2(t) \right]^{1/2}, \quad (4.2)$$

where $\hat{d}_i(t)$ is the Hilbert transform of $d_i(t)$. The corresponding envelope misfit function is

$$E_2(\mathbf{m}) = \frac{1}{2A_2} \sum_{i=1}^N w_i \int [e(d_i(t)) - e(s_i(t, \mathbf{m}))]^2 dt, \quad (4.3)$$

where $A_2 = \frac{1}{2} \sum_{i=1}^N w_i \int e(d_i(t))^2 dt$, and $E_2(\mathbf{m})$ usually takes values between zero and one. The disadvantage of the envelope misfit function (4.3) is that it does not provide any phase information, but when it is combined with the waveform misfit function (4.1), cycle skips may be avoided.

It is often desirable to subject the inversion to certain constraints. For example, if we require the moment tensor to have zero trace, i.e., the source involves no change in volume, then the inversion needs to be subject to the constraint $C_1(\mathbf{m}) = m_1 + m_2 + m_3 = 0$. If we further require that the earthquake is representable by a double-couple source mechanism, the moment tensor is subject to the additional constraint $C_2(\mathbf{m}) = \det(\mathbf{M}) = 0$.

Combining all these ingredients, we minimize the following misfit function:

$$E(\mathbf{m}, \mu_1, \mu_2) = \lambda E_1(\mathbf{m}) + (1 - \lambda) E_2(\mathbf{m}) + \mu_1 C_1(\mathbf{m}) + \mu_2 C_2(\mathbf{m}), \quad (4.4)$$

where λ and $1 - \lambda$ denote the relative weights of the waveform misfit function $E_1(\mathbf{m})$ and the envelope misfit function $E_2(\mathbf{m})$, and μ_1 and μ_2 are the Lagrange multipliers associated with the constraints $C_1(\mathbf{m})$ and $C_2(\mathbf{m})$. The optimal source parameters \mathbf{m}^{new} may be determined by solving the set of n equations

$$\frac{\partial E}{\partial m_j}(\mathbf{m}, \mu_1, \mu_2) = 0, \quad j = 1, \dots, n, \quad (4.5)$$

subject to the constraints $C_1(\mathbf{m}) = 0$ when μ_1 is non-zero and $C_2(\mathbf{m}) = 0$ when μ_2 is

non-zero. Under certain assumptions, equation (4.5) can always be reduced to a linear system. For the detailed formulation of both linear and non-linear inversions, please refer to Appendix B.

4.4 Time Shifts

The theory presented in the previous section is based on the assumption that we have perfect knowledge of the velocity structure in the region. Therefore, the differences between data and synthetics for given source parameters are completely due to errors in these parameters. Obviously this assumption does not hold in general. Source and structural effects are always entangled, and the difference between data and synthetics is due to both errors in the source mechanism and the inadequate 3-D model. 3-D velocity variations cause advances or delays in the arrival times of seismic phases as well as changes in the associated waveforms. Introducing the envelope misfit function (4.3) may help prevent cycle skip problems caused by relatively modest traveltimes anomalies; but for large velocity variations, such as those in southern California, the envelope function may not be sufficiently effective.

Komatitsch et al. [2004] illustrate that differences between the true 3-D velocity structure and our 3-D model mainly affect the arrival times of seismic phases but do not significantly alter the associated waveforms at periods of 6 seconds and longer. Therefore, we can shift the synthetics with respect to the data to obtain the optimal correlation between the two, and then attribute differences in waveforms between the data and synthetics to the source parameters of the earthquake. This approach was demonstrated to be quite useful in accommodating effects due to 3-D heterogeneity for the purpose of source mechanism inversions by *Zhao and Helmberger* [1994]. However, it may not be applicable to certain stations if the velocity variations between the source and the stations are not well described by our 3-D velocity model, e.g., when waveforms are severely distorted. Therefore, careful selection of stations is required to select paths for which our 3-D velocity model produces decent waveforms compared to the data. Of course the whole procedure requires an initial source model \mathbf{m}^0 for which we calculate the synthetics for our 3-D model. By correlating

the data $d_i(t)$ and synthetics $s_i(t, \mathbf{m}^0)$ we obtain the time lag δt_i between the two that gives:

$$\max_{\delta t_i} \frac{\int d_i(t) s_i(t - \delta t_i, \mathbf{m}^0) dt}{\left[\int d_i(t)^2 dt \int s_i(t - \delta t_i, \mathbf{m}^0)^2 dt \right]^{1/2}}. \quad (4.6)$$

We then apply the same time lag to all the Fréchet derivatives, and invert for source model parameters by minimizing a waveform misfit function with time lags:

$$E_s(\mathbf{m}) = \frac{1}{2A_1} \sum_{i=1}^N w_i \int \left[d_i(t) - s_i(t - \delta t_i, \mathbf{m}^0) - \sum_j \frac{\partial s_i}{\partial m_j}(t - \delta t_i, \mathbf{m}^0)(m_j - m_j^0) \right]^2 dt. \quad (4.7)$$

This will be the preferred misfit function in our inversion procedure. Another advantage of using equation (4.7) is that it is not affected by small timing errors associated with either the instrument or the origin of the event. We will evaluate how well our synthetics for the new source parameters \mathbf{m}^{new} fit the data by calculating the variance reduction

$$\sigma_i = \frac{\int [d_i(t) - s_i(t - \delta t_i, \mathbf{m}^{\text{new}})]^2 dt}{\int d_i(t)^2 dt}. \quad (4.8)$$

Alternatively, we may determine how much the misfit function (4.7) has been reduced after the inversion by calculating the misfit reduction:

$$\chi = [E_s(\mathbf{m}^0) - E_s(\mathbf{m}^{\text{new}})] / E_s(\mathbf{m}^0). \quad (4.9)$$

Fortunately, for most earthquakes with $M_w \geq 3.5$ we can use source-mechanism estimations from other methods as initial solutions. However, there are cases in which we do not have other solutions to start with, meaning we do not have any reference upon which to either select the data or obtain time lags to shift synthetics and Fréchet derivatives. Nevertheless, we can always determine our own initial source mechanism based upon a preliminary moment-tensor-only linear inversion in which we select data traces solely based upon their signal-to-noise ratio. As we will show in later examples, this inversion procedure proves to be quite robust for determining the source mechanisms of small- and moderate-size earthquakes in southern California, and can be easily implemented automatically whenever an

earthquake occurs.

4.5 Numerical Tests

In order to solve (4.5), we need to calculate the derivatives of the synthetics with respect to the source parameters to obtain the linear system (B.5) or (B.17). In our case, this implies numerically differentiating synthetics with respect to the source parameters. Because the synthetics $s_j(t, \mathbf{m})$ are linear combinations of the moment-tensor elements, the derivatives $\frac{\partial s_j}{\partial m_i}(t), i = 1, \dots, 6$, can be easily obtained by running forward calculations for moment tensors that have a non-zero i th element while all the other elements are zero. However, $s_j(t, \mathbf{m})$ is a non-linear function of depth, latitude, and longitude, $\{m_i, i = 7, 8, 9\}$, therefore, in order to approximate the derivatives for these three parameters with finite differences, appropriate intervals for the finite-difference length should be chosen such that the derivatives of the misfit function remain relatively constant within these intervals. Furthermore, in order to understand the functional dependence of the misfit function on depth, latitude, and longitude for (4.1) as well as (4.3) and (4.7), we perform several synthetic tests for the September 9, 2001, $M_w = 4.2$ Hollywood event (see Figure 3.1 for its location and mechanism). We adopt the event location and mechanism obtained by the *Thio and Kanamori* [1995] method (34.0745°N latitude, 118.3792°W longitude, and 5.4 km depth; please refer to Table 4.1 for further details), calculate synthetics at all available SCSN stations inside our 3-D model, and regard these as "synthetic data". We then run simulations for a series of "test" source models, generate synthetics at SCSN stations, and evaluate the misfit function between the "synthetic data" and the synthetics for each test source model. The misfit values have all been normalized by the amplitude of the data such that a misfit value of 1 means that all the synthetics fit the data on average as poorly as a straight line would fit the data, while a misfit value of 0 means that all the synthetics fit the data perfectly. For all the synthetic tests and inversions presented in this paper, both data and synthetics have been bandpass filtered between 6 seconds and 40 seconds.

4.5.1 Synthetic Test: Latitude

In order to examine how the waveform misfit function without time shifts (4.1) and the waveform misfit function with shifts (4.7) vary as a function of latitude, we fix all remaining 8 source parameters, varying only the latitude from 34.06°N to 34.08°N by 0.0025° increments, and calculate the synthetics for each case. We then evaluate the misfit functions (4.1) and (4.7). In Figure 4.1(a), the dashed and the solid lines, respectively, show how the waveform misfit function with and without time shifts varies with latitude. The functions have both symmetry and convexity centered on the "true" latitude 34.0745°N . Notice that the waveform misfit function with time shifts (dashed line) has values that are about one-third of those for the waveform misfit function without time shifts (solid line). The difference becomes larger when the latitude moves farther away from the "true" latitude. This suggests that the difference in waveform measured by (4.7) only takes up a small portion of the difference between the data and synthetics measured by (4.1) in the case of an incorrect latitude, and thus the arrival time difference should be responsible for a large portion of the waveform misfit without time shifts. Therefore, the sensitivity of the misfit function (4.7) to latitudinal variations will be smaller compared to (4.1). All the above statements hold true for longitudinal variations as well. Notice that within a latitudinal interval of 0.0025° , the derivative of the misfit function remains about constant, therefore we are confident that this should give a good approximation to the latitudinal derivatives at the initial solution.

4.5.2 Synthetic Test: Depth

We vary the depth from 4 km to 9.2 km by 0.4 km increments, and fix all other 8 parameters to study the functional dependence of the misfit functions (4.1) and (4.7) on depth. As one can see from Figure 4.1(b), both misfit functions change significantly faster at shallower depths compared to greater depths, suggesting that both misfit functions are more sensitive to shallower earthquakes, which may be due to the more rapidly varying shallow structure. Notice that the values for a waveform misfit function with time shifts (dashed line) are only slightly smaller than those for a waveform misfit function without time shifts (solid line),

showing that the depth variation of earthquakes mainly gives rise to changes in waveform but not in the traveltimes for waves at periods of 6 seconds and longer. This means that we should have good resolution for event depth when we use the waveform misfit function with time shifts (4.7) to determine the optimal source parameters. Also notice that within an interval of 0.4 km, the derivative of the misfit function remains almost constant; therefore it should be a good approximation to obtain the depth derivatives at the initial solution.

4.5.3 Synthetics Test: Envelope Misfit Versus Waveform Misfit Without Time Shifts

As mentioned earlier, the existence of structural heterogeneity that is not part of our 3-D velocity model is the most severe problem that we need to overcome in our source inversions. One way of doing this, in the case of relatively small 3-D heterogeneity, is to use a combination of waveform misfit without time shifts and the envelope misfit, as in (4.4). To demonstrate that the envelope misfit function does help accommodate the effects of heterogeneity not included in the 3-D model, we assign random time shifts of up to 3 seconds to the "synthetic data" to mimic the effect of 3-D heterogeneity and regard the shifted "synthetic data" as our "data". We evaluate both the waveform misfit without time shifts (4.1) and the envelope misfit (4.3) between the "data" and synthetics for each depth calculated in Section 4.5.2. We show the functional dependence of both misfit functions on depth in Figure 4.1(c). Notice that because of the "3-D heterogeneity", the waveform misfit function without time shifts, which calculates directly the difference between data and synthetics, does not have its minimum centered on the "true" depth of 5.4 km. However, the envelope misfit function, which is not very sensitive to small arrival-time anomalies, still has its minimum at 5.4 km. This suggests that in the case of relatively small unmodeled heterogeneity (for example, in global moment-tensor inversions), we may use a combination of the envelope misfit function (4.3) and the waveform misfit function without time shifts (4.1) to recover source mechanisms. However, as shown by *Komatitsch et al.* [2004], the traveltime anomalies for surface waves at periods of 6 s and longer may be as large as 5 s for stations at distances of 250 km and larger in our 3-D southern California model.

Therefore, allowing time shifts for the waveform misfit function is a better choice to invert for earthquake source parameters in southern California. Of course as soon as we have an improved shear-velocity model for the region we should be able to dispense with these time shifts.

4.5.4 Data Test: Depth

Next, we obtain actual data recorded at various SCSN stations for the September 9, 2001, Hollywood event using the STP program provided by the Southern California Earthquake Data Center (SCEDC). We calculate the waveform misfit function with and without time shifts for selected data traces and corresponding synthetics for different depths (please refer to section 4.7 for details about the selection criteria). The result is shown in Figure 4.1(d). Global minima are achieved by both misfit functions at around 5.5 km–6.5 km, however, the misfit function without time shifts (solid line) is flatter around the minimum compared to the misfit function with time shifts (dashed line), which has a distinct minimum at about 5.6 km. We suspect that unmodeled 3-D effects may have obscured the depth minimum of the misfit function without time shifts. Also notice that the value of the misfit function with time shifts is typically 50% smaller than the value of the misfit function without time shifts, suggesting again that the arrival-time anomaly takes up a significant portion of the misfit between data and synthetics. Of course the situation can be more complicated when all the moment-tensor elements and location parameters vary. Nevertheless, the waveform misfit function with time shifts has very good sensitivity to the depth of the earthquake and will be used in our inversion process.

4.6 Inversion Procedures and Technical Details

4.6.1 Calculation and Processing of Synthetics

If we are able to obtain an initial source solution for an earthquake from either first-motion or other inversion methods, we can calculate the Fréchet derivatives of the synthetics with respect to all nine source parameters, including the six moment-tensor elements, depth,

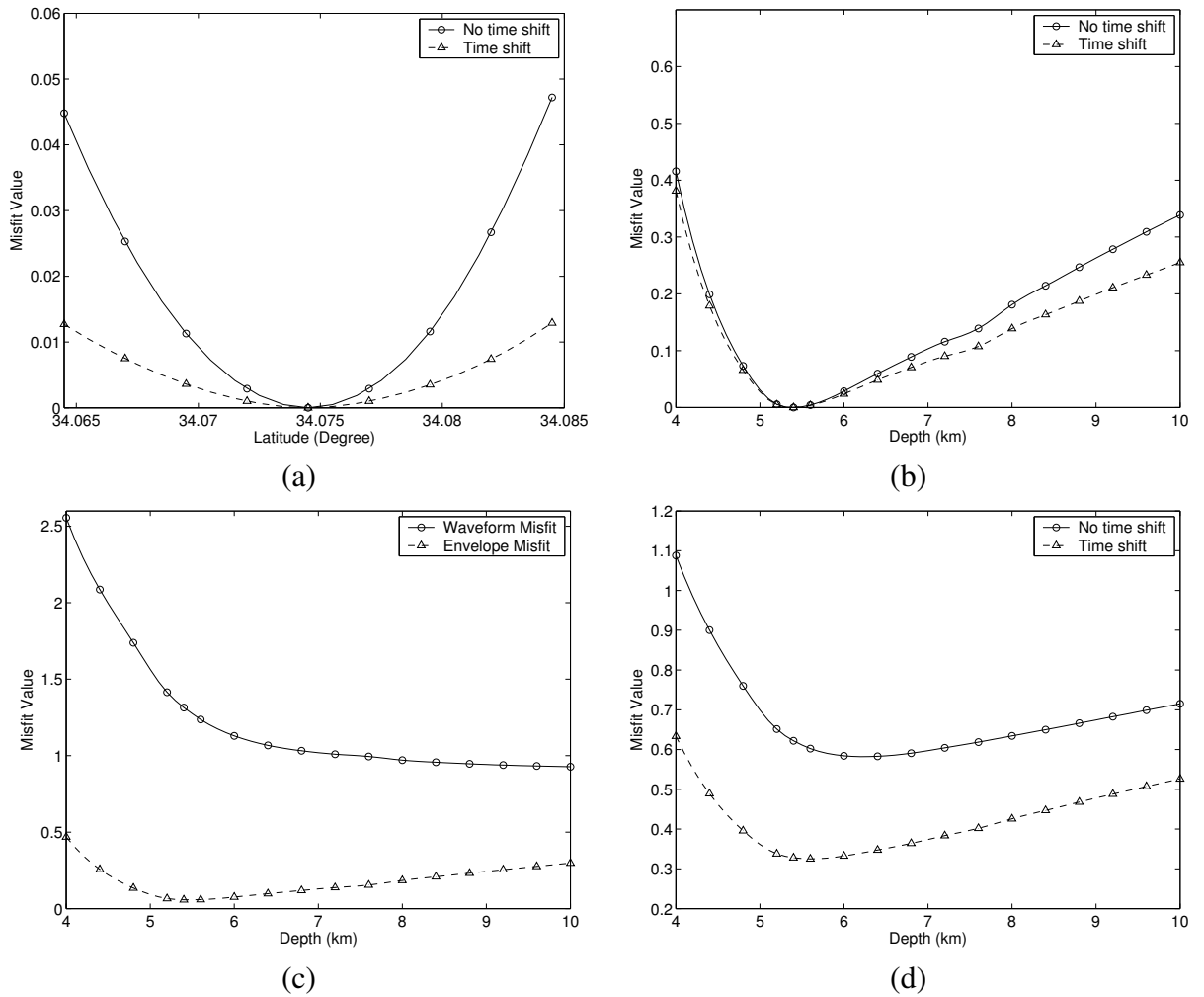


Figure 4.1: (a) Waveform misfit with or without time shifts between the "synthetic data" and the synthetics as a function of latitude. (b) Waveform misfit with or without time shifts between the "synthetic data" and the synthetics as a function of depth. (c) Waveform misfit without time shifts and envelope misfit between the "shifted synthetic data" and the synthetics as a function of depth. (d) Waveform misfit with and without time shifts between actual data and the synthetics as a function of depth.

latitude, and longitude of the earthquake. The calculations generate a mesh with a total of 45.4 million grid points (i.e., 136 million degrees of freedom), and require 14 gigabytes of distributed memory. Therefore it takes in total about one day to calculate all the derivatives needed for our source inversion. To save computation time, we combine the moment-tensor Fréchet derivatives to obtain the synthetics for the initial solution using (B.1). We then pad up to 20 seconds of zeroes before the first arrival to all the synthetics and Fréchet derivatives and bandpass filter them between 6 seconds and 40 seconds.

4.6.2 Data Processing

We collect all the broadband BH component data for the earthquake using the STP program, as mentioned earlier. We correct the raw data for the instrument response to obtain displacement records using the poles-and-zeroes files provided by the SCSN network, and then apply the same filters as applied to the synthetics.

4.6.3 Inversion

We select those data traces whose corresponding synthetics have relatively decent waveform fits to perform our source inversion. As will be discussed in more detail later, we require that the time shift between data and synthetics δt_i that satisfies (4.6) be less than a prescribed value that varies with distance, while the best correlation value itself be large enough to guarantee good similarity between data and synthetics. Since the correlation values are always normalized by the amplitude of the data and synthetics, we also require that the amplitude ratio of the data and synthetics be smaller than a certain value, to avoid problems that are associated with the stations themselves.

Regional P_{nl} waves have been shown to be relatively insensitive to 3-D heterogeneity in the crust [*Helmberger and Engen, 1980*], therefore the P_{nl} phase is a good candidate to invert for the source. We split the vertical and radial components of the data and synthetic traces into a P_{nl} part and a Rayleigh surface-wave parts for stations at distances larger than 150 km. Since the P_{nl} waves are relatively small in amplitude compared to the surface waves, we also implement a weighting factor for P_{nl} waves to balance their contribution in

the inversion.

We add the contributions of all the selected traces to form the matrix equation (B.5) or (B.17), which has dimension less than or equal to 11. We can simply solve this linear system by Gaussian elimination with partial pivoting [Conte and Boor, 1980] to obtain the new source parameters \mathbf{m}^{new} . In general, the initial source parameters are close to the true source parameters; therefore, we do not iterate the process. We evaluate the variance reduction (4.8) of the synthetics for the new source parameters \mathbf{m}^{new} relative to the data at each station to examine how well the new synthetics fit the data. We also evaluate how much the waveform misfit function with time shifts has been reduced by \mathbf{m}^{new} compared to the initial source parameters \mathbf{m}^0 by going over all the selected data and synthetics again, and calculating the misfit reduction (4.9). We are also interested in evaluating (4.9) for each trace, whether it has been used in the inversion or not, to give us some idea of how much the waveform fits have been improved by changing the source parameters.

In the following three sections we analyze three recent southern California earthquakes and determine their source parameters by the procedures outlined above. We also further discuss the effects of velocity heterogeneity that is not part of our 3-D model on the determination of earthquake source parameters.

4.7 Hollywood Earthquake

We first consider the September 9, 2001, $M_l = 4.2$ Hollywood Earthquake (Figure 3.1). This well-recorded event is small enough such that the finite size of the source can be safely ignored. We obtain data from 125 stations within our model domain, with the nearest station at a distance of 7.7 km and the farthest at a distance of 348 km. We use initial source parameters determined by the method of *Thio and Kanamori* [1995] (the second from last entry in Table 4.1). To show that this initial solution is accurate enough to start with, we plot in Figure 4.2 the correlation value (4.6) between the vertical components of the data and the synthetics for this initial solution at most of the SCSN broadband stations. Stations with high correlation values are shown in warm colors, while stations with low correlation values are shown in cold colors. We see overall very good waveform fits throughout

southern California, as also observed by *Komatitsch et al.* [2004]. We find that, in general, with increasing distance the correlation values decrease because of the smaller signal-to-noise ratio at more remote stations. As discussed in *Komatitsch et al.* [2004], the Salton Trough, Santa Barbara, and Ventura Basin regions are filled with thick sediments. However, they are outside the *Süss and Shaw* [2003] model range, and the background tomographic model [*Hauksson, 2000*] is not particularly good at describing the surface geology in these locations. Therefore, stations in these regions generally have small correlation values (Figure 4.2, station NSS, BTC, FIG, SYP, etc). In any event, this initial solution is good enough to help us select the good data traces to further fine-tune the source parameters. We use the whole seismogram (up to 3 minutes) and cut the vertical and radial traces into P_{nl} waves (from $-5.0 + \text{dist}(\text{km})/7.8$ seconds to $t_s + 2.0$ seconds, where "dist" denotes the distance of the station from the epicenter of the earthquake, and t_s denotes the predicted arrival time for S waves using the standard southern California model [*Hadley and Kanamori, 1977; Dreger and Helmberger, 1990*]), and surface waves (from $-15.0 + \text{dist}(\text{km})/3.5$ seconds to $35.0 + \text{dist}(\text{km})/3.1$ seconds) at distances larger than 150 km. We then shift the synthetics and Fréchet derivatives with respect to the corresponding data traces or segments to obtain the best correlation. We select P_{nl} data traces by requiring that the time shift between data and the initial synthetics be less than $3.0 + \text{dist}(\text{km})/80.0$ seconds, the amplitude ratio be less than 2.5, and the correlation be larger than 0.70. We select Love- and Rayleigh-wave data by requiring that the time shift between data and the initial synthetics be less than $3.0 + \text{dist}(\text{km})/50.0$ seconds, the amplitude ratio be less than 2.5, and the correlation be larger than 0.65. These values for cutting and selecting are empirical, but they have proved to be robust and appropriate for earthquakes in southern California. In total we selected for our inversion 28 P_{nl} traces from vertical and radial seismograms, and 209 surface-wave traces from all three components, including 71 vertical traces, 93 transverse traces, and 53 radial traces. We set a relative weight of 3 : 1 for P_{nl} and surface-wave traces to account for their amplitude difference. We taper the data, synthetics, and Fréchet derivatives by a trapezoidal function centered around the maximum value of the synthetics. We also multiply weights $w = N_a^{-a}(\text{dist})^{-d}N_c^{-c}$ to every single trace, where N_a represents the number of traces selected in the same azimuthal bin as the trace under consideration;

a is an exponent we set to express how much we want to emphasize differences in station azimuth; d is set to be 1.13 for P_{nt} traces, 0.74 for Rayleigh waves, and 0.55 for Love waves to accommodate an empirical amplitude decay with distance for southern California [Zhu and Helmberger, 1996]; N_c is the number of traces selected with the same component as the current trace; and c is a factor that controls the relative weights of the three components. Since the original SCSN station distribution is not uniform in azimuth, there may be directions in which the stations are sparse while the model is not well constrained, in which case a too large value of a would contaminate the final inversion result with contributions from stations with less reliable 3-D Green's functions. Therefore we generally set a to a small value. Because typically the transverse and vertical components of the synthetics are fitted better than the radial component, we set c to a smaller value for the radial component than the other two components.

To check the robustness of our inversion procedures, we usually invert with several combinations of parameters. Table 4.1 shows the inversion results for the following cases: 1) invert for both moment tensor and location; 2) invert for both moment tensor and depth; 3) invert for moment tensor only; 4) same as 1) but constrain the moment tensor to be a double couple; 5) same as 1) except that only surface-wave traces are used in the inversion; 6) same as 1) except that no weights are applied to the traces. For comparison, the solution obtained by the methods of *Thio and Kanamori* [1995] and *Zhu and Helmberger* [1996] are also listed in the table. We consider case 2) separately from 1), because typical mislocations of earthquakes in southern California are about 1–2 km in both the latitudinal and longitudinal directions, while they can be as large as 3–4 km in depth [Hauksson, 2000], and we are more concerned with obtaining a better depth. Also, because our misfit function has greater sensitivity to depth than to latitude and longitude, it is desirable to invert only for the moment tensor and depth to check for the stability of the inversion process. Comparing the inversion results in Table 4.1, all the inversion results are similar to the initial solution, with the strike and rake within 2° , and the dip within 9° . Although the presence of a non-double-couple component in zero-trace inversions may reduce the misfit, we believe it is not necessarily well-resolved, and may just reflect imperfection in our 3-D Green's function. The inversion results show that the depth of this earthquake is about 5.6 ± 0.2

Hollywood Earthquake









Type	M_w	Strike/Dip/Rake	Depth	ϵ	Mech.
moment+location	4.08	173/74/182	5.6	0.13	
moment+depth	4.08	171/72/179	5.5	0.13	
moment only	4.08	171/73/179	5.4	0.13	
double couple	4.08	172/73/178	5.5	0.00	
surface wave only	4.08	173/74/182	5.6	0.13	
equal weight	4.09	174/75/182	5.6	0.14	
Thio et al. (1995)	4.07	171/66/183	5.4	0.04	
Zhu et al. (1996)	4.17	168/74/159	6.0	0.00	

Table 4.1: Source parameters of the Hollywood event for various inversion methods or parameter settings as explained in the main text. The first six rows are the mechanisms determined by our inversions, and the last two rows are the inversion results based upon the methods of *Thio and Kanamori* [1995] and *Zhu and Helmberger* [1996]. For simplicity, only one of the fault plane solutions is listed in each row. Strike, dip, and rake are in degrees, and depth has the unit of km. The parameter ϵ represents the magnitude of the non-double-couple component in a moment tensor [*Dziewonski et al.*, 1981; *Lay and Wallace*, 1995].

km. Notice that this depth is a little bit shallower than that obtained by the method of *Zhu and Helmberger* [1996]; therefore, the focal mechanism and the moment magnitude are slightly different from those obtained by their method as well. Our evaluation of misfit reduction (4.9) is about 13% for the selected P_{nl} traces, and 7% for all selected surface-wave traces.

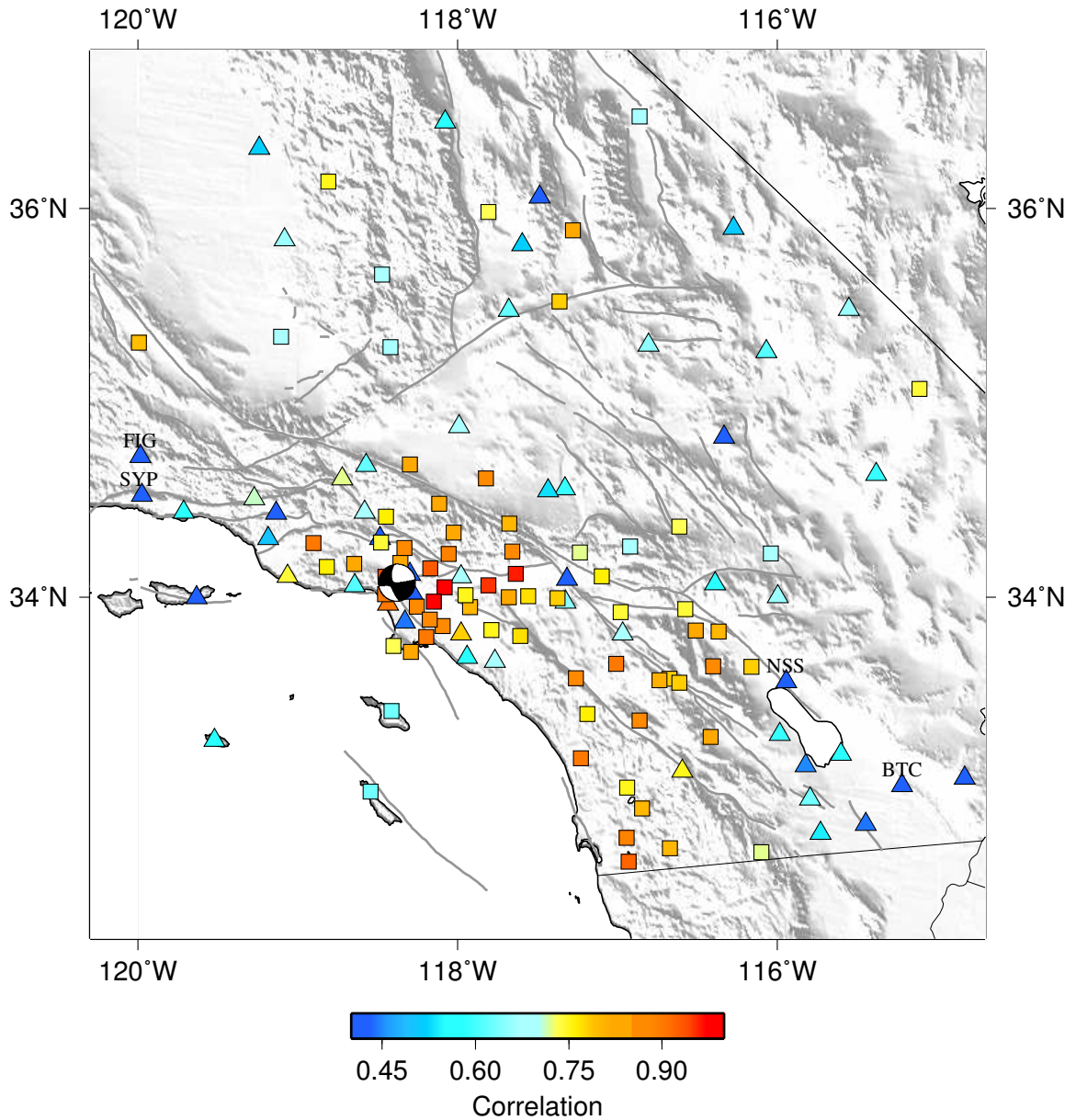


Figure 4.2: Correlation of the vertical component data and synthetics for the initial solution at most of the SCSN broadband stations for the Hollywood event. Warm colors indicate high similarity of the waveforms, while cold colors may indicate regions where our 3-D velocity model does not describe the true structure adequately. Squares indicate stations whose vertical component is used in the source inversion. All the data and synthetics have been filtered between 6 seconds and 40 seconds. The epicenter and source mechanism of the Hollywood earthquake are indicated by the beachball.

4.8 Big Bear Main Shock

We also study the February 22, 2003, $M_l = 5.4$ Big Bear main shock (Figure 3.1). This is the biggest event in southern California since the year 2000. It occurred in the San Bernardino Mountains, which are surrounded by complicated fault systems (Figure 3.1). We apply the same data selection criteria as in the previous section, except that we allow the time shift between the data and synthetics to be larger to accommodate the possible offset in origin time. Also, because this is a larger earthquake than the Hollywood event, we set the correlation criteria values to be a little larger than in the previous section. Data are available for 137 stations in our model region. From this data set, we selected 51 P_{nl} traces and 169 surface wave traces, which includes 66 vertical traces, 89 transverse traces, and 63 radial traces. The fact that we use fewer surface-wave records than for the Hollywood event is not surprising, since this event occurred outside the model of *Süss and Shaw* [2003], and we expect the overall waveform fits to be worse than for the Hollywood event. Although our integrated model has the highest resolution in the Los Angeles basin, the waves might have already developed complications before propagating into the basin.

The inversion results are summarized in Table 4.2. All the inversion results are consistent, and are within 5° of the strike, dip, and rake of the initial solution. Figure 4.3 shows the variance reduction given by (4.8) for the surface-wave part on the transverse component at each station to demonstrate how well the synthetics for the inversion result (moment tensor and location) fit the data. Obviously the variance reduction for each trace is related to both the correlation value as well as the amplitude difference between the data and the synthetics. Figure 4.3 shows that most stations have quite large variance reduction (50%). By comparing the waveform fits of the surface waves for some of the typical stations in southern California (Figure 4.4), it is clear that stations in the nodal directions of the transverse component tend to have smaller variance reductions because of the relatively small signal-to-noise ratio. However, our misfit function (4.7) recognizes the absolute amplitude information in the data; therefore, the stations in the nodal directions also provide critical information in estimating the source parameters as long as the noise level is not too high. Notice that we still have low variance reductions in the Salton Trough, Santa Barbara, and

Big Bear Earthquake









Type	M_w	Strike/Dip/Rake	Depth	ϵ	Mech.
moment+location	4.92	134/71/187	6.4	0.03	
moment+depth	4.92	134/71/187	6.4	0.03	
moment only	4.92	134/71/187	6.3	0.02	
double couple	4.92	134/72/186	6.4	0.00	
surface wave only	4.92	134/71/187	6.4	0.04	
equal weight	4.93	134/71/188	6.4	0.02	
Thio et al. (1995)	4.93	320/75/200	6.3	0.00	
Zhu et al. (1996)	4.96	135/72/193	6.5	0.00	

Table 4.2: Same as Table 4.1 but for the Big Bear main shock.

Ventura Basin area. Our variance reduction in the Los Angeles basin is relatively small, which may be due to complications in the waveforms developed along the propagation path. Overall, we achieve significant variance reduction for traces that have been used in the inversion (indicated by squares). The difference in variance reduction at different stations may also reflect the fact that a significant portion of the waveform misfit is due to complicated 3-D velocity structure. Different inversion schemes show that the depth of the earthquake is about 6.4 ± 0.3 km.

4.9 Diamond Bar Earthquake

The December 14, 2001, $M_l = 4.0$ Diamond Bar event (Figure 3.1) is quite a unique event in that it occurred close to the Los Angeles basin and had a deep hypocenter of 14 km

Big Bear Earthquake

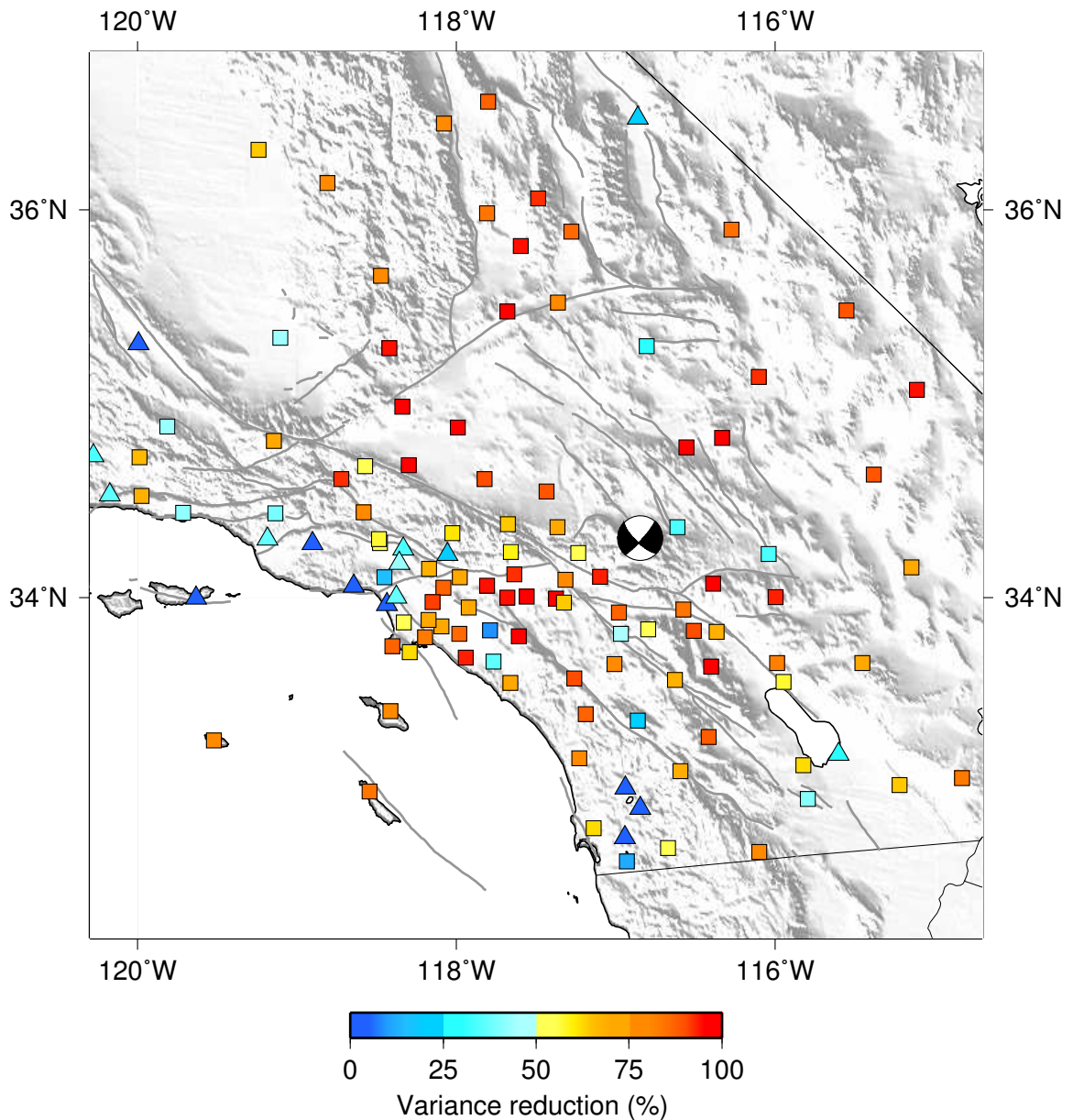


Figure 4.3: Variance reduction for transverse component synthetics calculated based upon the inverted source parameters at most of the SCSN broadband stations for the 2003 Big Bear main shock. Warm colors indicate large variance reductions, where the new synthetics generally fit the data better than at stations indicated by the cold colors. Squares indicate stations whose transverse component is used in the source inversion. Only stations with a correlation larger than 0.4 are plotted. The epicenter and source mechanism of the earthquake are denoted by the beachball.

Big Bear Earthquake

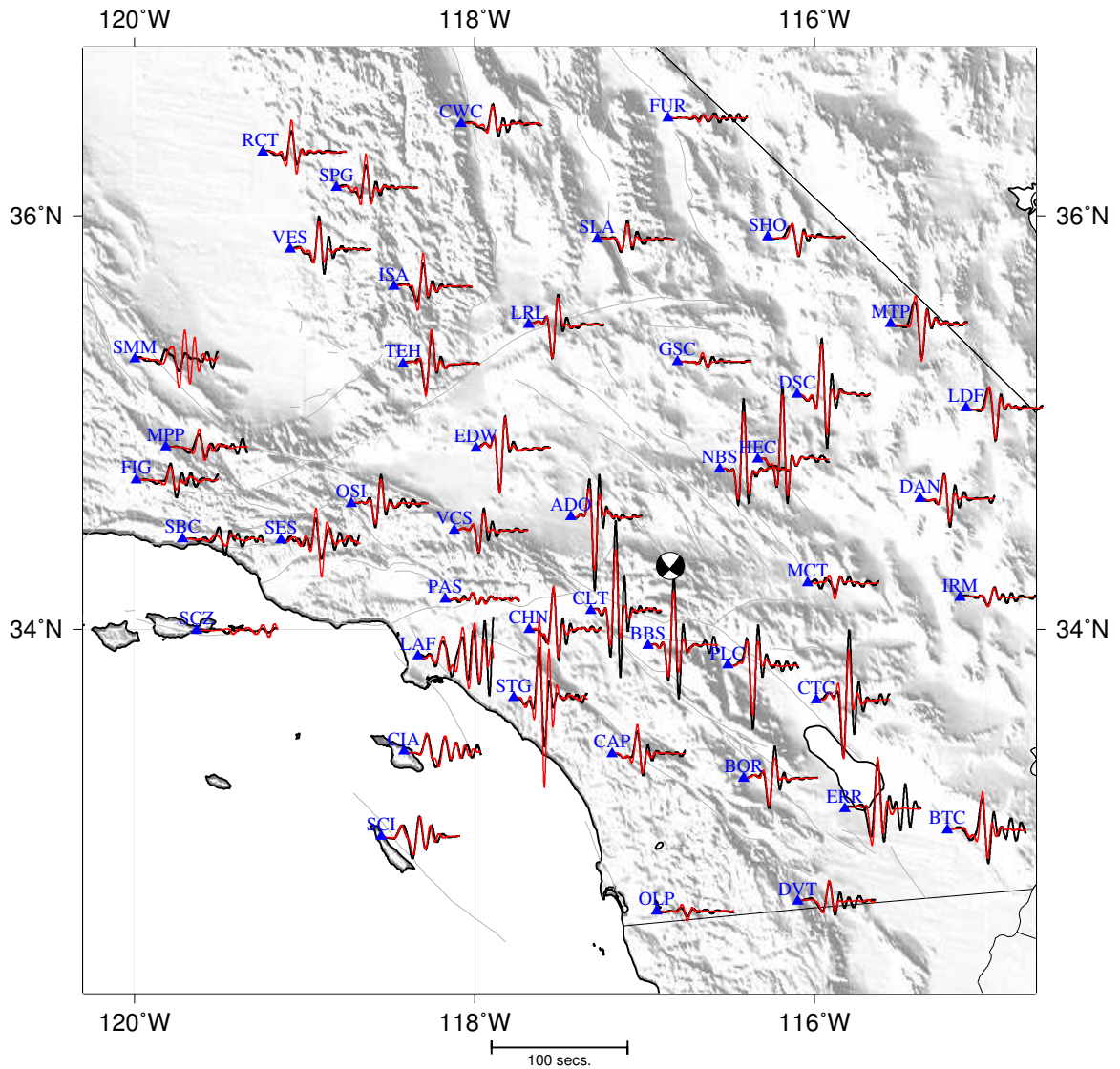


Figure 4.4: Selected waveform fits for transverse component data (black) and synthetics (red) for the inverted source parameters of the Big Bear main shock. The synthetics have been shifted with respect to the data to obtain the best correlation. Combined with Figure 4.3, it illustrates the general fits between data and synthetics throughout southern California. More examples of waveform fits can be found in *Komatitsch et al.* [2004].

(from the SCSN earthquake catalog). Besides its significant depth, its small magnitude also makes the data rather noisy to invert for a stable mechanism. The method of *Thio and Kanamori* [1995] is not able to produce a stable solution because their method mostly uses stations outside the basin, which have relatively smaller signal-to-noise ratios for this earthquake. It is desirable that we produce an initial solution by our own preliminary inversion. We select data traces based upon the noise level before the origin time, i.e., the signal-to-noise ratio. Since we have a rough estimate of the magnitude of this earthquake, we require the amplitude of the data to be below a certain value to avoid problems such as station clipping or long-period noise. Based upon these criteria, we selected 123 traces for the inversion, including 31 vertical traces, 60 transverse traces, and 32 radial traces. Figure 4.5 shows the station distribution of the selected traces (both red and green triangles) on the transverse component. Notice that they are well distributed in both distance and azimuth, showing the good quality of the data as well as the robustness of the data selection. As a preliminary inversion, we only use the surface-wave part of the traces. We apply no shift to the derivatives, because we do not have synthetics to obtain the δt_i value given by (4.6) for each trace, which means that we minimize the waveform misfit function without time shifts (4.1). We take the network event depth, 14 km, and invert only for a zero-trace moment tensor, which is a linear inversion that does not require any initial source mechanism. The inversion result gives us a source mechanism of 324° strike, 88° dip, and 178° rake; a moment magnitude of 3.64; and a variance reduction of the new synthetics (without time shifts) with respect to the selected data traces of 27%. Obviously, when no time shifts are allowed, the quality of the inversion might have been compromised due to problems such as cycle skips.

We next use the preliminary inversion result as an initial solution, calculate the location derivatives with respect to this initial solution as described in Section 4.5, and perform the standard source inversion to minimize the waveform misfit function with time shifts. The stations for which transverse component data were selected are plotted in Figure 4.5 (both red and blue triangles). Notice more than half of the selected data overlap with the data selected in the preliminary inversion, showing the suitability of the initial selection procedures. The well-distributed coverage of the selected data traces also gives us confidence

Diamond Bar Earthquake








Type	M_w	Strike/Dip/Rake	Depth	ϵ	Mech.
moment+location	3.69	320/90/187	12.0	0.15	
moment+depth	3.69	320/88/186	11.9	0.15	
moment only	3.72	320/89/183	14.0	0.15	
double couple	3.70	321/84/188	11.5	0.00	
surface wave only	3.69	320/90/187	12.0	0.15	
equal weight	3.69	320/88/189	12.0	0.14	
Preliminary Inversion	3.64	324/88/178	14.0	0.16	

Table 4.3: Same as Table 4.1 but for the Diamond Bar event. Our preliminary inversion result has been listed in the bottom row as the initial solution used in the second inversion.

in the inversion results we present in Table 4.3. Different inversion schemes show that the event depth changed from 14 km to about 12.0 ± 0.5 km. Figure 4.6 demonstrates the misfit reduction for stations with correlations higher than 0.4. Most of the stations that were used (denoted by squares) have a misfit reduction of at least 20%. Even some stations that are not used in the inversion have positive misfit reduction values, showing that the second inversion has significantly reduced the misfit function compared to the initial result. We achieve a total misfit reduction of about 18% for the selected P_{nl} traces and 25% for the surface-wave traces.

4.10 Discussion

We typically use about 200 traces to invert for the source parameters of earthquakes with $M_w \geq 4.0$, and about 120 traces for a $M_w = 3.7$ earthquake. Therefore, we may be able to extend our inversion method to earthquakes with $M_w \geq 3.5$ for southern California. In the Los Angeles basin, where we have the highest model resolution, we may be able to push

Diamond Bar Earthquake

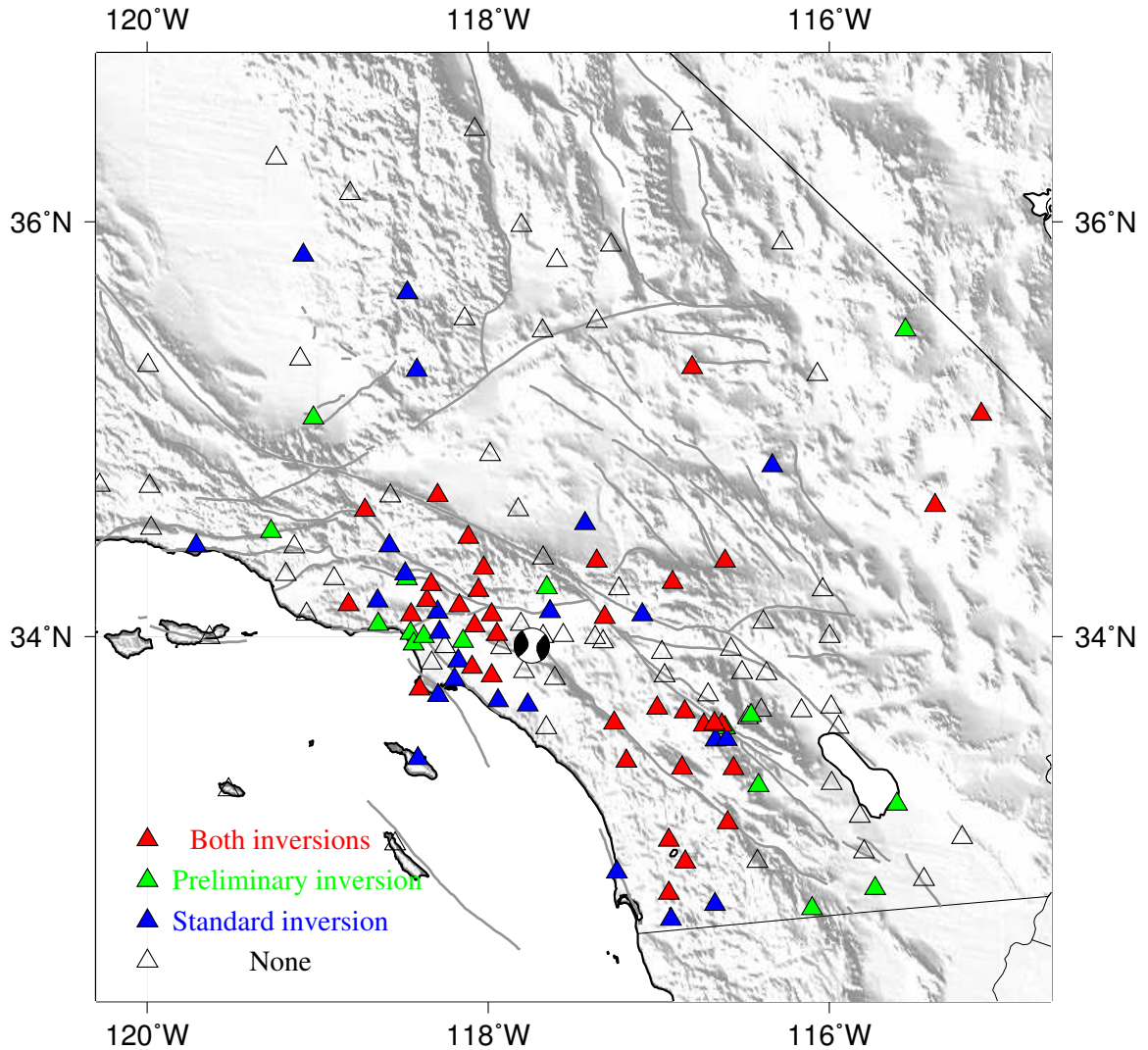


Figure 4.5: Stations whose transverse component has been used in the preliminary source inversion and the standard source inversions for the Diamond Bar event. Red triangles indicate stations that are used in both inversions, green triangles indicate stations used only in the preliminary inversion, blue triangles indicate stations used only in the second inversion, and white triangles indicate stations used in neither inversion. The epicenter and source mechanism of the earthquake are indicated by the beachball.

Diamond Bar Earthquake

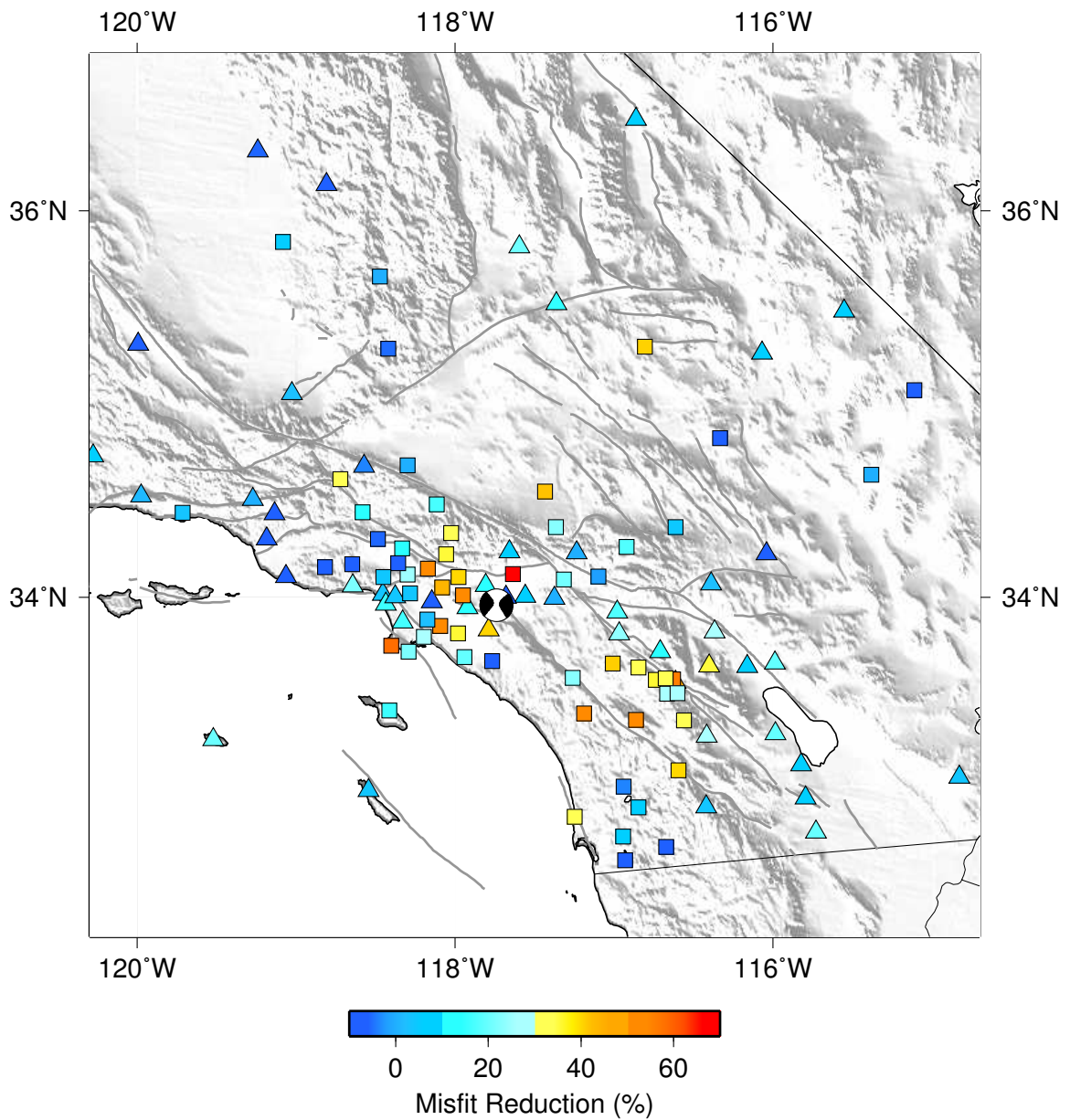


Figure 4.6: Misfit reduction for transverse component seismograms by the source parameters from the standard inversion compared to the ones determined by a preliminary inversion at most of the SCSN stations for the Diamond Bar event. Only stations with correlations larger than 0.4 are plotted. Squares indicate stations whose transverse component is used in the source inversion. Warm colors indicate a large misfit reduction, where the improvement in the source mechanism helps to explain the misfit between the data and the synthetics for the initial solution, whereas cold colors indicate locations where the waveform misfit is mainly due to an imperfect 3-D Green's function.

the limit to even smaller earthquakes ($M_w = 3.3$). The inversion procedure described for the Hollywood, Big Bear, and Diamond Bar earthquakes can be very easily automated for events with $M_w \geq 3.5$ in southern California. The point source mechanism should be a quite good representation for these small- and moderate-sized earthquakes. By minimizing the waveform misfit function with time shifts, we can accommodate the effects of an imperfect 3-D Green's function on the source inversion reasonably well. Future improvements to the 3-D velocity model should render such time shifts unnecessary.

4.11 Conclusion

We determined source parameters for three southern California earthquakes based upon spectral-element simulations of regional seismic wave propagation in an integrated 3-D southern California velocity model. By minimizing the waveform misfit function with time shifts, which accommodates the imperfect 3-D Green's function, we were able to determine moment tensors and locations for these earthquakes. The inversion results agree very well with estimates from either body- or surface-wave inversion methods for the Hollywood and Big Bear earthquakes. The inversion procedures can be performed with or without an initial solution, as shown for the Diamond Bar earthquake. This enables us to perform automatic source parameter inversions for $M_w \geq 3.5$ earthquakes in southern California. Since there are about 40 earthquakes with $M_l \geq 3.5$ in southern California every year, we will be able to build and maintain an earthquake source mechanism database, which should be of importance for understanding local fault systems and tectonic processes.

Chapter 5

Finite-frequency Kernels Based upon Adjoint Method for Regional Wave Simulations

5.1 Abstract

We derive the adjoint equations associated with the calculation of Fréchet derivatives for tomographic inversions based upon a Lagrange multiplier method. The Fréchet derivative of an objective function $\chi(m)$, where m denotes the earth model, may be written in the generic form $\delta\chi = \int K_m(\mathbf{x}) \delta \ln m(\mathbf{x}) d^3\mathbf{x}$, where $\delta \ln m = \delta m/m$ denotes the relative model perturbation and K_m the associated 3-D sensitivity or Fréchet kernel. Complications due to artificial absorbing boundaries for regional simulations as well as finite sources are accommodated. We construct the 3-D finite-frequency "banana-doughnut" kernel K_m by simultaneously computing the so-called "adjoint" wavefield forward in time and reconstructing the regular wavefield backward in time. The adjoint wavefield is produced by using time-reversed signals at the receivers as fictitious, simultaneous sources, while the regular wavefield is reconstructed on the fly by propagating the last frame of the wavefield saved by a previous forward simulation backward in time. The approach is based upon the spectral-element method, and only two simulations are needed to produce the 3-D finite-frequency sensitivity kernels. The method is applied to 1-D and 3-D regional models. Various 3-D shear- and compressional-wave sensitivity kernels are presented for different regional body- and surface-wave arrivals in the seismograms. These kernels illustrate the

sensitivity of the observations to the structural parameters and form the basis of fully 3-D tomographic inversions. Most of the results from this chapter are presented in *Liu and Tromp* [2006a].

5.2 Introduction

Seismic tomography is transitioning from classical ray-based tomography to finite-frequency tomography. The new approach incorporates traveltimes effects associated with wavefront healing and recognizes the inherent frequency dependence of the body-wave traveltimes or surface-wave phase. For layer cake or spherically symmetric Earth models, sensitivity or *Fréchet* kernels may be calculated based upon surface-wave Green's functions [*Marquering et al.*, 1999], normal modes [*Zhao et al.*, 2000], or asymptotic, ray-based methods [*Dahlen et al.*, 2000; *Hung et al.*, 2000; *Zhou et al.*, 2004]. Simple 3-D traveltimes kernels for phases like P and S are shaped like bananas with a doughnut-like cross-section, and thus the kernels are commonly referred to as "banana-doughnut" kernels. Such kernels were recently implemented for compressional-wave tomography by *Montelli et al.* [2004].

To go beyond 1-D reference models, i.e., allowing for laterally heterogeneous reference Earth models, requires fully 3-D numerical simulations. *Zhao et al.* [2005] demonstrate that 3-D finite-frequency sensitivity kernels for 3-D reference models may be obtained by calculating and storing 3-D Green's functions for all earthquakes and stations of interest. An advantage of this approach is that it gives access to both the gradient and the Hessian of the misfit function in the tomographic inverse problem. A disadvantage is the formidable storage requirements associated with saving the entire Green's function as a function of space and time for all sources and receivers. Alternatively, *Tromp et al.* [2005] demonstrate that the gradient of a misfit function may be obtained based upon just two numerical simulations for each earthquake: one calculation for the current model and a second "adjoint" calculation that uses time-reversed signals at the receivers as simultaneous, fictitious sources. The main benefit of the adjoint approach is that the *Fréchet* derivatives of the misfit function may be obtained based upon two 3-D simulations for each earthquake. Because one needs simultaneous access to both the regular wavefield and the adjoint wavefield during the con-

struction of the kernel, the approach doubles the memory requirements, but there is no need to store wavefields as a function of space and time. A disadvantage is the fact that the Hessian is unavailable, which leads to the use of iterative, e.g., conjugate-gradient, methods in the inverse problem. In this paper we introduce a Lagrange multiplier method from which the adjoint wave equations and the related finite-frequency kernel expressions follow naturally. We apply the adjoint method to regional phases, first, for educational purposes, to a simple homogeneous half-space model, and then to a realistic 3-D integrated southern California model. The method can be readily extended to include numerous phases per seismogram, as well as many seismograms per earthquake. This facilitates the rapid calculation of the gradient of a very general misfit function with respect to the model parameters.

5.3 Lagrange Multiplier Method

In this section we use a Lagrange multiplier method to derive the adjoint seismic wave equations and the associated finite-frequency sensitivity kernels. This approach complements the results obtained by *Tromp et al.* [2005] and clearly demonstrates the origin of the adjoint seismic wave equation and the related sensitivities to perturbations in structure and source parameters.

Suppose we seek to minimize the least-squares waveform misfit function

$$\chi = \frac{1}{2} \sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]^2 dt, \quad (5.1)$$

where the interval $[0, T]$ denotes the time series of interest, $\mathbf{s}(\mathbf{x}_r, t)$ denotes the synthetic displacement at receiver location \mathbf{x}_r as a function of time t , and $\mathbf{d}(\mathbf{x}_r, t)$ denotes the observed 3-component displacement vector. In practice, both the data \mathbf{d} and the synthetics \mathbf{s} will be windowed, filtered, and possibly weighted on the time interval $[0, T]$. In what follows we will implicitly assume that such filtering operations have been performed, i.e., the symbols \mathbf{d} and \mathbf{s} will denote processed data and synthetics, respectively. As demonstrated by *Tromp et al.* [2005], one may choose to minimize any number of misfit functions, e.g., cross-correlation traveltime measurements or surface-wave phase anomalies, but for the

purpose of this discussion we will use the waveform misfit function (5.1). Different measures of misfit simply give rise to different adjoint sources. We will seek to minimize the misfit function (5.1) subject to the constraint that the synthetic displacement field \mathbf{s} satisfies the seismic wave equation, as we discuss next.

Let us consider an Earth model with volume Ω and outer free surface $\partial\Omega$. In Appendix C we will consider the complications associated with regional Earth models, which have both a free surface and an artificial boundary on which energy needs to be absorbed. The synthetic wavefield $\mathbf{s}(\mathbf{x}, t)$ in (5.1) is determined by the seismic wave equation

$$\rho \partial_t^2 \mathbf{s} - \nabla \cdot \mathbf{T} = \mathbf{f}, \quad (5.2)$$

where ρ denotes the distribution of density. In an elastic medium, the stress \mathbf{T} is related to the displacement gradient through Hooke's law:

$$\mathbf{T} = \mathbf{c} : \nabla \mathbf{s}, \quad (5.3)$$

where \mathbf{c} denotes the elastic tensor. On the Earth's free surface $\partial\Omega$ the traction must vanish:

$$\hat{\mathbf{n}} \cdot \mathbf{T} = \mathbf{0} \quad \text{on } \partial\Omega, \quad (5.4)$$

where $\hat{\mathbf{n}}$ denotes the unit outward normal on the surface. In addition to the boundary condition (5.4), the seismic wave equation (5.2) must be solved subject to the initial conditions

$$\mathbf{s}(\mathbf{x}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}(\mathbf{x}, 0) = \mathbf{0}. \quad (5.5)$$

Finally, the force \mathbf{f} in (5.2) represents the earthquake. In the case of a simple point source it may be written in terms of the moment tensor \mathbf{M} as

$$\mathbf{f} = -\mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) S(t), \quad (5.6)$$

where the location of the point source is denoted by \mathbf{x}_s , $\delta(\mathbf{x} - \mathbf{x}_s)$ denotes the Dirac delta

distribution located at \mathbf{x}_s , and $S(t)$ denotes the source-time function. The complications associated with a finite source are discussed in Appendix C.2.

Our objective is to minimize the misfit function (5.1) subject to the constraint that the synthetic displacement field \mathbf{s} satisfies the seismic wave equation (5.2). Mathematically, this implies the minimization of the constrained action

$$\chi = \frac{1}{2} \sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]^2 dt - \int_0^T \int_{\Omega} \lambda \cdot (\rho \partial_t^2 \mathbf{s} - \nabla \cdot \mathbf{T} - \mathbf{f}) d^3 \mathbf{x} dt, \quad (5.7)$$

where the vector Lagrange multiplier $\lambda(\mathbf{x}, t)$ remains to be determined. Upon taking the variation of the action (5.7), using Hooke's law (5.3), we obtain

$$\begin{aligned} \delta\chi &= \int_0^T \int_{\Omega} \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta \mathbf{s}(\mathbf{x}, t) d^3 \mathbf{x} dt \\ &\quad - \int_0^T \int_{\Omega} \lambda \cdot [\delta \rho \partial_t^2 \mathbf{s} - \nabla \cdot (\delta \mathbf{c} : \nabla \mathbf{s}) - \delta \mathbf{f}] d^3 \mathbf{x} dt \\ &\quad - \int_0^T \int_{\Omega} \lambda \cdot [\rho \partial_t^2 \delta \mathbf{s} - \nabla \cdot (\mathbf{c} : \nabla \delta \mathbf{s})] d^3 \mathbf{x} dt. \end{aligned} \quad (5.8)$$

Upon integrating the terms involving spatial and temporal derivatives of the variation $\delta \mathbf{s}$ by parts, we obtain after some algebra

$$\begin{aligned} \delta\chi &= \int_0^T \int_{\Omega} \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta \mathbf{s}(\mathbf{x}, t) d^3 \mathbf{x} dt \\ &\quad - \int_0^T \int_{\Omega} (\delta \rho \lambda \cdot \partial_t^2 \mathbf{s} + \nabla \lambda : \delta \mathbf{c} : \nabla \mathbf{s} - \lambda \cdot \delta \mathbf{f}) d^3 \mathbf{x} dt \\ &\quad - \int_0^T \int_{\Omega} [\rho \partial_t^2 \lambda - \nabla \cdot (\mathbf{c} : \nabla \lambda)] \cdot \delta \mathbf{s} d^3 \mathbf{x} dt - \int_{\Omega} [\rho(\lambda \cdot \partial_t \delta \mathbf{s} - \partial_t \lambda \cdot \delta \mathbf{s})]_0^T d^3 \mathbf{x} \\ &\quad + \int_0^T \int_{\partial \Omega} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta \mathbf{c} : \nabla \mathbf{s} + \mathbf{c} : \nabla \delta \mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla \lambda) \cdot \delta \mathbf{s} d^2 \mathbf{x} dt, \end{aligned} \quad (5.9)$$

where the notation $[f]_0^T$ means $f(T) - f(0)$, for any function f . Perturbing the free surface boundary condition (5.4) implies $\hat{\mathbf{n}} \cdot (\delta \mathbf{c} : \nabla \mathbf{s} + \mathbf{c} : \nabla \delta \mathbf{s}) = \mathbf{0}$ on $\partial \Omega$, and perturbing the

initial conditions (5.5) implies that $\delta\mathbf{s}(\mathbf{x}, 0) = \mathbf{0}$ and $\partial_t\delta\mathbf{s}(\mathbf{x}, 0) = \mathbf{0}$. Thus we obtain

$$\begin{aligned} \delta\chi &= \int_0^T \int_{\Omega} \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta\mathbf{s}(\mathbf{x}, t) d^3\mathbf{x} dt \\ &\quad - \int_0^T \int_{\Omega} (\delta\rho\lambda \cdot \partial_t^2\mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda \cdot \delta\mathbf{f}) d^3\mathbf{x} dt - \int_0^T \int_{\Omega} [\rho\partial_t^2\lambda - \nabla \cdot (\mathbf{c} : \nabla\lambda)] \cdot \delta\mathbf{s} d^3\mathbf{x} dt \\ &\quad - \int_{\Omega} [\rho(\lambda \cdot \partial_t\delta\mathbf{s} - \partial_t\lambda \cdot \delta\mathbf{s})]_T d^3\mathbf{x} - \int_0^T \int_{\partial\Omega} \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} d^2\mathbf{x} dt, \end{aligned} \quad (5.10)$$

where the notation $[f]_T$ means $f(T)$.

In the absence of perturbations in the model parameters $\delta\rho$, $\delta\mathbf{c}$, and $\delta\mathbf{f}$, the variation in the action (5.10) is stationary with respect to perturbations $\delta\mathbf{s}$ provided the Lagrange multiplier λ satisfies the equation

$$\rho\partial_t^2\lambda - \nabla \cdot (\mathbf{c} : \nabla\lambda) = \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r), \quad (5.11)$$

subject to the free surface boundary condition

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = \mathbf{0} \quad \text{on } \partial\Omega, \quad (5.12)$$

and the end conditions

$$\lambda(\mathbf{x}, T) = \mathbf{0}, \quad \partial_t\lambda(\mathbf{x}, T) = \mathbf{0}. \quad (5.13)$$

More generally, provided the Lagrange multiplier λ is determined by equations (5.11)–(5.13), the variation in the action (5.10) reduces to

$$\delta\chi = - \int_0^T \int_{\Omega} (\delta\rho\lambda \cdot \partial_t^2\mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda \cdot \delta\mathbf{f}) d^3\mathbf{x} dt. \quad (5.14)$$

This equation tells us the change in the misfit function $\delta\chi$ due to changes in the model parameters $\delta\rho$, $\delta\mathbf{c}$, and $\delta\mathbf{f}$ in terms of the original wavefield \mathbf{s} determined by (5.2)–(5.5) and the Lagrange multiplier wavefield λ determined by (5.11)–(5.13).

To appreciate the nature of the Lagrange multiplier wavefield, let us define the *adjoint*

wavefield \mathbf{s}^\dagger in terms of the Lagrange multiplier wavefield λ by

$$\mathbf{s}^\dagger(\mathbf{x}, t) \equiv \lambda(\mathbf{x}, T - t), \quad (5.15)$$

i.e., the adjoint wavefield is the time-reversed Lagrange multiplier wavefield λ . Then the adjoint wavefield \mathbf{s}^\dagger is determined by the set of equations

$$\rho \partial_t^2 \mathbf{s}^\dagger - \nabla \cdot \mathbf{T}^\dagger = \sum_r [\mathbf{s}(\mathbf{x}_r, T - t) - \mathbf{d}(\mathbf{x}_r, T - t)] \delta(\mathbf{x} - \mathbf{x}_r), \quad (5.16)$$

where we have defined the *adjoint stress* in terms of the gradient of the adjoint displacement by

$$\mathbf{T}^\dagger = \mathbf{c} : \nabla \mathbf{s}^\dagger. \quad (5.17)$$

The adjoint wave equation (5.16) is subject to the free surface boundary condition

$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{0} \quad \text{on } \partial\Omega, \quad (5.18)$$

and the initial conditions

$$\mathbf{s}^\dagger(\mathbf{x}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}^\dagger(\mathbf{x}, 0) = \mathbf{0}. \quad (5.19)$$

Upon comparing (5.16)–(5.19) with (5.2)–(5.5), we see that the adjoint wavefield \mathbf{s}^\dagger is determined by exactly the same wave equation, boundary conditions, and initial conditions as the regular wavefield, with the exception of the source term: The regular wavefield is determined by the source \mathbf{f} , whereas the adjoint wavefield is generated by using the time-reversed differences between the synthetics \mathbf{s} and the data \mathbf{d} at the receivers as simultaneous sources.

In terms of the adjoint wavefield \mathbf{s}^\dagger , the gradient of the misfit function (5.14) may be rewritten in the form

$$\delta\chi = \int_{\Omega} (\delta\rho K_\rho + \delta\mathbf{c} :: \mathbf{K}_c) d^3\mathbf{x} + \int_0^T \int_{\Omega} \mathbf{s}^\dagger \cdot \delta\mathbf{f} d^3\mathbf{x} dt, \quad (5.20)$$

where we have defined the kernels

$$K_\rho(\mathbf{x}) = - \int_0^T \mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) dt, \quad (5.21)$$

$$\mathbf{K}_c(\mathbf{x}) = - \int_0^T \nabla \mathbf{s}^\dagger(\mathbf{x}, T-t) \nabla \mathbf{s}(\mathbf{x}, t) dt. \quad (5.22)$$

Realizing that δc and \mathbf{K}_c are both fourth-order tensors, we use the notation $\delta c :: \mathbf{K}_c = \delta c_{ijkl} K_{c_{ijkl}}$ in (5.20).

The perturbation to the point source (5.6) may be written in the form

$$\delta \mathbf{f} = -\delta \mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) S(t) - \mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s - \delta \mathbf{x}_s) S(t) - \mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) \delta S(t), \quad (5.23)$$

where $\delta \mathbf{M}$ denotes the perturbed moment tensor, $\delta \mathbf{x}_s$ the perturbed point source location, and $\delta S(t)$ the perturbed source-time function. Upon substituting (5.23) into the gradient of the misfit function (5.20), using the properties of the Dirac delta distribution, we obtain

$$\begin{aligned} \delta \chi = & \int_\Omega (\delta \rho K_\rho + \delta c :: \mathbf{K}_c) d^3 \mathbf{x} + \int_0^T \delta \mathbf{M} : \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) S(t) dt \\ & + \int_0^T \mathbf{M} : (\delta \mathbf{x}_s \cdot \nabla_s) \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) S(t) dt + \int_0^T \mathbf{M} : \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) \delta S(t) dt, \end{aligned} \quad (5.24)$$

where $\boldsymbol{\epsilon}^\dagger = \frac{1}{2}[\nabla \mathbf{s}^\dagger + (\nabla \mathbf{s}^\dagger)^T]$ denotes the adjoint strain tensor and a superscript T denotes the transpose.

In an isotropic Earth model we have $c_{ijklm} = (\kappa - \frac{2}{3}\mu)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, where μ and κ denote the shear and bulk moduli, respectively. Thus we may write

$$\delta c :: \mathbf{K}_c = \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa, \quad (5.25)$$

where the isotropic kernels K_κ and K_μ represent Fréchet derivatives with respect to relative bulk and shear moduli perturbations $\delta \ln \kappa = \delta \kappa / \kappa$ and $\delta \ln \mu = \delta \mu / \mu$, respectively. These isotropic kernels are given by

$$K_\mu(\mathbf{x}) = - \int_0^T 2\mu(\mathbf{x}) \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) dt, \quad (5.26)$$

$$K_\kappa(\mathbf{x}) = - \int_0^T \kappa(\mathbf{x}) [\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)] [\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \quad (5.27)$$

where

$$\mathbf{D} = \frac{1}{2}[\nabla \mathbf{s} + (\nabla \mathbf{s})^T] - \frac{1}{3}(\nabla \cdot \mathbf{s})\mathbf{I}, \quad \mathbf{D}^\dagger = \frac{1}{2}[\nabla \mathbf{s}^\dagger + (\nabla \mathbf{s}^\dagger)^T] - \frac{1}{3}(\nabla \cdot \mathbf{s}^\dagger)\mathbf{I}, \quad (5.28)$$

denote the traceless strain deviator and its adjoint, respectively.

Finally, we may express the Fréchet derivatives in an isotropic Earth model in terms of relative variations in density $\delta \ln \rho$, shear-wave speed $\delta \ln \beta$, and compressional-wave speed $\delta \ln \alpha$ based upon the relationship

$$\delta \ln \rho K_\rho + \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa = \delta \ln \rho K_{\rho'} + \delta \ln \beta K_\beta + \delta \ln \alpha K_\alpha, \quad (5.29)$$

where

$$K_{\rho'} = K_\rho + K_\kappa + K_\mu, \quad (5.30)$$

$$K_\beta = 2 \left(K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right), \quad (5.31)$$

$$K_\alpha = 2 \left(\frac{\kappa + \frac{4}{3}\mu}{\kappa} \right) K_\kappa. \quad (5.32)$$

In later sections we will see numerous examples of shear- and compressional-wave kernels for various body- and surface-wave arrivals.

5.4 Spectral Element Method

We simulate wave propagation in both a homogeneous velocity model and a realistic southern California velocity model by the spectral element method described in Chapter 2 and Chapter 3. The combination of a detailed crustal model and an accurate numerical technique results in generally good fits between data and synthetic seismograms on all three components at most stations in the Southern California Seismic Network at periods of 6 s and longer [Komatitsch *et al.*, 2004]. This provides us with a good starting point for further improvement of the 3-D wave speed model. A typical 3-D simulation for 3-minutes long

seismograms takes approximately 40 minutes on a 72-node PC cluster. Time shifts of up to 5 s are needed to align the data and the synthetics on the transverse component, suggesting significant deviations of the model shear-wave speed from reality, i.e., significant improvements may be achieved through further inversions. The adjoint methods discussed in this paper makes 3-D inversions based upon highly heterogeneous initial models feasible.

5.5 Numerical Implementation of the Adjoint Method

From the kernel expressions (5.30), (5.31), and (5.32), it is obvious that in order to perform the time integration, simultaneous access to the forward wavefield \mathbf{s} at time t and the adjoint wavefield \mathbf{s}^\dagger at time $T - t$ is required, as illustrated in Figure 5.1.

This rules out the possibility of carrying both the forward and the adjoint simulation simultaneously in the spectral-element simulation, where both wavefields would only be available at a given time t . One apparent solution is to run the forward simulation, save the whole forward wavefield as a function of space and time, and then launch the adjoint simulation while performing the time integration by accessing the time t slice of the adjoint wavefield and reading back the corresponding $T - t$ slice of the forward wavefield stored on the hard disk. However, this poses a serious storage problem because the complete forward field $\mathbf{s}(\mathbf{x}, t)$ can be very large when saved at every time step and every grid point, especially when the problem is large enough that parallel computing is involved. One remedy might be to introduce a highly efficient compression scheme to reduce the storage requirements.

In the absence of attenuation, an alternative approach is to introduce the backward wave equation, i.e., to reconstruct the forward wavefield backwards in time from the displacement and velocity wavefield at the end of the simulation. The backward wavefield is determined by

$$\rho \partial_t^2 \mathbf{s} = \nabla \cdot (\mathbf{c} : \nabla \mathbf{s}) + \mathbf{f} \quad \text{in } V, \quad (5.33)$$

$$\mathbf{s}(\mathbf{x}, T) \quad \text{and} \quad \partial_t \mathbf{s}(\mathbf{x}, T) \quad \text{given}, \quad (5.34)$$

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla \mathbf{s}) = \mathbf{0} \quad \text{on } \Omega. \quad (5.35)$$

This initial and boundary value problem can be solved to reconstruct $s(\mathbf{x}, t)$ for $T \geq t \geq 0$ the same way the forward wave equation is solved. Technically, the only difference between solving the backward wave equation versus solving the forward wave equation is a change in the sign of the time step parameter Δt .¹

If we carry both the backward and the adjoint simulation simultaneously in memory during the spectral-element simulation (Figure 5.1), we have access to the forward wavefield at time t and the adjoint wavefield at time $T - t$, which is exactly what we need to perform the time integration involved in the construction of the kernels (5.30), (5.31), and (5.32). A great advantage of this approach is that only the wavefield at the last time step of the forward simulation needs to be stored and read back for the reconstruction of $s(\mathbf{x}, t)$ and the construction of the kernels.

For regional spectral-element simulations, because of the limited size of the computational domain, absorbing boundary conditions are applied to mimic wave propagation in a semi-infinite medium. By saving the forward wavefield on the absorbing boundaries at every time step, we can add back the absorbed wavefield in the backward simulation that follows the forward simulation. Therefore, by saving the wavefield on the absorbing boundaries and the entire wavefield at the end of the forward simulation, we can reconstruct the forward wavefield in reverse time by solving the backwards wave equation, reinjecting the absorbed wavefield as we go along. In the parallel simulation, only those mesh slices that involve a part of an absorbing boundary need to access the absorbed field, and obviously the storage requirements are still relatively modest compared to saving the entire forward field at every time step. For more details about the implementation of absorbing boundary conditions in the adjoint method, please refer to Appendix C.

5.6 Application to a 3-D Homogeneous Model

For educational purposes, we first implement our adjoint method for a 3-D homogeneous model, as shown in Figure 5.2. The adjoint experiments presented in this section are the

¹In an attenuating medium, solving the backward wave equation is numerically challenging because one needs to "undo" the effects of attenuation. We are currently experimenting with a number of implementations.

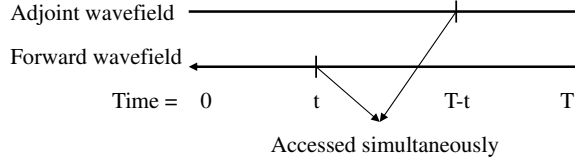


Figure 5.1: During the construction of the finite-frequency sensitivity kernels (5.30), (5.31), and (5.32), one needs simultaneous access to the forward wavefield at time t and the adjoint wavefield at time $T - t$, where T denotes the duration of the numerical simulation. In our implementation this is accomplished by reconstructing the forward simulation backwards in time by solving the backward wave equation (5.33).

3-D complement to some of the 2-D experiments discussed in *Tromp et al.* [2005].

The 3-D model box has dimensions of $500 \text{ km} \times 500 \text{ km}$ on the surface and 60 km in the vertical direction. A hypothetical source and receiver are located at a subsurface depth of 40 km at a mutual distance of 100 km . For simplicity, we use a point force with a Ricker wavelet source time function with a half-duration of 2 s . Synthetic seismograms, obtained from a spectral-element simulation accurate to a shortest period of 2 s , are recorded at the receiver, as illustrated in Figure 5.3(a) for an SH (transverse) source and in Figure 5.3(b) for a P-SV (vertical) source. During the forward simulation, absorbing boundary contributions are saved to the hard disk, and at the end of the forward simulation the displacement and velocity of the last time frame are also recorded on the disk. Next, the adjoint simulation is launched, and its source is created by cutting the arrival of interest out of the recorded seismogram and time-reversing it, as illustrated in Figure 5.3. The backward equation is solved simultaneously with the adjoint simulation, starting from the last frames of displacement and velocity that were saved, and reinserting the absorbing boundary contribution from the appropriate time step. The time integration involved in the construction of the finite-frequency sensitivity kernels (5.30), (5.31), and (5.32) is performed on the fly, based upon simultaneous access to the forward wavefield $s(\mathbf{x}, t)$ and the adjoint wavefield $s^\dagger(\mathbf{x}, T - t)$.

Based upon this approach, we have computed Fréchet kernels relating finite-frequency traveltime anomalies of P, S, SS, and PS+SP arrivals to structural perturbations. These kernels are discussed individually in the next few sections.

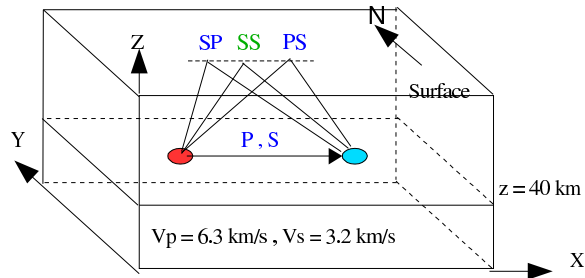


Figure 5.2: Geometry of the experiment for a homogeneous half-space with a compressional-wave speed of 6.3 km/s and a shear-wave speed of 3.2 km/s. For educational purposes, the source and the receiver are located at a depth of 40 km at a mutual distance of 100 km. The direct P and S rays as well as the surface-reflected SS, SP and PS rays are labeled for reference.

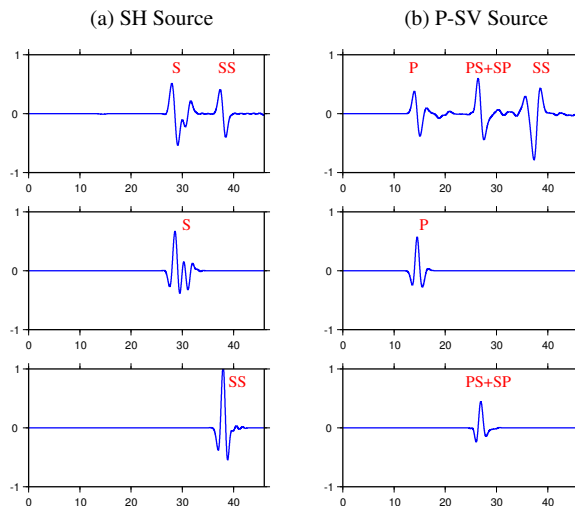


Figure 5.3: (a) Left column. Top: synthetic transverse seismogram for the source-receiver geometry shown in Figure 5.2. The direct S and the surface-reflected SS arrival are indicated. Middle: adjoint source for the S arrival. This adjoint source is obtained by differentiating and time-reversing the S arrival in the top seismogram. Bottom: adjoint source for the SS arrival. (b) Right column. Synthetic vertical component seismogram in which the P, PS+SP, and SS arrivals are labeled. Middle: adjoint source for the P arrival obtained by differentiating and time-reversing the P arrival in the top seismogram. Bottom: adjoint source for the PS+SP arrival.

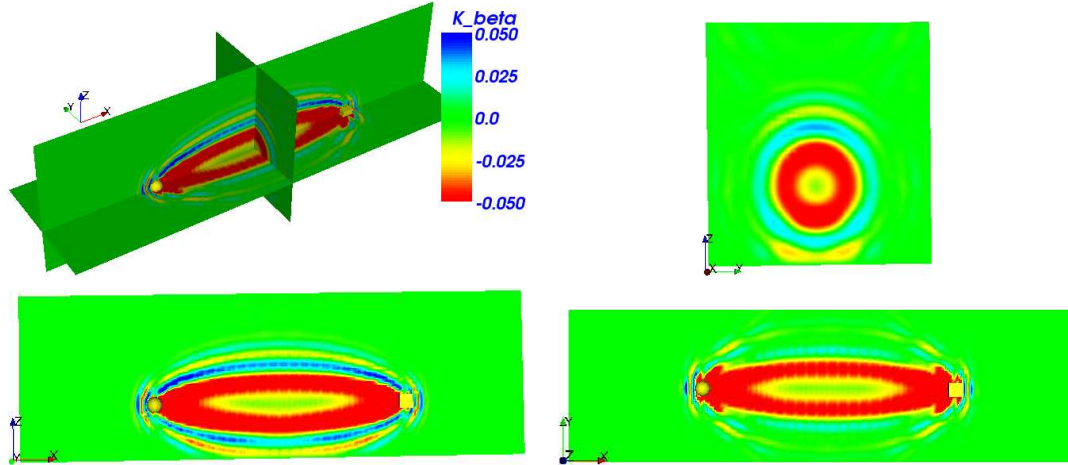


Figure 5.4: S sensitivity kernel K_β for the ray geometry shown in Figure 5.2. Top left: combined vertical and horizontal cross-sections through the source and receiver illustrating the hollow cigar-shaped kernel. Top right: vertical cross-section perpendicular to the middle of the source-receiver line illustrating the "doughnut hole" in the middle of the kernel. Bottom left: vertical cross-section through the source and receiver showing the cigar shape of the kernel. Bottom right: horizontal cross-section through the source and receiver.

5.6.1 S Kernel

The S-wave Fréchet kernel K_β , given by (5.31), exhibits a distinctive hollow cigar shape, as shown in Figure 5.4. Note the minimal sensitivity along the geometrical ray path as dictated by the hole in the cigar. Note also that the kernel has negative sensitivity some distance off the ray path, as indicated by the blue ring, implying that a positive anomaly off the ray path can cause an actual advance in the finite-frequency (cross-correlation) traveltime. The width of the related first Fresnel zone is approximately given by $\sqrt{\lambda L}$ [Dahlen *et al.*, 2000], where λ denotes the wavelength and L the distance between the source and the receiver.

5.6.2 SS Kernel

Similar to the S phase, the K_β Fréchet kernel for the SS arrival delineates its geometrical ray path, as illustrated in Figure 5.5. The kernel displays nearly zero sensitivity kernels along the geometrical ray path, except near the surface reflection point where the two legs of SS fold on top of each other. Besides the expected sensitivity along the geometric ray path, an ellipsoidal locus of points of diffraction shows up faintly in the source-receiver

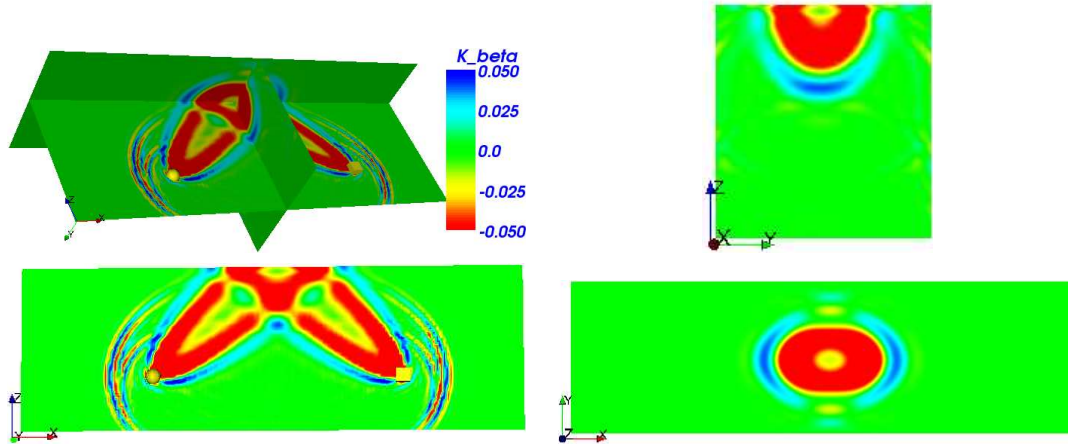


Figure 5.5: SS sensitivity kernel K_β for the ray geometry shown in Figure 5.2. Top left: combined vertical and horizontal cross-sections through the source and receiver looking up to the free surface. Top right: vertical cross-section perpendicular to the middle of the source-receiver line illustrating the "doughnut hole" in the middle of the kernel. Bottom left: vertical cross-section showing the "folded-over cigar" shape of the kernel. Bottom right: horizontal cross-section at the surface.

vertical cross-section, delineating the points that have the same traveltime as the SS arrival. This illustrates one of the differences between calculating sensitivity kernels based upon an adjoint method versus an asymptotic ray-based calculation: The adjoint kernels involve all possible regions of sensitivity that contribute to the arrival of interest, while a ray-based sensitivity kernel calculation will only pick out the sensitivities along the geometrical ray path. Note, however, that the kernel oscillates rapidly in the locus of diffraction points, and this tends to average out the associated traveltime anomalies.

5.6.3 P Kernel

The Fréchet kernel K_α for the P phase, given by (5.32) and shown in Figure 5.6, looks very similar to that of the S phase (see Figure 5.4), except, because of the longer wavelength of the P phase, the width of the Fresnel zone is larger than that of the S phase, in accordance with the scaling relation $\text{width} \sim \sqrt{\lambda L}$.

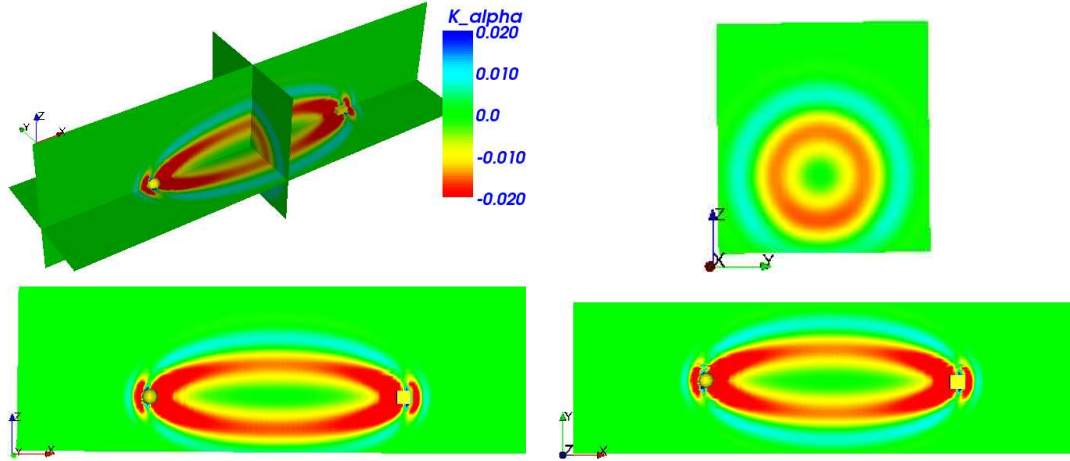


Figure 5.6: P sensitivity kernel K_α for the ray geometry shown in Figure 5.2. Top left: combined vertical and horizontal cross-section through the source and receiver. Top right: vertical cross-section perpendicular to the middle of the source-receiver line illustrating the doughnut hole in the middle of the kernel. Bottom left: vertical cross-section showing the cigar shape of the kernel. Bottom right: horizontal cross-section through the source and receiver.

5.6.4 PS+SP Kernel

The left column in Figure 5.7 shows the Fréchet kernels for the PS phase in terms of density K_ρ , shear modulus K_μ , and bulk modulus K_κ , whereas the right column in the figure shows the Fréchet kernels parameterized terms of density K_ρ , S-wave speed K_β , and P-wave speed K_α . Notice that although the density kernel K_ρ defined by (5.21) shows a strong negative sensitivity when we use a model parameterization in terms of density, shear modulus μ , and bulk modulus κ , the density kernel $K_{\rho'}$ given by (5.30) — corresponding to a parameterization in terms of density, shear-wave speed β , and compressional-wave speed α — is practically zero. This reflects the fact that the traveltime is controlled by the wave speed and not the density. Notice that the Fréchet kernel for the P-wave speed is most pronounced along the P legs of the PS+SP ray path, while the kernel for the S-wave speed is mainly sensitive around the S legs of the PS+SP ray path. This example illustrates that the adjoint approach can be used to invert waveforms that consist of multiple arrivals, because the resulting kernel will clearly reflect the main contributions to the waveform.

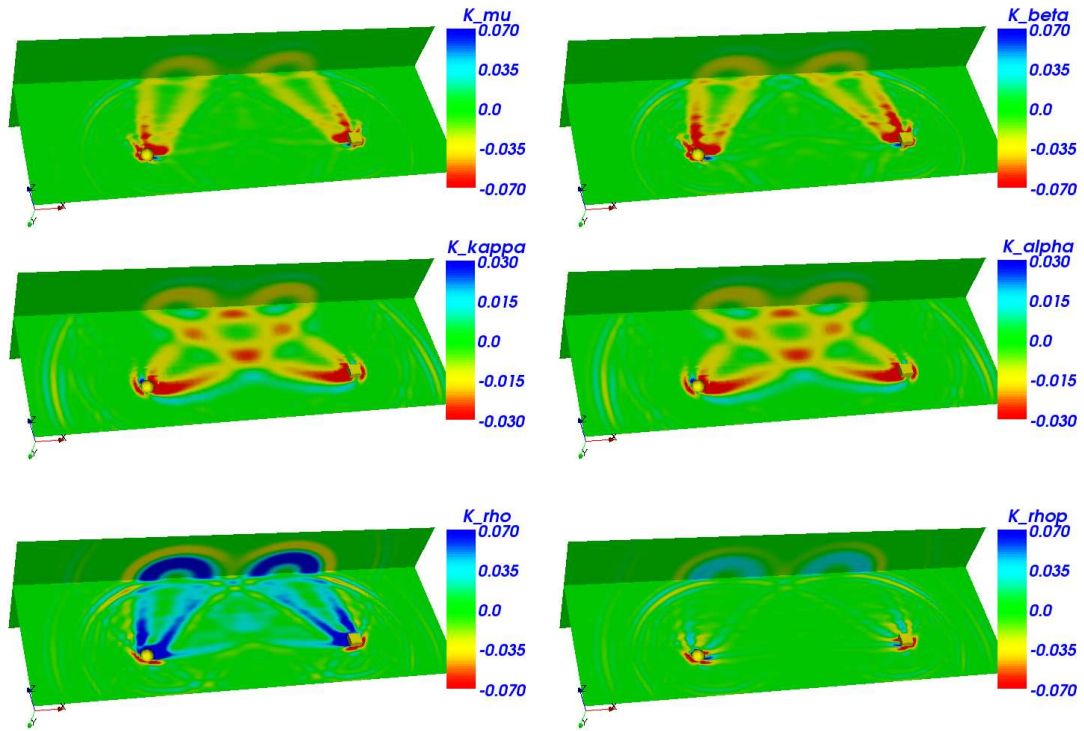


Figure 5.7: PS+SP sensitivity kernels for the ray geometry shown in Figure 5.2. Top left: K_μ sensitivity kernel defined by (5.26). This kernel reflects the S legs of the PS+SP arrival. Middle left: K_κ sensitivity kernel defined by (5.27). This kernel reflects the P legs of the PS+SP arrival. Bottom left: K_ρ sensitivity kernel defined by (5.21). This kernel shows large negative values related to the model parameterization in terms of density, shear modulus μ , and bulk modulus κ . Top right: K_β sensitivity kernel defined by (5.31). This kernel reflects the S legs of the PS+SP arrival. Middle right: K_α sensitivity kernel defined by (5.32). This kernel reflects the P legs of the PS+SP arrival. Bottom right: $K_{\rho'}$ sensitivity kernel defined by (5.30). This kernel shows hardly any sensitivity to density perturbations because we are measuring traveltime anomalies and the model is parameterized in terms of density, shear-wave speed β , and compressional-wave speed α . The nearly-zero $K_{\rho'}$ reflects the fact that the traveltime is not affected by density but rather by wave speed.

5.7 Application to a 3-D Southern California Model

In this section we use the adjoint method to calculate finite-frequency sensitivity kernels for a 3-D integrated southern California model. The model consists of a detailed Los Angeles basin model developed by *Süss and Shaw* [2003] embedded in the *Hauksson* [2000] regional tomographic background model. This model was evaluated extensively by *Komatitsch et al.* [2004] and is currently used for centroid moment-tensor inversions based upon the 3-D source inversion technique introduced by *Liu et al.* [2004]. The synthetics produced by this model generally fit the data reasonably well at periods of 5 s and longer throughout the entire region, and up to 2 s and longer within the Los Angeles basin.

We will be studying Fréchet kernels for body- and surface-wave arrivals generated by the September 3, 2002, Yorba Linda earthquake, which occurred at a depth of 6.8 km [*Liu et al.*, 2004]. Figure 3.1 shows a topographic map of southern California with major late Quaternary faults indicated by black lines. The blue boxes indicate the detailed Los Angeles basin model developed by *Süss and Shaw* [2003] and the Salton Trough model developed by *Lovely et al.* [2006], which are embedded in the *Hauksson* [2000] regional model. Red triangles indicate the stations for which body-wave and surface-wave kernels are presented later.

5.7.1 P Kernels

3-D P-wave Fréchet kernels and corresponding model cross-sections are shown in Figure 5.8 for stations DLA and OLP. The P arrival at station DLA, which is about 32 km from the epicenter, consists primarily of the direct P wave, and therefore its sensitivity kernel shows the characteristic, simple banana-doughnut shape with some minor variations caused by 3-D heterogeneity. In stark contrast, the P arrival at station OLP, at an epicentral distance of 165 km, is the P_{nl} wave train, i.e., the combination of the P_n and PL phases [*Helmberger and Engen*, 1980]. Most of the P_{nl} sensitivity is along the P_n ray path, which dives down from the source to the Moho, runs along the Moho, and then comes up to the receiver. The magnitude of the sensitivity kernel is largest along the upgoing and downgoing legs of the ray path and relatively small along the refracted portion the ray path. Notice

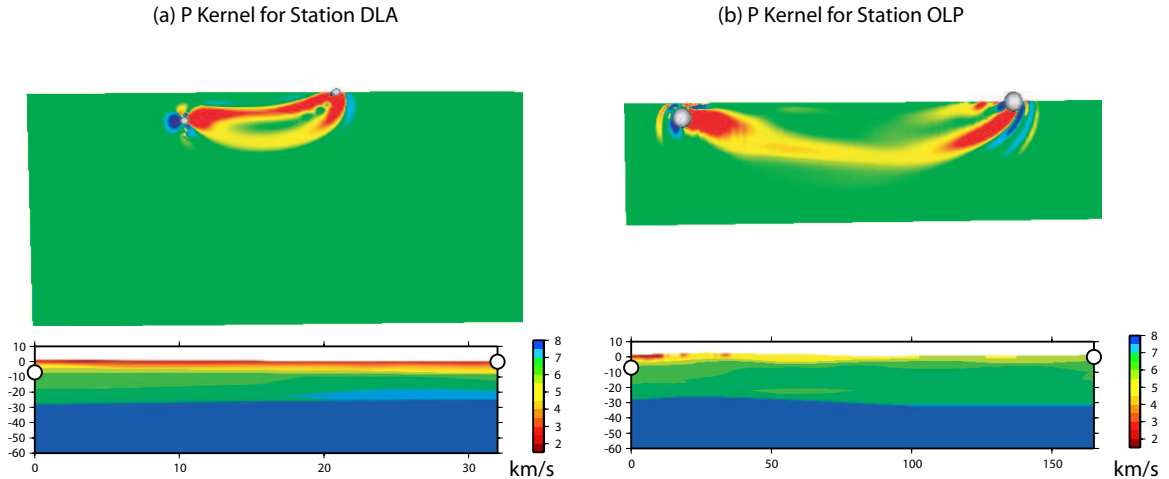


Figure 5.8: (a) Top Panel: vertical source-receiver cross-section of the P-wave finite-frequency sensitivity kernel K_α at station DLA station at an epicentral distance of 32 km from the September 3, 2002, Yorba Linda earthquake. Red colors denote negative sensitivity while blue colors denote positive sensitivity. The locations of the source and receiver are indicated by white circles. At this relatively short epicentral distance the kernel looks like a classical "banana-doughnut" kernel. Lower Panel: vertical source-receiver cross-section of the 3-D P-wave velocity model used for the spectral-element simulations [Komatitsch *et al.*, 2004]. The locations of the source and receiver are indicated by the white circles. (b) The same as (a) but for station OLP at an epicentral distance of 165 km. At this relatively larger distance the P-wave kernel reflects a combination of the P_n and PL waves.

in the model cross-section that the Moho slopes towards the receiver, which is reflected in the sloping sensitivity kernel. Another noticeable feature is near-surface sensitivity to the left of the receiver, indicating body-to-Rayleigh-wave conversions. This example illustrates that fully 3-D numerical methods must be used in the construction of finite-frequency sensitivity kernels for complicated Earth models.

5.7.2 S Kernels

Three-dimensional S-wave sensitivity kernels for stations GSC and HEC are shown in Figure 5.9. Because it is more difficult to isolate a clean S arrival for the adjoint source compared to the P wave, the S kernels are generally not as clean and sharp as those for the P wave. At station GSC, at an epicentral distance of 176 km, the Moho-reflected SmS phase and the Moho-refracted Sn phase arrive very close to each other. Therefore, the kernel for

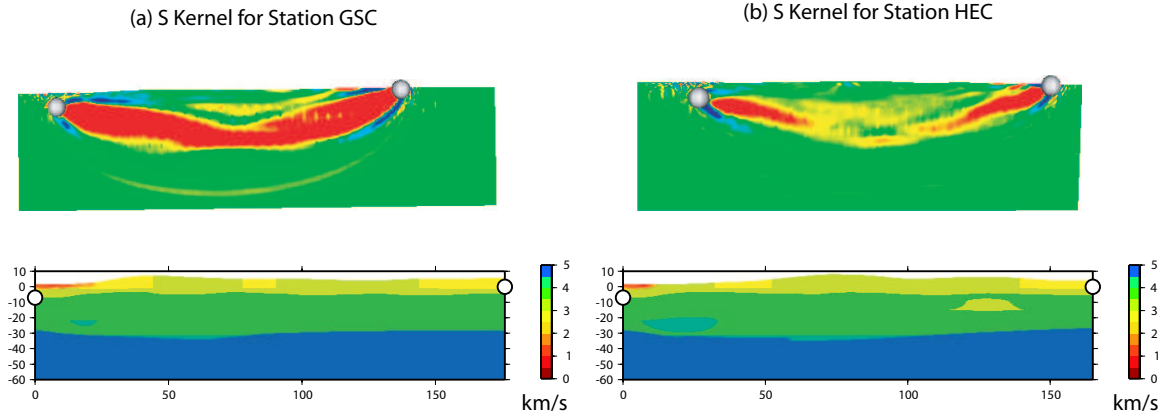


Figure 5.9: (a) Top Panel: vertical source-receiver cross-section of the S-wave finite-frequency sensitivity kernel K_β for station GSC at an epicentral distance of 176 km from the September 3, 2002, Yorba Linda earthquake. Lower Panel: vertical source-receiver cross-section of the 3-D S-wave velocity model used for the spectral-element simulations [Komatitsch *et al.*, 2004]. (b) The same as (a) but for station HEC at an epicentral distance of 165 km.

the S wave includes contributions from both phases, which cannot be separated from each other. At station HEC, at an epicentral distance of 165 km, the S-wave kernel is mainly composed of the S_n phase plus some reflected phases from internal layers between the surface and the Moho. This kernel serves as an illustration of one of the main advantages of computing sensitivity kernels using adjoint methods: By simply selecting a waveform of interest in the seismogram as the adjoint source we automatically determine all the structural parameters that affected it without prior knowledge of the contributing phases or their ray paths.

5.7.3 Surface-wave Kernels

Unlike the homogeneous model with buried source and receiver shown in Figure 5.2, the 3-D southern California model generates surface waves along the free surface. These waves are mostly sensitive to near-surface structure between the source and the receiver, as illustrated in Figure 5.10 for Rayleigh and Love waves recorded at station HEC at an epicentral distance of 165 km from the September 3, 2002, Yorba Linda earthquake. Notice the large 3-D variations on the surface along the path connecting the surface projections

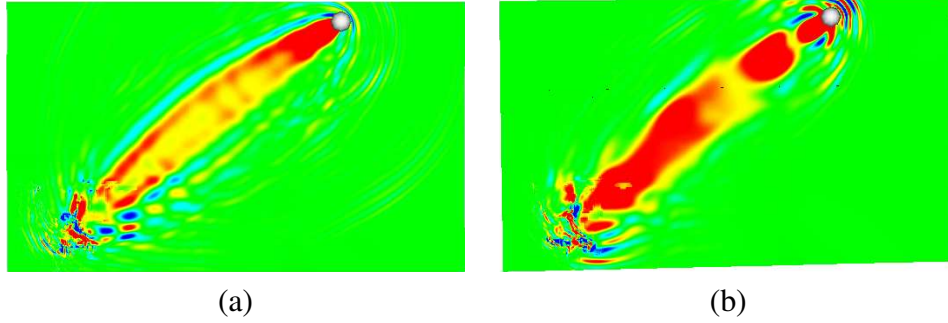


Figure 5.10: (a) Map view of the finite-frequency K_α kernel for the Rayleigh wave at station HEC (indicated by the white circle) generated by the September 3, 2002, Yorba Linda earthquake. (b) Map view of the finite-frequency K_β kernel for the Love wave at HEC for the same earthquake.

of the source and receiver due to both topographic and wave-speed variations. Although the kernels shown in Figure 5.10 are simple finite-frequency surface-wave kernels, it is straightforward to make frequency-dependent phase and amplitude measurement for these wave trains and generate Fréchet sensitivity kernels for individual phase measurements, as discussed by *Zhou et al.* [2004].

5.8 Tomographic Inversions

The adjoint approach introduced in this paper may be used to relate changes in a misfit function, $\delta\chi$, to relative model perturbations, $\delta \ln m = \delta m/m$, through a general relationship of the form

$$\delta\chi = \int K_m(\mathbf{x}) \delta \ln m(\mathbf{x}) d^3\mathbf{x}, \quad (5.36)$$

where K_m denotes the associated 3-D sensitivity or Fréchet kernel, which is also weighted banana-doughnut kernels, with weights determined by the measurements, e.g., cross-correlation traveltime anomalies [*Tromp et al.*, 2005].

It is important to recognize that in the adjoint approach we do not need to calculate individual banana-doughnut kernels for each measurement. If N_{events} denotes the number of earthquakes, N_{stations} the number of stations, and N_{picks} the number of measurements at that station, such an approach would require $N_{\text{events}} \times N_{\text{stations}} \times N_{\text{picks}}$ simulations, i.e., one

simulation for each banana-doughnut kernel corresponding to one pick. For a given earthquake, the adjoint approach is to measure as many arrivals as possible on three components at all available stations. Ideally, every component at every station will have a number of arrivals suitable for measurement, for example in terms of frequency-dependent phase and amplitude anomalies. During the adjoint simulation, each component of every receiver will transmit its measurements in reverse time, and the interaction of the so-generated adjoint wavefield with the forward wavefield results in a misfit kernel for that particular event. This "event kernel" is essentially a sum of weighted banana-doughnut kernels, with weights determined by the traveltimes anomaly, and is obtained based upon only two 3-D simulations. By summing these event kernels one obtains the "summed event kernel", which highlights where the current 3-D model is inadequate and enables one to obtain an improved Earth model, for example based upon a conjugate-gradient approach. The number of 3-D simulations at each conjugate-gradient step scales linearly with the number of earthquakes N_{events} but is independent of the number of receivers N_{stations} or the number of measurements N_{picks} . Every iteration of the conjugate-gradient method requires one forward and one adjoint calculation for the current model to obtain the value of the misfit function and its gradient, and one forward simulation for a "trial" model, i.e., a model in the direction of the gradient, to obtain the value of the misfit function at this trial location. A quadratic polynomial may then be used to determine the minimum of the misfit function in the search direction, which forms the starting point of the next iteration. Thus one conjugate-gradient iteration requires a total of three spectral-element simulations. Alternatively, at the cost of one more adjoint simulation, one may choose to evaluate both the misfit function and its gradient at the trial location. In that case a cubic polynomial may be used to determine the minimum of the misfit function in the search direction, and a total of four spectral-element simulations is required. Both conjugate-gradient approaches are discussed in detail for 2-D problems in *Tape et al.* [2006].

5.9 Conclusions

Based upon a Lagrange multiplier technique, we have developed and implemented an adjoint method for the calculation of finite-frequency sensitivity kernels for 3-D reference Earth models. We have demonstrated that such Fréchet kernels may be obtained based upon just two 3-D simulations: one forward simulation to determine the current fit of the synthetic seismograms to the data, and a second, adjoint simulation in which a measurement of the remaining differences between the data and the synthetics is used in reverse time to generate a wavefield that originates at the receiver(s). The interaction between the regular and adjoint wavefields determines the sensitivity kernels.

The main advantages of our adjoint approach are five-fold. First, the kernels are calculated on-the-fly by carrying the adjoint wavefield and the regular wavefield in memory at the same time. This doubles the memory requirements for the simulation but avoids the storage of Green's functions for all events and stations as a function of space and time. One only has to store the final frame of the forward simulation plus the wavefield that is absorbed on the artificial boundaries of the domain. (At the scale of the globe there are no absorbing boundaries and thus one only needs to store the final frame of the forward simulation.) Second, the kernels can be calculated for fully 3-D reference models, something that is critical in highly heterogeneous settings, e.g., in regional seismology or exploration geophysics. Third, the approach scales linearly with the number of earthquakes but is independent of the number of receivers and the number of arrivals that is used in the inversion. Thus one should use all available stations and make as many measurements as possible. Fourth, any time segment where the data and the synthetics have significant amplitudes and match reasonably well is suitable for a measurement. One does not need to be able to label the phase, e.g., identify it as SS or SSS, because the adjoint simulation will reveal how this particular measurement "sees" the Earth model, and the resulting 3-D sensitivity kernel will reflect this view. Finally, the cost of the simulation is independent of the number of model parameters, i.e., one can consider fully anisotropic Earth models with 21 elastic parameters for practically the same numerical cost as an isotropic simulation involving just two parameters.

In the near future we plan to use the adjoint method developed and implemented here to perform 3-D tomographic inversions.

Chapter 6

Finite-frequency Kernels Based upon Adjoint Method for Global Wave Simulations

6.1 Abstract

We derive the complete set of adjoint equations and the expressions of the Fréchet sensitivity kernels based upon the Lagrange multiplier methods. We start from the global equations of motion for a rotating, self-gravitating Earth model initially in hydrostatic equilibrium, and deduce the corresponding adjoint equations that involve motions on an Earth model that rotates in the opposite direction. The variations of the misfit function χ can be expressed as $\delta\chi = \int_V K_m(\mathbf{x}) \delta \ln m(\mathbf{x}) d^3\mathbf{x} + \int_\Sigma K_h(\mathbf{x}) \delta h(\mathbf{x}) d^2\mathbf{x} + \int_{\Sigma_{FS}} \mathbf{K}_h(\mathbf{x}) \cdot \nabla^\Sigma \delta h(\mathbf{x}) d^2\mathbf{x}$, where $\delta \ln m = \delta m/m$ denotes the relative model perturbations in the earth volume V , Σ denotes the solid-solid or fluid-solid boundaries, K_h denotes the kernels associated with topography variations on Σ , while Σ_{FS} denotes the fluid-solid boundary, ∇^Σ denotes the surface gradient, and \mathbf{K}_h denotes the associated 2-D kernel. Anelasticity also fits in well with the framework of the adjoint methods and the Lagrange multiplier methods. We calculate these 3-D and 2-D finite-frequency kernels by simultaneously computing the adjoint wavefield \mathbf{s}^\dagger forward in time and reconstructing the regular wavefield \mathbf{s} backward in time. Both the forward and adjoint simulations of wave propagation are accomplished by the spectral element methods. We apply the adjoint methods to generate finite-frequency travel time Fréchet kernels for global seismic phases (P, PKP, S, SKS, depth phases, surface-reflected

phases, etc.) in both the 1-D PREM model and 3-D tomographic models. These kernels generally agree well with the results obtained from ray-based finite frequency methods; however, the adjoint methods do not have the same theoretical limitations, and can produce sensitivity kernels for any given phase in any 3-D initial velocity model. The Fréchet kernels illustrate the sensitivity of the observations to the structural parameters and the topography of the internal discontinuities, which forms the basis of full 3-D tomographic inversions. Most of the results from this chapter is collected in *Liu and Tromp* [2006b].

6.2 Introduction

Global tomography studies have been around for more than twenty years, during which more and more images of the velocity structures of the earth's crust and mantle have been produced with the expansion of global and regional seismic networks and increasing number of seismic records [e.g., *Woodhouse and Dziewonski*, 1984; *Su et al.*, 1994; *Masters et al.*, 1996; *Grand*, 1994; *Li and Romanowica*, 1996; *der Hilst et al.*, 1997; *Ritsema and Van Heijst*, 2000]. These images of the earth's deep interior provide important information for the understanding of the earth's internal geodynamic processes. Recently more efforts have been made to further improve the theoretical basis of the tomographic inversions by replacing the simple Fréchet sensitivity kernels calculated from either the conventional ray path or other asymptotically theories with the 3-D volumetric Fréchet kernels that relate the observations to the 3-D volumetric perturbations of the structural model parameters. These include sensitivity kernels that are produced based upon the surface-wave Green's function [*Marquering et al.*, 1999], body-wave ray theory (also known as the banana-doughnut kernels, *Dahlen et al.* [2000]; *Hung et al.* [2000]; *Zhou et al.* [2004]), surface-wave ray theory [*Zhou et al.*, 2004], and normal-model theory [*Zhao et al.*, 2000; *Zhao and Jordan*, 2006]. Global tomographic inversions based upon these 3-D sensitivity kernels have also been attempted [*Montelli et al.*, 2004; *Zhou et al.*, 2005].

Using the existing 3-D tomography models as initial models to further invert and update the earth's velocity structure is a big challenge in seismology since it involves fully 3-D numerical simulation of seismic wave propagation. *Tromp et al.* [2005] introduced

the adjoint formulations of the Fréchet kernels, which involves the interactions between the regular forward wave field and the adjoint wave field produced by time-reversed signals at the receivers as simultaneous, fictitious sources. The main advantage is that only two simulations are needed to compute the gradient of the misfit function for each event. Because it is computationally costly to calculate the Hessian of the misfit function, a conjugate gradient algorithm is introduced to optimize the misfit function and obtain the best velocity model. The application of the adjoint method to calculate the 3-D Fréchet sensitivity kernels for the regional 3-D velocity models has been demonstrated in *Liu and Tromp* [2006a].

Following *Liu and Tromp* [2006a], in this paper, we derive the complete set of adjoint equations and expressions of the Fréchet kernels for the global seismic wave propagation problem based upon the Lagrange multiplier method. We apply the adjoint methods to numerically compute the 3-D travel-time sensitivity kernels for typical seismic phases of both 1-D and 3-D global velocity models. For a full 3-D tomographic inversion, one event kernel is computed for all the travel-time measurements made at all the receivers for this particular event. The summed event kernels then form the gradient of the misfit function, and they will be used in the successive velocity model updates based upon the conjugate gradient method.

6.3 Equation of Motion

We begin by introducing the equations of motion for a rotating, self-gravitating Earth model that is initially in hydrostatic equilibrium. In such an Earth model, pressure gradients are balanced by gradients of the geopotential, which is the sum of the gravitational and centrifugal potentials. As a consequence, the equilibrium model is an ellipsoid in which surfaces of constant pressure, density, and the geopotential coincide. The results in this section are well known [e.g., *Dahlen and Tromp*, 1998], but serve the purpose of introducing the necessary framework and the related notation.

The gravitational potential Φ is determined by Poisson's equation:

$$\nabla^2\Phi = 4\pi G\rho, \quad (6.1)$$

where ρ denotes the density distribution and G the gravitational constant. For mathematical and notational convenience the density distribution is defined to be zero outside of the Earth model, where the gravitational potential satisfies Laplace's equation. The Earth model contains internal solid-solid discontinuities, collectively denoted by Σ_{SS} , and internal fluid-solid discontinuities, collectively denoted by Σ_{FS} . The collection of all internal solid-solid and fluid-solid discontinuities plus the surface of the model volume ∂V is denoted by $\Sigma = \partial V \cup \Sigma_{SS} \cup \Sigma_{FS}$. Poisson's equation (6.1) must be solved subject to the boundary conditions

$$[\Phi]_{\pm}^{\pm} = 0, \quad [\hat{\mathbf{n}} \cdot \nabla\Phi]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma, \quad (6.2)$$

where the notation $[\cdot]_{\pm}^{\pm}$ denotes the jump in the enclosed quantity when going from the outward (+) side to the inward (−) side of the discontinuity. The unit normal $\hat{\mathbf{n}}$ points outward from the − side to the + side of a discontinuity. The solution to the boundary value problem (6.1)–(6.2) is

$$\Phi(\mathbf{r}) = -G \int_V \frac{\rho(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|} d^3\mathbf{r}', \quad (6.3)$$

where V denotes the volume of the Earth model and \mathbf{r} the position vector.

The centrifugal potential ψ is defined in terms of the Earth's angular rotation vector $\boldsymbol{\Omega}$ by

$$\psi = -\frac{1}{2}[\Omega^2 r^2 - (\boldsymbol{\Omega} \cdot \mathbf{r})^2], \quad (6.4)$$

where $r = \|\mathbf{r}\|$ denotes the radius and $\Omega = \|\boldsymbol{\Omega}\|$ is the angular rotation rate.

The displacement field \mathbf{s} in a rotating, self-gravitating Earth model is governed by the equation of motion

$$\rho(\partial_t^2\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t\mathbf{s}) - \nabla \cdot \mathbf{T} + \nabla[\rho\mathbf{s} \cdot \nabla(\Phi + \psi)] + \rho\nabla\phi - \nabla \cdot (\rho\mathbf{s})\nabla(\Phi + \psi) = \mathbf{f}. \quad (6.5)$$

In an anisotropic, elastic Earth model the stress \mathbf{T} is defined in terms of the displacement gradient by

$$\mathbf{T} = \mathbf{c} : \nabla \mathbf{s}, \quad (6.6)$$

where \mathbf{c} denotes the elastic tensor. The complications associated with an anelastic Earth model will be addressed in section 6.5. On all boundaries the traction $\hat{\mathbf{n}} \cdot \mathbf{T}$ needs to be continuous:

$$[\hat{\mathbf{n}} \cdot \mathbf{T}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma. \quad (6.7)$$

For mathematical convenience, the Earth model parameters ρ and \mathbf{c} are defined to be zero outside of the model volume V , such that at the free surface ∂V the boundary condition (6.7) reduces to the vanishing of the traction: $\hat{\mathbf{n}} \cdot \mathbf{T} = \mathbf{0}$. On a fluid-solid boundary the traction is normal: $\hat{\mathbf{n}} \cdot \mathbf{T} = (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. On all solid-solid boundaries the displacement \mathbf{s} must be continuous:

$$[\mathbf{s}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma_{\text{SS}}, \quad (6.8)$$

whereas on fluid-solid discontinuities there may be slip:

$$[\hat{\mathbf{n}} \cdot \mathbf{s}]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma_{\text{FS}}. \quad (6.9)$$

The displacement is subject to the initial conditions

$$\mathbf{s}(\mathbf{r}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}(\mathbf{r}, 0) = \mathbf{0}. \quad (6.10)$$

The perturbation in the gravitational potential ϕ induced by the displacement \mathbf{s} is determined by

$$\nabla \cdot \xi = 0, \quad (6.11)$$

$$[\phi]_{\pm}^{\pm} = 0, \quad [\hat{\mathbf{n}} \cdot \xi]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma, \quad (6.12)$$

where

$$\xi = (4\pi G)^{-1} \nabla \phi + \rho \mathbf{s}. \quad (6.13)$$

The solution to the boundary-value problem (6.11)–(6.12) is

$$\phi(\mathbf{r}, t) = -G \int_V \rho(\mathbf{r}') \frac{\mathbf{s}(\mathbf{r}', t) \cdot (\mathbf{r} - \mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|^3} d^3\mathbf{r}'. \quad (6.14)$$

6.4 Adjoint Equations

In this section we derive the adjoint equations that arise in the context of tomographic inverse problems. Our objective will be to minimize the difference between observed three-component seismograms and the corresponding synthetic seismograms. This may be accomplished based upon a wide variety of misfit criteria, e.g., waveform or cross-correlation traveltimes differences. In this paper we consider the minimization of a waveform misfit function χ subject to the constraint that the synthetic seismograms are determined by the equations of motion (6.5)–(6.12):

$$\begin{aligned} \chi = & \frac{1}{2} \sum_r \int_0^T [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)]^2 dt \\ & - \int_0^T \int_V \boldsymbol{\lambda} \cdot \{ \rho(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) - \nabla \cdot \mathbf{T} \nabla [\rho \mathbf{s} \cdot \nabla(\Phi + \psi)] + \rho \nabla \phi \\ & \quad - \nabla \cdot (\rho \mathbf{s}) \nabla(\Phi + \psi) - \mathbf{f} \} d^3\mathbf{r} dt \\ & + \int_0^T \int_{\circlearrowleft} \mu(\nabla \cdot \boldsymbol{\xi}) d^3\mathbf{r} dt. \end{aligned} \quad (6.15)$$

Here \circlearrowleft denotes all of space, and the vector Lagrange multiplier $\boldsymbol{\lambda}(\mathbf{r}, t)$ and the scalar Lagrange multiplier $\mu(\mathbf{r}, t)$ remain to be determined. The interval $[0, T]$ denotes the time series of interest, $\mathbf{s}(\mathbf{r}_r, t)$ denotes the synthetic displacement field at receiver location \mathbf{r}_r , and $\mathbf{d}(\mathbf{r}_r, t)$ denotes the corresponding observed three-component displacement vector. In practice, both the data \mathbf{d} and the synthetics \mathbf{s} will be windowed, filtered, and possibly weighted on the time interval $[0, T]$. In what follows we will implicitly assume that such operations have been performed. We note that different measures of misfit simply give rise to different adjoint sources [Tromp *et al.*, 2005].

Upon taking the variation of the action (6.15) we obtain

$$\begin{aligned}
\delta\chi = & \int_0^T \int_V \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)] \delta(\mathbf{r} - \mathbf{r}_r) \cdot \delta\mathbf{s}(\mathbf{r}, t) \, d\mathbf{r}^3 \, dt \\
& - \int_0^T \int_V \boldsymbol{\lambda} \cdot \{ \rho(\partial_t^2 \delta\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \delta\mathbf{s}) + \delta\rho(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) - \nabla \cdot \delta\mathbf{T} \\
& \quad + \nabla[\rho \delta\mathbf{s} \cdot \nabla(\Phi + \psi)] + \nabla[\delta\rho \mathbf{s} \cdot \nabla(\Phi + \psi)] + \nabla[\rho \mathbf{s} \cdot \nabla(\delta\Phi)] + \rho \nabla \delta\phi + \delta\rho \nabla \phi \\
& \quad - \nabla \cdot (\rho \delta\mathbf{s}) \nabla(\Phi + \psi) - \nabla \cdot (\delta\rho \mathbf{s}) \nabla(\Phi + \psi) - \nabla \cdot (\rho \mathbf{s}) \nabla(\delta\Phi) - \delta\mathbf{f} \} \, d^3\mathbf{r} \, dt \\
& + \int_0^T \int_{\mathcal{O}} \mu(\nabla \cdot \delta\xi) \, d^3\mathbf{r} \, dt, \tag{6.16}
\end{aligned}$$

The perturbation in stress $\delta\mathbf{T}$ may be obtained by perturbing Hooke's law (6.6):

$$\delta\mathbf{T} = \mathbf{c} : \nabla \delta\mathbf{s} + \delta\mathbf{c} : \nabla \mathbf{s}, \tag{6.17}$$

and the perturbation $\delta\xi$ is obtained by perturbing (6.13):

$$\delta\xi = (4\pi G)^{-1} \nabla \delta\phi + \rho \delta\mathbf{s} + \delta\rho \mathbf{s}. \tag{6.18}$$

The perturbation in the gravitational potential $\delta\Phi$ associated with a perturbation in density $\delta\rho$ and a displacement δd of the discontinuities Σ is determined by [Dahlen and Tromp, 1998]

$$\nabla^2(\delta\Phi) = 4\pi G \delta\rho, \tag{6.19}$$

$$[\delta\Phi]_{\pm}^{\pm} = 0, \quad [\hat{\mathbf{n}} \cdot \nabla(\delta\Phi) + 4\pi G \rho \delta d]_{\pm}^{\pm} \quad \text{on } \Sigma, \tag{6.20}$$

where the density perturbation $\delta\rho$ is defined to be zero outside of the volume of the Earth model V . The solution to the boundary-value problem (6.19)–(6.20) is

$$\delta\Phi(\mathbf{r}) = -G \int_V \frac{\delta\rho(\mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|} \, d^3\mathbf{r}' + G \int_{\Sigma} \frac{\delta d(\mathbf{r}') [\rho(\mathbf{r}')]_{\pm}^{\pm}}{\|\mathbf{r} - \mathbf{r}'\|} \, d^2\mathbf{r}', \tag{6.21}$$

i.e., the perturbation in the gravitational potential $\delta\Phi$ is completely determined by the perturbations in density $\delta\rho$ and boundary topography δd .

The next step in determining the adjoint equations involves manipulating the variation

of the action (6.16) by performing integrations by parts over time and repeatedly using Gauss's theorem for a volume with discontinuities:

$$\int_V \nabla \cdot \mathbf{v} \, d^3\mathbf{r} = - \int_{\Sigma} [\hat{\mathbf{n}} \cdot \mathbf{v}]_{\pm}^{\pm} \, d^2\mathbf{r}, \quad (6.22)$$

for any tensor \mathbf{v} . In an Earth model without any internal discontinuities this reduces to the familiar form

$$\int_V \nabla \cdot \mathbf{v} \, d^3\mathbf{r} = \int_{\partial V} \hat{\mathbf{n}} \cdot \mathbf{v} \, d^2\mathbf{r}. \quad (6.23)$$

After a significant amount of tedious algebra we can rewrite the variation of the action (6.16) in the form

$$\begin{aligned} \delta\chi = & \int_0^T \int_V \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)] \delta(\mathbf{r} - \mathbf{r}_r) \cdot \delta\mathbf{s}(\mathbf{r}, t) \, d\mathbf{r}^3 \, dt \\ & - \int_0^T \int_V \delta\mathbf{s} \cdot \{ \rho(\partial_t^2 \boldsymbol{\lambda} + 2\boldsymbol{\Omega} \times \partial_t \boldsymbol{\lambda}) - \nabla \cdot \boldsymbol{\sigma} + \nabla[\rho \boldsymbol{\lambda} \cdot \nabla(\Phi + \psi)] \\ & \quad + \rho \nabla \mu - \nabla \cdot (\rho \boldsymbol{\lambda}) \nabla(\Phi + \psi) \} \, d^3\mathbf{r} \, dt + \int_0^T \int_{\mathcal{O}} \delta\phi(\nabla \cdot \boldsymbol{\zeta}) \, d^3\mathbf{r} \, dt \\ & - \int_V \rho [\boldsymbol{\lambda} \cdot \partial_t \delta\mathbf{s} - \partial_t \boldsymbol{\lambda} \cdot \delta\mathbf{s} + 2\boldsymbol{\lambda} \cdot \boldsymbol{\Omega} \times \delta\mathbf{s}]_0^T \, d^3\mathbf{r} \\ & - \int_0^T \int_V \{ \delta\rho [\boldsymbol{\lambda} \cdot (\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) + \boldsymbol{\lambda} \cdot \nabla\phi + \mathbf{s} \cdot \nabla\mu + \mathbf{s} \cdot \nabla\nabla(\Phi + \psi) \cdot \boldsymbol{\lambda} \\ & \quad + \nabla(\Phi + \psi) \cdot (\mathbf{s} \cdot \nabla\boldsymbol{\lambda} - \mathbf{s}\nabla \cdot \boldsymbol{\lambda})] + \nabla\boldsymbol{\lambda} : \delta\mathbf{c} : \nabla\mathbf{s} + \rho\mathbf{s} \cdot \nabla\nabla(\delta\Phi) \cdot \boldsymbol{\lambda} \\ & \quad + \rho\nabla(\delta\Phi) \cdot (\mathbf{s} \cdot \nabla\boldsymbol{\lambda} - \mathbf{s}\nabla \cdot \boldsymbol{\lambda}) - \boldsymbol{\lambda} \cdot \delta\mathbf{f} \} \, d^3\mathbf{r} \, dt \\ & - \int_0^T \int_{\Sigma} [\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \boldsymbol{\lambda} - \hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \cdot \delta\mathbf{s} + \hat{\mathbf{n}} \cdot \delta\xi \mu - \hat{\mathbf{n}} \cdot \boldsymbol{\zeta} \delta\phi]_{\pm}^{\pm} \, d^2\mathbf{r} \, dt, \end{aligned} \quad (6.24)$$

where we have introduced the notation $[f]_0^T = f(T) - f(0)$, for any function f , and where we have defined

$$\boldsymbol{\sigma} = \mathbf{c} : \nabla\boldsymbol{\lambda}, \quad (6.25)$$

and

$$\boldsymbol{\zeta} = (4\pi G)^{-1} \nabla\mu + \rho\boldsymbol{\lambda}. \quad (6.26)$$

In the absence of model parameter perturbations $\delta\rho$, $\delta\mathbf{c}$, δd , and $\delta\mathbf{f}$, the variation in

the action (6.24) is stationary with respect to permissible (i.e., honoring the appropriate boundary and initial conditions) perturbations $\delta\mathbf{s}$ and $\delta\phi$ provided the vector Lagrange multiplier $\boldsymbol{\lambda}$ satisfies the equation

$$\begin{aligned} & \rho(\partial_t^2 \boldsymbol{\lambda} + 2\boldsymbol{\Omega} \times \partial_t \boldsymbol{\lambda}) - \nabla \cdot \boldsymbol{\sigma} + \nabla[\rho \boldsymbol{\lambda} \cdot \nabla(\Phi + \psi)] + \rho \nabla \mu - \nabla \cdot (\rho \boldsymbol{\lambda}) \nabla(\Phi + \psi) \\ &= \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)] \delta(\mathbf{r} - \mathbf{r}_r), \end{aligned} \quad (6.27)$$

subject to the boundary conditions

$$[\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma, \quad (6.28)$$

$$[\boldsymbol{\lambda}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma_{\text{SS}}, \quad (6.29)$$

$$[\hat{\mathbf{n}} \cdot \boldsymbol{\lambda}]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma_{\text{FS}}, \quad (6.30)$$

and the end conditions

$$\boldsymbol{\lambda}(\mathbf{r}, T) = \mathbf{0}, \quad \partial_t \boldsymbol{\lambda}(\mathbf{r}, T) = \mathbf{0}. \quad (6.31)$$

On fluid-solid boundaries the quantity $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}$ is normal: $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} = (\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. The scalar Lagrange multiplier μ is determined by

$$\nabla \cdot \boldsymbol{\zeta} = 0, \quad (6.32)$$

subject to the boundary conditions

$$[\mu]_{\pm}^{\pm} = 0, \quad [\hat{\mathbf{n}} \cdot \boldsymbol{\zeta}]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma. \quad (6.33)$$

At this point let us define the *adjoint wavefield* \mathbf{s}^{\dagger} in terms of the vector Lagrange multiplier wavefield $\boldsymbol{\lambda}$ by

$$\mathbf{s}^{\dagger}(\mathbf{r}, t) \equiv \boldsymbol{\lambda}(\mathbf{r}, T - t), \quad (6.34)$$

i.e., the adjoint wavefield is the time-reversed Lagrange multiplier wavefield $\boldsymbol{\lambda}$. Similarly, we define the *adjoint perturbed gravitational potential* ϕ^{\dagger} in terms of the scalar Lagrange

multiplier μ by

$$\phi^\dagger(\mathbf{r}, t) \equiv \mu(\mathbf{r}, T - t). \quad (6.35)$$

Then the adjoint wavefield \mathbf{s}^\dagger is determined by the set of equations

$$\begin{aligned} \rho(\partial_t^2 \mathbf{s}^\dagger - 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}^\dagger) - \nabla \cdot \mathbf{T}^\dagger + \nabla[\rho \mathbf{s}^\dagger \cdot \nabla(\Phi + \psi)] + \rho \nabla \phi^\dagger - \nabla \cdot (\rho \mathbf{s}^\dagger) \nabla(\Phi + \psi) \\ = \sum_r [\mathbf{s}(\mathbf{r}_r, T - t) - \mathbf{d}(\mathbf{r}_r, T - t)] \delta(\mathbf{r} - \mathbf{r}_r), \end{aligned} \quad (6.36)$$

where the adjoint stress is determined in terms of the adjoint displacement gradient by

$$\mathbf{T}^\dagger = \mathbf{c} : \nabla \mathbf{s}^\dagger. \quad (6.37)$$

On all boundaries the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ needs to be continuous:

$$[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma, \quad (6.38)$$

and on fluid-solid boundaries the adjoint traction is normal: $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. On all solid-solid boundaries the adjoint displacement \mathbf{s}^\dagger must be continuous

$$[\mathbf{s}^\dagger]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma_{SS}, \quad (6.39)$$

whereas on fluid-solid discontinuities

$$[\hat{\mathbf{n}} \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma_{FS}. \quad (6.40)$$

The adjoint displacement is subject to the initial conditions

$$\mathbf{s}^\dagger(\mathbf{r}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}^\dagger(\mathbf{r}, 0) = \mathbf{0}. \quad (6.41)$$

The adjoint perturbation in the gravitational potential ϕ^\dagger induced by the adjoint displace-

ment \mathbf{s}^\dagger is determined by the boundary-value problem

$$\nabla \cdot \xi^\dagger = 0, \quad (6.42)$$

$$[\phi^\dagger]_-^+ = 0, \quad [\hat{\mathbf{n}} \cdot \xi^\dagger]_-^+ = 0 \quad \text{on } \Sigma, \quad (6.43)$$

where

$$\xi^\dagger = (4\pi G)^{-1} \nabla \phi^\dagger + \rho \mathbf{s}^\dagger. \quad (6.44)$$

Upon comparing the equations for the regular wavefield (6.5)–(6.13) with those for the adjoint wavefield (6.36)–(6.44) we see that they are identical with two exceptions: First, the source term for the adjoint wave equation (6.36) is $\sum_r [\mathbf{s}(\mathbf{r}_r, T - t) - \mathbf{d}(\mathbf{r}_r, T - t)] \delta(\mathbf{r} - \mathbf{r}_r)$ and involves the time-reversed differences between the synthetics and the data at all receivers; and, second, the adjoint wavefield must be calculated in an Earth model with the opposite sense of rotation, as reflected in the change in the sign of the Coriolis term in (6.36). Thus the adjoint wavefield may be generated with the same software that generates the regular wavefield by reversing the sense of rotation of the Earth model and by using a source that consists of the time-reversed differences between all available data and the corresponding synthetics.

With the regular and adjoint wavefields now defined, the change in the misfit function (6.24) reduces to

$$\begin{aligned} \delta\chi = & - \int_0^T \int_V \{ \delta\rho[\mathbf{s}^\dagger \cdot (\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) + \mathbf{s}^\dagger \cdot \nabla \phi + \mathbf{s} \cdot \nabla \phi^\dagger + \mathbf{s} \cdot \nabla \nabla (\Phi + \psi) \cdot \mathbf{s}^\dagger \\ & + \nabla (\Phi + \psi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s} \nabla \cdot \mathbf{s}^\dagger)] + \nabla \mathbf{s}^\dagger : \delta \mathbf{c} : \nabla \mathbf{s} + \rho \mathbf{s} \cdot \nabla \nabla (\delta \Phi) \cdot \mathbf{s}^\dagger \\ & + \rho \nabla (\delta \Phi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s} \nabla \cdot \mathbf{s}^\dagger) - \mathbf{s}^\dagger \cdot \delta \mathbf{f} \} d^3 \mathbf{r} dt \\ & - \int_0^T \int_\Sigma [\hat{\mathbf{n}} \cdot \delta \mathbf{T} \cdot \mathbf{s}^\dagger - \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta \mathbf{s} + \hat{\mathbf{n}} \cdot \delta \xi \phi^\dagger - \hat{\mathbf{n}} \cdot \xi^\dagger \delta \phi]_-^+ d^2 \mathbf{r} dt, \end{aligned} \quad (6.45)$$

where it is implied that when the regular wavefield \mathbf{s} is evaluated at time t , the adjoint wavefield is evaluated at time $T - t$.

Because we are allowing perturbations δd in the location of all discontinuities Σ , we need to carefully consider the boundary integral in (6.45) because such perturbations render

the variations $\delta \mathbf{s}$ and $\delta \phi$ inadmissible. Because the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ is continuous on Σ and normal on Σ_{FS} we have [Dahlen and Tromp, 1998, eqn. 13.35]

$$[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta \mathbf{s}]_{\pm}^{\pm} = -\delta d[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \partial_n \mathbf{s}]_{\pm}^{\pm} + \nabla^{\Sigma}(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}}) \mathbf{s}]_{\pm}^{\pm}. \quad (6.46)$$

Here $\partial_n = \hat{\mathbf{n}} \cdot \nabla$ denotes the partial derivative in the direction of the outward normal $\hat{\mathbf{n}}$, and $\nabla^{\Sigma} = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla$ denotes the surface gradient on Σ , where \mathbf{I} is the identity tensor. Thus we may express the gradient operator as $\nabla = \hat{\mathbf{n}}\partial_n + \nabla^{\Sigma}$. Note that on a solid-solid boundary both \mathbf{s} and $\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}}$ are continuous, which means that the last term in (6.46) contributes only on a fluid-solid boundary. Next, using the fact that $[\phi^\dagger]_{\pm}^{\pm} = 0$ [and Dahlen and Tromp, 1998, eqn. 13.62] we have

$$[\hat{\mathbf{n}} \cdot \delta \xi \phi^\dagger]_{\pm}^{\pm} = -\delta d[\phi^\dagger \hat{\mathbf{n}} \cdot \partial_n \xi]_{\pm}^{\pm} + \nabla^{\Sigma}(\delta d) \cdot [\phi^\dagger \xi]_{\pm}^{\pm} = \nabla^{\Sigma} \cdot [\delta d \phi^\dagger \xi]_{\pm}^{\pm} - \delta d[\xi \cdot \nabla \phi^\dagger]_{\pm}^{\pm} + \delta d[(\hat{\mathbf{n}} \cdot \xi) \partial_n \phi^\dagger]_{\pm}^{\pm}, \quad (6.47)$$

where we have used (6.11). Because $\nabla^{\Sigma}(\delta d)$ is a tangent vector to Σ , i.e., $\hat{\mathbf{n}} \cdot \nabla^{\Sigma}(\delta d) = 0$, any vector dotted with $\nabla^{\Sigma}(\delta d)$ retains only its tangent components. Thus the integration by parts associated with the second equality involves only the tangent components of ξ ; this becomes important when we use the 2D version of Gauss's theorem in the surface integral in (6.45). Using the continuity $[\hat{\mathbf{n}} \cdot \xi^\dagger]_{\pm}^{\pm}$ [and Dahlen and Tromp, 1998, eqn. 13.36] we have

$$[\hat{\mathbf{n}} \cdot \xi^\dagger \delta \phi]_{\pm}^{\pm} = -\delta d[\hat{\mathbf{n}} \cdot \xi^\dagger \partial_n \phi]_{\pm}^{\pm}. \quad (6.48)$$

Finally, on solid-solid boundaries $[\mathbf{s}^\dagger]_{\pm}^{\pm} = \mathbf{0}$, which implies that [using Dahlen and Tromp, 1998, eqn. 13.61]

$$\begin{aligned} [\hat{\mathbf{n}} \cdot \delta \mathbf{T} \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} &= -\delta d[\hat{\mathbf{n}} \cdot \partial_n \mathbf{T} \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} + \nabla^{\Sigma}(\delta d) \cdot [\mathbf{T} \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} \\ &= -\delta d[(\nabla \cdot \mathbf{T}) \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} + \nabla^{\Sigma} \cdot [\delta d \mathbf{T} \cdot \mathbf{s}^\dagger]_{\pm}^{\pm} - \delta d[\mathbf{T} : \nabla \mathbf{s}^\dagger]_{\pm}^{\pm} + \delta d[\mathbf{T} : \hat{\mathbf{n}} \partial_n \mathbf{s}^\dagger]_{\pm}^{\pm}, \end{aligned} \quad (6.49)$$

where we note, again in anticipation of using the 2D version of Gauss's theorem in (6.45), that in the second equality only the tangent part of $\mathbf{T} \cdot \mathbf{s}^\dagger$ is involved in the horizontal

divergence. It may be shown that on a fluid-solid boundary (6.50) acquires one extra term:

$$[\hat{\mathbf{n}} \cdot \delta \mathbf{T} \cdot \mathbf{s}^\dagger]^\pm = -\delta d[(\nabla \cdot \mathbf{T}) \cdot \mathbf{s}^\dagger]^\pm + \nabla^\Sigma \cdot [\delta d \mathbf{T} \cdot \mathbf{s}^\dagger]^\pm - \delta d[\mathbf{T} : \nabla \mathbf{s}^\dagger]^\pm + \delta d[\mathbf{T} : \hat{\mathbf{n}} \partial_n \mathbf{s}^\dagger]^\pm - \nabla^\Sigma(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}}) \mathbf{s}^\dagger]^\pm. \quad (6.50)$$

Thus we find that we may rewrite the surface integral in (6.45) in the form

$$\begin{aligned} & \int_0^T \int_\Sigma [\hat{\mathbf{n}} \cdot \delta \mathbf{T} \cdot \mathbf{s}^\dagger - \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta \mathbf{s} + \hat{\mathbf{n}} \cdot \delta \xi \phi^\dagger - \hat{\mathbf{n}} \cdot \xi^\dagger \delta \phi]^\pm d^2 \mathbf{r} dt \\ &= \int_0^T \int_\Sigma \delta d [\rho \{ -(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) \cdot \mathbf{s}^\dagger - \mathbf{s}^\dagger \cdot \nabla \phi - \mathbf{s} \cdot \nabla \phi^\dagger - \mathbf{s} \cdot \nabla \nabla (\Phi + \psi) \cdot \mathbf{s}^\dagger \\ &\quad - \nabla (\Phi + \psi) \cdot (\mathbf{s}^\dagger \cdot \nabla \mathbf{s} - \mathbf{s}^\dagger \nabla \cdot \mathbf{s}) \} - \nabla \mathbf{s}^\dagger : \mathbf{c} : \nabla \mathbf{s} - (4\pi G)^{-1} (\nabla \phi \cdot \nabla \phi^\dagger) + (\hat{\mathbf{n}} \cdot \mathbf{T}) \cdot \partial_n \mathbf{s}^\dagger \\ &\quad + (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger) \cdot \partial_n \mathbf{s} + (\hat{\mathbf{n}} \cdot \xi) \partial_n \phi^\dagger + (\hat{\mathbf{n}} \cdot \xi^\dagger) \partial_n \phi]^\pm d^2 \mathbf{r} dt \\ &\quad - \int_0^T \int_{\Sigma_{\text{FS}}} \nabla^\Sigma(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}}) \mathbf{s} + (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}}) \mathbf{s}^\dagger]^\pm d^2 \mathbf{r} dt, \end{aligned} \quad (6.51)$$

where we have used the equation of motion (6.5) and the 2D version of Gauss's theorem on a closed surface [Dahlen and Tromp, 1998, eqn. A.79]:

$$\int_\Sigma \nabla^\Sigma \cdot \mathbf{v}^\Sigma d^2 \mathbf{r} = 0, \quad (6.52)$$

for any tangent vector \mathbf{v}^Σ .

Upon combining (6.45) and (6.51) we may write the variation in the misfit function in the form

$$\delta \chi = \int_V \delta \oplus K_\oplus d^3 \mathbf{r} + \int_\Sigma \delta d K_d d^2 \mathbf{r} + \int_{\Sigma_{\text{FS}}} \nabla^\Sigma(\delta d) \cdot \mathbf{K}_d d^2 \mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T-t) \cdot \delta \mathbf{f}(t) d^3 \mathbf{r} dt, \quad (6.53)$$

where the symbol \oplus denotes the Earth model parameters ρ and \mathbf{c} , and the kernels K_\oplus , K_d , and \mathbf{K}_d are defined by

$$\begin{aligned} \delta \oplus K_\oplus = & - \int_0^T \{ \delta \rho [\mathbf{s}^\dagger \cdot (\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) + \mathbf{s}^\dagger \cdot \nabla \phi + \mathbf{s} \cdot \nabla \phi^\dagger + \mathbf{s} \cdot \nabla \nabla (\Phi + \psi) \cdot \mathbf{s}^\dagger \\ & + \nabla (\Phi + \psi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s}^\dagger \nabla \cdot \mathbf{s})] + \nabla \mathbf{s}^\dagger : \delta \mathbf{c} : \nabla \mathbf{s} + \rho \mathbf{s} \cdot \nabla \nabla (\delta \Phi) \cdot \mathbf{s}^\dagger \\ & + \rho \nabla (\delta \Phi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s}^\dagger \nabla \cdot \mathbf{s}) \} dt, \end{aligned} \quad (6.54)$$

$$\begin{aligned}
K_d = & - \int_0^T [\rho \{ -(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) \cdot \mathbf{s}^\dagger - \mathbf{s}^\dagger \cdot \nabla \phi - \mathbf{s} \cdot \nabla \phi^\dagger - \mathbf{s} \cdot \nabla \nabla (\Phi + \psi) \cdot \mathbf{s}^\dagger \\
& - \nabla (\Phi + \psi) \cdot (\mathbf{s}^\dagger \cdot \nabla \mathbf{s} - \mathbf{s}^\dagger \nabla \cdot \mathbf{s}) \} - \nabla \mathbf{s}^\dagger : \mathbf{c} : \nabla \mathbf{s} - (4\pi G)^{-1} (\nabla \phi \cdot \nabla \phi^\dagger) \\
& + (\hat{\mathbf{n}} \cdot \mathbf{T}) \cdot \partial_n \mathbf{s}^\dagger + (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger) \cdot \partial_n \mathbf{s} + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi}) \partial_n \phi^\dagger + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger) \partial_n \phi]_+^- dt, \quad (6.55)
\end{aligned}$$

$$\mathbf{K}_d = \int_0^T [(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}}) \mathbf{s} + (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}}) \mathbf{s}^\dagger]_+^- dt, \quad (6.56)$$

where it should be remembered that in these kernel expressions the adjoint field is evaluated at time $T - t$, whereas the regular wavefield is evaluated at time t .

If we ignore rotation and self-gravitation, the gradient (6.53) may be rewritten as

$$\delta \chi = \int_V (\delta \rho K_\rho + \delta \mathbf{c} :: \mathbf{K}_c) d^3 \mathbf{r} + \int_\Sigma \delta d K_d d^2 \mathbf{r} + \int_{\Sigma_{\text{FS}}} \nabla^\Sigma (\delta d) \cdot \mathbf{K}_d d^2 \mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T-t) \cdot \delta \mathbf{f}(t) d^3 \mathbf{r} dt, \quad (6.57)$$

where kernels K_ρ , \mathbf{K}_c , K_d and \mathbf{K}_d are defined by [Tromp *et al.*, 2005]

$$K_\rho = - \int_0^T \mathbf{s}^\dagger(T-t) \cdot \partial_t^2 \mathbf{s}(t) dt, \quad (6.58)$$

$$\mathbf{K}_c = - \int_0^T \nabla \mathbf{s}^\dagger(T-t) \nabla \mathbf{s}(t) dt, \quad (6.59)$$

$$K_d = - \int_0^T [-\rho \mathbf{s}^\dagger(T-t) \cdot \partial_t^2 \mathbf{s}(t) - \nabla \mathbf{s}^\dagger(T-t) : \mathbf{c} : \nabla \mathbf{s}(t) + \hat{\mathbf{n}} \cdot \mathbf{T}(t) \cdot \partial_n \mathbf{s}^\dagger(T-t) + \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger(T-t) \cdot \partial_n \mathbf{s}(t)]_+^- dt, \quad (6.60)$$

$$\mathbf{K}_d = \int_0^T [\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger(T-t) \cdot \hat{\mathbf{n}} \mathbf{s}(t) + \hat{\mathbf{n}} \cdot \mathbf{T}(t) \cdot \hat{\mathbf{n}} \mathbf{s}^\dagger(T-t)]_+^- dt. \quad (6.61)$$

In an isotropic Earth model we have $c_{jklm} = (\kappa - \frac{2}{3}\mu) \delta_{jk} \delta_{lm} + \mu (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl})$, where μ and κ denote the shear and bulk moduli, respectively. Note that the shear modulus μ should not be confused with the scalar Lagrange multiplier used earlier. Thus we may write

$$\delta \mathbf{c} :: \mathbf{K}_c = \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa, \quad (6.62)$$

where the isotropic kernels K_κ and K_μ represent Fréchet derivatives with respect to relative bulk and shear moduli perturbations $\delta \ln \kappa = \delta \kappa / \kappa$ and $\delta \ln \mu = \delta \mu / \mu$, respectively. These

isotropic kernels are given by

$$K_\mu(\mathbf{r}) = - \int_0^T 2\mu(\mathbf{r}) \mathbf{D}^\dagger(\mathbf{r}, T-t) : \mathbf{D}(\mathbf{r}, t) dt, \quad (6.63)$$

$$K_\kappa(\mathbf{r}) = - \int_0^T \kappa(\mathbf{r}) [\nabla \cdot \mathbf{s}^\dagger(\mathbf{r}, T-t)] [\nabla \cdot \mathbf{s}(\mathbf{r}, t)] dt, \quad (6.64)$$

where \mathbf{D} and \mathbf{D}^\dagger denote the traceless strain deviator and its adjoint, respectively.

Similarly, in an isotropic Earth model (6.60) and (6.61) reduce to

$$\begin{aligned} K_d(\mathbf{r}) = & \int_0^T [\rho(\mathbf{r}) \mathbf{s}^\dagger(\mathbf{r}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{r}, t) + \kappa(\mathbf{r}) \nabla \cdot \mathbf{s}^\dagger(\mathbf{r}, T-t) \nabla \cdot \mathbf{s}(\mathbf{r}, t) \\ & + 2\mu(\mathbf{r}) \mathbf{D}^\dagger(\mathbf{r}, T-t) : \mathbf{D}(\mathbf{r}, t) - \kappa(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \cdot \partial_n \mathbf{s}^\dagger(\mathbf{r}, T-t) \nabla \cdot \mathbf{s}(\mathbf{r}, t) \\ & - 2\mu(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \partial_n \mathbf{s}^\dagger(\mathbf{r}, T-t) : \mathbf{D}(\mathbf{r}, t) - \kappa(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \cdot \partial_n \mathbf{s}(\mathbf{r}, t) \nabla \cdot \mathbf{s}^\dagger(\mathbf{r}, T-t) \\ & - 2\mu(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \partial_n \mathbf{s}(\mathbf{r}, t) : \mathbf{D}^\dagger(\mathbf{r}, T-t)]_+^- dt, \end{aligned} \quad (6.65)$$

$$\begin{aligned} \mathbf{K}_d(\mathbf{r}) = & \int_0^T [\mathbf{s}^\dagger(\mathbf{r}, T-t) [\kappa(\mathbf{r}) \nabla \cdot \mathbf{s}(\mathbf{r}, t) + 2\mu(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \cdot \mathbf{D}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}(\mathbf{r})] \\ & + \mathbf{s}(\mathbf{r}, t) [\kappa(\mathbf{r}) \nabla \cdot \mathbf{s}^\dagger(\mathbf{r}, T-t) + 2\mu(\mathbf{r}) \hat{\mathbf{n}}(\mathbf{r}) \cdot \mathbf{D}^\dagger(\mathbf{r}, T-t) \cdot \hat{\mathbf{n}}(\mathbf{r})]]_+^- dt. \end{aligned} \quad (6.66)$$

Finally, we may express the Fréchet derivatives in an isotropic Earth model in terms of relative variations in density $\ln \rho$, shear-wave speed $\ln \beta$, and compressional-wave speed $\ln \alpha$ based upon the relationship

$$\delta \ln \rho K_\rho + \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa = \delta \ln \rho K_{\rho'} + \delta \ln \beta K_\beta + \delta \ln \alpha K_\alpha, \quad (6.67)$$

where

$$K_{\rho'} = K_\rho + K_\kappa + K_\mu, \quad K_\beta = 2 \left(K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right), \quad K_\alpha = 2 \left(\frac{\kappa + \frac{4}{3} \mu}{\kappa} \right) K_\kappa. \quad (6.68)$$

6.5 Anelasticity

In an anelastic Earth model, Hooke's law (6.6) needs to be replaced by [e.g., *Aki and Richards*, 1980; *Dahlen and Tromp*, 1998]

$$\mathbf{T}(t) = \int_0^t \mathbf{c}(t-t') : \partial_{t'} \nabla \mathbf{s}(t') dt', \quad (6.69)$$

i.e., the tensor \mathbf{c} becomes time dependent and the stress at time t depends on the entire strain history. To avoid clutter we have suppressed the spatial dependence. In this section we will determine the implications of this more general, anelastic constitutive relationship for the adjoint equations. We begin by manipulating the term of concern in (6.15):

$$\begin{aligned} \int_0^T \int_V \boldsymbol{\lambda} \cdot (\nabla \cdot \mathbf{T}) d^3\mathbf{r} dt &= - \int_0^T \int_V \nabla \boldsymbol{\lambda} : \mathbf{T} d^3\mathbf{r} dt \\ &= - \int_0^T \int_V \nabla \boldsymbol{\lambda}(t) : \left[\int_0^t \mathbf{c}(t-t') : \partial_{t'} \nabla \mathbf{s}(t') dt' \right] d^3\mathbf{r} dt, \end{aligned} \quad (6.70)$$

where we have used the boundary condition (6.7) and imposed the (anticipated) continuity conditions (6.29) and (6.30). Upon integrating by parts over t' , using the initial condition $\mathbf{s}(0) = \mathbf{0}$ and the relationship $\partial_{t'} \mathbf{c}(t-t') = -\partial_t \mathbf{c}(t-t')$, changing the order of integration over t and t' , integrating by parts over t , imposing the (anticipated) end condition (6.31), and the symmetries of the tensor \mathbf{c} , we obtain

$$\int_0^T \int_V \boldsymbol{\lambda} \cdot (\nabla \cdot \mathbf{T}) d^3\mathbf{r} dt = \int_V \int_0^T \left[\int_{t'}^T \mathbf{c}(t-t') : \partial_t \nabla \boldsymbol{\lambda}(t) dt \right] : \nabla \mathbf{s}(t') dt' d^3\mathbf{r}. \quad (6.71)$$

If we define the second-order tensor

$$\boldsymbol{\sigma}(t) = - \int_t^T \mathbf{c}(t'-t) : \partial_{t'} \nabla \boldsymbol{\lambda}(t') dt', \quad (6.72)$$

then it is straightforward to show based upon varying (6.71) that in an anelastic medium the Lagrange multiplier field $\boldsymbol{\lambda}$ is determined by the wave equation (6.27), provided we use the definition (6.72) for $\boldsymbol{\sigma}$, rather than the elastic form (6.25).

As in the elastic case, we define the adjoint wavefield \mathbf{s}^\dagger in terms of the Lagrange

multiplier field λ by (6.34). By analogy, let us define the anelastic adjoint stress \mathbf{T}^\dagger in terms of the second order tensor $\boldsymbol{\sigma}$ by

$$\mathbf{T}^\dagger(\mathbf{r}, t) = \boldsymbol{\sigma}(\mathbf{r}, T - t). \quad (6.73)$$

Based upon (6.72) and (6.73) it is now straightforward to show that in an anelastic medium the adjoint wavefield \mathbf{s}^\dagger is determined by (6.36) provided the adjoint stress \mathbf{T}^\dagger is defined in terms of the gradient of the adjoint displacement by

$$\mathbf{T}^\dagger(t) = \int_0^t \mathbf{c}(t - t') : \partial_{t'} \nabla \mathbf{s}^\dagger(t') dt'. \quad (6.74)$$

We see that in an anelastic Earth model, just as in the elastic case, the adjoint wavefield is determined by exactly the same equations as the regular wavefield, with the exception of the source term and the sense of rotation.

In the context of the tomographic inverse problem, the gradient of the misfit function becomes

$$\begin{aligned} \delta\chi = & \int_V [\delta\rho K_\rho - \int_0^T \delta\mathbf{T}^\dagger(T - t) : \nabla \mathbf{s}(t) dt] d^3\mathbf{r} \\ & + \int_\Sigma \delta d K_d d^2\mathbf{r} + \int_{\Sigma_{\text{FS}}} \nabla^\Sigma(\delta d) \cdot \mathbf{K}_d d^2\mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T - t) \cdot \delta\mathbf{f}(t) d^3\mathbf{r} dt, \end{aligned} \quad (6.75)$$

where K_ρ is defined by (6.58) and where we have defined

$$\delta\mathbf{T}^\dagger(t) = \int_0^t \delta\mathbf{c}(t - t') : \partial_{t'} \nabla \mathbf{s}^\dagger(t') dt'. \quad (6.76)$$

Note that for purely elastic perturbations, when the perturbation $\delta\mathbf{c}$ does not depend on time, the gradient (6.75) reduces to (6.57). This implies that a tomographic inversion for elastic perturbations superimposed upon an anelastic reference model may be accomplished based upon the gradient (6.57), provided the regular and adjoint wavefields are calculated in the anelastic reference model.

6.6 Global Seismic Wave Propagation Based upon the Spectral Element Method

As demonstrated in Section 6.4, the expressions for the Fréchet derivatives of volumetric model parameters, e.g., (6.58), (6.59), (6.63) and (6.64), and topographic variations on internal discontinuities, e.g., (6.60) and (6.61), involve simultaneous access to the forward wavefield $\mathbf{s}(\mathbf{x}, t)$ and the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{x}, T - t)$. We calculate both the forward and adjoint wavefield based upon the spectral-element method (SEM), which has been successfully applied to simulate wave propagation on both global and regional scales [e.g., *Komatitsch and Vilotte*, 1998; *Komatitsch and Tromp*, 1999, 2002a,b; *Chaljub et al.*, 2003; *Komatitsch et al.*, 2004]. The method combines the geometrical flexibility of the finite-element method with an accurate representation of the wavefield in terms of high-order Lagrange polynomials. The global spectral-element mesh is designed to honor topography and bathymetry on the Earth's surface and velocity discontinuities inside the model, such as the Moho, the upper mantle discontinuities, the core-mantle boundary (CMB), and the inner-core boundary (ICB). The choice of Lagrange interpolants in combination with Gauss-Lobatto-Legendre quadrature produces an exactly diagonal mass matrix, which makes it straightforward to implement the method on parallel computers. A simple predictor-corrector scheme is used to time-march the wavefield [*Komatitsch and Tromp*, 2002a]. The detailed numerical implementation and numerous benchmarking results are discussed in *Komatitsch and Tromp* [2002a,b]. In general, use of the SEM to simulate wave propagation in 3D global Earth models results in better traveltimes and waveform fits compared to synthetics generated based upon semi-analytical methods for 1D Earth models [e.g., *Ritsema et al.*, 2002; *Tsuboi et al.*, 2003; *Ji et al.*, 2005]. A typical 3D global simulation for 40-minute long synthetic seismograms accurate to a shortest period of 18 s takes approximately 4 hours on a 75-node PC cluster.

6.7 Numerical Implementation of the Adjoint Method

Although it is straightforward to compute the forward and adjoint wavefields by solving the regular wave equation (6.5), and the adjoint wave equation (6.36) using the SEM, calculation of the kernel expressions (6.58), (6.59), (6.60) and (6.61) involves simultaneous access to the forward wavefield $\mathbf{s}(\mathbf{x}, t)$ and the adjoint wavefield in reverse time $\mathbf{s}^\dagger(\mathbf{x}, T - t)$. This makes it difficult to carry both the forward and adjoint wavefield simultaneously in the SEM, because at any given time t we would only be able to access the forward wavefield $\mathbf{s}^\dagger(\mathbf{x}, t)$ and the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{x}, t)$, while the time integration involved in constructing the kernels requires access to the time-reversed adjoint field $\mathbf{s}^\dagger(\mathbf{x}, T - t)$. One apparent solution is to run the forward simulation and save the entire regular wavefield $\mathbf{s}(\mathbf{x}, t)$ as a function of both time and space, followed by an adjoint simulation during which at any given time t we read back the forward wavefield at time $T - t$. This scheme is easy to understand and theoretically simple to implement; however, realistic global SEM simulations accurate down to 18 s generally require approximately 12 GB of storage space per time step to save the forward wavefield. A typical simulation for 40-minutes-long seismograms involves a number of time steps on the order of 10^4 . Therefore, the total storage requirements involve hundreds of terabytes of disk space, and the related I/O requirements may significantly increase the total simulation time. One possible solution may be to employ an efficient data compression scheme to reduce the storage requirements and overhead time associated with assessing the time slices of the forward wavefield.

In this paper we employ an approach first introduced by *Liu and Tromp* [2006a] for regional wave propagation problems, in which, ignoring anelasticity effects, we reconstruct the forward wavefield by solving the equation of motion (6.5) backwards in time:

$$\rho(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) - \nabla \cdot \mathbf{T} + \nabla[\rho \mathbf{s} \cdot \nabla(\Phi + \psi)] + \rho \nabla \phi - \nabla \cdot (\rho \mathbf{s}) \nabla(\Phi + \psi) = \mathbf{f}, \quad (6.77)$$

$$[\hat{\mathbf{n}} \cdot \mathbf{T}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma, \quad (6.78)$$

$$[\mathbf{s}]_{\pm}^{\pm} = \mathbf{0} \quad \text{on } \Sigma_{\text{SS}}, \quad [\hat{\mathbf{n}} \cdot \mathbf{s}]_{\pm}^{\pm} = 0 \quad \text{on } \Sigma_{\text{FS}}, \quad (6.79)$$

$$\mathbf{s}(\mathbf{r}, T) \quad \text{and} \quad \partial_t \mathbf{s}(\mathbf{r}, T) \quad \text{given.} \quad (6.80)$$

This initial and boundary value problem can be solved to reconstruct $s(\mathbf{x}, t)$ for $T \geq t \geq 0$ the same way the forward wave equation (6.5) is solved subject to the boundary conditions (6.7), (6.8) and (6.9), and the initial conditions (6.10). Actually, we simply need to replace the time step parameter Δt in the forward simulation by $-\Delta t$ in the backward simulation. Notice that we need to have the ‘ending conditions’ $s(\mathbf{r}, T)$ and $\partial_t s(\mathbf{r}, T)$ to start the reconstruction of $s(\mathbf{x}, t)$, and therefore we need to save on disk the final displacement and velocity fields of a previous forward simulation. By carrying the backward simulation (6.77) and the adjoint simulation (6.36) simultaneously, at any given time t we have access to both the regular wavefield $s(\mathbf{x}, T - t)$ and the adjoint wavefield $s^\dagger(\mathbf{x}, t)$, making it feasible to compute on the fly the contributions to the time integrations involved in the construction of the finite-frequency kernels (6.58), (6.59), (6.60) and (6.61).

In summary, sensitivity kernels may be calculated based upon just two spectral-element simulations. One forward simulation provides the synthetic seismograms that may be used for a cross-correlation traveltimes measurement. During the second simulation we calculate the adjoint wavefield while simultaneously reconstructing the forward wavefield based upon the last snapshots from the first simulation, thereby enabling the on-the-fly calculation of the sensitivity kernels. The second simulation requires about twice the memory and run time of the first simulation but avoids overwhelming storage requirements and I/O demands.

As discussed by *Liu and Tromp* [2006a], in regional spectral-element simulations one needs to save the wavefield absorbed on the fictitious boundaries of the model domain as well as a snapshot of the entire final wavefield. Since the global simulations involve no absorbing boundaries, only the final snapshot needs to be saved. This requires very little storage space compared to the absorbing contributions that need to be saved at every time step during the regional forward simulations.

6.8 Applications to Spherically Symmetric Models

In this section we calculate cross-correlation traveltimes sensitivity kernels for typical global arrivals in the 1D isotropic Preliminary Reference Earth Model (PREM) [*Dziewonski and*

Anderson, 1981]. For a given component of the seismogram, we may express the cross-correlation traveltime perturbation as [Luo and Schuster, 1990; Marquering et al., 1999; Dahlen et al., 2000]

$$\delta T = \int_V \delta \oplus K_{\oplus} d^3 \mathbf{r} + \int_{\Sigma} \delta d K_d d^2 \mathbf{r} + \int_{\Sigma_{\text{FS}}} \nabla^{\Sigma}(\delta d) \cdot \mathbf{K}_d d^2 \mathbf{r}. \quad (6.81)$$

The expressions for the volumetric kernels K_{\oplus} and the boundary kernels K_d and \mathbf{K}_d are the same as equations (6.54), (6.55), and (6.56), except that for a cross-correlation traveltime measurement the associated adjoint wavefield \mathbf{s}^{\dagger} is generated by the adjoint source [Tromp et al., 2005]

$$\mathbf{f}^{\dagger} = \frac{1}{N} w(T-t) \partial_t s(\mathbf{x}_r, T-t) \delta(\mathbf{x} - \mathbf{x}_r), \quad (6.82)$$

where \mathbf{x}_r denotes the receiver location, $w(t)$ is the cross-correlation window, and the normalization factor N is given by

$$N = \int_0^T w(t) s(\mathbf{x}_r, t) \partial_t^2 s(\mathbf{x}_r, t) dt. \quad (6.83)$$

Notice that both the adjoint source \mathbf{f}^{\dagger} and the adjoint wavefield \mathbf{s}^{\dagger} are determined by the synthetic seismogram $s(t)$, and therefore the sensitivity kernels can be computed based upon only this information.

We simulate three-component seismograms for the June 9, 1994, Bolivian Earthquake. This earthquake occurred at a depth of 647 km and is one of the largest deep events in modern recording history. We use the event location and the centroid-moment tensor (CMT) solution from the Harvard CMT catalog (www.seismology.harvard.edu) for our simulation, and we modify the half duration of the CMT solution to be compatible with the resolution of the spectral-element simulation. We follow the procedures outlined in Section 6.7 to calculate sensitivity kernels for typical global phases at various distances. Most of the kernels are generated for a hypothetical station at a source azimuth of 65° , unless otherwise noted. We will consider the sensitivity kernels of the following phases: direct P and S at an epicentral distance of 60° , the core-reflected phase ScS at a distance of 60° , the core phases PKPab and PKPdf at a distance of 170° and SKS at a distance of 115° , the

core-diffracted phase P_{diff} at a distance of 103° , the depth phases sS at a distance of 40° and sP at a distance of 90° , as well as the surface-reflected phases PP at a distance of 115° and SS at a distance of 90° . These arrivals are chosen because they are readily identified in the seismograms in this particular azimuth at these specific distances. Figure 6.1 illustrates the ray geometry of these arrivals for the PREM model.

One of the advantages of our fully numerical approach using the adjoint method is that we need not be able to ‘label’ a particular pulse in the seismogram, i.e., we need not have any knowledge of the raypath associated with this particular pulse. By performing the forward and adjoint simulations, combining the forward and adjoint wavefields during the construction of the kernels, we automatically obtain the 3D sensitivity associated with this particular pulse in the seismogram. Sometimes the kernels may be readily identified with a particular geometrical raypath, but frequently the sensitivity kernels are much richer. Another advantage of the adjoint approach is that it may be used for fully 3D reference Earth models, as we discuss further in Section 6.9.

6.8.1 P Kernel

We generate 3D compressional wave-speed sensitivity kernels K_α for the P phase at an epicentral distance of 60° . Figure 6.2(a), (b) and (c) show source-receiver cross-sections of the kernels calculated using the adjoint method for simulations accurate down to periods of 27 s, 18 s and 9 s, respectively. The vertical component seismograms corresponding to the simulations with different resolution are shown in Figure 6.2(d). All kernels show very nice banana-doughnut shapes in good agreement with results obtained based upon ray-based finite-frequency kernels [Dahlen *et al.*, 2000; Hung *et al.*, 2000; Zhou *et al.*, 2004] as well as those obtained based upon normal-mode methods [Zhao *et al.*, 2005; Zhao and Jordan, 2006]. Notice that the size of the doughnut holes decreases with increasing resolution, in accordance with the scaling relation $\text{width} \sim \sqrt{\lambda L}$, where L denotes the length of the raypath and λ the wavelength. At very high frequencies the sensitivity kernel will asymptotically collapse onto the P raypath shown in Figure 6.1(a). Notice that the absolute amplitude of the P kernel at 9 s is about 5 times larger than the amplitude of the P

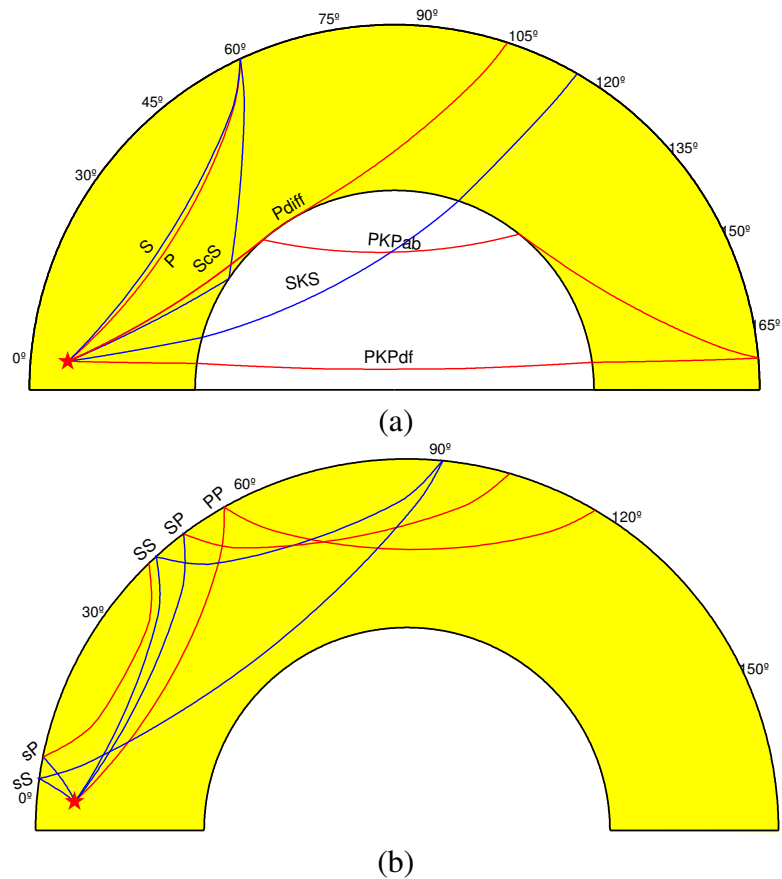


Figure 6.1: (a) PREM ray geometry for the following arrivals: P, S, ScS, P_{diff} , SKS, PKPab and PKPdf. The S legs of the raypaths are colored blue, while the P legs of the raypaths are colored red. The hypocenter location of the June 9, 1994, Bolivian earthquake at a depth of 647 km is denoted by a red star. (b) PREM ray geometry for surface-reflected phases such as PP, SS, and SP, and the depth phases sS and sP.

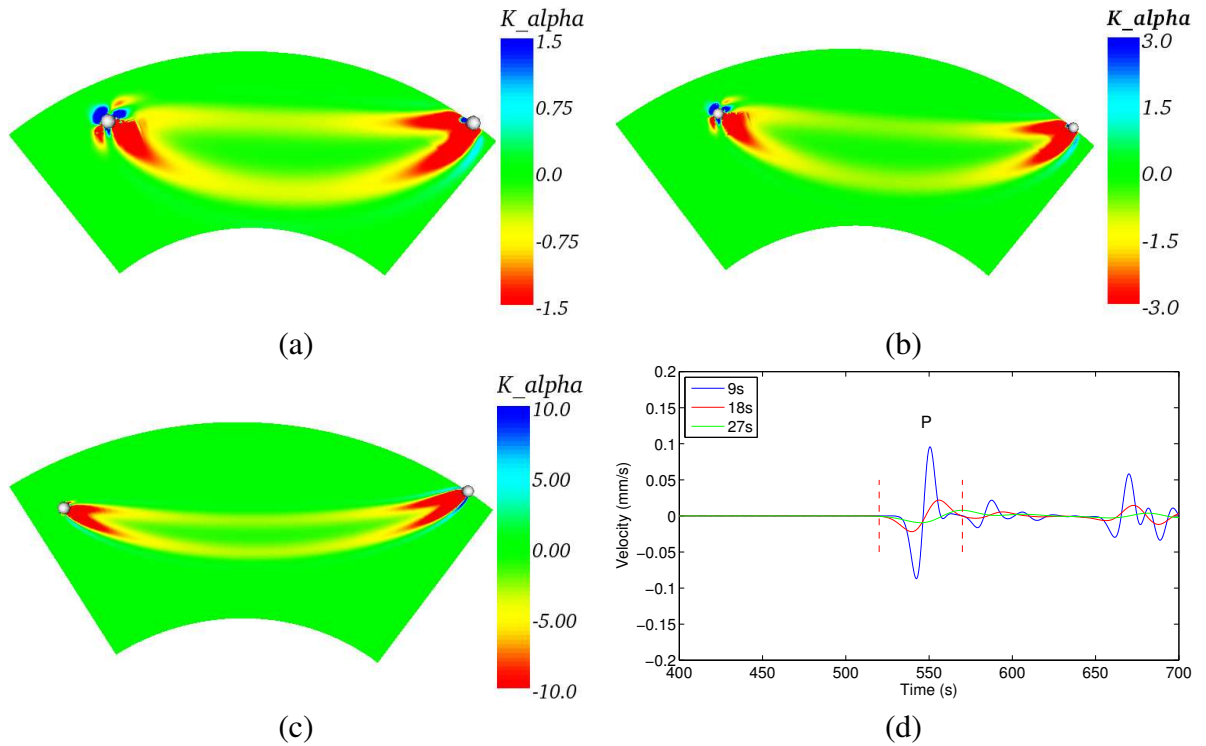


Figure 6.2: (a) Source-receiver cross-section of the K_α kernel for a 27 s P wave recorded at a station at an epicentral distance of 60° . The source and receiver locations are denoted with two small white circles. The unit of the sensitivity kernels is 10^7 s/km³ throughout this paper. (b) K_α kernel for an 18 s P wave recorded at a station at an epicentral distance of 60° . (c) K_α kernel for a 9 s P wave recorded at a station at an epicentral distance of 60° . (d) Vertical component of the synthetic velocity seismograms recorded at an epicentral distance of 60 for simulations accurate down to periods of 27 s, 18 s and 9 s, respectively.

kernel at 27 s. The source radiation patterns clearly show up near the hypocenter (denoted by a white circle in Figures 6.2(a),(b) and (c)).

6.8.2 P_{diff} Kernel

Because of the sudden drop in compressional-wave speed across the CMB towards the outer core, a shadow zone exists beyond an epicentral distance of approximately 103° for the direct P wave. Seismic energy enters the shadow zone in the form of waves diffracted along the CMB. The raypath for the P_{diff} phase at an epicentral distance of 103° is shown in Figure 6.1(a). The P_{diff} phase, because of its diffracted nature, generally has a lower frequency content relative to the direct P wave at closer distances. Figure 6.3(a) shows a vertical

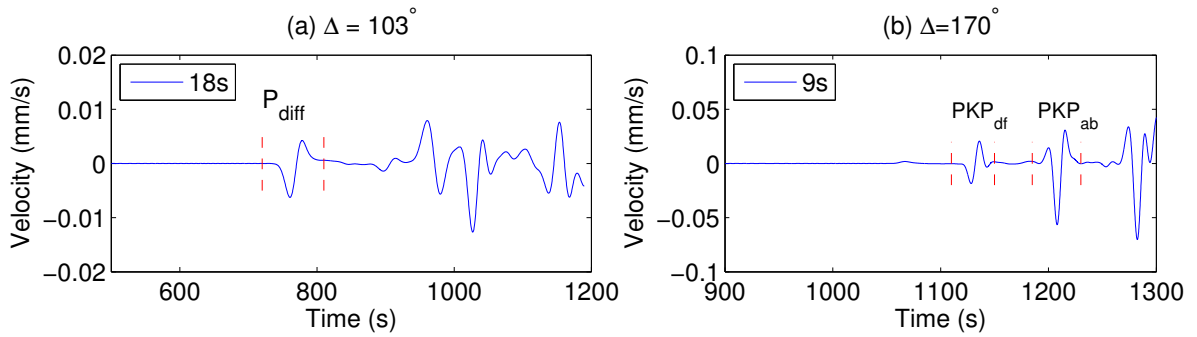


Figure 6.3: (a) Vertical component synthetic velocity seismograms recorded at a station at an epicentral distance of 103° . The spectral-element simulation is accurate down to periods of 18 s. (b) Vertical component synthetic velocity seismograms recorded at a station at an epicentral distance of 170° . The spectral-element simulation is accurate down to periods of 9 s.

velocity seismogram recorded at a station at an epicentral distance of 103° . Figure 6.4(a) shows a source-receiver cross-section of the corresponding K_α kernel. It is similar to a regular banana-doughnut kernel except that it has been cut through at the bottom by the CMB. Figure 6.4(b) shows the sensitivity of the P_{diff} phase along the mantle side of the CMB. Due to the source radiation pattern the sensitivity kernel on the CMB is asymmetric with respect to the source-receiver plane. The P_{diff} phase is frequently assumed to be mainly sensitivity to the diffracted portion of the raypath along the CMB; however, the sensitivity kernel shows significant sensitive away from the raypath, e.g., to regions more than 500 km above the CMB.

6.8.3 PKP Kernel

For stations beyond the shadow zone, the first arrival is usually the P_{diff} phase. However, at epicentral distances larger than approximately 130° the PKP phases that arrive about 60–100 s after the diffracted P phase are usually the first clearly identifiable arrivals in vertical component seismograms. Figure 6.5 displays the traveltime curve for the various PKP branches, the most noticeable of which are PKP_{ab} and PKP_{df} , whose raypaths are shown in Figure 6.1(a). Notice that the PKP_{ab} phase generally samples the shallow part of the outer core, while the PKP_{df} phase samples the deeper parts of the core. Figure 6.3 shows

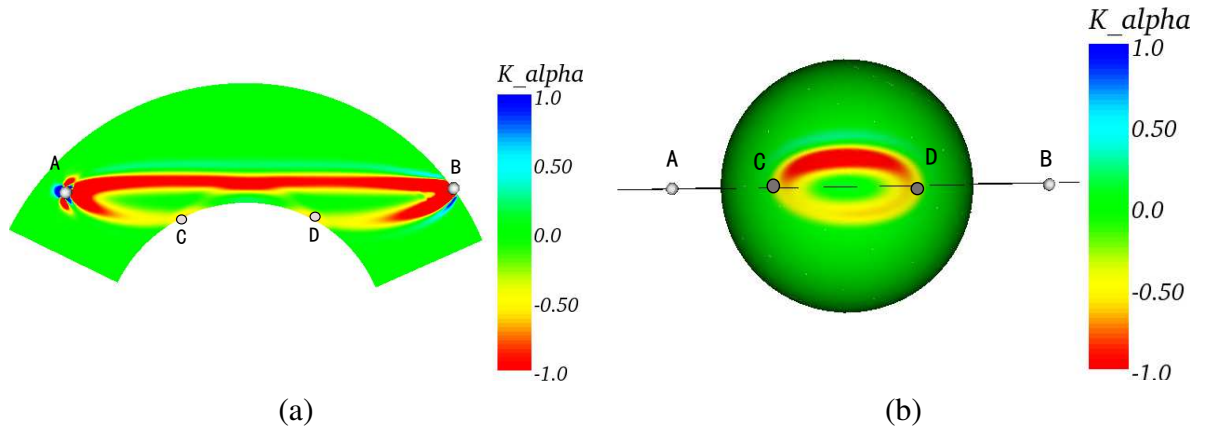


Figure 6.4: (a) Source-receiver cross-section of the K_α kernel for an 18 s P_{diff} phase at an epicentral distance of 103° ; (b) Spherical cross-section on the CMB of the K_α kernel for an 18 s P_{diff} phase at an epicentral distance of 103° .

a vertical displacement seismogram recorded at an epicentral distance of 170° for a simulation accurate down to 9 s. The PKP_{df} and PKP_{ab} arrivals can be clearly identified in the seismogram. We pick an epicentral distance of 170° because the traveltime difference between the PKP_{df} and PKP_{ab} branches can be as large as 70 s at this distance, which makes it very easy to identify and separate them. Figure 6.6(a) and 6.6(b) show the finite-frequency traveltime sensitivity kernels K_α for these two PKP branches. The PKP_{df} kernel clearly follows the raypath shown in Figure 6.1 while exhibiting the typical banana-doughnut shape. The upper part of the kernel has relatively smaller sensitivity compared to the lower part, which is partly a result of the source radiation pattern of the Bolivian earthquake. The PKP_{ab} kernel not only follows the conventional raypath, but also shows sensitivity along a path that goes from the source to the receiver in the major-arc direction. This is a result of the finite-frequency nature of the cross-correlation traveltime measurements: points along the unconventional path represent scatterers that produce arrivals within the PKP_{ab} cross-correlation traveltime window. Note, however, that the oscillatory nature of the kernel along this major-arc path tends to average out longer wavelength heterogeneity.

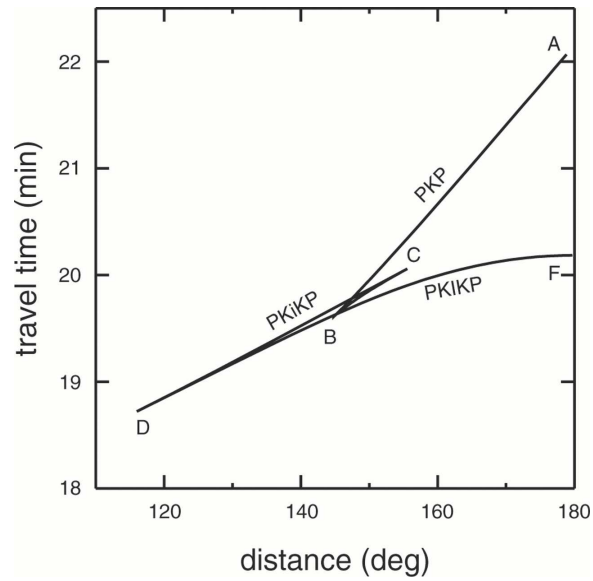


Figure 6.5: Traveltime curves of the various PKP branches for the PREM model.

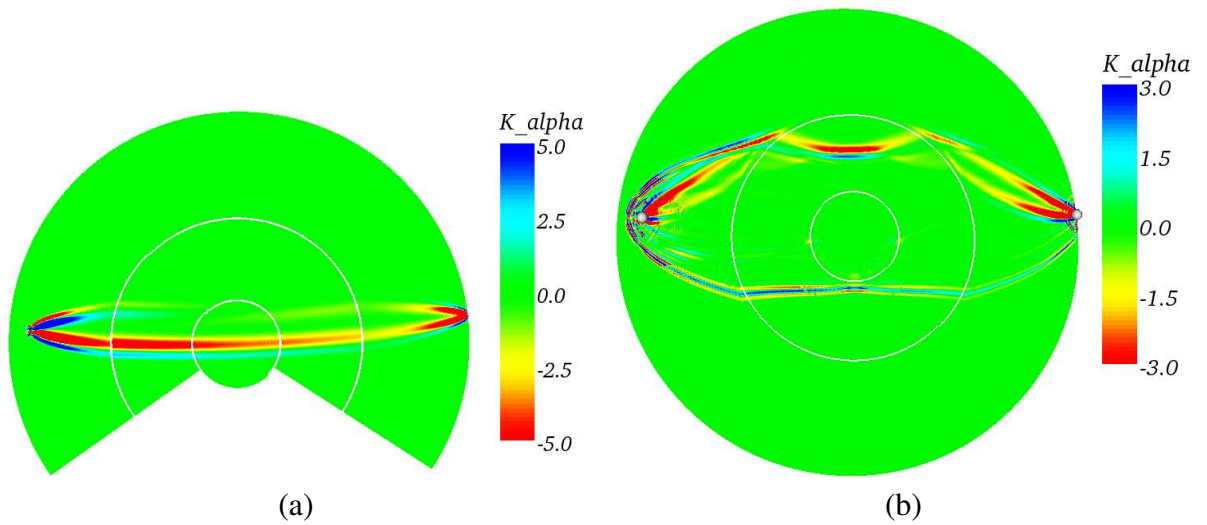


Figure 6.6: (a) K_α kernel for 9-second PKPdf phase at 170° ; (b) K_α kernel for 9-second PKPab phase at 170° .

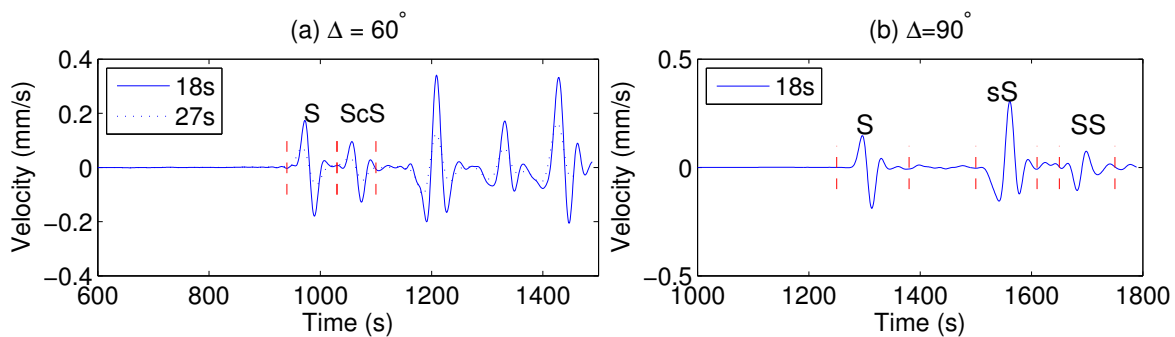


Figure 6.7: (a) Transverse velocity seismograms recorded at an epicentral distance of 60° for spectral-element simulations accurate down to periods of 18 s and 27 s, respectively. The S and ScS phases are labeled. (b) Transverse velocity seismograms recorded at an epicentral distance of 90° for spectral-element simulations accurate down to periods of 18 s. The S, sS and SS phases are labeled.

6.8.4 S and ScS Kernels

In this section we calculate the K_β traveltimes sensitivity kernels for the S and ScS phases at 60° for the Bolivian earthquake. These two phases are identified on transverse velocity seismograms as illustrated in Figure 6.7(a) for synthetics with a shortest period of 27 s and 18 s, respectively. The K_β Fréchet kernels corresponding to these phases are compared in Figure 6.8. Notice that they all nicely follow their respective theoretical raypaths (Figure 6.1a) and exhibit clear banana-doughnut shapes, except near the core-reflection point of the ScS phase where the two S legs come together, producing a complicated pattern on and right above the CMB. The kernels corresponding to an accuracy of 18 s and longer are generally sharper and have larger amplitudes than the kernels at periods of 27 s and longer, in agreement with the results obtained for the P kernels discussed in Section 6.8.1.

6.8.5 SKS Kernel

Next, we generate traveltimes sensitivity kernels for an 18 s SKS phase at an epicentral distance of 115° . Because the outer core is liquid, the SKS phase can only be observed on the vertical and radial components of PREM seismograms. In Figure 6.9(a) we identify the SKS phase on the North component. Its mantle K_β kernel and core K_α kernel are jointly

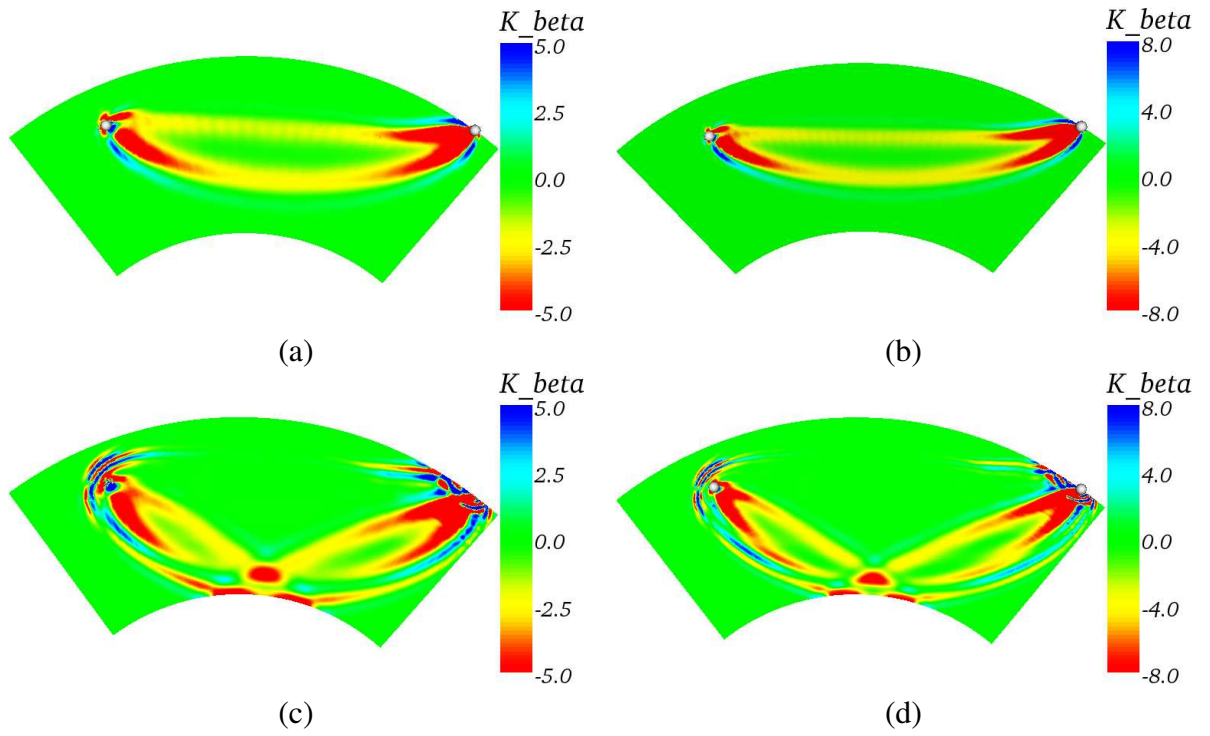


Figure 6.8: (a) K_β kernel for a 27 s S phase at 60° . (b) K_β kernel for an 18 s S phase at 60° . (c) K_β kernel for a 27 s ScS phase at 60° . (d) K_β kernel for an 18 s ScS phase at 60° .

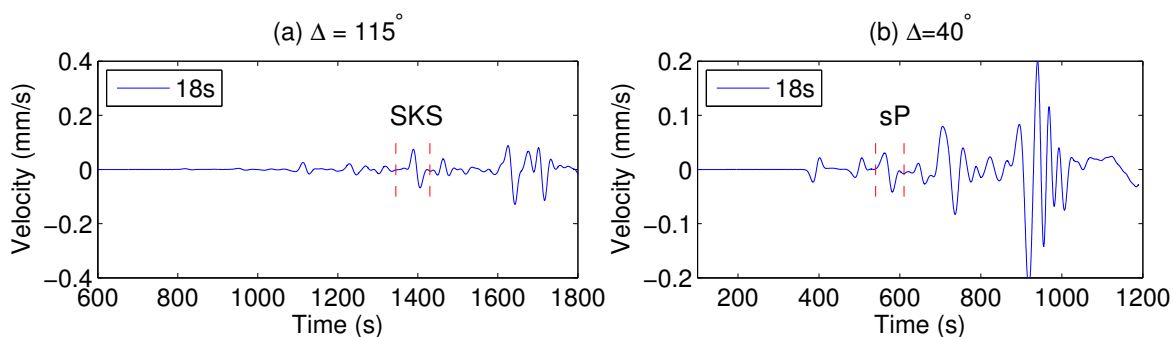


Figure 6.9: (a) North component of a velocity seismogram recorded for a station at an epicentral distance of 115° for a simulation accurate at periods of 18 s and longer. The prominent SKS phase is labeled. (b) Vertical component of the velocity seismogram recorded for a station at an epicentral distance of 40° for a simulation accurate at periods of 18 s and longer. The sP phase is labeled.

plotted in Figure 6.10. Overall the sensitivity kernel follows the SKS raypath, and due to the source radiation pattern the upper part of the banana-doughnut kernel shows much stronger sensitivity than the lower part of the kernel. Because the SKS phase is not the first arrival in the seismogram, mantle reverberated waves caused by reflections and refractions from both the Earth's surface and internal discontinuities may arrive in the SKS cross-correlation time window, and hence the sensitivity kernel has more structure than for a simple direct arrival such as P or P_{diff} . This is true in general for later arriving phases, especially when observed on the radial or vertical components, which involve many S-to-P and P-to-S conversions. As remarked earlier, the associated rapid oscillations in the amplitude of the kernel provide no constraints on relatively smooth structure.

6.8.6 Depth Phases

Since the Bolivian earthquake occurred at a depth of 647 km, we can clearly identify the depth phases on the three-component seismograms. In this section we pick the sP phase identified in a vertical velocity seismogram recorded at a distance of 40° and the sS phase in a transverse velocity seismogram recorded at a distance of 90° (Figures 6.9(b) and 6.7(b)). Figure 6.11(a) shows the K_α sensitivity kernel for the sP phase, while Figure 6.11(b) shows the corresponding K_β sensitivity kernel. Clearly the K_α kernel reflects the P leg of the sP

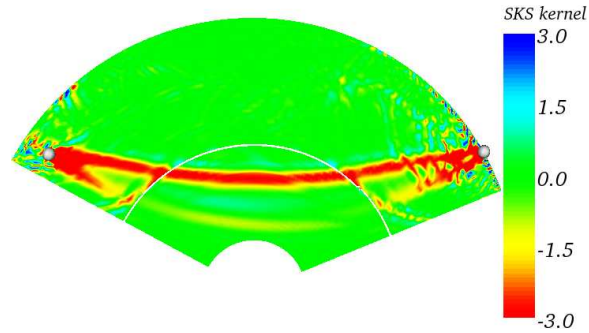


Figure 6.10: Source-receiver cross-section of the K_β kernel for an 18 s SKS phase recorded at an epicentral distance of 115° .

phase, while the K_β kernel represents the upgoing S leg of the sP phase. Note that the amplitude of the K_β kernel is much larger than the K_α kernel. Figure 6.11(c) shows the K_β sensitivity kernel for the sS arrival, which clearly follows the geometrical raypath and exhibits the characteristic banana-doughnut shape along the S leg. Notice that although the raypath of the sS arrival is at least 500 km above the CMB, as shown in Figure 6.1(b), the lower part of the banana-doughnut kernel actually touches the CMB and leaves a small elliptical area of sensitivity on the CMB, as shown in Figure 6.11(d).

6.8.7 Kernels for Surface-reflected Phases

Surface-reflected phases generally become distinct at epicentral distances $> 40^\circ$ and dominate the body-wave signals at a distance of approximately 100° . These arrivals are mostly sensitive to velocity structure in the upper mantle, in particular in the transition zone. We compute K_β sensitivity kernels for the SS phase recorded on the transverse component at a distance of 90° , the SP phase recorded on the vertical component at a distance of 100° , and the PP phase recorded on the vertical component at a distance of 115° (Figure 6.12). The K_β kernel for the SS phase follows the raypath shown in Figure 6.1 and displays a complicated folded pattern near the surface reflection point. Similar patterns are also observed in the K_α kernel for the PP phase. Figures 6.13(c) and 6.13(d) show the K_β and K_α kernels for the SP phase, and clearly the K_β kernel mainly reflects the S leg of the SP raypath, while the K_α corresponds to the P leg.

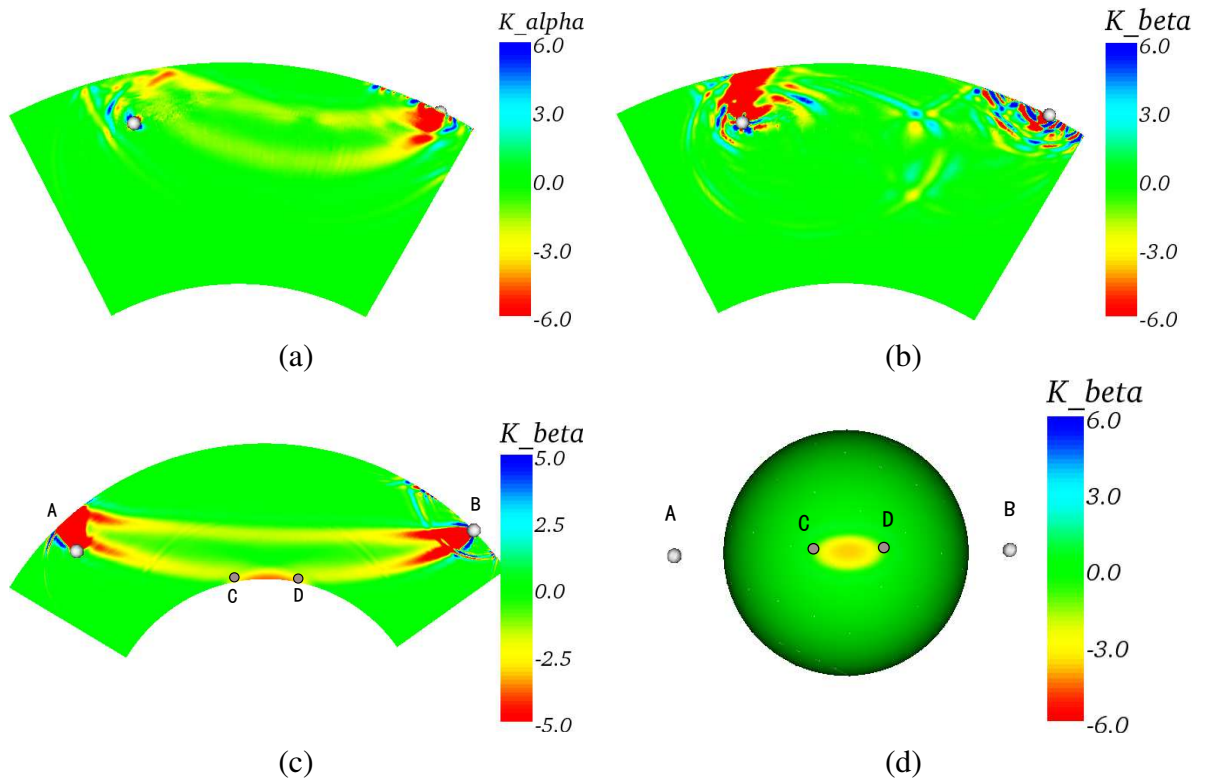


Figure 6.11: (a) K_α kernel for an 18 s sP phase recorded at a distance of 40°. (b) K_β kernel for an 18 s sP phase recorded at a distance of 40°. (c) Source-receiver cross-section of the K_β kernel for an 18 s sS arrival recorded at a distance of 90°. (d) Spherical cross-section on the CMB of the K_β Kernel for an 18 s sS arrival recorded at a distance of 90°.

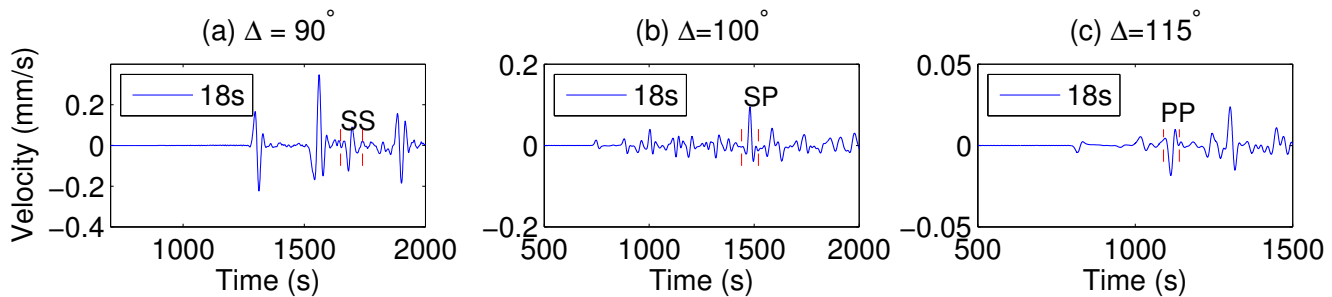


Figure 6.12: (a) Transverse component velocity seismogram recorded at a station at an epicentral distance of 90° for a spectral-element simulation accurate at periods of 18 s and longer. The emergent SS arrival is labeled. (b) Vertical component velocity seismogram recorded at a station at a distance of 100°. The SP phase is labeled. (c) Vertical component velocity seismogram recorded at 115°. The PP phase is denoted.

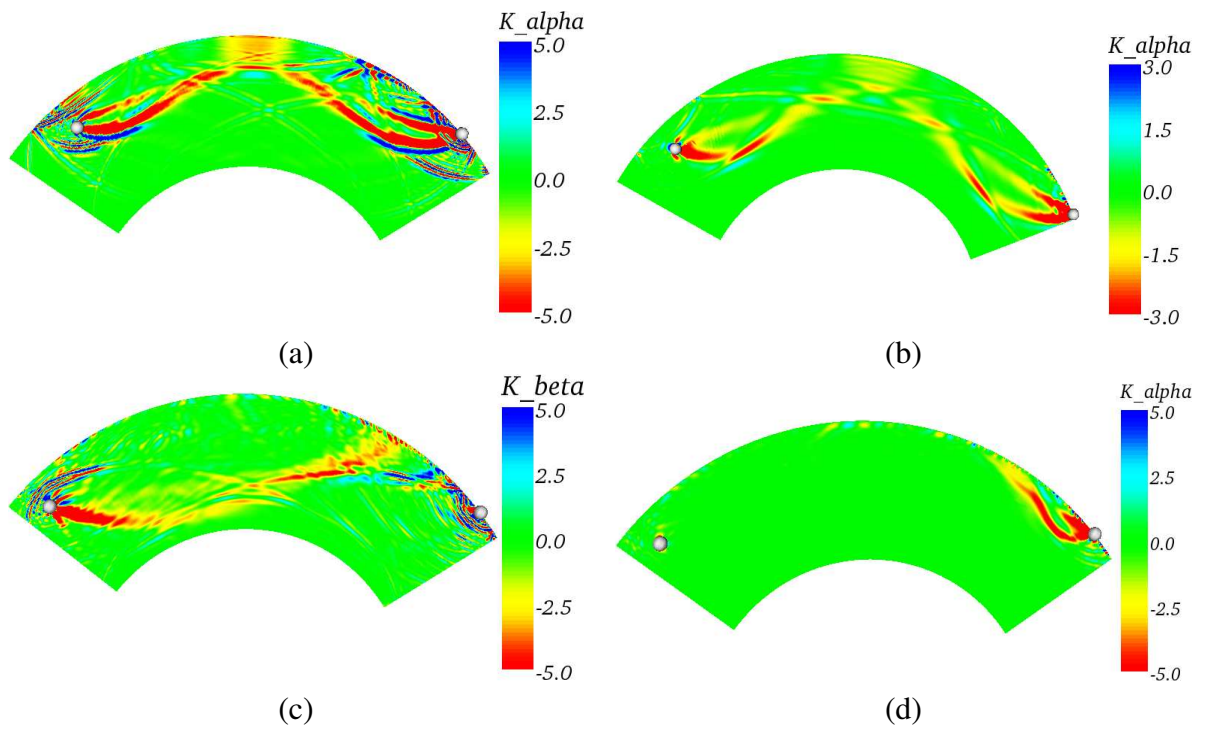


Figure 6.13: (a) K_β kernel for an 18 s SS phase recorded at a distance of 90° . (b) K_α kernel for an 18 s PP phase recorded at a distance of 115° . (c) K_β kernel for an 18 s SP phase recorded at a distance of 100° . (d) K_α kernel for an 18 s SP phase recorded at a distance of 100° .

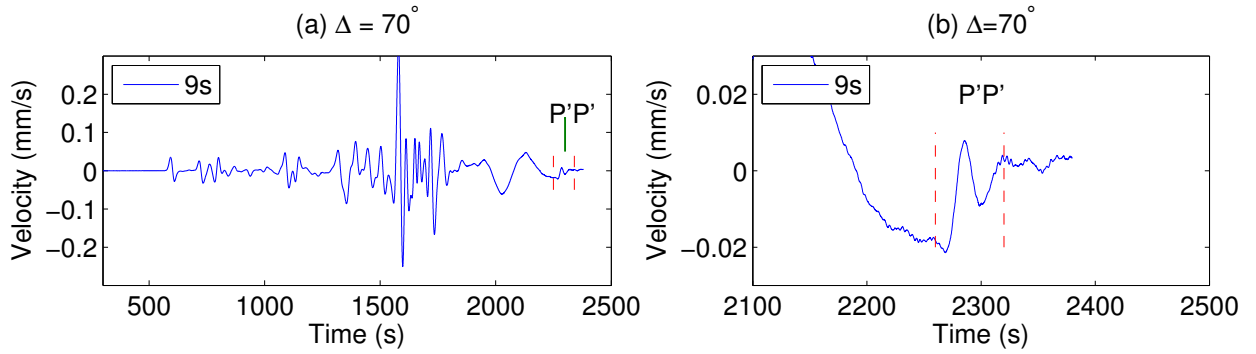


Figure 6.14: (a) Vertical component velocity seismogram recorded at a distance of 70° . (b) Last 300 s of the seismogram shown in (a) magnified by a factor of 10. The P, P, phase is labeled.

6.8.8 P, P, Kernel

The adjoint approach can readily be used to look at more exotic arrivals. As an example, we generate an 9 s K_α sensitivity kernels for the P, P, arrival (also known as PKPPKP): a PKP phase reflected off the Earth's surface. This phase, which arrives around 2300 s, is nearly unidentifiable in the complete vertical seismogram, (Figure 6.14(a)). However, when the last few hundred seconds of the seismograms are magnified by a factor of 10, we can identify the P, P, phase clearly at the tail of the surface waves. Its corresponding K_α kernel (Figure 6.15) clearly follows the PKPPKP raypath. Since this pulse resides in the tail of the surface waves in the very late part of the seismogram, the kernel contains 'noise' due to mantle and core reverberations.

6.9 Conjugate-gradient Method for Tomography Inversions

The Fréchet kernels discussed in section 6.8 are only the so-called "banana-doughnut kernels". One such kernel relates the perturbations of one particular traveltime delay to the perturbations of model parameters. For a finite-frequency travel-time tomographic inversion, we need to minimize the following misfit function:

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^N [T_i^{\text{obs}} - T_i(\mathbf{m})]^2, \quad (6.84)$$

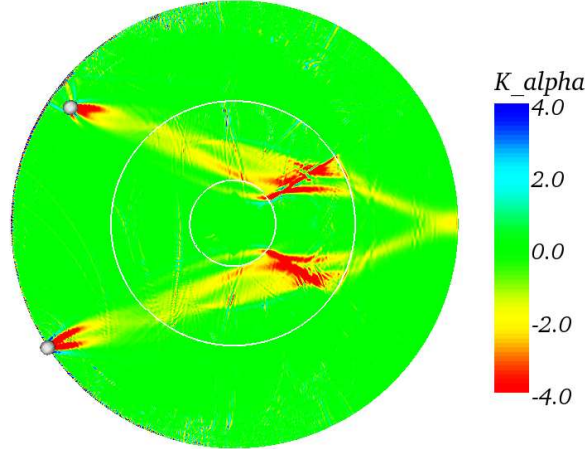


Figure 6.15: K_α kernel for a 9 s P, P , phase recorded at a distance of 70° .

where T_i^{obs} denotes the observed traveltimes for the i th source-receiver combination, $T_i(m)$ the predicted traveltimes based upon the current model m , and N the number of traveltimes measurements. We define the traveltimes anomaly $\Delta T_i = T_i^{\text{obs}} - T_i(m)$. Then the variation of the misfit function (6.84) becomes

$$\delta\chi = - \sum_{i=1}^N \Delta T_i \delta T_i, \quad (6.85)$$

Upon ignoring the boundary kernels in equation (6.53), we can relate the perturbations in the traveltimes to the fractional perturbation of the volumetric model parameters by

$$\delta T_i = \int_V K_i \delta \ln m \, d^3\mathbf{x}, \quad (6.86)$$

and if we define the misfit kernel as

$$K(\mathbf{x}) = - \sum_{i=1}^N \Delta T_i K_i(\mathbf{x}), \quad (6.87)$$

and expand our fractional phase-speed perturbations, $\delta \ln m(\mathbf{x})$, into basis functions $B(\mathbf{x})$:

$$\delta \ln m(\mathbf{x}) = \sum_{k=1}^M \delta m_k B_k(\mathbf{x}), \quad (6.88)$$

where δm_k , $k = 1, \dots, M$, represent the perturbed model coefficients, then the perturbation of the misfit function can be rewritten as

$$\delta\chi = \sum_k^M g_k \delta m_k \quad (6.89)$$

where the gradient of the misfit function χ with respect to the k th discrete model parameter m_k is defined as

$$g_k = \frac{\partial\chi}{\partial m_k} = \int_V K B_k d^3\mathbf{x}, \quad k = 1, \dots, M. \quad (6.90)$$

Notice that to obtain g_k , we only need to compute the misfit kernel $K(\mathbf{x})$, not the individual banana-doughnut kernels $K_i(\mathbf{x})$. Therefore, for each event, we generate the event kernel by putting the adjoint sources

$$f^\dagger(\mathbf{x}, t) = - \sum_{r=1}^{N_r} \Delta T_r \frac{1}{M_r} w_r(T-t) \partial_t s(\mathbf{x}_r, T-t) \delta(\mathbf{x} - \mathbf{x}_r) \quad (6.91)$$

simultaneously at the individual receivers and computing the adjoint wavefield through 3-D wave simulations. This greatly reduces the number of spectral-element simulations required to compute the gradient of the misfit function, to two simulations per event. Although the expression for the Hessian of the misfit function can be obtained in a similar manner as shown in *Tape et al.* [2006], individual $G_{ik} = \int_V K_i B_k d^3\mathbf{x}$ has to be computed, meaning every K_i has to be computed and stored separately, which makes it computationally prohibitive to calculate the Hessian when a large number of events, source-receiver pairs, and traveltime measurements are involved [*Tape et al.*, 2006]. Now a feasible tomographic inversion problem can be posed as the following: Given a set of discrete model parameters $\mathbf{m} = [m_k], k = 1, \dots, M$ that describes the velocity structure of the earth, a misfit function $\chi(\mathbf{m})$ that measures the travel-time delays between the observed seismograms and the synthetic seismograms for this set of model parameters \mathbf{m} , and the ability to only compute the gradient of the misfit function g_k , how can one iteratively update the velocity model? The conjugate gradient methods with approximate updates of the search directions have been proposed to solve this inversion problem, and have been successfully

applied to both source and structural inversions in synthetic problems [*Tromp et al.*, 2005; *Tape et al.*, 2006].

6.10 Discussion

As shown in the previous section, we compute the gradient of the misfit function using simultaneous time-reversed signals (equation (6.91)) at the receivers as the adjoint sources. Therefore, it is very crucial for us to make the measurements of travel-time delay ΔT_i 's and their associated error bars carefully due to the significant computation time required to generate the misfit kernels $K(\mathbf{x})$. On the other hand, we will collect both the cross-correlation travel-time measurements for the body wave phases, and the frequency-dependent phase-delay measurements for the surface waves. With these datasets ready, we can compute the gradient of the misfit function for any given 3-D velocity model, and by iteratively updating the velocity model and remake these travel-time measurements, we can greatly improve our knowledge of the 3-D velocity structures of the earth's interior.

Appendix A

Attenuation by Standard Linear Solids

A.1 Complex Wave Number

We define the complex wave number as a function of angular frequency ω :

$$k(\omega) = \frac{\omega}{V_p(\omega)} - i\alpha(\omega), \quad (\text{A.1})$$

where $V_p(\omega)$ is the phase velocity and $\alpha(\omega)$ is the attenuation factor.

A.2 Theory of viscoelasticity

Boltzmann's after-effect equation (superposition principle and causality principle), which describes the relationship between the stress and the strain in a viscoelastic medium,

$$\epsilon(t) = \int_0^{\infty} \dot{\sigma}(t - \theta)\phi(\theta) d\theta \quad (\text{A.2})$$

$$= \dot{\sigma} * \phi, \quad (\text{A.3})$$

states that the strain at time t , $\epsilon(t)$, is caused linearly by the total history of stress $\sigma(\tau)$ up to time t . We can define the **creep function** as $\sigma = \text{const}$, $\epsilon = \epsilon(t)$, and the **relaxation function** as $\epsilon = \text{const}$, $\sigma = \sigma(t)$. Therefore when $\sigma(t) = H(t)$, the Heaviside function, the strain is given by $\epsilon(t) = \phi(t)$.

A.3 Standard Linear Solids

A single standard linear solid (Figure A.1) satisfies the following first order differential equation

$$\sigma + \tau_\sigma \dot{\sigma} = M_R(\epsilon + \tau_\epsilon \dot{\epsilon}), \quad (\text{A.4})$$

where $M_R = k_0$ is the deformation (relaxed elastic) modulus; $\tau_\sigma = \frac{\eta}{k_1}$ is the stress relaxation time at constant ϵ ; and $\tau_\epsilon = (1 + \frac{k_1}{k_0})\tau_\sigma$ is the strain relaxation time at constant σ . For unit constant stress $\sigma(t) = H(t)$, we have

$$\begin{aligned} \epsilon(t) &= \frac{1}{M_R} \left[1 - \left(1 - \frac{\tau_\sigma}{\tau_\epsilon}\right) \exp(-t/\tau_\epsilon) \right] H(t) = \phi(t) \\ t = 0, \epsilon(t) &= \frac{1}{M_R} \frac{\tau_\sigma}{\tau_\epsilon} \sigma(t) = \frac{1}{M_U} \sigma(t) \\ t = \infty, \epsilon(t) &= \frac{1}{M_R} \sigma(t), \end{aligned} \quad (\text{A.5})$$

where the relaxed modulus $M_U = \frac{\tau_\epsilon}{\tau_\sigma} M_R$. Therefore, for arbitrary stress $\sigma(t)$,

$$\epsilon(t) = \sigma(t) * \phi(t) = \frac{\sigma(t)}{M_R} - \frac{1}{M_R} \left(1 - \frac{\tau_\sigma}{\tau_\epsilon}\right) \int_0^\infty e^{-\theta/\tau_\epsilon} \dot{\sigma}(t - \theta) d\theta. \quad (\text{A.6})$$

For sinusoidal stress $\sigma(t) = e^{i\omega t}$, the strain becomes

$$\begin{aligned} \epsilon(t) &= \frac{\sigma(t)}{M_R} \left[1 - \left(1 - \frac{\tau_\sigma}{\tau_\epsilon}\right) \int_0^\infty i\omega e^{(\frac{1}{\tau_\epsilon} - i\omega)\theta} d\theta \right] \\ &= \frac{\sigma(t)}{M_R} (A(\omega) - iB(\omega)) = \frac{\sigma(t)}{M_c(\omega)}, \end{aligned} \quad (\text{A.7})$$

where

$$A(\omega) = 1 - \frac{\omega^2 \tau_\epsilon^2}{1 + \omega^2 \tau_\epsilon^2} \left(1 - \frac{\tau_\sigma}{\tau_\epsilon}\right) \quad (\text{A.8})$$

$$B(\omega) = \frac{\omega(\tau_\epsilon - \tau_\sigma)}{1 + \omega^2 \tau_\epsilon^2}. \quad (\text{A.9})$$

If we define $x^2 = \frac{\sqrt{A^2+B^2}+A}{2}$, $y^2 = \frac{\sqrt{A^2+B^2}-A}{2}$, we have $A - iB = (x + iy)^2$. If we define the elastic velocity V_e such that $M_R = \rho V_e^2$, and the complex velocity $V(\omega)$ such

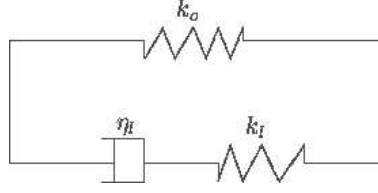


Figure A.1: Single standard linear solid

that $M(\omega) = \rho V(\omega)^2$, then the complex wave number becomes

$$k(\omega) = \frac{\omega}{V(\omega)} = \frac{\omega}{V_e} (A - iB)^{1/2} = \frac{\omega}{V_p(\omega)} - i\alpha(\omega). \quad (\text{A.10})$$

Then the expressions for the phase velocity $V_p(\omega)$ and the amplification factor $\alpha(\omega)$ are given by

$$V_p^2(\omega) = \frac{2V_e^2}{B^2} \Omega(\omega) \quad (\text{A.11})$$

$$\alpha^2(\omega) = \frac{\omega^2}{2V_e^2} \Omega(\omega), \quad (\text{A.12})$$

where $\Omega(\omega) = \sqrt{A^2 + B^2} - A$. Therefore, for a constant stress, the phase velocity $V_p^2(\omega = 0) = \frac{M_R}{\rho}$, and at very high frequency, the phase velocity becomes $V_p^2(\omega = \infty) = \frac{M_R}{\rho} \frac{\tau_\epsilon}{\tau_\sigma} > \frac{M_R}{\rho}$. We define the internal friction coefficient $Q^{-1} = B/A = \frac{\omega(\tau_\epsilon - \tau_\sigma)}{1 + \omega^2 \tau_\epsilon \tau_\sigma}$, which satisfies that $Q^{-1}(\omega = 0, \infty) = 0$, and $Q^{-1}(\omega = \frac{1}{\sqrt{\tau_\epsilon \tau_\sigma}}) = \frac{\tau_\epsilon - \tau_\sigma}{2\sqrt{\tau_\epsilon \tau_\sigma}}$.

A.4 Multiple Standard Linear Solid

We use a series of standard linear solids (Figure A.2) to mimic the constant Q over a range of frequency bands in seismology, in which case equation (A.5) becomes

$$\phi(t) = \frac{1}{M_R} \left[1 - \sum_{k=1}^N \left(1 - \frac{\tau_{\sigma k}}{\tau_{\epsilon k}} \right) \exp(-t/\tau_{\epsilon k}) \right] H(t) \quad (\text{A.13})$$

For sinusoid stress $\sigma(t) = e^{i\omega t}$, equation (A.7) becomes

$$\epsilon(t) = \frac{\sigma(t)}{M_R}(A(\omega) - iB(\omega)), \quad (\text{A.14})$$

Where

$$A(\omega) = 1 - \sum_{k=1}^N \frac{\omega^2 \tau_{\epsilon k} (\tau_{\epsilon k} - \tau_{\sigma k})}{1 + \omega^2 \tau_{\epsilon k}^2} \quad (\text{A.15})$$

$$B(\omega) = \sum_{k=1}^N \frac{\omega (\tau_{\epsilon k} - \tau_{\sigma k})}{1 + \omega^2 \tau_{\epsilon k}^2} \quad (\text{A.16})$$

and the unrelaxed modulus is given by

$$Q^{-1} = B/A \quad (\text{A.17})$$

$$M_U = M(\omega = \infty) = M_R / [1 - \sum \frac{\tau_{\epsilon k} - \tau_{\sigma k}}{\tau_{\epsilon k}}]. \quad (\text{A.18})$$

The expressions for the complex phase velocity $V_p(\omega)$, and the attenuation factor $\alpha(\omega)$, i.e. equation (A.12), still hold, and when Q^{-1} is small, $B \ll A$, then $V_p = \frac{V_e}{\sqrt{A}}$.

A.5 Constant Q

In seismology, given constant Q over a wide frequency band, the complex velocities at two different angular frequencies ω_1 and ω_2 are related by

$$\frac{V_p(\omega_2)}{V_p(\omega_1)} = 1 + \frac{1}{\pi Q} \ln \frac{\omega_2}{\omega_1} \quad (\text{A.19})$$

For example, in the Preliminary Reference Earth Model (PREM) model, all the elastic moduli are given at the reference period of 1 second, and the elastic modulus at any arbitrary angular frequency ω may be computed similarly to equation (A.19) [Liu *et al.*, 1976; Dziewonski and Anderson, 1981].

Now the question becomes how to pick $\tau_{\sigma k}, \tau_{\epsilon k}, k = 1, \dots, N$ in order to approximate the constant Q ($Q = Q_0$) effect over a range of frequency we are interested in? We usually

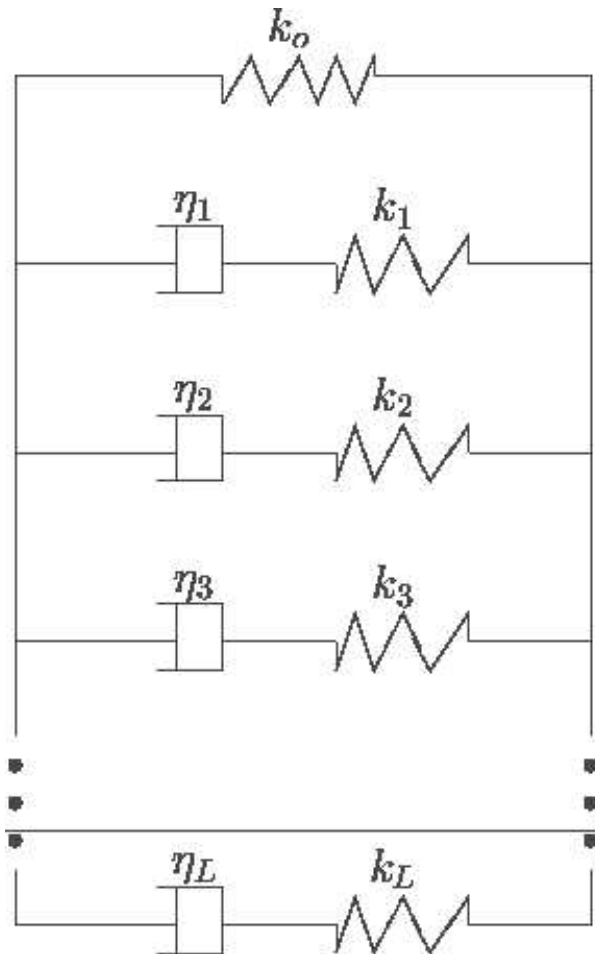


Figure A.2: Multiple standard linear solids

choose three standard linear solids, i.e., $N = 3$, and pick N_f frequencies ω_p , $p = 1, \dots, N_f$, to minimize the following misfit function:

$$\chi(\tau_{\sigma k}, \tau_{\epsilon k}) = \sum_{p=1}^{N_f} [Q^{-1}(\tau_{\sigma k}, \tau_{\epsilon k}, \omega_p) - Q_0]^2, \quad (\text{A.20})$$

where the expressions for $Q^{-1}(\tau_{\sigma k}, \tau_{\epsilon k}, \omega_p)$ are given by equation (A.17). In practice, $\tau_{\sigma k}$'s are chosen first, and the misfit function is reduced to only a function of $\tau_{\epsilon k}$, $k = 1, \dots, N$. Because only a very small number of variables is involved in this optimization problem, the simplex method can be applied to obtain the best set of $\tau_{\epsilon k}$'s.

A.6 Implementations of Anelasticity in SEM

The stress-strain relation in an anelastic medium can be expressed in its most general form by

$$T(t) = \int_{-\infty}^t \partial_t \mathbf{C}(t - t') : \nabla \mathbf{S}(t') dt', \quad (\text{A.21})$$

where \mathbf{C} is the fourth-order anelastic tensor, and $bS(t')$ is the displacement at time $t = t'$. Upon taking the Laplace transform of both sides, and inserting the expressions for anelastic modulus (A.13), we obtain

$$\begin{aligned} \bar{T} &= s\bar{\mathbf{C}} : \nabla \bar{\mathbf{S}} \\ &= sC_{ijkl}^R \left[\frac{1}{s} - \sum_{n=1}^N \left(1 - \frac{\tau_{ijkl}^{\epsilon n}}{\tau^{\sigma n}}\right) \frac{1}{s + \frac{1}{\tau^{\epsilon n}}} \right] \partial_k \bar{S}_l \\ &= C_{ijkl}^R \left[1 - \sum_{n=1}^N \left(1 - \frac{\tau_{ijkl}^{\epsilon n}}{\tau^{\sigma n}}\right) \right] \partial_k \bar{S}_l - \sum_{n=1}^N \bar{\mathbf{R}}_{ij}^n \\ &= C_{ijkl}^U \partial_k \bar{S}_l - \sum_{n=1}^N \bar{\mathbf{R}}_{ij}^n, \end{aligned} \quad (\text{A.22})$$

where the unrelaxed modulus \mathbf{C}^U is related to the relaxed modulus by

$$C_{ijkl}^U = C_{ijkl}^R \left[1 - \sum_{n=1}^N \left(1 - \frac{\tau_{ijkl}^{\epsilon n}}{\tau^{\sigma n}}\right) \right], \quad (\text{A.23})$$

while the memory variables are defined by

$$\bar{\mathbf{R}}_{ij}^n = -C_{ijkl}^R \left(1 - \frac{\tau_{ijkl}^{\epsilon n}}{\tau^{\sigma n}}\right) \frac{\tau^{\epsilon n}}{s + \frac{1}{\tau^{\epsilon n}}} \quad (\text{A.24})$$

or inverse Laplace transform back to the time domain as a first-order differential equation

$$\partial_t \mathbf{R}_{ij}^n = -\mathbf{R}_{ij}^n / \tau^{\sigma n} + \delta C_{ijkl} \nabla S_{kl} / \tau^{\sigma n}, \quad (\text{A.25})$$

which is solved to update the memory variable at every time step.

Appendix B

More on Spectral-Element Moment-Tensor Inversion Theory

Starting from the misfit function (4.4), we can either perform linear or non-linear source inversions depending on our choice of source parameters and constraints. The waveform misfit function with time shifts follows the same formulation, except that a time lag must be applied to both the synthetics and the Fréchet derivatives. For simplicity, we set the normalization factors A_1 and A_2 to 1.

B.1 Linear inversions

If we set $\lambda = 1$ (i.e., we do not include the envelope misfit function), $\mu_2 = 0$ (i.e., we do not require a double-couple solution), and invert only for the six elements of the moment tensor by holding the earthquake location fixed, then the problem becomes linear. In this case the synthetics $s_i(t, \mathbf{m})$ can be expressed as linear combinations of the Fréchet derivatives with respect to moment-tensor elements $\frac{\partial s_i}{\partial m_k}(t)$:

$$s_i(t, \mathbf{m}) = \sum_{k=1}^6 \frac{\partial s_i}{\partial m_k}(t) m_k. \quad (\text{B.1})$$

Substitution of (B.1) into (4.4) yields

$$\frac{\partial E_1}{\partial m_j} + \mu_1 \frac{\partial C_1}{\partial m_j} = - \sum_{i=1}^N w_i \int \left[d_i(t) - \sum_{k=1}^6 \frac{\partial s_i}{\partial m_k}(t) m_k \right] \frac{\partial s_i}{\partial m_j}(t) dt + \mu_1 \frac{\partial C_1}{\partial m_j} = 0. \quad (\text{B.2})$$

If we define

$$H_{jk} = \sum_{i=1}^N w_i \int \frac{\partial s_i}{\partial m_k}(t) \frac{\partial s_i}{\partial m_j}(t) dt, \quad G_j = \sum_{i=1}^N w_i \int d_i(t) \frac{\partial s_i}{\partial m_j}(t) dt, \quad (\text{B.3})$$

then (B.2) becomes

$$\sum_{k=1}^6 H_{jk} m_k - G_j + \mu_1 \frac{\partial C_1}{\partial m_j} = 0. \quad (\text{B.4})$$

Let $\mathbf{U} = \{\partial C_1 / \partial m_j; j = 1, \dots, 6\} = [1, 1, 1, 0, 0, 0]^T$, where a superscript T denotes the transpose, then (B.4) becomes the following matrix equation:

$$\begin{bmatrix} \ddots & & \ddots & \vdots \\ & H_{jk} & U_j & \\ \ddots & & \ddots & \vdots \\ \dots & U_k & \dots & 0 \end{bmatrix} \begin{bmatrix} \vdots \\ m_k \\ \vdots \\ \mu_1 \end{bmatrix} = \begin{bmatrix} \vdots \\ G_j \\ \vdots \\ 0 \end{bmatrix}. \quad (\text{B.5})$$

This equation can be solved for the six elements of the moment tensor. Notice that (H_{jk}) is a positive-definite symmetric matrix. Therefore, when the zero-trace constraint is not invoked, a unique solution is guaranteed to exist.

B.2 Non-linear inversions

When $\lambda \neq 1$ or $\mu_2 \neq 0$, equation (4.4) becomes non-linear with respect to \mathbf{m} , either because we need to invert for the event location or because the derivative of the envelope misfit function is non-linear. In any event, an initial solution \mathbf{m}^0 is required to solve (4.4). In our case we use point source parameters inverted using alternative methods [Thio and Kanamori, 1995; Zhu and Helmberger, 1996] as the starting solution. The synthetics can be linearized with respect to the initial source parameters:

$$s_i(t, \mathbf{m}) = s_i(t, \mathbf{m}^0) + \sum_{k=1}^n \frac{\partial s_i}{\partial m_k}(t, \mathbf{m}^0) (m_k - m_k^0), \quad (\text{B.6})$$

and (4.5) becomes

$$\lambda \frac{\partial E_1}{\partial m_j}(\mathbf{m}^0) + (1 - \lambda) \frac{\partial E_2}{\partial m_j}(\mathbf{m}^0) + \mu_1 \frac{\partial C_1}{\partial m_j}(\mathbf{m}^0) + \mu_2 \frac{\partial C_2}{\partial m_j}(\mathbf{m}^0) = 0, \quad j = 1, \dots, n. \quad (\text{B.7})$$

In the following we will consider the case of a double-couple source inversion in detail; other non-linear inversions have similar characteristics. Using (B.6) and assuming that the initial solution is sufficiently close to the true solution, we obtain

$$\frac{\partial E_1}{\partial m_j}(\mathbf{m}^0) = - \sum_{i=1}^N w_i \int \left[d_i(t) - s_i(t, \mathbf{m}^0) - \sum_{k=1}^n \frac{\partial s_i}{\partial m_k}(t, \mathbf{m}^0)(m_k - m_k^0) \right] \frac{\partial s_i}{\partial m_j}(t, \mathbf{m}^0) dt, \quad (\text{B.8})$$

$$\frac{\partial E_2}{\partial m_j}(\mathbf{m}^0) = - \sum_{i=1}^N w_i \int \left[e(d_i(t)) - e(s_i(t, \mathbf{m}^0)) - \sum_{k=1}^n \frac{\partial e(s_i)}{\partial m_k}(t, \mathbf{m}^0)(m_k - m_k^0) \right] \frac{\partial e(s_i)}{\partial m_j}(t, \mathbf{m}^0) dt,$$

where

$$\frac{\partial e(s_i)}{\partial m_j}(t, \mathbf{m}^0) = e(s_i(t, \mathbf{m}^0))^{-1/2} \left[s_i(t, \mathbf{m}^0) \frac{\partial s_i}{\partial m_j}(t, \mathbf{m}^0) + \hat{s}_i(t, \mathbf{m}^0) \frac{\partial \hat{s}_i}{\partial m_j}(t, \mathbf{m}^0) \right], \quad (\text{B.9})$$

and $\hat{s}(t, \mathbf{m})$ denotes the Hilbert transform of $s(t, \mathbf{m})$. Let

$$H_{jk} = \sum_{i=1}^N w_i \int \left[\lambda \frac{\partial s_i}{\partial m_k}(t, \mathbf{m}^0) \frac{\partial s_i}{\partial m_j}(t, \mathbf{m}^0) + (1 - \lambda) \frac{\partial e(s_i)}{\partial m_k}(t, \mathbf{m}^0) \frac{\partial e(s_i)}{\partial m_j}(t, \mathbf{m}^0) \right] dt, \quad (\text{B.10})$$

$$G_j = \sum_{i=1}^N w_i \int \left\{ \lambda [d_i(t) - s_i(t, \mathbf{m}^0)] \frac{\partial s_i}{\partial m_j}(t, \mathbf{m}^0) + (1 - \lambda) [e(d_i(t)) - e(s_i(t, \mathbf{m}^0))] \frac{\partial e(s_i)}{\partial m_j}(t, \mathbf{m}^0) \right\} dt, \quad (\text{B.11})$$

and rewrite (B.7) as

$$f_j(\mathbf{m}, \mu_1, \mu_2) = \sum_{k=1}^n H_{jk}(m_k - m_k^0) - G_j + \mu_1 \frac{\partial C_1}{\partial m_j}(\mathbf{m}) + \mu_2 \frac{\partial C_2}{\partial m_j}(\mathbf{m}) = 0, \quad j = 1, \dots, n, \quad (\text{B.12})$$

with constraints $f_{n+1}(\mathbf{m}) = C_1(\mathbf{m}) = 0$ and $f_{n+2}(\mathbf{m}) = C_2(\mathbf{m}) = 0$. To solve the non-linear equations (B.12), we use the non-linear Newton's iterative solver [Conte and Boor, 1980]. Let us define a vector $\mathbf{x} = [\mathbf{m}, \mu_1, \mu_2]$ and an initial solution vector $\mathbf{x}^0 = [\mathbf{m}^0, \mu_1^0, \mu_2^0]$ (we usually start with $\mu_1^0 = \mu_2^0 = 0$), then the $(i + 1)$ th iteration can be expressed as

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}^i)(\mathbf{x}^{i+1} - \mathbf{x}^i) = -\mathbf{f}(\mathbf{x}^i). \quad (\text{B.13})$$

If we define

$$\mathbf{Y} = \left(\frac{\partial^2 C_2}{\partial m_j \partial m_k} \right) = \begin{bmatrix} 0 & m_3 & m_2 & 0 & 0 & -2m_6 \\ m_3 & 0 & m_1 & 0 & -2m_5 & 0 \\ m_2 & m_1 & 0 & -2m_4 & 0 & 0 \\ 0 & 0 & -2m_4 & -2m_3 & 2m_6 & 2m_5 \\ 0 & -2m_5 & 0 & 2m_6 & -2m_2 & 2m_5 \\ -2m_6 & 0 & 0 & 2m_5 & 2m_4 & -2m_1 \end{bmatrix} \quad j, k = 1, \dots, n, \quad (\text{B.14})$$

$$\mathbf{V} = \{\partial C_2 / \partial m_j; j = 1, \dots, 6\} \quad (\text{B.15})$$

$$= [m_2 m_3 - m_6^2, m_1 m_3 - m_5^2, m_1 m_2 - m_4^2, -2(m_3 m_4 - m_5 m_6), \\ -2(m_2 m_5 - m_4 m_6), -2(m_1 m_6 - m_4 m_5)]^T,$$

and

$$\mathbf{D} = \left(\frac{\partial f_j}{\partial x_k} \right) = \begin{bmatrix} \ddots & & \ddots & \vdots & \vdots \\ & H_{jk} + \mu_2 Y_{jk} & & U_j & V_j \\ \ddots & & \ddots & \vdots & \vdots \\ \dots & U_k & \dots & 0 & 0 \\ \dots & V_k & \dots & 0 & 0 \end{bmatrix} \quad j, k = 1, \dots, n, \quad (\text{B.16})$$

then (B.13) becomes a linear system

$$\mathbf{D}^i(\mathbf{x}^{i+1} - \mathbf{x}^i) = -\mathbf{f}(\mathbf{x}^i), \quad (\text{B.17})$$

and the source model parameters \mathbf{m} are determined by iterating this linear system with respect to \mathbf{x} .

Appendix C

More on the Adjoint Method

C.1 Absorbing Boundaries

A regional Earth model has both a free surface Σ and an artificial boundary Γ , such that the model volume Ω has a boundary $\partial\Omega = \Sigma \cup \Gamma$. On the artificial boundary Γ , energy needs to be absorbed in order to mimic a semi-infinite medium. In an isotropic medium, this may be accomplished based upon a paraxial equation [*Clayton and Engquist, 1977; Quarteroni et al., 1998*]:

$$\hat{\mathbf{n}} \cdot \mathbf{T} = \rho [\alpha(\hat{\mathbf{n}}\hat{\mathbf{n}}) + \beta(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}})] \cdot \partial_t \mathbf{s} \equiv \mathbf{B} \cdot \partial_t \mathbf{s} \quad \text{on } \Gamma, \quad (\text{C.1})$$

thereby defining the tensor \mathbf{B} . The unit outward normal to the absorbing boundary is denoted by $\hat{\mathbf{n}}$, α denotes the P-wave speed, β denotes the S-wave speed, and \mathbf{I} denotes the 3×3 identity tensor. The absorbing boundary condition (C.1) perfectly absorbs waves impinging at a right angle to the boundary, but is less effective for waves that graze the boundary [*Clayton and Engquist, 1977*]. A much more effective absorbing boundary may be obtained based upon the perfectly matched layer (PML) methodology [*Bérenger, 1994; Collino and Tsogka, 2001; Komatitsch and Tromp, 2003; Festa and Vilotte, 2005*]. The PML approach amounts to solving an alternative wave equation in a thin shell surrounding the artificial boundary Γ that perfectly absorbs energy leaving the model domain Ω . One can obtain the adjoint equations associated with the PML region, but this is beyond the scope of this appendix. For the purposes of the present discussion the simpler one-way

condition (C.1) will suffice.

In the variation of the action (5.9) the boundary integral represented by the last term needs to be split in terms of contributions from the free surface Σ and the absorbing boundary Γ :

$$\begin{aligned} & \int_0^T \int_{\partial\Omega} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt \\ &= \int_0^T \int_{\Sigma} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt \\ &+ \int_0^T \int_{\Gamma} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt, \quad (\text{C.2}) \end{aligned}$$

Perturbing the free surface boundary condition (5.4) implies $\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s}) = \mathbf{0}$ on Σ , and perturbing the absorbing boundary condition (C.1) implies $\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s}) = \delta\mathbf{B} \cdot \partial_t\delta\mathbf{s} + \mathbf{B} \cdot \partial_t\delta\mathbf{s}$ on Γ . Without loss of generality, we are of course free to choose our artificial boundary Γ such that the perturbation $\delta\mathbf{B}$ vanishes: $\delta\mathbf{B} = \mathbf{0}$, e.g., by tapering the perturbed model parameters to zero. Upon integrating the temporal integration on the absorbing boundary Γ by parts we obtain

$$\begin{aligned} & \int_0^T \int_{\partial\Omega} \lambda \cdot [\hat{\mathbf{n}} \cdot (\delta\mathbf{c} : \nabla\mathbf{s} + \mathbf{c} : \nabla\delta\mathbf{s})] - \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt = - \int_0^T \int_{\Sigma} \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt \\ &+ \int_{\Gamma} [\lambda \cdot \mathbf{B} \cdot \delta\mathbf{s}]_T \, d^2\mathbf{x} - \int_0^T \int_{\Gamma} [\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) + \mathbf{B} \cdot \partial_t\lambda] \cdot \delta\mathbf{s} \, d^2\mathbf{x} \, dt. \quad (\text{C.3}) \end{aligned}$$

Thus we see that for the action (5.10) to make the Lagrange multiplier field vanish is subject to the free surface condition

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = \mathbf{0} \quad \text{on } \Sigma, \quad (\text{C.4})$$

and the absorbing boundary condition

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = -\mathbf{B} \cdot \partial_t\lambda \quad \text{on } \Gamma, \quad (\text{C.5})$$

where we have used the end conditions (5.13). This implies that the adjoint wave equation

(5.16) is subject to the free surface boundary conditions

$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{0} \quad \text{on } \Sigma, \quad (\text{C.6})$$

and the absorbing boundary condition

$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{B} \cdot \partial_t \mathbf{s}^\dagger \quad \text{on } \Gamma. \quad (\text{C.7})$$

We deduce that the adjoint wavefield is determined by exactly the same equations as the regular wavefield, with the exception of the source term.

C.2 Finite Source

In the case of a finite fault plane Σ_s , the source term may be written in terms of the moment-density tensor \mathbf{m} as

$$\mathbf{f} = -\mathbf{m}(\mathbf{x}_s, t) \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) \quad \text{on } \Sigma_s. \quad (\text{C.8})$$

The perturbation to the finite source (C.8) may be written in the form

$$\delta \mathbf{f} = -\delta \mathbf{m} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) - \delta h \mathbf{m} \cdot [\hat{\boldsymbol{\nu}} \cdot \nabla \nabla \delta(\mathbf{x} - \mathbf{x}_s)], \quad (\text{C.9})$$

where $\delta \mathbf{m}$ denotes the perturbed moment-density tensor and $\delta h \hat{\boldsymbol{\nu}}$ the fault plane mislocation δh in the direction of the fault normal $\hat{\boldsymbol{\nu}}$. Upon substituting (C.9) into the gradient of the misfit function (5.20), using the properties of the Dirac delta distribution, we obtain

$$\begin{aligned} \delta \chi = & \int_{\Omega} (\delta \rho K_\rho + \delta \mathbf{c} :: \mathbf{K}_c) d^3 \mathbf{x} + \int_0^T \int_{\Sigma_s} \delta \mathbf{m}(\mathbf{x}, t) : \boldsymbol{\epsilon}^\dagger(\mathbf{x}, T - t) d\mathbf{x}^2 dt \\ & + \int_0^T \int_{\Sigma_s} \delta h(\mathbf{x}) \mathbf{m}(\mathbf{x}, t) : [\hat{\boldsymbol{\nu}}(\mathbf{x}) \cdot \nabla_s] \boldsymbol{\epsilon}^\dagger(\mathbf{x}, T - t) d\mathbf{x}^2 dt. \end{aligned} \quad (\text{C.10})$$

This result may be used to improve finite fault models of large earthquakes. Note that if one is not concerned about perturbations in Earth structure, one only needs to track the time dependence of the adjoint strain $\boldsymbol{\epsilon}^\dagger$ in the vicinity of the fault plane.

Bibliography

- Aki, K., and P. G. Richards, *Quantitative Seismology, Theory and Methods*, W. H. Freeman, San Francisco, ISBN: 0716710595, 1980.
- Alford, R., K. Kelly, and D. Boore, Accuracy of finite difference modeling of the acoustic wave equation, *Geophysics*, 39, 834–842, 1974.
- Bao, H., J. Bielak, O. Ghattas, L. F. Kallivokas, D. R. O’Hallaron, J. R. Shewchuk, and J. Xu, Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers, *Comput. Methods Appl. Mech. Engrg.*, 152, 85–102, 1998.
- Bérenger, J. P., A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.*, 114, 185–200, 1994.
- Chaljub, E., Y. Capdeville, and J. P. Vilotte, Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, *J. Comput. Phys.*, 187, 457–491, 2003.
- Clayton, R., and B. Engquist, Absorbing boundary conditions for acoustic and elastic wave equations, *Bull. Seism. Soc. Am.*, 67, 1529–1540, 1977.
- Collino, F., and C. Tsogka, Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media, *Geophysics*, 66(1), 294–307, 2001.
- Conte, S. D., and C. Boor, *Elementary Numerical Analysis: An Algorithm Approach*, third ed., McGraw-Hill Book Company, 1980.

- Dahlen, F. A., and J. Tromp, *Theoretical Global Seismology*, Princeton University Press, Princeton, 1998.
- Dahlen, F. A., G. Nolet, and S.-H. Hung, Fréchet kernels for finite-frequency traveltime – I. Theory, *Geophys. J. Int.*, *141*, 157–174, 2000.
- der Hilst, R. D. V., R. D. Widiantoro, and E. R. Engdahl, Evidence for deep mantle circulation from global tomography, *Nature*, *386*, 1997.
- Dreger, D. S., and D. Helmberger, Source parameters of the Sierra Madre Earthquake from regional and local body waves, *Geophys. Res. Lett.*, *18*, 2015–2018, 1991.
- Dreger, D. S., and D. V. Helmberger, Broadband modeling of local earthquakes, *Bull. Seism. Soc. Am.*, *80*, 1162–1179, 1990.
- Dziewonski, A. M., and D. L. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356, 1981.
- Dziewonski, A. M., T. A. Chou, and J. H. Woodhouse, Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.*, *86*, 2825–2852, 1981.
- Festa, G., and J. P. Vilotte, The Newmark scheme as a velocity-stress time triggering: An efficient PML for spectral-element simulation of elastodynamics, *Geophys. J. Int.*, *161*, 789–812, 2005.
- Frankel, A., Three-dimensional simulations of ground motions in the San Bernardino valley, California, for hypothetical earthquakes on the San Andreas fault, *Bull. Seism. Soc. Am.*, *83*, 1020–1041, 1993.
- Furumura, T., B. L. N. Kennett, and M. Furumura, Seismic wavefield calculation for laterally heterogeneous whole Earth models using the pseudospectral method, *Geophys. J. Int.*, *135*(3), 845–860, 1998.
- Gazdag, J., Modeling of the acoustic wave equation with transform methods, *Geophysics*, *46*, 854–859, 1981.

- Geller, R. J., and T. Ohminato, 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.
- Gilbert, F., and A. M. Dziewonski, An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra, *Phil. Trans. R. Soc. London, Ser. A*, 278, 187–269, 1975.
- Grand, S., Mantle shear structure beneath the Americas and surrounding oceans, *J. Geophys. Res.*, 99, 1994.
- Graves, R. W., Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. Seism. Soc. Am.*, 86(4), 1091–1106, 1996.
- Hadley, D., and H. Kanamori, Seismic structure of the Transverse Ranges, California, *Geol. Soc. Am. Bull.*, 88, 1469–1478, 1977.
- Hauksson, E., Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in Southern California, *J. Geophys. Res.*, 105, 13,875–13,903, 2000.
- Hauksson, E., L. Jones, S. Perry, and K. Hutton, Emerging from the stress shadow of the 1992 M_w 7.3 Landers southern California earthquake? A preliminary assessment, *Seism. Res. Lett.*, 73, 33–38, 2002.
- Helmberger, D. V., and G. R. Engen, Modeling the long-period body waves from shallow earthquakes at regional ranges, *Bull. Seism. Soc. Am.*, 70, 1699–1714, 1980.
- Hung, S.-H., F. A. Dahlen, and G. Nolet, Fréchet kernels for finite-frequency traveltime – II. Examples, *Geophys. J. Int.*, 141, 175–203, 2000.
- Igel, H., and M. Weber, P-SV wave propagation in the whole mantle using high-order finite differences: application to lowermost mantle structure, *Geophys. Res. Lett.*, 23, 415–418, 1996.

- Jennings, P., Fault map of California with volcanoes, thermal springs and thermal wells at 1:750,000 scale, in *Geological Data Map 1*, California Division of Mines and Geology, Sacramento, California, 1975.
- Ji, C., S. Tsuboi, D. Komatitsch, and J. Tromp, Rayleigh-wave multipathing along the West coast of North America, *Bull. Seism. Soc. Am.*, 95, 2115–2124, 2005.
- Kelly, K. R., R. W. Ward, S. Treitel, and R. M. Alford, Synthetic seismograms: a finite difference approach, *Geophysics*, 41, 2–27, 1976.
- Komatitsch, D., and J. Tromp, Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.*, 139, 806–822, 1999.
- Komatitsch, D., and J. Tromp, Spectral-element simulations of global seismic wave propagation-I. validation, *Geophys. J. Int.*, 149, 390–412, 2002a.
- Komatitsch, D., and J. Tromp, Spectral-element simulations of global seismic wave propagation-II. 3-D models, oceans, rotation, and self-gravitation, *Geophys. J. Int.*, 150, 303–318, 2002b.
- Komatitsch, D., and J. Tromp, A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation, *Geophys. J. Int.*, 154, 146–153, 2003.
- Komatitsch, D., and J. P. Vilotte, The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. Seism. Soc. Am.*, 88, 368–392, 1998.
- Komatitsch, D., S. Tsuboi, C. Ji, and J. Tromp, A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the Earth Simulator, *Proceedings of the ACM/IEEE Supercomputing SC'2003 conference*, published on CD-ROM and at www.sc-conference.org/sc2003, 2003.
- Komatitsch, D., Q. Liu, J. Tromp, P. Süß, C. Stidham, and J. H. Shaw, Simulations of ground motion in the Los Angeles basin based upon the Spectral-Element method, *Bull. Seism. Soc. Am.*, 94, 187–206, 2004.

- Lay, T., and T. C. Wallace, *Modern Global Seismology*, Academic Press, 1995.
- Li, X.-D., and B. Romanowica, Global mantle shear velocity model developed using non-linear asymptotic coupling theory, *J. Geophys. Res.*, *101*, 1996.
- Liu, H. P., D. L. Anderson, and H. Kanamori, Velocity dispersion due to anelasticity: implications for seismology and mantle composition, *Geophys. J. R. Astro. Soc.*, *47*, 41–58, 1976.
- Liu, Q., and J. Tromp, Finite frequency Kernels based upon adjoint method, *Bull. Seism. Soc. Am.*, submitted, 2006a.
- Liu, Q., and J. Tromp, Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, in preparation, 2006b.
- Liu, Q., J. Polet, D. Komatitsch, and J. Tromp, Spectral-element moment tensor inversions for earthquakes in southern California, *Bull. Seism. Soc. Am.*, *94*, 1748–1761, 2004.
- Lovely, P., J. Shaw, Q. Liu, and J. Tromp, A structural (V_p) model of the Salton Trough, California, and its implication for seismic hazard, *Bull. Seism. Soc. Am.*, submitted, 2006.
- Luo, Y., and G. Schuster, Parsimonious staggered grid finite-differencing of the wave equation, *Geophys. Res. Lett.*, *17*, 155–158, 1990.
- Mallet, J. L., Discrete smooth interpolation in geometric modeling, *Computer-Aided Design*, *24*, 178–191, 1992.
- Marquering, H., F. Dahlen, and G. Nolet, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, *137*, 805–815, 1999.
- Masters, G., S. Johnson, and H. Bolton, A shear-velocity model of the mantle, *Phil. Trans. R. Soc. London, Ser. A*, *354*, 1996.
- Montelli, R., G. Nolet, F. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343, 2004.

- Olsen, K. B., and R. J. Archuleta, 3-D simulation of earthquakes on the Los Angeles fault system, *Bull. Seism. Soc. Am.*, 86(3), 575–596, 1996.
- Quarteroni, A., A. Tagliani, and E. Zampieri, Generalized Galerkin approximations of elastic waves with absorbing boundary conditions, *Comput. Methods Appl. Mech. Engrg.*, 163, 323–341, 1998.
- Ritsema, J., and T. Lay, Rapid source mechanism determination of large ($M_w \geq 5$) earthquakes in the Western United States, *Geophys. Res. Lett.*, 20, 1611–1614, 1993.
- Ritsema, J., and H. J. Van Heijst, Seismic imaging of structural heterogeneity in Earth's mantle: Evidence for large-scale mantle flow, *Science Progress*, 83, 243–259, 2000.
- Ritsema, J., L. A. Rivera, D. Komatitsch, J. Tromp, and H. J. van Heijst, The effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios, *Geophys. Res. Lett.*, 29, 72–1–72–4, 2002.
- Robertsson, J. O. A., A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography, *Geophysics*, 61, 1921–1934, 1996.
- Stacey, R., Improved transparent boundary formulations for the elastic wave equation, *Bull. Seism. Soc. Am.*, 78(6), 2089–2097, 1988.
- Su, W.-J., R. Woodward, and A. M. Dziewonski, Degree-12 model of shear-velocity heterogeneity in the mantle, *J. Geophys. Res.*, 99, 1994.
- Süss, M. P., and J. H. Shaw, P-wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *J. Geophys. Res.*, 108, 2170, doi:10.1029/2001JB001628, 2003.
- Tape, C. H., Q. Liu, and J. Tromp, Finite-frequency tomography using adjoint methods – methodology and examples using membrane surface waves, *Geophys. J. Int.*, submitted, 2006.
- Tessmer, E., and D. Kosloff, 3-D elastic modeling with surface topography by a Chebyshev spectral method, *Geophysics*, 59(3), 464–473, 1994.

- Thio, H. K., and H. Kanamori, Moment-tensor inversions for local earthquakes using surface waves recorded at TERRAScope, *Bull. Seism. Soc. Am.*, *95*, 1021–1038, 1995.
- Tromp, J., C. H. Tape, and Q. Liu, Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216, 2005.
- Tsuboi, S., D. Komatitsch, C. Ji, and J. Tromp, Spectral-element simulations of the November 3, 2002, Denali, Alaska earthquake on the Earth Simulator, *Phys. Earth Planet. Inter.*, *139*, 305–313, 2003.
- Virieux, J., *P-SV* wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, *51*, 889–901, 1986.
- Wallace, T. C., and D. V. Helmberger, A technique for the inversion of regional data in source parameter studies, *J. Geophys. Res.*, *86*, 1679–1685, 1981.
- Woodhouse, J. H., and A. M. Dziewonski, Mapping the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*, 5953–5986, 1984.
- Zhao, L., and D. V. Helmberger, Source Estimation from Broadband Regional Seismograms, *Bull. Seism. Soc. Am.*, *84*, 91–104, 1994.
- Zhao, L., and T. H. Jordan, Structure sensitivities of finite-frequency seismic waves: A full-wave approach, *Geophys. J. Int.*, submitted, 2006.
- Zhao, L., T. H. Jordan, and C. H. Chapman, Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, *141*, 558–576, 2000.
- Zhao, L., T. H. Jordan, K. B. Olsen, and P. Chen, Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Geophys. J. Int.*, *95*, 2066–2080, 2005.
- Zhou, Y., F.A.Dahlen, and G. Nolet, 3-D sensitivity kernels for surface-wave observables, *Geophys. J. Int.*, *158*, 142–168, 2004.

Zhou, Y., G. Nolet, F. A. Dahlen, and G. Laske, Global upper-mantle structure from finite-frequency surface-wave tomography, *J. Geophys. Res.*, *111*, doi:10.1029/2005JB003677, 2005.

Zhu, L., and D. V. Helmberger, Advancement in source estimation techniques using broadband regional seismograms, *Bull. Seism. Soc. Am.*, *86*, 1634–1641, 1996.

Zhu, L., and H. Kanamori, Moho depth variation in southern California from teleseismic receiver functions, *J. Geophys. Res.*, *105*, 2969–2980, 2000.