

Springer Geophysics

Po Chen
En-Jui Lee

Full-3D Seismic Waveform Inversion

Theory, Software, and Practice

 Springer

Springer Geophysics

The Springer Geophysics series seeks to publish a broad portfolio of scientific books, aiming at researchers, students, and everyone interested in geophysics. The series includes peer-reviewed monographs, edited volumes, textbooks, and conference proceedings. It covers the entire research area including, but not limited to, applied geophysics, computational geophysics, electrical and electromagnetic geophysics, geodesy, geodynamics, geomagnetism, gravity, lithosphere research, paleomagnetism, planetology, tectonophysics, thermal geophysics, and seismology.

More information about this series at <http://www.springer.com/series/10173>

Po Chen • En-Jui Lee

Full-3D Seismic Waveform Inversion

Theory, Software and Practice

 Springer

Po Chen
Department of Geology & Geophysics
University of Wyoming
Laramie
Wyoming
USA

En-Jui Lee
Department of Earth Sciences
National Cheng Kung University
Tainan
Taiwan

Springer Geophysics

ISBN 978-3-319-16603-2

ISBN 978-3-319-16604-9 (eBook)

DOI 10.1007/978-3-319-16604-9

Library of Congress Control Number: 2015938450

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

*Dedicated to my father Chen, Zhujun;
my mother Bu, Ping
and my wife Yao, Xiaojia*

Preface

Full-3D seismic waveform inversion (F3DWI) refers to inversions that seek to minimize the discrepancies between the observed and synthetic seismic waveforms, wiggle for wiggle, by solving the three-dimensional acoustic or (visco) elastic wave equations. Its development is important both for the theoretical foundations of modern quantitative seismology and for the practical applications of seismological methods in exploring the Earth's interior. Driven by the rapid advances in high-performance computing technology and efficient numerical methods for solving 3D wave equations, significant progresses in F3DWI have been made in the past decade, especially in large-scale structural studies that use passive sources. This book is derived from what I have learned in the past 10 years. F3DWI is by its very nature both a theoretical and a practical subject. It requires a certain level of understanding of the underlying mathematical formulation, a collection of parallelized software tools and a certain amount of practice. In this book, I try to give an integrated treatment of all three.

In Chap. 1, I give a brief introduction about the subject of this book and some discussions that motivates the development of F3DWI. Throughout this book, a parallelized finite-difference (visco)elastic wave-equation solver is used for demonstration purposes. The mathematical formulation and detailed instructions about how to set up and run this wave-equation solver for F3DWI purposes are summarized in Chap. 2.

The theoretical framework for F3DWI developed in Chaps. 3–5 is quite general and encompasses both the adjoint method (F3DT-AW), which back-propagates the misfits from the receivers to image structures, and the scattering-integral method (F3DT-SI), which sets up the Gauss-Newton normal equation by calculating and storing the sensitivity (Fréchet) kernel for each misfit.

The derivation of F3DT-SI in our previous publications requires the use of the reciprocity principle and the receiver-side Green's tensor (RGT). In Chap. 3, I generalize the formulation of F3DT-SI through adjoint analysis and show that the requirement on reciprocity can be removed by replacing the RGT with the time-reversed adjoint Green's tensor. This result may open up the possibility of applying the "scattering-integral-type" methods based on Green's functions to a larger class of inverse problems, in which the reciprocity principle may not hold. In Chap. 3, I

also derive the adjoint representation theorem and its application in Chap. 5 simplifies the derivation of the adjoint method for constructing the gradient of the objective function and also the Hessian-vector product.

In our previous publications of F3DT-SI, the data sensitivity (Fréchet) kernels were derived for the broadband cross-correlation delay time and amplitude misfits with respect to isotropic, elastic model parameters. In Chap. 4, I extend the kernel formulation to arbitrary misfits defined for individual receivers, receiver arrays, pressure and rotational sensors. In Chap. 4, I also derive the kernels for anisotropy and anelastic attenuation.

In Chap. 5, I examine both F3DT-SI and F3DT-AW in the context of numerical optimization. Possible extensions and some practical aspects of both types of methods are discussed briefly. Some more detailed discussions about the practical aspects can be found in Chap. 6, in which I give detailed descriptions about our 3D model for the crustal structure in southern California, CVM-S4.26, as well as the inversion that produced this model.

I started working on this model back in 2007 when I was a postdoc at the Lamont-Doherty Earth Observatory (LDEO) of Columbia University. After I moved to University of Wyoming in 2008, I kept working on it by myself until En-Jui Lee joined my research group in 2009. In the next 5–6 years, both En-Jui and I spent a lot of time developing this model. As shown in Chaps. 2 and 6 and also in our journal publications, CVM-S4.26 can fit a quite extensive collection of seismograms in southern California (more than 50,000 earthquake seismograms and ambient-noise Green's functions) from the first-arriving P-wave to about 30-sec after the surface wave, almost wiggle by wiggle. CVM-S4.26 reveals small-scale crustal heterogeneities that were not well imaged in previous crustal-scale tomography studies. Some of those small-scale features in CVM-S4.26 are highly consistent with geology and results from previous localized ray-theoretic travel-time tomography studies. In Chap. 6, I show many cross-sections throughout the entire model and discuss possible correlations with other independent geologic and seismic evidences. Materials in Chap. 6 complement our journal publications about this model.

Readers who are not particularly interested in the intricacies of the theoretical aspects of F3DWI can skip the mathematical formulations and start with descriptions of the software package, also named F3DWI. Those descriptions, instructions and worked examples of F3DWI are grouped into sections with the title “Software” throughout this book. I wrote most of the Fortran and C/C++ codes for F3DWI when I was doing my postdoc at LDEO in 2006–2007. The parallel finite-difference wave-equation solver, AWP-ODC, was provided by Yifeng Cui and Kim Olsen at the Southern California Earthquake Center (SCEC). I modified it to work with the rest of my kernel and inversion codes. During the development of CVM-S4.26, I wrote a command-line user interface using python to simplify and streamline the entire inversion process based on F3DWI. The scalable, parallel LSQR code, SPLSQR, for solving very large sparse linear systems associated with the Gauss-Newton normal equation in F3DT-SI was developed in collaboration with Liqiang Wang and his students and collaborators at the Computer Science Department in University of Wyoming. The software and other materials related to the book can

be requested from <http://pochenfullwave.ddns.net>. Like many other research codes, it takes some effort and practice to fully understand the behavior of these codes and to be able to effectively utilize these codes in realistic applications. Your comments, suggestions (of references), questions, corrections about the codes and the book are very much appreciated and please email them to pochengeophysics@gmail.com or pchen@uwo.edu.

I would like to thank Tom Jordan, Li Zhao and Xiaofei Chen for teaching me so many things about seismology and geophysics when I was a student. I would also like to thank all my students, En-Jui Lee, Dawei Mu, Xiaofeng Zhang, Jing Xu, and Wei Wang, for trusting in me and working with me. I would like to thank Jeroen Tromp for opening my eyes to the adjoint method during the 2003 Caltech-USC reading class on waveform tomography. The last chapter of this book benefited from constructive discussions with Tom Jordan, David Okaya, John Platt, Ken Dueker and Derek Schutt. I would like to thank all my collaborators in the past few years, Phillip Maechling, Marine Denolle, Greg Beroza, Liqiang Wang, He Huang, John Dennis, Youlin Chen, Zhiguo Xu, Daniel O'Connell, Pengcheng Liu and Satish Pulammanappallil. The Department of Geology and Geophysics and the School of Energy Resources (SER) at University of Wyoming have been very supportive on my research. Grants from the NSF, USGS, AFRL, SCEC, SER, Anadarko, Fugro Consultants Inc and Optim Software LLC are very much appreciated.

01/19/2015

Po Chen

Contents

1	Introduction	1
1.1	Advantages of the Seismic Ray Method	2
1.2	Drawbacks of the Seismic Ray Method	3
1.3	Numerical Solutions of the Seismic Wave Equation	4
1.4	A Brief History of HPC Technology	7
	References	11
2	Anelastic Wave Propagation (AWP)	15
2.1	Formulation	17
2.1.1	Elastic Media	17
2.1.2	Viscoelastic Media	18
2.1.3	Discretization	31
2.1.4	Free-Surface Boundary	34
2.1.5	PML Boundary	36
2.1.6	Internal Welded Material Interfaces	40
2.1.7	Source Representation	42
2.2	Software	43
2.2.1	Mesh Generation	45
2.2.2	Structural Model	49
2.2.3	Parameter File	53
2.2.4	Sources and Receivers	57
2.2.5	Recording Grid Points	64
2.2.6	Wavefield Simulation	66
2.2.7	Synthetic Seismograms	70
2.2.8	Half-Space Example Summary	73
2.3	Application	77
2.3.1	SCEC Community Velocity Models (CVMs)	77
2.3.2	Earthquake Source Models	78
2.3.3	Validation of SCEC CVMs	79
	References	86

3	Green's Functions	91
3.1	Mathematical Preliminary	91
3.1.1	Vector Spaces	91
3.1.2	Inner Product	92
3.1.3	Linear Transformations	97
3.1.4	Hilbert Spaces	101
3.1.5	Functions As Vectors	105
3.1.6	Dirac Delta Function	106
3.1.7	Fourier Transforms	111
3.1.8	Operator Algebra	113
3.1.9	Fréchet Derivative	119
3.2	The Adjoint of a Differential Operator	123
3.2.1	Ordinary Differential Operators	125
3.2.2	Partial Differential Operators	127
3.2.3	Scalar Wave Equations	129
3.2.4	Gradient, Divergence and Curl	131
3.2.5	Viscoelastic Equation of Motion	133
3.2.6	Adjoint Initial and Boundary Conditions	138
3.3	Properties of Green's Functions	146
3.3.1	Adjoint Green's Functions	147
3.3.2	Reciprocity	149
3.3.3	Inhomogeneous Initial/Boundary Conditions	154
3.4	Receiver Green's Tensor (RGT)	162
3.4.1	Time-Reversed Adjoint Green's Tensor	164
3.4.2	Receiver-Side Strain Green's Tensor (RSGT)	165
3.4.3	RSGTs in the Half-Space Example	168
3.5	Born Series	174
3.5.1	The Scalar Wave Equation	176
3.5.2	The Anisotropic Viscoelastic Wave Equation	186
	References	189
4	Data Sensitivity Kernels	191
4.1	Definition	192
4.2	Data Functionals	194
4.2.1	Data Functionals for $\nabla \times \mathbf{u}$ and $\nabla \cdot \mathbf{u}$	195
4.2.2	Filtered RGT	196
4.2.3	The Chain Rule	198
4.2.4	Time-Domain Waveform Differences	200
4.2.5	Frequency-Domain Waveform Differences	203
4.2.6	Broadband Cross-Correlation Measurements	206
4.2.7	Frequency-Dependent Phase and Amplitude Misfits	213
4.2.8	Rytov Approximation	225
4.2.9	"Banana-Doughnut" Structure of Rytov Phase Kernels	231
4.3	Model Parameterization	234
4.3.1	Bulk Modulus	234

4.3.2	Bulk and Shear Moduli	236
4.3.3	Lamé Parameters	239
4.3.4	P- and S-wave Velocities	243
4.3.5	Love’s Elastic Parameters (Hexagonal Symmetry)	247
4.3.6	Asymptotic Parameters (Triclinic Symmetry)	252
4.3.7	Anelastic Attenuation Quality Factors	257
4.3.8	Finite Moment Tensor (FMT)	266
4.4	Software	273
4.4.1	WPK File	275
4.4.2	Kernel Calculation	277
4.4.3	Kernel Output and Visualization	282
4.4.4	More Kernel Examples	285
	References	306
5	Optimization Algorithms	311
5.1	The Scattering-Integral (SI) Method	312
5.1.1	The Gauss–Newton Algorithm	313
5.1.2	F3DT-SI	316
5.1.3	Scalable Parallel LSQR (SPLSQR)	318
5.1.4	The Roughing Operator	319
5.1.5	Annihilator Matrix	321
5.2	The Adjoint-Wavefield (AW) Method	323
5.2.1	The Gradient	325
5.2.2	F3DT-AW	329
5.2.3	Hessian-Vector Product	331
5.2.4	Software	333
	References	340
6	CVM-S4.26	345
6.1	Background	345
6.1.1	Convergence Rate	346
6.1.2	Data Functionals	348
6.1.3	Model Evaluation	350
6.2	Model Description	351
6.2.1	CVM-S4.26 and CVM-S4 Comparison	352
6.2.2	CVM-S4.26 and CVM-H11.9 Comparison	352
6.2.3	Southern Walker Lane Belt	392
6.2.4	Southern Sierra Nevada	411
6.2.5	Southern San Joaquin Basin	427
6.2.6	Southern Coast Ranges	436
6.2.7	Transverse Ranges	446
6.2.8	Inner Continental Borderland	464
6.2.9	Mojave Desert Region	476
6.2.10	Peninsular Ranges—Salton Trough Region	488
	References	498
	Index	511

List of Figures

Fig. 2.1 Solution of the memory variable obtained using the exponential integrator method Eq. 2.21 (*solid lines*) and that obtained using the finite-difference method Eq. 2.22 (*dotted solid lines*) for different relaxation time τ_l . The initial condition is set as $\zeta_l(0) = 0$, the weight coefficient $w_l = 1$, the strain is set to a constant $\epsilon = 1$, the time-step length $\Delta t = 0.01$ s. 21

Fig. 2.2 Examples of the Debye function for $\omega_l = 0.2\pi, 2\pi, 20\pi$ and $Y_l = 1$ 25

Fig. 2.3 Comparisons of the frequency-independent Q model obtained by solving Eq. 2.62 using a least-squares method (*solid lines*) and the more simplistic approach of (Day and Bradley 2001) (*dash lines*) for $N = 3, 5, 8$. The target Q model is set at $\tilde{Q}^{-1} = 0.05$ 29

Fig. 2.4 Geometry of the staggered grid and the locations of the different components of the particle velocity, the stress tensor and the memory variables used in AWP-ODC. The *lower-right* plot shows a 3D array of unit cells. A unit cell is a cube with edge-length $h/2$, where h is the grid spacing. A blow-up view of the four unit cells inside the *dash-line box* is shown in the *upper-left plot*. The *three black arrows* on the *lower-left cell* indicate the coordinate system, in which the vertical axis x_3 points upward. 32

Fig. 2.5 The PML damping profile as a function of the distance from the Interior domain-PML interface located at $x_1 = 0$. The width of the PML is n_b and the maximum damping value d_0 is reached at $x_1 = n_b$. The damping profile shown here was generated using the equation $d_I = d_0 (I/n_b)^p$ (Marcinkovich and Olsen 2003), where I is the number of grid points counted from the Interior domain-PML interface, $p = 2$ and the PML width is $n_b = 12$ 37

Fig. 2.6 The effective material model for two Hooke bodies connected in serial. M_1 and M_2 are the moduli of the two different Hooke bodies in *model 1*. We seek an effective material model, *model 2*, which is composed of two identical Hooke bodies with modulus M_A connected in serial and has the same stress-strain relation as *model 1*. The stress and strain in the two Hooke bodies in model 1 are σ_1 and ϵ_1 and σ_2 and ϵ_2 , respectively. The total stress and strain both in *model 1* and *model 2* are σ and ϵ 40

Fig. 2.7 The source-time-function described by Eq. 2.122 in the time domain (*left*) and its amplitude spectrum in the frequency domain (*right*). 44

Fig. 2.8 The *top view* of the rectangular modeling region and the box coordinate system used in the forward wave-propagation simulations. *Left*: the two vertices on the bottom lie at the same longitude; *right*: an arbitrarily oriented rectangular modeling region. The directions of the x_1 and x_2 axes are marked with *arrows*. The x_3 axis points out of the page towards the viewer and forms a right-hand system with x_1 and x_2 45

Fig. 2.9 The sub-meshes for the configuration with 2 MPI processes in each dimension. The x_1 , x_2 and x_3 axes of the box coordinate system are shown as three *black arrows*. The 3-number index of each sub-mesh is shown in the *parentheses* in *white*. The sub-mesh with the index (1,1,0) is located underneath the sub-mesh with the index (1,1,1) and is not visible from the viewing angle used in this figure. 46

Fig. 2.10 The geographic location of the half-space example. The *black box* shows the rectangular modeling area and its four vertices are numbered. *Black circle*: epicenter of the seismic source; *black triangle*: surface projection of the receiver; *gray-scale background*: topography. 49

Fig. 2.11 The source-time function for the three body-force sources in the time domain (*left*) and its amplitude spectrum in the frequency domain (*right*). 61

Fig. 2.12 Verifying the reciprocity principle for the half-space example. *Thin black solid lines*: synthetics in IN.RC01.rgt.rot after the convolution with the source-time function shown in Fig. 2.7; *thick gray dash-lines*: synthetics in IN.RC01.rot after the convolution with the source-time function shown in Fig. 2.11; *horizontal axis*: time in seconds; *vertical axis*: particle velocity in m/s. From *left to right, top to bottom*, the sub-plots (a)–(g) show synthetics in the x_1 , x_2 directions of the box coordinate, the East and North directions, the Radial and Transverse directions and the vertical direction. 74

Fig. 2.13 *Black box*: our modeling domain; *black triangles*: three-component broadband stations used in our validation study; *beach balls*: source mechanisms for the Encino earthquake (*smaller beach ball*) and the La Habra earthquake (*larger beach ball*); *gray solid lines*: major faults; *gray-scale background*: topography. 80

Fig. 2.14 Examples of observed and synthetic seismograms for the Encino earthquake. *Black solid lines*: synthetic seismograms; *gray solid lines*: observed seismograms. In the sub-plot for each station, *left column*: CVM-S4.26 synthetics; *center column*: CVM-S4 synthetics; *right column*: CVM-H11.9 synthetics; *top row*: radial components; *center row*: transverse components; *bottom row*: vertical components. In the map, *beach ball*: source mechanism of the earthquake at its epicenter; *black triangles*: stations whose seismograms are shown; *gray and black solid lines*: faults and coast lines. 81

Fig. 2.15 Examples of observed and synthetic seismograms for the La Habra earthquake. The format is identical to that in Fig. 2.14. 82

Fig. 2.16 Histograms of the *RWMs* for all synthetics (*top row*), radial-component synthetics (*second row from top*), transverse-component synthetics (*third row from top*) and vertical-component synthetics (*bottom row*) computed using CVM-S4.26 (*left column*), CVM-S4 (*center column*) and CVM-H11.9 (*right column*). *Vertical black dash lines*: locations of the *RWM* median values (*mRWMs*). 83

Fig. 2.17 Spatial distributions of station misfits for all synthetics (*top row*), radial-component synthetics (*second row from top*), transverse-component synthetics (*third row from top*) and vertical-component synthetics (*bottom row*) computed using CVM-S4.26 (*left column*), CVM-S4 (*center column*) and CVM-H11.9 (*right column*). Sizes of the *black circles* are proportional to the station misfits and the component-dependent station misfits and a scale is shown on the *top-right corner* of each sub-plot. The *circles* are centered at station locations. 85

Fig. 3.1 The boundaries of the space-time domain for functions satisfying the scalar wave-equation defined in Eq. 3.182 and the normal directions on each boundary of the space-time domain. 142

Fig. 3.2 Comparisons of vertical (*top row*), radial (*center row*) and transverse (*bottom row*) synthetic velocity seismograms for the La Habra earthquake (Sect. 2.3.3) to the station CLC (*left column*) and BOR (*right column*) calculated by forward finite-difference simulation (*thick gray dash lines*) and by reciprocity using the RSGTs for CLC and BOR (*thin black solid lines*). The seismic structural model used in the forward simulations and the RSGT calculations is the three-dimensional CVM-S4.26 (Sect. 2.3.3). Figure 2.15 shows the source-receiver paths. 171

Fig. 3.3 A cross-section view of the recording grid points perpendicular to the source-receiver plane $i_y=40$. The explosive point source 100001 (*black circle*) is located at global grid index $(i_x, i_y, i_z) = (210, 40, 60)$, the receiver IN.RC01 (*black triangle*) is located at global grid index $(40, 40, 60)$, the receiver SF.RC02 (*black triangle*) is at $(40, 40, 180)$ and the receiver SF.RC03 (*black triangle*) is at $(210, 40, 180)$. Both SF.RC02 and SF.RC03 are located at the free surface. Locations marked with “+” are recording grid points generated by `set_ker_grd.py`. Locations marked with “x” are recording grid points generated by `set_sr_grd.py`. The “x” markers at the source and receiver locations are covered by the black circle and the black triangles, therefore not visible. The “x” markers for recording grid points of the source stencil can be seen around the hypocenter of 100001. 172

Fig. 3.4 Synthetic seismograms generated through the forward simulation and by using RSGTs. The source is 100001 and the receiver is SF.RC02. The format of this figure is identical to the format of Fig. 2.12. 174

Fig. 3.5 Synthetics for receiver SF.RC03. The format of this figure is identical to the format of Fig. 2.12. 175

Fig. 3.6 The contributions to the full propagation from the source \mathbf{x}_s to the receiver \mathbf{x} in the media $\alpha(\mathbf{x})$ from the zeroth-order (*left*), first-order (*center*) and second-order (*right*) scattering terms. Left: free propagation from the source to the receiver in the reference media $\tilde{\alpha}$; center: a free propagation from the source to a scatterer located at \mathbf{x}' in the reference media and a free propagation from the scatterer to the receiver in the reference media; right: free propagations in the reference media that involve two scatterers located at \mathbf{x}'_1 and \mathbf{x}'_2 . Locations of the source and the receiver are shown as *black dots* and locations of the scatterers are shown as *gray dots*. 179

Fig. 3.7 Cross-section views of the frequency-domain Born kernels for the whole-space structural model. *Left column*: imaginary part of the kernel; *right column*: real part of the kernel. *Top row*: the kernel at 5 Hz; *bottom row*: the kernel at 10 Hz. Black star: source location; black triangle: receiver location. The wave velocity $\tilde{\alpha} = 6.0$ km/s. The source is located at coordinate (4 km, 0 km, 5 km) and the receiver is located at coordinate (11 km, 0 km, 5 km). The cross-section views are perpendicular to the source-receiver $x - z$ plane. *Gray scale* shows the value (darker: positive; lighter: negative) of the real or imaginary part of the kernel at every spatial location \mathbf{x} on the cross-section. 183

Fig. 3.8 Cross-section views of the time-domain Born kernels for the whole-space structural model. From *top* to *bottom*, *left* to *right*, are time-domain kernel amplitudes at $t_i = 1.25$ s, 1.5 s, 1.75 s and 2.0 s. Black star: source location; black triangle: receiver location. Gray scale shows kernel amplitude (lighter: positive; darker: negative) at every spatial location \mathbf{x} on the cross-section. Note that in 3D the kernel is shaped as an ellipsoid with the source and the receiver located at its two focal points. 184

Fig. 3.9 An incident plane wave in the whole-space structure model $\hat{u}(\mathbf{x}, \omega)$, a cylindrical object with radius a and the wavefield inside the cylinder $\hat{u}(\mathbf{x}, \omega)$ 185

Fig. 4.1 Cross-section views of the data sensitivity kernels for the time-domain waveform differences at three different time samples on the x_1 component of the half-space example. The location of the cross-section in this and the following half-space kernel figures is identical to that in Fig. 3.3. *First column*: synthetic seismogram (*black solid lines*) and the location of the time sample (*vertical dash lines* and title of the plot); *second column*: data sensitivity kernel with respect to the relative perturbation in P-velocity $\delta\alpha/\alpha$; *third column*: data sensitivity kernel with respect to the relative perturbation in S-velocity $\delta\beta/\beta$; *fourth column*: data sensitivity kernel with respect to the relative perturbation in density $\delta\rho/\rho$. The locations of the source 100001 and the receiver IN.RC01 are the same as those shown in Fig. 3.3 and the cross-section views are perpendicular to the source-receiver plane at $y=40$. The synthetic seismogram shown on the first column is identical to that in Fig. 2.12a. 202

Fig. 4.2 The WPK, as given in Eq. 4.45, for the direct-arriving P-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01. *Solid line*: the real part; *dash line*: the imaginary part. The time window $W_n(t)$ is a Tukey window and the frequency $\omega_n = 4\pi$ 205

Fig. 4.3 Cross-section views of the real and imaginary parts of the data sensitivity kernels for the frequency-domain waveform differences measured at two different frequencies on the P-wave of the x_1 -component synthetic velocity seismogram from the source 100001 to the receiver IN.RC01. *First row*: the real-part kernels at 1 Hz; *second row*: the imaginary-part kernels at 1 Hz; *third row*: the real-part kernels at 2 Hz; *bottom row*: the imaginary-part kernels at 2 Hz. The *first column* shows the time-domain synthetic seismogram (row 1 and 3) and the amplitude spectrum of the selected waveform (row 2 and 4). The frequencies at which the kernels were computed are indicated with the vertical lines on the amplitude spectrum. The amplitudes of the kernels shown in Figs. 4.3, 4.4 and 4.5 have been normalized by the frequency and the maximum amplitudes of the selected waveforms. 206

Fig. 4.4 The format is identical to that in Fig. 4.3. The data sensitivity kernels shown here are for the PP-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01. 207

Fig. 4.5 The format is identical to that in Fig. 4.3. The data sensitivity kernels shown here are for the PS-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01. 208

Fig. 4.6 *Upper plot*: the reference synthetic velocity seismogram (*top*) and the observed velocity seismogram (*bottom*) and the waveforms selected for making the cross-correlation measurement (*black thick lines*). *Lower plot*: the auto-correlation of the selected waveform on the reference synthetic (*top*) and the cross-correlation between the selected waveforms on the reference synthetic and the observed waveforms (*bottom*). The maximum on the auto-correlation, which is located at $\Delta t = 0$ s, is marked using a *black vertical dash line*. The maximum on the cross-correlation, which is at around $\Delta t = -0.31532$ s, is marked using a *black vertical solid line*. 209

Fig. 4.7 The WPKs for the broadband cross-correlation delay-time (*middle row*) and the broadband cross-correlation amplitude anomaly (*bottom row*) for the direct-arriving P-wave (*left column*), the PP-wave (*middle column*) and the PS-wave (*right column*) on the x_1 -component velocity synthetic seismogram from 100001 to IN.RC01 in the half-space example. The waveforms used for computing the WPKs of each column are plotted in *thick black lines* on the *top row*. The synthetic seismograms shown on the *top row* are identical to that shown in Fig. 2.12a. 213

Fig. 4.8 The data sensitivity kernels of the broadband cross-correlation delay time (row 1, 3, 5) and the broadband cross-correlation amplitude anomaly (row 2, 4, 6) for the direct-arriving P-wave (row 1-2), the PP-wave (row 3-4) and the PS-wave (row 5-6) on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example. The synthetic seismogram and the WPKs used for computing these kernels are shown in Fig. 4.7. The kernels shown here do not have the numerical artifacts discussed in (Zhao et al. 2005). 214

Fig. 4.9 Comparisons of data sensitivity kernels for the broadband cross-correlation delay time measured for the P-wave on the velocity seismograms (*top* and *center* rows) and the displacement seismogram (*bottom* row), filtered with a low-pass Butterworth filter with corner frequency at 1.5 Hz (*top* row) and 3.5 Hz (*center* and *bottom* rows). From *left* to *right*: kernels with respect to relative perturbations in P- and S-velocity and density. The source-receiver setup is identical to that in Fig. 4.8. 215

Fig. 4.10 An example of measuring frequency-dependent phase-delay time $\Delta\tau_p(\omega)$, group-delay time $\Delta\tau_g(\omega)$ and amplitude-reduction time $\Delta\tau_q(\omega)$ using the x_1 -component seismogram from 100001 to IN.RC01 in the half-space example. **a** Reference synthetic seismogram (*upper*) and observed seismogram (*lower*), the waveforms selected for making the measurements are plotted in *thick black lines*. **b** The upper two traces show the fit between the selected reference synthetic waveform (*dash* line) and the selected observed waveform (*solid* line), the lower two traces show the fit between the selected observed waveform (*solid* line) and the corrected synthetic waveform (*dash* line) obtained by applying the optimized phase and amplitude corrections $\Delta\phi(\omega)$ and $\Delta\psi(\omega)$ to the reference synthetic waveform. **c** The amplitude spectrum of the observed (*dash*) and reference synthetic (*solid*) waveforms. The two *vertical dash* lines in (**c**), (**d**) and (**e**) indicate the frequencies at which either the observed or the reference synthetic amplitude falls below 20% of its peak amplitude. **d** The phase-delay (*solid*) and group-delay (*dash*) time measurements. **e** The amplitude-reduction time measurements. 218

Fig. 4.11 Examples of the WPKs of the $\Delta\tau_{p,q,g}$ measurements at 0.8 and 1.6 Hz for the three arrivals on the x_1 -component reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The second and third rows are WPKs for $\Delta\tau_p$, the fourth and fifth rows are WPKs for $\Delta\tau_g$ and the sixth and seventh rows are WPKs for $\Delta\tau_q$. The second, fourth and sixth rows are WPKs at 0.8 Hz and the third, fifth and seventh rows are WPKs at 1.6 Hz. 221

Fig. 4.12 Data sensitivity kernels of $\Delta\tau_p$ (first and fourth rows), $\Delta\tau_g$ (second and fifth rows) and $\Delta\tau_q$ (third and sixth rows) at 0.8 Hz (first to third row) and 1.6 Hz (fourth to sixth row) for the direct-arriving P-wave on the x_1 -component reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The WPKs for these kernels are shown on the first column of Fig. 4.11. 222

Fig. 4.13 Data sensitivity kernels for the PP-wave on the same seismogram as that in Fig. 4.12. The format is identical to that in Fig. 4.12. The WPKs for these kernels are shown on the second column of Fig. 4.11. 223

Fig. 4.14 Data sensitivity kernels for the PS-wave on the same seismogram as that in Fig. 4.12. The format is identical to that in Fig. 4.12. The WPKs for these kernels are shown on the third column of Fig. 4.11. 224

Fig. 4.15 Locations of the source \mathbf{x}_s , the receiver \mathbf{x}_r and a scatterer \mathbf{x} . The source is located at the origin. The receiver is located on the x -axis of the coordinate system at a distance d from the source. The scatterer is located at (x, y, z) . The source-receiver distance is therefore d . The distance between the source and the scatterer is $|\mathbf{x} - \mathbf{x}_s| = \sqrt{x^2 + y^2 + z^2}$ and the distance between the receiver and the scatterer is $|\mathbf{x} - \mathbf{x}_r| = \sqrt{(d - x)^2 + y^2 + z^2}$. . . 233

Fig. 4.16 Cross-section views of the frequency-domain Rytov kernels for the scalar wave equation in the whole-space structure model. *Left column:* Rytov phase kernels (Eq. 4.127) at 5 Hz (*top row*) and 10 Hz (*bottom row*); *right column:* Rytov amplitude kernels (Eq. 4.128) at 5 Hz (*top row*) and 10 Hz (*bottom row*). The source is located at coordinate $(0, 0, 0)$ and the receiver is located at $(11, 0, 0)$. The cross-section view is perpendicular to the source-receiver $x - z$ plane. 233

Fig. 4.17 Data sensitivity kernels of the broadband cross-correlation delay time with respect to relative perturbations in the bulk modulus κ (*left column, top row*) (Eq. 4.145), shear modulus μ (*center column, top row*) (Eq. 4.151) and density ρ (*right column, top row*) (Eq. 4.178), the first Lamé parameter λ (*left column, center row*) (Eq. 4.167), the second Lamé parameter μ (*center column, center row*) (Eq. 4.168), the density ρ (*right column, center row*) (Eq. 4.178), the P-wave velocity α (*left column, bottom row*) (Eq. 4.185), the S-wave velocity β (*center column, bottom row*) (Eq. 4.186) and the density ρ (*right column, bottom row*) (Eq. 4.175). The data functional is applied on the P-wave on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example. 240

Fig. 4.18	The format is identical to that in Fig. 4.17, except that the data functional is applied on the PP-wave.	241
Fig. 4.19	The format is identical to that in Fig. 4.17, except that the data functional is applied on the PS wave.	242
Fig. 4.20	Data sensitivity kernels of the broadband cross-correlation delay time with respect to the perturbations in the 21 asymptotic parameters. The data functional is applied on the P-wave on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example.	258
Fig. 4.21	The format is identical to that in Fig. 4.20, except that the data functional is applied on the PP wave.	259
Fig. 4.22	The format is identical to that in Fig. 4.20, except that the data functional is applied on the PS wave.	260
Fig. 4.23	<i>Dash lines</i> : the WPKs of the broadband cross-correlation amplitude anomalies measured on the P (<i>left</i>), PP (<i>center</i>) and PS (<i>right</i>) waves on the x_1 -component of the synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example; <i>solid lines</i> : the WPKs for the anelastic attenuation kernels obtained by convolving the WPKs shown in <i>dash lines</i> with the function $\psi'(t)$ using Eq. 4.296.	263
Fig. 4.24	Data sensitivity kernels of the broadband cross-correlation amplitude anomaly with respect to Q_κ^{-1} (<i>left column</i>) and Q_μ^{-1} (<i>right column</i>) for the P (<i>top row</i>), PP (<i>center row</i>) and PS (<i>bottom row</i>) waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example.	264
Fig. 4.25	Spatial variations of the partial derivatives of the frequency-dependent phase-delay time $\Delta\tau_p(\omega)$ (<i>left column</i>) and amplitude-reduction time $\Delta\tau_q(\omega)$ (<i>right column</i>) with respect to $\mu_1^{(1,0)}$ (<i>top row</i>), $\mu_1^{(1,1)}$ (<i>second row</i>), $\mu_{11}^{(2,0)}$ (<i>third row</i>), $\mu_{12}^{(2,0)}$ (<i>fourth row</i>), $\mu_{13}^{(2,0)}$ (<i>bottom row</i>).	271
Fig. 4.26	The time window used for selecting waveforms from the complete synthetic seismogram. At time t_1 and t_4 , the window tapers to zero and at time t_2 and t_3 , the window increases to 1. The tapering parts of the window between t_1 and t_2 and between t_3 and t_4 follow a \cos^2 function.	275
Fig. 4.27	Three-component synthetic velocity seismograms from the double-couple source 100002 to IN.RC01 in our half-space example with the larger mesh. <i>Top trace</i> : x_1 component; <i>center trace</i> : x_2 component; <i>bottom trace</i> : x_3 component. The S-wave on the x_2 component (<i>black thick line on center trace</i>) is used for computing the WPK.	288

Fig. 4.28	The WPKs for the broadband cross-correlation delay time (<i>left</i>) and amplitude anomaly (<i>right</i>) computed from the selected S-wave shown in Fig. 4.27.	290
Fig. 4.29	Cross-section views of the data sensitivity kernel for the broadband cross-correlation delay time (<i>left</i>) and amplitude anomaly (<i>right</i>) with respect to relative perturbations in S-velocity for the S-wave on the x_2 -component synthetic velocity seismogram from 100002 to IN.RC01 in our half-space example. The cross-section lies in the source-receiver plane perpendicular to the x_2 direction.	291
Fig. 4.30	Three-component synthetic velocity seismograms from the vertical body-force source applied at SF.RC03 to the receiver SF.RC02 in the half-space model. <i>Top trace</i> : x_1 component; <i>center trace</i> : x_2 component; <i>bottom trace</i> : x_3 component. The thick <i>black solid line</i> on the bottom trace shows the waveform used for computing the WPKs.	293
Fig. 4.31	The WPKs for frequency-dependent phase-delay time (<i>top row</i>), group-delay time (<i>center row</i>) and amplitude-reduction time (<i>bottom row</i>) at 0.5 Hz (<i>left column</i>) and 2 Hz (<i>right column</i>). The synthetic velocity waveform used for computing these WPKs is shown in thick <i>black solid line</i> in Fig. 4.30.	294
Fig. 4.32	Data sensitivity kernels of the frequency-dependent phase and amplitude misfits with respect to the asymptotic parameter L for the vertical-component Rayleigh wave shown in Fig. 4.30. <i>Left column</i> : kernels at 0.5 Hz; <i>right column</i> : kernels at 2.0 Hz. a–d kernels for frequency-dependent phase-delay time $\delta\tau_p$; e–h kernels for frequency-dependent group-delay time $\delta\tau_g$; i–l kernels for frequency-dependent amplitude-reduction time $\delta\tau_q$. a, b, e, f, i, j map-views of the kernels at the depths of maximum sensitivity, which is about 1.6 km at 0.5 Hz and about 400 m at 2.0 Hz; c, d, g, h, k, l cross-section views perpendicular to the source-receiver plane.	296
Fig. 4.33	Map-views of the data sensitivity kernels for the 2-Hz phase-delay time with respect to the perturbations in the 21 asymptotic parameters. The map-views are at the depths of the kernel's maximum amplitudes.	297

Fig. 4.34 The layer-over-half-space seismic velocity model and the source-receiver geometry used for computing the data sensitivity kernels for head waves. The spatial dimension is 112 km in x_1 , 120 km in x_2 and 50 km in x_3 . This figure shows the cross-section view in the source-receiver plane perpendicular to the x_2 direction. The velocity interface is at 21.4-km depth. The P- and S-velocity and density are 5207 m/s, 3189 m/s and 2950 kg/m³, respectively, for the *top* layer, and 9058 m/s, 5307 m/s and 3992 kg/m³, respectively, for the *bottom* layer. The source 100003 (*black star*) is located at 18-km depth and the receiver IN.RC04 (*black circle*) is located at 6.4-km depth. Both the source and the receiver are located at $x_2=60$ km. The source-receiver horizontal distance is 88 km. 298

Fig. 4.35 Head waves on the vertical-component synthetic velocity seismograms from 100003 to IN.RC04. *Black dash line*: synthetic seismogram generated using the layer-over-half-space model shown in Fig. 4.34; *red solid line*: synthetic seismogram generated using the structural model whose P- and S-velocity in the *bottom* layer have a positive gradient of 0.02 km/s/km with depth. 301

Fig. 4.36 The WPKs of the broadband cross-correlation delay time (*left*) and amplitude anomaly (*right*) for the head wave generated using the layer-over-half-space model (*black dash lines*) and the head wave generated using the model with the bottom layer having a positive gradient of 0.02 km/s/km with depth. 302

Fig. 4.37 Cross-section views of the data sensitivity kernels of the broadband cross-correlation delay time (*upper*) and amplitude anomaly (*lower*) for the head wave shown as *black dash lines* in Fig. 4.35 with respect to the relative perturbation in P-velocity. *Left column*: cross-sections in the source-receiver plane perpendicular to the x_2 axis; *right column*: cross-sections mid-way between source and receiver ($x_1 = 56$ km) perpendicular to the x_1 axis. The *black dash lines* indicate the velocity interface at 21.4 km depth (Fig. 4.34). 303

Fig. 4.38 The P- and S-velocity as functions of depth in the “layer-over-gradient” model. The top layer has constant P- and S-velocity with values identical to those shown in Fig. 4.34 and the bottom layer has a constant positive gradient of 0.02 km/s/km with depth. The velocities in the bottom layer increase from 9058 m/s and 5307 m/s at the interface to 9650 m/s and 5899 m/s at the bottom of our model. 304

Fig. 4.39 Cross-sections of the data sensitivity kernels for the broadband cross-correlation delay time (*top row*) and amplitude anomaly (*bottom row*) with respect to P-velocity. *Left column*: cross-sections on the source-receiver plane perpendicular to the x_2 axis; *center column*: cross-sections mid-way between source and receiver and perpendicular to the x_1 axis; *right column*: sensitivity along the *vertical dash lines* shown on the center column for the “layer-over-half-space” structural model (*black dash lines*) and the “layer-over-gradient” structural model (*red solid lines*), the vertical axis is depth and the horizontal axis is the value of the data sensitivity kernel. The *horizontal black dash lines* on the *left* and *center* columns show the location of the velocity interface at 21.4 km depth. 305

Fig. 4.40 Cross-sections of the data sensitivity kernels for the frequency-dependent phase-delay time with respect to P-velocity at 1.0 Hz (*top row*) and 2.0 Hz (*bottom row*). *Left column*: cross-section at the source-receiver plane perpendicular to the x_2 axis; *right column*: cross-section at the plane mid-way between source and receiver and perpendicular to the x_1 axis. 305

Fig. 5.1 Event kernels of broadband cross-correlation delay-time with respect to relative perturbations in κ , μ , ρ (*top row*), λ , μ , ρ (*center row*) and α , β , ρ (*bottom row*) for 100001 in the half-space example). 338

Fig. 5.2 Anisotropic event kernels of broadband cross-correlation delay time with respect to the 21 asymptotic parameters for 100001 in the half-space example. 339

Fig. 6.1 Map-view of S-velocity at 2-km depth in CVM-S4 (*top*), CVM-S4.26 (*center*) and CVM-H11.9 (*bottom*). *Warm colors* indicate lower velocities and *cold colors* indicate higher velocities. *Black solid lines*: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages. 353

Fig. 6.2 Map-view of P-velocity at 2-km depth in CVM-S4 (*top*), CVM-S4.26 (*center*) and CVM-H11.9 (*bottom*). *Warm colors* indicate lower velocities and *cold colors* indicate higher velocities. *Black solid lines*: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages. 365

Fig. 6.3 Map-view of Poisson’s ratio at 2 km depth in CVM-S4 (*top*), CVM-S4.26 (*center*) and CVM-H11.9 (*bottom*). Warm colors indicate higher Poisson’s ratio and cold colors indicate lower Poisson’s ratio. *Black solid lines*: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages. 377

Fig. 6.4 S-velocity in CVM-S4.26 (*upper-left*), CVM-S4 (*lower-left*) and CVM-H11.9 (*lower-right*) compared with the geologic map (*upper-right*) around the Owens Valley and Death Valley region. The color range is identical to that in Fig. 6.1. 393

Fig. 6.5 Cross-sections of S-velocity in CVM-S4.26 along 13 profiles (A-M) across the Walker Lane belt. The map-view at *upper-left* corner shows S-velocity at 2-km depth and the color-scale is identical to the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown in the *lower-left corner*. 395

Fig. 6.6 Cross-sections of P-velocity in CVM-S4.26 along the 13 profiles (A-M) across the Walker Lane belt. The format is identical to that in 6.5. 396

Fig. 6.7 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 13 profiles (A-M) across the Walker Lane belt. The format is identical to that in 6.5. 397

Fig. 6.8 Cross-sections of S-velocity in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.5. 398

Fig. 6.9 Cross-sections of P-velocity in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.6. 399

Fig. 6.10 Cross-sections of the Poisson’s ratio in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.7. 400

Fig. 6.11 Cross-sections of S-velocity in CVM-H11.9 along the 13 profiles. The format is identical to that in 6.5. 401

Fig. 6.12 Cross-sections of P-velocity in CVM-H11.9 along the 13 profiles. The format is identical to that in Fig. 6.6. 402

Fig. 6.13 Cross-sections of Poisson’s ratio in CVM-H11.9 along the 13 profiles. The format is identical to that in Fig. 6.7. 403

Fig. 6.14 The cross-section views show P- (*top row*) and S-velocity (*center row*) and Poisson’s ratio (*bottom row*) in CVM-S4.26 along A-A’ (*left column*), B-B’ (*center column*) and C-C’ (*right column*). The map-view plot at the *upper-left corner* shows the S-velocity in CVM-S4.26 at 2-km depth. 405

Fig. 6.15 The same cross-sections for CVM-S4 and CVM-H11.9. The format and color-scale are identical to those in Fig. 6.14. 406

Fig. 6.16 The map-view plots (*top row*) show the S-velocity (*left*), P-velocity (*center*) and Poisson’s ratio (*right*) at 10-km depth beneath the Coso geothermal area in CVM-S4.26. Locations of the two cross-sections, A-A’ and B-B’, are plotted on the S-velocity map-view. Cross-section plots show the S-velocity (*top row*), P-velocity (*center row*) and Poisson’s ratio (*bottom row*) along A-A’ (*left column*) and B-B’ (*right column*). 408

Fig. 6.17 Map-views and cross-sections extracted from CVM-H11.9. The format and color-scale are identical to those in Fig. 6.16. 409

Fig. 6.18 Map-views and cross-sections extracted from CVM-S4. The format and color-scale are identical to those in Fig. 6.16. 410

Fig. 6.19 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in and around the southern Sierra Nevada. The color scale is identical to that in the 2-km map-view in Fig. 6.1. 412

Fig. 6.20 Cross-section views of the S-velocity in CVM-S4.26 along 9 parallel profiles (A-I). The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the color scale identical to that in the 2-km map-view of Fig. 6.1. The color scale for the cross-sections is displayed in the *lower-left corner* and is identical to that used in Fig. 6.5. Black dots on the cross-sections indicate the approximate lateral extent of the high-velocity body. *Black dash lines* on the map-view are obtained by connecting the *black dots* on the cross-sections. 414

Fig. 6.21 Cross-sections of P-velocity in CVM-S4.26. The format is identical to that in Fig. 6.20. 415

Fig. 6.22 Cross-sections of Poisson’s ratio in CVM-S4.26. The format is identical to that in Fig. 6.20. 416

Fig. 6.23 Cross-sections of S-velocity in CVM-S4. The format is identical to that in Fig. 6.20. 417

Fig. 6.24 Cross-sections of P-velocity in CVM-S4. The format is identical to that in Fig. 6.20. 418

Fig. 6.25 Cross-sections of Poisson’s ratio in CVM-S4. The format is identical to that in Fig. 6.20. 419

Fig. 6.26 Cross-sections of S-velocity in CVM-H11.9. The format is identical to that in Fig. 6.20. 420

Fig. 6.27 Cross-sections of P-velocity in CVM-H11.9. The format is identical to that in Fig. 6.20. 421

Fig. 6.28 Cross-sections of Poisson’s ratio in CVM-H11.9. The format is identical to that in Fig. 6.20. 422

Fig. 6.29 Cross-sections of S-velocity in CVM-S4.26 along the 7 profiles (A-G). The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of the surface traces of faults. The black dots at horizontal distance of about 65 km are for the Garlock Fault. The black dots at horizontal distance of about 45 km on profile B-D are for the White Wolf Fault. The black dots left of the Garlock fault on profile E-G are for the Kern Canyon Fault. Locations of the faults are shown on Fig. 6.19. 423

Fig. 6.30 Cross-sections of P-velocity in CVM-S4.26 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 424

Fig. 6.31 Cross-sections of Poisson’s ratio in CVM-S4.26 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 425

Fig. 6.32 Cross-sections of S-velocity in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 426

Fig. 6.33 Cross-sections of P-velocity in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 427

Fig. 6.34 Cross-sections of Poisson’s ratio in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 428

Fig. 6.35 Cross-sections of S-velocity in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 429

Fig. 6.36 Cross-sections of P-velocity in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 430

Fig. 6.37 Cross-sections of Poisson’s ratio in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29. 431

Fig. 6.38 S-velocity at 2-km depth in CVM-S4.16 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in the southern San Joaquin Basin. The color scale is identical to that in the 2-km map-view in Fig. 6.1. 432

Fig. 6.39 Cross-sections of the S-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of the surface traces of San Andreas Fault. 433

Fig. 6.40 Cross-sections of the P-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 434

Fig. 6.41 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 435

Fig. 6.42 Cross-sections of the S-velocity in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 436

Fig. 6.43 Cross-sections of the P-velocity in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 437

Fig. 6.44 Cross-sections of the Poisson’s ratio in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 438

Fig. 6.45 Cross-sections of the S-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 439

Fig. 6.46 Cross-sections of the P-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 440

Fig. 6.47 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39. 441

Fig. 6.48 Map-views of the S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in and around the southern Coast Ranges. The color scale is identical to that in the 2-km map-view in Fig. 6.1. 442

Fig. 6.49 Cross-sections of the S-velocity in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of the surface traces of San Andreas Fault (SAF), the Rinconada Fault (RF) and the Hosgri Fault Zone (HFZ). Locations of the faults are marked on Fig. 6.48. 443

Fig. 6.50 Cross-sections of the P-velocity in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 444

Fig. 6.51 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 445

Fig. 6.52 Cross-sections of the S-velocity in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 446

Fig. 6.53 Cross-sections of the P-velocity in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 447

Fig. 6.54 Cross-sections of the Poisson’s ratio in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 448

Fig. 6.55 Cross-sections of the S-velocity in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 449

Fig. 6.56 Cross-sections of the P-velocity in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 450

Fig. 6.57 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49. 451

Fig. 6.58 S-velocity at 2-km depth in CVM-S4.26 (b), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (a) of the western Transverse Ranges region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. 452

Fig. 6.59 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) of the eastern Transverse Ranges region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. Black stars indicate the epicenters of several major earthquakes in this region. 453

Fig. 6.60 Cross-sections of the S-velocity in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. 454

Fig. 6.61 Cross-sections of the P-velocity in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 455

Fig. 6.62 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 456

Fig. 6.63 Cross-sections of the S-velocity in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 457

Fig. 6.64 Cross-sections of the P-velocity in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 458

Fig. 6.65 Cross-sections of the Poisson’s ratio in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 459

Fig. 6.66 Cross-sections of the S-velocity in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 460

Fig. 6.67 Cross-sections of the P-velocity in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 461

Fig. 6.68 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60. 462

Fig. 6.69 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) of the Inner Continental Borderland region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. 465

Fig. 6.70 Cross-sections of the S-velocity in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-right corner*. Major faults are shown as *black dots* on the cross-sections. Elsinore Fault (*EF*), Newport-Inglewood Fault (*NIF*), Palos Verdes Fault (*PVF*), Rose Canyon Fault (*RCF*), Santa Cruz Island Fault (*SCIF*), Santa Cruz—Santa Catalina Ridge Fault (*SCSCRF*). 466

Fig. 6.71 Cross-sections of the P-velocity in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 467

Fig. 6.72 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 468

Fig. 6.73 Cross-sections of the S-velocity in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 469

Fig. 6.74 Cross-sections of the P-velocity in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 470

Fig. 6.75 Cross-sections of the Poisson’s ratio in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 471

Fig. 6.76 Cross-sections of the S-velocity in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 472

Fig. 6.77 Cross-sections of the P-velocity in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 473

Fig. 6.78 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70. 474

Fig. 6.79 S-velocity at 2-km depth in CVM-S4.26 **(b)**, CVM-S4 **(c)** and CVM-H11.9 **(d)** compared with the geologic map **(a)** of the Mojave Desert region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. 477

Fig. 6.80 Cross-sections of the S-velocity in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-right corner*. Major faults are shown as black dots on the cross-sections. Garlock Fault (*GF*), Lockhart–Lenwood Fault (*LF*), San Andreas Fault (*SAF*), San Gabriel Fault (*SGF*). 479

Fig. 6.81 Cross-sections of the P-velocity in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 480

Fig. 6.82 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 481

Fig. 6.83 Cross-sections of the S-velocity in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 482

Fig. 6.84 Cross-sections of the P-velocity in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 483

Fig. 6.85 Cross-sections of the Poisson’s ratio in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 484

Fig. 6.86 Cross-sections of the S-velocity in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 485

Fig. 6.87 Cross-sections of the P-velocity in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 486

Fig. 6.88 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80. 487

Fig. 6.89 S-velocity at 2-km depth in CVM-S4.26 **(a)**, CVM-S4 **(c)** and CVM-H11.9 **(d)** compared with the geologic map **(b)** of the Peninsular Ranges–Salton Trough region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. 489

Fig. 6.90 Cross-sections of the S-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. Major faults are shown as black dots on the cross-sections. Elsinore Fault (*EF*); San Jacinto Fault (*SJF*); San Andreas Fault (*SAF*). 490

Fig. 6.91 Cross-sections of the P-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 491

Fig. 6.92 Cross-sections of the Poisson’s ratio in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 492

Fig. 6.93 Cross-sections of the S-velocity in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 493

Fig. 6.94 Cross-sections of the P-velocity in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 494

Fig. 6.95 Cross-sections of the Poisson’s ratio in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 495

Fig. 6.96 Cross-sections of the S-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 496

Fig. 6.97 Cross-sections of the P-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 497

Fig. 6.98 Cross-sections of the Poisson’s ratio in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90. 498

Chapter 1

Introduction

The subject of this book is the solution of the seismic inverse problem. The corresponding “forward problem” deals with the solution of the seismic wave equation (i.e., the elastodynamic equation with specified initial and boundary conditions). In the forward problem, the mathematical models for the seismic sources and the physical properties of the geological media are given and we solve the seismic wave equation for the spatial-temporal evolution of the displacement-stress wavefields¹. In the inverse problem, accurate knowledge about the seismic sources and the geological media is usually unavailable and we try to make inferences about them from discrete observations of the wavefields (i.e., seismograms and other types of ground-motion observations). Techniques for constructing mathematical models of the physical properties of the geological media using information extracted from observed seismograms are often referred to as “seismic tomography”, which is the emphasis of this book. For controlled-source seismic surveys, it is sometimes possible to build mathematical models of the seismic sources based on the procedures used for generating the sources. For problems involving natural earthquakes, we need to invert for earthquake source models, sometimes jointly with models of the geological media, using observed seismograms. I will also discuss seismic source inversion problems in this book.

A general method for solving seismic inverse problems, and also inverse problems in many other disciplines, is to take an iterative approach. In seismic tomography we can initiate the iteration with an assumed model (i.e., the starting model) for the geological media and solve the forward problem to obtain the model-predicted (i.e., synthetic) observations. The assumed model is likely to be different from the true model, therefore the synthetic observations are likely to be different from the real observations. The discrepancies between the synthetic and the corresponding real observations provide us some feedback that can be used to make corrections to the assumed model. The corrected model can then be used as the reference model for the next iteration and the process can be repeated until a pre-specified convergence

¹ In practice, we may solve for the particle velocity wavefield instead of the displacement wavefield.

criterion is met. Such an iterative approach is often formulated using the mathematical optimization theory. We can define an objective function in terms of the discrepancies between the synthetic and the corresponding real observations and we search for an optimal model that can minimize the objective function. Depending upon the particular optimization algorithm adopted for solving the inverse problem, as well as the complexity of the inverse problem, the forward problem often needs to be solved for many times through the entire optimization process. Therefore our capability for solving the seismic inverse problem depends very much upon our ability of solving the forward problem accurately and efficiently.

Our ability for solving the forward problem depends in turn upon the computing technology. In the past, when computers were slow (by modern standards), techniques based on the seismic ray theory (e.g., Babich 1956, 1961; Červený 1972; Chapman 2002; Gjøystdal et al. 2002; Červený 2005) were in fact the only practically applicable solutions for the seismic wave equation in three-dimensional geological structure models². Analytic or semi-analytic solutions developed for one-dimensional geological structure models (e.g., Thomson 1950; Haskell 1953; Dunkin 1965; Fuchs and Muller 1971; Takeuchi and Saito 1972; Chapman 1973, 1978; Bouchon 1981; Harvey 1981; Kerry 1981; Saikia 1994; Wang 1999; Kawai et al. 2006; Kennett 2009) are usually not directly applicable to complex two-dimensional or three-dimensional structural models. Unlike the analytic or semi-analytic methods, which provide exact solutions of the elastodynamic equation for one-dimensional models, the seismic ray method provides useful approximate solutions of satisfactory accuracy for three-dimensional structural models. The ray method has its own advantages and drawbacks. Its drawbacks have been motivating the development of full-3D modeling and inversion methods based on purely numerical solutions of the elastodynamic equation in the past few decades.

1.1 Advantages of the Seismic Ray Method

The roots of the advantages, as well as the drawbacks, of the seismic ray theory mainly lie in the asymptotic ray series representation of the seismic wavefield. The formal ray series solution of the elastodynamic equation for the displacement wavefield can be represented as an asymptotic series in inverse powers of the angular frequency ω ,

$$\mathbf{u}(\mathbf{x}, t) = \exp\{-i\omega[t - \tau(\mathbf{x})]\} \times \left[\mathbf{U}^{(0)}(\mathbf{x}) + \frac{\mathbf{U}^{(1)}(\mathbf{x})}{(-i\omega)} + \frac{\mathbf{U}^{(2)}(\mathbf{x})}{(-i\omega)^2} + \dots \right] \quad (1.1)$$

² By “three-dimensional” I mean that the physical properties of the geological media vary with three spatial coordinates. If the physical properties do not vary with one (two) spatial coordinate(s), we often say that they are two-dimensional (one-dimensional) geological media. If we apply a three-dimensional method on two-dimensional geological media, such a calculation is sometimes called as 2.5-dimensional.

Here \mathbf{x} and t are spatial and temporal coordinates, $\tau(\mathbf{x})$ is the real-valued travel time and $\mathbf{U}^{(k)}(\mathbf{x})$, $k = 0, 1, 2, \dots$ are complex-valued vectorial amplitude coefficients. If we fix the number of terms in the asymptotic expansion 1.1 (one usually uses only the leading term or just the first few terms sometimes), the error of the approximation can be made arbitrarily small, provided the angular frequency ω being sufficiently high.

Inserting 1.1 into the elastodynamic equation yields a system of equations, which includes the Eikonal equation for the travel time $\tau(\mathbf{x})$,

$$|\nabla\tau(\mathbf{x})| = \frac{1}{v(\mathbf{x})} \quad (1.2)$$

where v is the P- or S-wave velocity (depending upon the type of the wave that is being solved), and successively the equations for the amplitude coefficients $\mathbf{U}^{(k)}(\mathbf{x})$. The Eikonal Eq. 1.2 can be solved very efficiently using the method of characteristics, which provides not only the model-predicted travel-time observations but also the trajectories in the structural model through which the energy propagates (i.e., the ray paths), an aspect highly important in ray-theoretic travel-time tomography. The transport equation, which describes the zeroth-order amplitudes of the waves $\mathbf{U}^{(0)}(\mathbf{x})$, depends primarily upon the geometrical spreading effect, which can be evaluated very efficiently by solving the dynamic (or paraxial) ray tracing system (e.g., Chapman 2004). The dynamic ray tracing system is composed of a set of linear first-order ordinary differential equations along the central ray and the computational cost for solving this system is relatively low.

In addition to its computational efficiency, especially for three-dimensional structural models, the seismic ray method also has several other benefits. For instance, in the seismic ray method, the wavefield is separated into individual elementary waves. The set of rays belonging to a certain elementary wave corresponds to a continuous range of initial conditions, which are often parametrized using the ray parameters. The physical meaning of the ray parameters may vary. In the case of a point source, the ray parameters may represent the initial slowness vector defined in terms of the take-off angles from the source. Such separation of the wavefield allows identification of individual waves and provides clear physical insight into the wave propagation process, which can be very useful in solving inverse problems.

1.2 Drawbacks of the Seismic Ray Method

The limitations of the seismic ray method can also be attributed to the asymptotic ray series representation 1.1 to some extent. The wavefield provided by the seismic ray method contains only those waves that conform to the asymptotic series representation, which are mainly the far-field body waves. It does not include any near-field terms, evanescent waves (e.g., surface waves, head waves) or diffracted waves. The seismic ray method only provides a high-frequency approximation of the wavefield,

which is, in principle, incomplete. The “high-frequency” requirement for the seismic ray method is often translated into a “smoothness” constraint on the geological structure model. The characteristic length of the structural heterogeneity must be considerably larger than the dominant wavelength of the considered wave for the seismic ray method to provide satisfactory accuracy.

The seismic ray method is capable of handling complex geological structure models, but if the heterogeneity of the structural model exceeds a certain degree, the rays may exhibit chaotic behavior (Keers et al. 1997). The rays with slightly different initial conditions may diverge exponentially at large distances from the source. The ray chaos makes two-point ray tracing practically impossible in overly complex structural models. The exponential divergence of neighboring rays with respect to the complexity of the structural model can be quantified using the Lyapunov exponents, which can be estimated from the second-order derivatives of the seismic velocity model prior to ray tracing (Klimeš 2002). For models with abnormal Lyapunov exponents, smoothing of the model is needed to make it suitable for ray tracing.

The seismic ray method may yield highly distorted solutions or fail to provide a solution in certain singular regions or in certain singular directions in anisotropic models. The singular regions may include shadow zones (i.e., the boundary between illuminated and shadow regions), caustics (i.e., regions where the ray amplitude becomes infinite) and critical regions where rays are tangential to material interfaces. In anisotropic models, the ray method may break down in certain singular directions characterized by the coupling between the quasi-shear waves.

Some of the limitations of the zeroth-order seismic ray method can be alleviated, at least in part, using various extensions of the ray method (e.g., Červený and Ravindra 1971; Starnes 1986; Thomson 1989; Ayzenberg et al. 2007; Kravtsov et al. 2011). These extensions are usually designed to remedy specific types of singularities. For problems with many singularities or for regions with multiple types of singularities, the ray method may become too cumbersome to apply. For singularities related to caustics, the Gaussian beam method provides a more general extension of the ray method (e.g., Červený et al. 1982; Popov 1982). It can be shown that by using infinitely broad Gaussian beams we recover the standard Maslov method (Klimeš 1984), another general extension of the ray method (Chapman and Drummond 1982; Thomson and Chapman 1985). These more general extensions of the ray method are still built up on the high-frequency asymptotic solution of the elastodynamic equation.

1.3 Numerical Solutions of the Seismic Wave Equation

In the past few decades seismologists have developed various numerical algorithms to solve the elastodynamic equation. Those numerical algorithms include the finite-difference, finite-element, pseudo-spectral, spectral-element, discontinuous-Galerkin methods. Compared with the seismic ray method, these numerical solutions can provide “complete” representations of the seismic wavefields in highly

complex 3D geological structure models. Of course, these numerical solutions are limited by the accuracy of the underlying numerical algorithms used for solving the elastodynamic equation. Nevertheless, they should contain all possible waves propagating in a given geological structure model and there is no inherent limit in improving the accuracy of the numerical solutions in any frequency range or time window in those numerical algorithms, given sufficient computational resources. The computational cost for obtaining numerical solutions is significantly higher than that for obtaining ray-theoretic solutions, especially for large models³ at high frequencies. However, our computing capability has been increasing exponentially, more than doubling every 2 years in the past 20 years (Fuller et al. 2011). The computing industry marched from the first teraflop machine (1 TFLOP⁴) to the first petaflop machine (1 PFLOP) in just 12 years and it is anticipated that the fastest computer will have a peak performance of about 1 EFLOP in around 2020 (Service 2013). The progress in the computing technology has opened up the possibilities for solving seismic inverse problems using numerical solutions of the three-dimensional elastodynamic equation, an approach that is called “Full-3D seismic Waveform Inversion (F3DWI)” in this book.

Among all the numerical algorithms that have been adopted to solve the seismic wave equation, several algorithms have been developed more substantially than others. In the 1970s and 1980s the finite-difference (FD) method was introduced to simulate SH and P-SV waves on regular, staggered-grid, two-dimensional meshes (Madariaga 1976; Virieux 1984, 1986). Extensions of the FD method to three spatial dimensions and other curvilinear coordinates and to account for anisotropic, visco-elastic material properties were carried out subsequently (e.g., Mora 1989; Igel et al. 1995; Tessmer 1995; Graves 1996; Moczo et al. 2002). The spatial derivatives in the FD method is approximated through differencing schemes and their accuracy is mainly controlled by the number of grid points required to accurately sample the wavelength. The pseudo-spectral (PS) method with Chebychev or Legendre polynomials (Carcione 1994; Tessmer and Kosloff 1994; Igel 1999) partially overcomes some limitations of the FD method and can substantially improve the accuracy of spatial derivatives. The PS method also has its drawbacks, mainly due to the global character of its spatial derivative operator, which makes it relatively cumbersome to account for irregular modeling geometries. And parallelizing the PS method on modern distributed-memory computer clusters in an efficient and scalable way is not as straightforward as parallelizing the FD method.

There are also numerical algorithms based on the weak (i.e., variational) form of the elastodynamic equation. The finite-element (FE) method (Lysmer and Drake 1972; Bao et al. 1998) and the spectral-element (SE) method (Komatitsch and Vilotte 1998; Komatitsch and Tromp 1999, 2002) have received wide attention in

³ By “large”, I mean that the propagation distance is much larger than the dominant wavelength.

⁴ 1 GFLOP = 10^9 FLOPS (floating point operations per second), 1 TFLOP = 10^{12} FLOPS, 1 PFLOP = 10^{15} FLOPS, 1 EFLOP = 10^{18} FLOPS.

the seismic modeling community. An important advantage of the weak form representation is that the free-surface boundary condition is naturally accounted for even for highly irregular surface topography.

In the SE method, each element is mapped to a reference cube using the Jacobian matrix. High-order Lagrange polynomial interpolants are used to represent functions on the reference cube. The control points needed in defining the Lagrange polynomials are chosen to be the classical Gauss-Lobatto-Legendre (GLL) quadrature points. Any smooth 3D function can then be interpolated in a 3D hexahedral element by triple products of Lagrange polynomials at the GLL points. The spatial derivatives of any functions can be obtained by computing the derivatives of the Lagrange polynomials and the integrals of any functions can be approximated using the GLL integration quadrature. A direct consequence of using Lagrange interpolants and GLL quadrature is that the resulting mass matrix of the system is diagonal and a fully explicit time-marching scheme can be adopted.

The discontinuous-Galerkin (DG) method was adopted to solve the elastodynamic equation very recently (Käser and Dumbser 2006; Dumbser and Käser 2006; Käser et al. 2007; de la Puente et al. 2007; Dumbser et al. 2007b). The advantages of the DG method can be summarized as follows.

- In three spatial dimensions, the DG method can achieve high-order accuracy on fully unstructured tetrahedral meshes, which are usually easier to generate than hexahedral meshes, especially for geological structure models with complex geometries associated with topography, faults and other types of structural discontinuities.
- Unlike conventional numerical schemes, which often adopt a relatively low-order time-stepping method such as the Newmark scheme or a high-order, multi-stage method that requires storing intermediate fields such as the Runge-Kutta method, the DG method uses a high-order, single-step, explicit time integration scheme, the Arbitrary high-order DERivatives (ADER) method (Toro 2009).
- The DG method allows for different polynomial degrees inside different elements (i.e., p -adaptivity) and also different time-step lengths for different elements (i.e., local time stepping). In conventional global time-stepping schemes, the element with the smallest time step length dictates its own time step length to all other elements. In the local time-stepping scheme, each element can use a different time-step length that is optimal according to its own local stability condition. This is particularly advantageous for meshes with strongly degenerate elements (i.e., slivers), as the time-step lengths for other elements are not restricted in any way by the small time-step lengths of the badly-shaped elements.
- The DG method is particularly suitable for modeling wave propagation in heterogeneous fluid-solid environment. The Riemann solver used for computing the numerical flux across element interfaces can automatically recognize discontinuities in material properties across highly irregular fluid-solid interfaces (Käser and Dumbser 2008). The acoustic fluid is characterized by simply setting the shear modulus to zero. The DG method also accounts for the convection of the fluid, which allows us to correctly model the effects of ocean currents on wave propagation in complex acoustic environment.

- The majority of the operators in the DG method are applied in an element-local way, with weak nearest-neighbor coupling between elements through numerical flux functions. The resulting locality allows the DG method to achieve very high parallel efficiency on single-instruction-multiple-data (SIMD) distributed-memory parallel computers.

The primary disadvantage of the DG method is its high computational cost. Preliminary results have shown that it is about 50–60 times more expensive than the SE method (Dumbser et al. 2007b). However, some recent results have shown that the DG method for solving the elastodynamic equation can be effectively accelerated using graphic processing units (GPUs) (Mu et al. 2013a, b). Compared with the CPU implementation, the GPU-CPU hybrid implementation can achieve a speedup of about 28 times with a moderate-level optimization of the GPU memory access pattern. This is highly encouraging because with the current computing capability, the GPU-accelerated DG algorithm can already be adopted for solving F3DWI problems of moderate size. With the continued development of the computing technology, the DG method will be applicable to much larger problems in the near future.

Numerical algorithms for solving the elastodynamic equation are not limited to those discussed above. I also mention the boundary-element method (BEM) (Bouchon and Sánchez-Sesma 2007), the finite-volume method (Dumbser et al. 2007a), the variable- and discontinuous-grid finite-difference method (Aoi and Fujiwara 1999; Pitarka 1999), the hybrid finite-difference-finite-element method (Galis et al. 2008). Different numerical algorithms have different advantages and drawbacks. For a more comprehensive review of the numerical algorithms used for solving the elastodynamic problems, please refer to some of the more recent books (Dmowska et al. 2006; Robertsson et al. 2012) and review articles (Takenaka et al. 1998; Moczo et al. 2007; Igel et al. 2009) and the references therein.

1.4 A Brief History of HPC Technology

F3DWI was enabled by recent advances in high-performance computing (HPC). Its wide adoption and continued development will depend upon sustained improvement in HPC technology in the future. Understanding the pathway through which HPC technology has been evolving will help us to design better F3DWI software that can take the full advantage of new HPC systems in the future.

The demand for more computing power has periodically inspired major innovations in HPC hardware. The famous Moore's law predicts that the density of the transistors would double every 2 years, a prediction that has been roughly consistent with observations since the 1970s. However, to translate the exponential growth in transistor density to exponential growth in actual computing power, which is often measured in FLOPS (floating-point-operations-per-second), still requires additional hardware innovations.

In the 1970s and 1980s, the performance improvements were mainly obtained from increasing the clock frequency (i.e., number of clock cycles per second) and engineering more sophisticated instruction scheduling in a single processing core. Many of the added transistors were used to double computer word sizes⁵, to increase the amount of caches⁶ and to add major core-level features, such as additional floating-point units, out-of-order execution of instructions⁷ and pipelining⁸. The dominant programming model for this type of hardware was the single-threaded sequential model. In the 1990s and early 2000s, the increasing number of transistors were still used to add relatively incremental core-level features. At this stage, the objective of the computing industry was to maximize the FLOPS of a single core.

By around 2005, the industry was facing a number of engineering difficulties associated with using a large number of transistors in a single core. One problem was related to energy. The power consumption, as well as the need for dissipating the heat, increases with clock frequency⁹. A second problem was that no more important core-level features could be added. As a consequence, the clock speeds stayed relatively constant at about 3 GHz or less and the core architecture was not changed too much and was expected to become simpler. To effectively utilize the added transistors, a reasonable step was to put more processing cores onto the same chip. The priority of the industry was shifted towards maximizing the FLOPS per unit area of silicon. Today most processors have multiple cores. An important advantage of the multi-core approach is that these cores are identical and have the same instruction set, therefore can be programmed using a relatively homogeneous programming model within the confines of a single processor. Most of the manufacturers currently implement the shared-memory model¹⁰ for multi-core processors and communication among different cores of the same processor can be implemented through memory accesses to a same memory address. This type of architecture is often

⁵ A “word” is the smallest unit of data processed by the CPU. Modern processors usually have a word size of 32-bit or 64-bit.

⁶ A CPU spends most of its time waiting for memory transactions (i.e., reading/writing data from/to the main memory). The cache is a smaller, faster memory on the CPU which stores temporary copies of frequently accessed main-memory data. The cache can help reduce the frequency of memory transactions that have to leave the CPU, thereby improving overall CPU performance.

⁷ An out-of-order execution allows later instructions to be executed before the earlier ones as long as there are no data dependencies among those instructions.

⁸ Each instruction is split into multiple steps and different steps from multiple instructions are executed concurrently on different parts of the same processor. Pipelining is usually used together with the superscalar design, which allows the processor to issue multiple instructions in a single clock cycle.

⁹ $P = C \times V^2 \times F$, where P is power, C is the capacitance being switched per clock cycle (proportional to the number of transistors whose inputs change), V is voltage, and F is the clock frequency. In reality, the CPU power consumption is more complicated than what this formula shows.

¹⁰ All cores on the same processor share a single view of the same memory space and can access data stored in the memory through a single address space.

called Symmetric Multi-processing (SMP), in which two or more identical cores are connected to a single, shared main memory. All cores are controlled by a single operating system instance, which treats all cores equally. Parallelism can be implemented using well-established Application Programming Interfaces (APIs) such as POSIX Threads (Butenhof 1997) and OpenMP (Rendell et al. 2013).

The computing power can also be increased by combining many processors using a network. Each processor runs its own instance of an operating system in its own private memory and communicates with each other by passing messages through a high-speed Local Area Network (LAN). Such a system is often called a “distributed-memory” system. Calculations can only operate on local data and if remote data are needed, they must be transported over the network through “message-passing”. Such a “computer cluster” can be built from inexpensive commodity components, in which case the cluster is often called a Beowulf cluster (Becker et al. 1995). The processors making up the cluster can be shared-memory multi-core processors, in which case the system is often called an SMP-cluster and each multi-core processor in the cluster is called a node. The number of nodes in a supercomputer has increased from a few thousands in the early 1990s to hundreds of thousands today. In systems with a very large number of nodes, the speed and flexibility of the interconnect become highly important. Many of the latest HPC systems, such as the IBM Blue Gene series, the Cray XT series and Fujitsu’s K computer, use multi-dimensional torus networks (Ajima et al. 2009) or other types of specialized interconnect networks. Such systems are often called Massively Parallel Processing (MPP) systems.

The *de facto* standard API for inter-node message exchanges is the Message Passing Interface (MPI), a collection of well-tested library routines for writing portable message-passing parallel programs in a variety of computer languages (Gropp et al. 1999). The MPI library can also be used for passing messages among different cores of the same SMP node, therefore a single-level, flat message-passing only programming model is applicable to clusters of SMP nodes. If the MPI library can be made SMP-aware¹¹ (Träff 2003), such a programming model is both relatively simple to implement and also capable of achieving high-performance on SMP-clusters. Another option for programming SMP-clusters is to use a two-level parallel programming model, in which POSIX Threads or OpenMP can be used for managing intra-node multi-core parallelism and the inter-node message-passing is performed through MPI (Smith and Bull 2001). Such a two-level programming model is more difficult to implement since it requires mastering both the shared-memory and the message-passing programming paradigms. In general this programming model should give better performance on SMP-clusters, since it can adapt better to the underlying hardware.

To increase the number of identical cores in a single processor to a very large number also faces various engineering difficulties related to the effective utilization

¹¹ Message-passing between different cores on the same processor can bypass the network, thereby reducing network performance bottlenecks.

of the ever increasing transistors. But a more pressing issue, especially when the industry is moving towards Exascale, is energy efficiency. Tianhe-2¹², the world's fastest supercomputer at the time of writing, runs at 33.86 PFLOPS and consumes 24 megawatts (MW) of power, which amounts to about 1.4 GFLOPS per watt, a substantial improvement in energy efficiency over previous-generation supercomputers. If an Exascale machine runs at the same efficiency, it would require 709 MW, enough energy to power nearly one million homes. For Exascale HPC systems to be viable, the energy efficiency needs to be improved by about 23 fold (Service 2013). The priority of the industry has now shifted towards maximizing the FLOPS per unit energy.

One approach that has been explored in Tianhe-2 and some other recent systems such as Titan¹³ and Blue Waters¹⁴ is to use heterogeneous many-core nodes. The different cores on the same node can be tailored to the particular calculations run on them so that the energy consumption is minimized while the performance is maximized. Hardware support can also be built into chips for adaptive control of the clock frequency and the voltage to maximize energy efficiency (Rahman 2013). At the current stage, examples of such specialized processing units, which are often called co-processors or accelerators, include the Cell Broadband Engine, Graphic Processing Units (GPUs), Field Programmable Gate Arrays (FPGAs) and the Intel Xeon Phi. These specialized processors usually have simpler architecture, lower power consumption and higher floating-point throughput performance than general-purpose CPUs and are often used alongside general-purpose CPUs to improve the energy efficiency of numerically intensive calculations in latest heterogeneous HPC systems.

At the current stage, there is no general consensus about how to program heterogeneous HPC systems to achieve optimal overall performance. Despite the success of the single-level, flat message-passing-only programming model, there are growing evidences that future parallel software developers will have to address the ever increasing node-level complexities by introducing another level of parallelism. Message-passing may continue to be the dominant model for managing inter-node parallelism, but it may not be sufficient for driving all the different types of cores on a single node for much longer. One possibility is to use threads for intra-node parallelism in conjunction with MPI. However, POSIX Threads and OpenMP, which have been widely adopted in programming homogeneous clusters of SMP nodes, do not provide fine-grained control over thread placement, scheduling and resource management, which are often needed in programming heterogeneous processors. For heterogeneous nodes with many-core accelerators

¹² Tianhe-2 is a hybrid supercomputer built using Intel Xeon E5 and Xeon Phi. Xeon Phi is a type of co-processor built on the Intel Many Integrated Core (MIC) architecture.

¹³ Titan is a GPU-CPU hybrid supercomputer built by Cray at the Oak Ridge National Laboratory. It ranks at the second place in the TOP500 benchmark at the time of writing.

¹⁴ Blue Waters is a GPU-CPU hybrid system built by Cray at the National Center for Supercomputing Applications (NCSA).

such as GPUs, vendor-provided APIs, such as ATI Stream Software Development Kit (Bayoumi et al. 2009) and NVIDIA Compute Unified Device Architecture (CUDA) (Cook 2013), provide high-level instructions for the launching, organization, resource assignment and scheduling of thousands of threads. The emerging industry standard Open Computing Language (OpenCL), which is a framework for programming heterogeneous nodes consisting of multi-core CPUs, GPUs and other types of co-processors (Gaster et al. 2013), specifies a parallel computing model that is very close to CUDA. It is likely that new programming models will emerge as the heterogeneous HPC hardware continues to evolve. A major design goal of our F3DWI software in the future should be the portability and performance scalability on heterogeneous HPC systems.

References

- Ajima, Y., Sumimoto, S., & Shimizu, T. (2009). Tofu: A 6D mesh/torus interconnect for exascale computers. *Computer*, 42(11), 36–40.
- Aoi, S., & Fujiwara, H. (1999). 3D finite-difference method using discontinuous grids. *Bulletin of the Seismological Society of America*, 89(4), 918–930.
- Ayzenberg, M., Aizenberg, A., Helle, H., Klem-Musatov, K., Pajchel, J., & Ursin, B. (2007). 3D diffraction modeling of singly scattered acoustic wavefields based on the combination of surface integral propagators and transmission operators. *Geophysics*, 72, SM19–SM34.
- Babich, V. M. (1956). Ray method of the computation of the intensity of wave fronts (in russian). *Doklady Akademii Nauk SSSR*, 110, 355–357.
- Babich, V. M. (1961). Ray method of the computation of the intensity of wave fronts in elastic inhomogeneous anisotropic medium. In G. I. Petrashen (Ed.), *Problems of the dynamic theory of propagation of seismic waves (in Russian)* (Vol. 5, pp. 36–46). Leningrad: Leningrad University Press.
- Bao, H., Bielak, J., Ghattas, O., Kallivokas, L., O’Hallaron, D., Shewchuk, J., & Xu, J. (1998). Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. *Computer Methods in Applied Mechanics and Engineering*, 152, 85–102.
- Bayoumi, A., Chu, M., Hanafy, Y., Harrell, P., & Refai-Ahmed, G. (2009). Scientific and engineering computing using ATI Stream technology. *Computing in Science & Engineering*, 11(6), 92–97.
- Becker, D. J., Sterling, T., Savarese, D., Dorband, J. E., Ranawak, U. A., & Packer, C. V. (1995). Beowulf: A parallel workstation for scientific computation. In *Proceedings, International Conference on Parallel Processing* (Vol. 95).
- Bouchon, M. (1981). A simple method to calculate Green’s functions for elastic layered media. *Bulletin of the Seismological Society of America*, 71, 959–971.
- Bouchon, M., & Sánchez-Sesma, F. J. (2007). Boundary integral equations and boundary elements methods in elastodynamics. *Advances in Geophysics*, 48, 157–189.
- Butenhof, D. (1997). *Programming with POSIX threads* (Addison-Wesley professional computing series). Addison-Wesley.
- Carcione, J. (1994). The wave equation in generalized coordinates. *Geophysics*, 59, 1911–1919.
- Červený, V. (1972). Seismic rays and ray intensities in inhomogeneous anisotropic media. *Geophysical Journal of the Royal Astronomical Society*, 29, 1–13.
- Červený, V. (2005). *Seismic ray theory*. Cambridge University Press.
- Červený, V., & Ravindra, R. (1971). *Theory of seismic head waves*. University of Toronto Press.

- Červený, V., Popov, M., & Pšenčík, I. (1982). Computation of wave fields in inhomogeneous media, Gaussian beam approach. *Geophysical Journal of the Royal Astronomical Society*, 4, 109–128.
- Chapman, C. H. (1973). The earth flattening transformation in body wave theory. *Geophysical Journal of the Royal Astronomical Society*, 35, 55–70.
- Chapman, C. H. (1978). A new method for computing synthetic seismograms. *Geophysical Journal of the Royal Astronomical Society*, 54, 481–518.
- Chapman, C. H. (2002). Seismic ray theory and finite frequency extensions. In W. H. K. Lee, H. Kanamori, & P. C. Jennings (Eds.), *International handbook of earthquake and engineering seismology* (Vol. Part A, pp. 103–123). Academic Press.
- Chapman, C. H. (2004). *Fundamentals of seismic wave propagation*. Cambridge University Press.
- Chapman, C. H., & Drummond, R. (1982). Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory. *Bulletin of the Seismological Society of America*, 72, S277–S317.
- Cook, S. (2013). *CUDA programming: A developer's guide to parallel computing with GPUs* (Applications of GPU computing series). Morgan Kaufmann.
- de la Puente, J., Käser, M., Dumbser, M., & Igel, H. (2007). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes IV. Anisotropy. *Geophysical Journal International*, 169, 1210–1228.
- Dmowska, R., Wu, R.-S., & Maupin, V. (2006). *Advances in geophysics: Advances in wave propagation in heterogeneous earth* (Vol. 48). Academic Press.
- Dumbser, M., & Käser, M. (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—II. The three-dimensional isotropic case. *Geophysical Journal International*, 167(1), 319–336.
- Dumbser, M., Käser, M., & De La Puente, J. (2007a). Arbitrary high-order finite volume schemes for seismic wave propagation on unstructured meshes in 2d and 3d. *Geophysical Journal International*, 171(2), 665–694.
- Dumbser, M., Käser, M., & Toro, E. F. (2007b). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—V. Local time stepping and p-adaptivity. *Geophysical Journal International*, 171(2), 695–717.
- Dunkin, J. W. (1965). Computations of modal solutions in layered elastic media at high frequencies. *Bulletin of the Seismological Society of America*, 55, 335–358.
- Fuchs, K., & Muller, G. (1971). Computation of synthetic seismograms with reflectivity method and comparison with observations. *Geophysical Journal of the Royal Astronomical Society*, 23, 417–433.
- Fuller, S., Millett, L., Performance, C., Board, C., Sciences, D., & Council, N. (2011). *The future of computing performance: Game over or next level?* National Academies Press.
- Galis, M., Moczo, P., & Kristek, J. (2008). A 3-D hybrid finite-difference finite-element viscoelastic modelling of seismic wave motion. *Geophysical Journal International*, 175(1), 153–184.
- Gaster, B., Howes, L., Kaeli, D., Mistry, P., & Schaa, D. (2013). *Heterogeneous computing with OpenCL*. Morgan Kaufmann.
- Gjøystdal, H., Iversen, E., Laurain, R., Lecomte, I., Vinje, V., & Åstebol, K. (2002). Review of ray theory applications in modelling and imaging of seismic data. *Studia Geophysica et Geodaetica*, 46, 113–164.
- Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America*, 86(4), 1091–1106.
- Gropp, W., Lusk, E., & Skjellum, A. (1999). *Using MPI: Portable parallel programming with the message-passing interface*. Scientific and engineering computation. MIT Press.
- Harvey, D. (1981). Seismogram synthesis using normal mode superposition: The locked mode approximation. *Geophysical Journal of the Royal Astronomical Society*, 66, 37–70.
- Haskell, N. A. (1953). The dispersion of surface waves on multilayered media. *Bulletin of the Seismological Society of America*, 43, 17–34.

- Igel, H. (1999). Wave propagation in three-dimensional spherical sections by the chebyshev spectral method. *Geophysical Journal International*, 136, 559–566.
- Igel, H., Mora, P., & Rioulet, B. (1995). Anisotropic wave propagation through finite-difference grids. *Geophysics*, 60, 1203–1216.
- Igel, H., Käser, M., & Stupazzini, M. (2009). Simulation of seismic wave propagation in media with complex geometries. In W. Lee (Ed.), *Encyclopedia of complexity and system science*. Springer Verlag.
- Käser, M., & Dumbser, M. (2006). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—I. The two-dimensional isotropic case with external source terms. *Geophysical Journal International*, 166(2), 855–877.
- Käser, M., & Dumbser, M. (2008). A highly accurate discontinuous Galerkin method for complex interfaces between solids and moving fluids. *Geophysics*, 73, T23–T35.
- Käser, M., Dumbser, M., de la Puente, J., & Igel, H. (2007). An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—III. Viscoelastic attenuation. *Geophysical Journal International*, 168, 224–242.
- Kawai, K., Takeuchi, N., & Geller, R. (2006). Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media. *Geophysical Journal International*, 164, 411–424.
- Keers, H., Dahlen, F., & Nolet, G. (1997). Chaotic ray behaviour in regional seismology. *Geophysical Journal International*, 131, 361–380.
- Kennett, B. (2009). *Seismic wave propagation in stratified media*. ANU E Press.
- Kerry, N. J. (1981). Synthesis of seismic surface wave. *Geophysical Journal of the Royal Astronomical Society*, 64, 425–446.
- Klimeš, L. (1984). The relation between Gaussian beams and Maslov asymptotic theory. *Studia Geophysica et Geodaetica*, 28, 237–247.
- Klimeš, L. (2002). Lyapunov exponents for 2-d ray tracing without interfaces. *Pure and Applied Geophysics*, 159, 1465–1485.
- Komatitsch, D., & Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, 139, 806–822.
- Komatitsch, D., & Tromp, J. (2002). Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation and self-gravitation. *Geophysical Journal International*, 150(1), 303–318.
- Komatitsch, D., & Vilotte, J. (1998). The spectral-element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, 88, 368–392.
- Kravtsov, Y., Orlov, Y., & Edelev, G. (2011). *Caustics, catastrophes and wave fields* (Springer series on wave phenomena). Springer London, Limited.
- Lysmer, J., & Drake, L. (1972). A finite-element method for seismology. In B. Alder, S. Fernbach, & B. Bolt (Eds.), *Methods in computational physics* (Vol. 11, Chap. 6, pp. 181–216). Academic Press.
- Madariaga, R. (1976). Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America*, 65, 163–182.
- Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R., et al. (2002). 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities. *Bulletin of the Seismological Society of America*, 92, 3042–3066.
- Moczo, P., Kristek, J., Galis, M., Pazak, P., & Balazovjeh, M. (2007). The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. *Acta Physica Slovaca. Reviews and Tutorials*, 57(2), 177–406.
- Mora, P. (1989). Modeling anisotropic seismic waves in 3-D. In 59th Ann. Int. Mtg. Exploration Geophysicists (pp. 1039–1043).
- Mu, D., Chen, P., & Wang, L. (2013a). Accelerating the discontinuous Galerkin method for seismic wave propagation simulations using multiple GPUs with CUDA and MPI. *Earthquake Science*, 26, 377–393.

- Mu, D., Chen, P., & Wang, L. (2013b). Accelerating the discontinuous Galerkin method for seismic wave propagation simulations using the graphic processing unit (GPU)—Single-GPU implementation. *Computers & Geosciences*, *51*, 282–292.
- Pitarka, A. (1999). 3D elastic finite-difference modeling of seismic motion using staggered grids with nonuniform spacing. *Bulletin of the Seismological Society of America*, *89*(1), 54–68.
- Popov, M. (1982). A new method of computation of wave fields using Gaussian beams. *Wave Motion*, *4*, 85–97.
- Rahman, R. (2013). *Intel Xeon Phi coprocessor architecture and tools: The guide for application developers*. Apress.
- Rendell, A., Chapman, B., & Müller, M. (2013). OpenMP in the era of low power devices and accelerators: 9th International Workshop on OpenMP, IWOMP 2013, Canberra, Australia, September 16–18, 2013, Proceedings. LNCS sublibrary: Programming and software engineering. Springer Berlin Heidelberg.
- Robertsson, J. O., Blanch, J. O., Nihei, K., Tromp, J., & No, S. G. R. S. (2012). *Numerical modeling of seismic wave propagation: Gridded two-way wave-equation methods*. Society of Exploration Geophysicists, the international society of applied geophysics.
- Saikia, C. K. (1994). Modified frequency-wavenumber algorithm for regional seismograms using Filon’s quadrature: Modelling of L_g waves in eastern north america. *Geophysical Journal International*, *118*, 142–158.
- Service, R. (2013). Who will step up to Exascale? *Science*, *339*(6117), 264–266.
- Smith, L., & Bull, M. (2001). Development of mixed mode MPI/OpenMP applications. *Scientific Programming*, *9*(2), 83–98.
- Stamnes, J. (1986). *Waves in focal regions: Propagation, diffraction and focusing of light, sound and water waves*. Series in optics and optoelectronics. Taylor & Francis.
- Takenaka, H., Furumura, T., & Fujiwara, H. (1998). Recent developments in numerical methods for ground motion simulation. *The Effects of Surface Geology on Seismic Motion*, *2*, 91–101.
- Takeuchi, H., & Saito, M. (1972). Seismic surface waves. *Methods in Computational Physics*, *II*, 217–295.
- Tessmer, E. (1995). 3-D seismic modelling of general material anisotropy in the presence of the free surface by a Chebyshev spectral method. *Geophysical Journal International*, *121*, 557–575.
- Tessmer, E., & Kosloff, D. (1994). 3-D elastic modeling with surface topography by a Chebyshev spectral method. *Geophysics*, *59*(3), 464–473.
- Thomson, W. T. (1950). Transmission of elastic waves through a stratified solid medium. *Journal of Applied Physics*, *21*, 89–93.
- Thomson, C. J. (1989). Corrections for grazing rays to 2-D seismic modelling. *Geophysical Journal International*, *96*, 415–446.
- Thomson, C. J., & Chapman, C. H. (1985). An introduction to Maslov’s asymptotic method. *Geophysical Journal of the Royal Astronomical Society*, *83*, 143–168.
- Toro, E. (2009). *Riemann solvers and numerical methods for fluid dynamics: A practical introduction*. Springer.
- Träff, J. L. (2003). SMP-aware message passing programming. In *Parallel and Distributed Processing Symposium, 2003*, 56–65. *Proceedings. International* (pp. 10–pp). IEEE.
- Virieux, J. (1984). SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, *49*, 1933–1942.
- Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, *51*, 889–901.
- Wang, R. (1999). A simple orthonormalization method for stable and efficient computation of Green’s functions. *Bulletin of the Seismological Society of America*, *89*(3), 733–741.

Chapter 2

Anelastic Wave Propagation (AWP)

Among all the numerical methods for solving the 3D elastodynamic equation, the finite-difference method is probably the most widely used. In this book, the numerical wave-equation solver adopted for solving F3DWI problems is the Anelastic Wave Propagation (AWP) code developed and maintained by Kim B. Olsen, Steven M. Day and Yifeng Cui (ODC) (Olsen 1994; Olsen et al. 1995, 2003, 2006, Cui et al. 2009). It implements the staggered-grid finite-difference scheme to solve the three-dimensional velocity-stress elastodynamic equation. The scheme is fourth-order accurate in space and second-order accurate in time. The AWP-ODC code implements the traction-free boundary condition for the free surface and can use either the Absorbing Boundary Conditions (ABC) of Cerjan (Cerjan et al. 1985) or the Perfectly Matched Layers (PML) of Berenger (Berenger 1994; Marcinkovich and Olsen 2003) for the four sides and the bottom of the computation domain. Representations of the seismic sources can be included in the staggered-grid system using the stress components for moment sources and/or the velocity components for body-force sources.

The code is written in Fortran 77/90. Message passing is implemented using MPI with domain decomposition. Each processor is responsible for updating the velocity and stress fields on its own portion of the mesh and also for handling the boundary conditions if some of its grid points are located on the external boundaries of the modeling volume. Two-grid-point-thick ghost layers containing the Latest velocity and stress fields are used for exchanging information with the neighboring sub-meshes. Some very large-scale simulations may take several days to execute therefore the AWP-ODC code includes checkpoint/restart capability. To ensure the integrity of the simulation output, the AWP-ODC code generates MD5 checksums in parallel for each mesh sub-array.

The accuracy of the AWP-ODC code has been extensively validated by researchers at the Southern California Earthquake Center (SCEC) in the SCEC 3D Numerical Simulation Code Validation Project (Day et al. 2001, 2003, 2005) using a hierarchy of test problems ranging from simple point-source problems in half-space or layered structural models (i.e., material properties are constant or vary only with depth) to finite-rupture sources in complex 3D structural models. Five different codes, including an earlier version of AWP-ODC, were compared in this validation

project. Four of them were finite-difference codes (Larsen and Schultz 1995; Graves 1996; Pitarka 1999) and one was a finite-element code (Akcelik et al. 2003). The validation was carried out mainly through cross-comparisons of the synthetic seismograms generated by the different numerical codes. For half-space and layered structural models, synthetic seismograms generated using a semi-analytic method were also included in the comparisons.

The AWP-ODC code has been extensively optimized for performance improvements on different types of HPC systems during the TeraShake project (Cui et al. 2009), an inter-disciplinary, collaborative project for simulating earthquake ruptures and seismic wave propagation in Southern California using the NSF-funded TeraGrid HPC resources¹. The result of those extensive optimization efforts is that the AWP-ODC code can now easily scale to tens of thousands of processors with very high parallel efficiency. In 2009 the AWP-ODC code was able to scale from 4096 to 40,960 cores with about 96% parallel efficiency on the Blue Gene/L system at the IBM TJ Watson Research Center in a strong scaling² experiment that used about 32 billion grid points. The weak scaling³ performance was also nearly perfect for up to 32,768 processors on the same system. In 2010, the AWP-ODC code was used for simulating seismic wavefields generated by a hypothetical magnitude-8 earthquake on southern San Andreas Fault⁴. The simulation was performed on 223,074 cores on the Jaguar Cray XT5 system at the National Center for Computational Sciences (NCCS) of the Oak Ridge National Laboratory in Tennessee and the AWP-ODC code achieved 220 TFLOPS of sustained performance. This study was selected as one of the year's Gordon Bell Prize⁵ finalists. Recently, optimization effort on the AWP-ODC code has been mainly focused on porting the code to GPUs and improving its efficiency on CPU-GPU hybrid heterogeneous HPC systems (Zhou et al. 2013).

¹ The TeraGrid was a grid-computing infrastructure combining HPC resources at 11 different partner sites. It was supported by the US National Science Foundation and operated from 2004 through 2011. The TeraGrid project was succeeded by the Extreme Science and Engineering Discovery Environment (XSEDE), a partnership involving HPC resources from 17 institutions, which has been supported by NSF since 2011.

² The overall problem size is fixed and the total number of processors is increased. Therefore the workload on each processor decreases with the number of processors.

³ The workload for each processor is constant and the overall problem size increases in proportion with the number of processors.

⁴ The "M8" earthquake simulation was the largest earthquake simulation coordinated by SCEC. It used a uniform mesh with 40-m grid spacing and the simulation volume was 810-km long, 405-km wide and 85-km deep. The total number of finite-difference grid points used in this simulation was about 436 billion. The full simulation took about 24 h and produced about 6 min of the 3D wavefields in the modeling volume.

⁵ The Gordon Bell prize is awarded by the Association of Computing Machinery (ACM) each year to recognize outstanding achievements in high-performance computing, especially in applying HPC technology to science and engineering applications.

2.1 Formulation

The AWP-ODC code solves the strong formulation⁶ of the velocity-stress equation of motion

$$\rho \dot{\mathbf{v}} = \nabla \cdot \boldsymbol{\sigma} + \mathbf{f} \quad (2.1)$$

where ρ is the density, $\dot{\mathbf{v}}$ is the time derivative of the particle-velocity vector (i.e., particle acceleration), $\boldsymbol{\sigma}$ is the stress tensor and \mathbf{f} is the body-force source. I use the over-dot (i.e., Newton's notation) to represent time derivatives in the following equations. I have dropped the dependence on the spatial and temporal coordinates \mathbf{x} and t of the variables to clarify the basic physics. Written in index notation, Eq. 2.1 can be expressed as

$$\rho \dot{v}_i = \sigma_{ij,j} + f_i \quad (2.2)$$

where $i, j, k = 1, 2, 3$, repeated indices imply summation (i.e., Einstein convention), and $\sigma_{ij,j}$ is the divergence of the stress tensor $\sum_{j=1}^3 \partial \sigma_{ij} / \partial x_j$.

2.1.1 Elastic Media

For three-dimensional isotropic elastic media, the stress-strain relation used in the code is

$$\dot{\sigma}_{ij} = \lambda \dot{\epsilon}_{kk} \delta_{ij} + 2\mu \dot{\epsilon}_{ij} - \dot{m}_{ij} \quad (2.3)$$

where λ and μ are the time-independent, but spatially-varying, Lamé parameters,

$$\dot{\epsilon}_{ij} = \frac{1}{2} (v_{i,j} + v_{j,i}) \quad (2.4)$$

is the strain rate tensor, δ_{ij} is the Kronecker delta, m_{ij} is the moment-density tensor of any moment source and \dot{m}_{ij} is its time derivative (i.e., moment-rate-density tensor). The strain rate tensor $\dot{\epsilon}_{ij}$ is defined in terms of the spatial gradient of the particle velocity vector $\partial v_i / \partial x_j$ and its transpose $\partial v_j / \partial x_i$. Denoting the particle displacement vector as \mathbf{u} , the symmetric strain tensor $\boldsymbol{\epsilon}$ can then be expressed as

$$\begin{aligned} \boldsymbol{\epsilon} &= \frac{1}{2} \left[\nabla \mathbf{u} + (\nabla \mathbf{u})^T \right] \\ &= \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \\ \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{\partial u_2}{\partial x_2} & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) \\ \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) & \frac{\partial u_3}{\partial x_3} \end{bmatrix}. \end{aligned} \quad (2.5)$$

⁶ The solution we are seeking must be sufficiently smooth and have regular second-order spatial derivatives. In the weak formulation, the solution we are seeking may have only first-order derivatives (Evans 2010).

This strain tensor will be playing an important role in constructing the Fréchet kernels in Chap. 4.

2.1.2 Viscoelastic Media

For three-dimensional isotropic viscoelastic media, the stress-strain relation can be defined as (Christensen 2003)

$$\sigma_{ij}(t) = \delta_{ij} \int_{-\infty}^t \lambda(t-\tau) \dot{\epsilon}_{kk}(\tau) d\tau + 2 \int_{-\infty}^t \mu(t-\tau) \dot{\epsilon}_{ij}(\tau) d\tau \quad (2.6)$$

where the Lamé parameters, $\lambda(t)$ and $\mu(t)$, are time-dependent relaxation functions. Here I have dropped the dependence on spatial coordinates \mathbf{x} of the functions to reduce clutter. A direct integration of Eq. 2.6 is numerically intractable due to unrealistic requirements on computer memory and computing time, approximations are therefore needed. The standard approach is to convert the convolutory stress-strain relation Eq. 2.6 into a differential form, which can be numerically solved more easily using the finite-difference method.

2.1.2.1 The Complex Modulus and its Approximation

To demonstrate the basic physics, I will use a simplified scalar notation and write a scalar stress-strain relation analogous to Eq. 2.6

$$\sigma(t) = \int_{-\infty}^t \psi(t-\tau) \dot{\epsilon}(\tau) d\tau = \psi(t) * \dot{\epsilon}(t) \quad (2.7)$$

where $*$ represents the temporal convolution. I will deal with the conversion back to index notations for stress and strain tensors afterwards. Here the stress relaxation function $\psi(t)$ is the stress response to a unit Heaviside function in strain. In the frequency domain, Eq. 2.7 can be expressed as

$$\hat{\sigma}(\omega) = \hat{\psi}(\omega)(i\omega)\hat{\epsilon}(\omega) = \hat{M}(\omega)\hat{\epsilon}(\omega) \quad (2.8)$$

where $\hat{\sigma}$, $\hat{\psi}$ and $\hat{\epsilon}$ are Fourier transforms of the stress, the stress relaxation function and the strain and ω is the angular frequency. Here I have introduced the frequency-domain complex modulus

$$\hat{M}(\omega) = (i\omega)\hat{\psi}(\omega) \quad (2.9)$$

which, in the time domain, can be expressed as

$$M(t) = \dot{\psi}(t) \quad (2.10)$$

and corresponds to the stress response to the Dirac function in strain. The stress-strain relation 2.8 can then be expressed as

$$\sigma(t) = M(t) * \epsilon(t) \quad (2.11)$$

in the time domain. The anelastic attenuation factor or the seismic quality factor $Q(\omega)$ is defined in terms of the complex modulus as

$$Q(\omega) = \frac{\Re[\hat{M}(\omega)]}{\Im[\hat{M}(\omega)]} = \frac{1}{\tan \angle \hat{M}(\omega)} \quad (2.12)$$

where $\angle \hat{M}(\omega)$ is the phase of the complex modulus. For Eq. 2.8 to be causal, $\hat{M}(\omega)$ must satisfy the Kramers-Kronig relation⁷, therefore $\hat{M}(\omega)$ is uniquely determined by a given quality factor $Q(\omega)$ and vice versa. Observations have shown that $Q(\omega)$ is nearly constant over the seismic frequency range (i.e., periods from about 0.01 s to 1 h) (McDonal et al. 1958; Liu et al. 1976; Spencer 1981; Murphy III 1982). Our goal is to find an approximation to the complex modulus $\hat{M}(\omega)$ such that the quality factor $Q(\omega)$ fits a given target frequency-dependence over a specified frequency range and also the stress-strain relation Eq. 2.11 can be easily solved through a time-domain finite-differencing scheme. An approximation that has been widely used in the literature has the form

$$\hat{M}(\omega) = M_U \left(1 - \sum_{l=1}^N \frac{w_l}{1 + i\omega\tau_l} \right) \quad (2.13)$$

where M_U is the unrelaxed modulus, i.e.,

$$M_U = \lim_{\omega \rightarrow \infty} \hat{M}(\omega) = \lim_{t \rightarrow 0} \psi(t) \quad (2.14)$$

and the two sets of parameters w_l and τ_l ($l = 1, 2, \dots, N$), as well as N , can be chosen in such a way that the corresponding quality factor (considering Eq. 2.12),

$$Q(\omega) = \frac{1 - \sum_{l=1}^N \frac{w_l}{1 + (\omega\tau_l)^2}}{\sum_{l=1}^N \frac{w_l\omega\tau_l}{1 + (\omega\tau_l)^2}} \quad (2.15)$$

can adequately fit a target frequency-dependent model (Day and Minster 1984; Emmerich and Korn 1987; Blanch et al. 1995; Liu and Archuleta 2006). The time-domain expression for the approximate modulus 2.13 can be obtained through inverse Fourier transform,

$$M(t) = M_U \delta(t) - M_U \sum_{l=1}^N \frac{w_l}{\tau_l} e^{-t/\tau_l} H(t), \quad (2.16)$$

⁷ Causality implies that the complex function is analytic. The Kramers-Kronig relation indicates that the real and imaginary parts of any analytic function are related through the Hilbert transform, therefore the full function can be reconstructed by knowing just one of its parts. A corollary is that an analytic function can be reconstructed by knowing only its phase.

where $\delta(t)$ is the Dirac function and $H(t)$ is the Heaviside function. Bring Eq. 2.16 into the stress-strain relation Eq. 2.11, we obtain

$$\sigma(t) = M_U \epsilon(t) - M_U \sum_{l=1}^N \frac{w_l}{\tau_l} \int_{-\infty}^t \epsilon(\tau) e^{-(t-\tau)/\tau_l} d\tau. \quad (2.17)$$

Introduce the memory variables⁸,

$$\zeta_l(t) = \frac{w_l}{\tau_l} \int_{-\infty}^t \epsilon(\tau) e^{-(t-\tau)/\tau_l} d\tau, \quad (2.18)$$

the stress can be expressed as

$$\sigma(t) = M_U \left[\epsilon(t) - \sum_{l=1}^N \zeta_l(t) \right]. \quad (2.19)$$

Taking the time derivative on both sides of Eq. 2.18, we obtain

$$\begin{aligned} \dot{\zeta}_l(t) &= -\frac{w_l}{\tau_l^2} \int_{-\infty}^t \epsilon(\tau) e^{-(t-\tau)/\tau_l} d\tau + \frac{w_l}{\tau_l} \int_{-\infty}^{+\infty} \dot{H}(t-\tau) \epsilon(\tau) e^{-(t-\tau)/\tau_l} d\tau \\ &= -\frac{1}{\tau_l} \zeta_l(t) + \frac{w_l}{\tau_l} \int_{-\infty}^{+\infty} \delta(t-\tau) \epsilon(\tau) e^{-(t-\tau)/\tau_l} d\tau \\ &= -\frac{1}{\tau_l} \zeta_l(t) + \frac{w_l}{\tau_l} \epsilon(t). \end{aligned}$$

Therefore the memory variables $\zeta_l(t)$ follow simple first-order ordinary differential equations (ODEs)

$$\tau_l \dot{\zeta}_l(t) + \zeta_l(t) = w_l \epsilon(t) \quad (2.20)$$

which has a time-stepping solution (Day 1998; Liu and Archuleta 2006) given by the first-order exponential integrator method⁹ (Certaine 1960; Pope 1963)

$$\zeta_l \left(t_n + \frac{\Delta t}{2} \right) = e^{-\frac{\Delta t}{\tau_l}} \zeta_l \left(t_n - \frac{\Delta t}{2} \right) + w_l \left[1 - e^{-\frac{\Delta t}{\tau_l}} \right] \epsilon(t_n) \quad (2.21)$$

and also a finite-difference solution

$$\zeta_l \left(t_n + \frac{\Delta t}{2} \right) = \frac{\left(\frac{\tau_l}{\Delta t} - \frac{1}{2} \right) \zeta_l \left(t_n - \frac{\Delta t}{2} \right) + w_l \epsilon(t_n)}{\frac{\tau_l}{\Delta t} + \frac{1}{2}} \quad (2.22)$$

⁸ In the literature these variables are also known as relaxation functions, anelastic functions, internal variables, additional functions, etc.

⁹ To obtain Eq. 2.21, multiply both sides of Eq. 2.20 with e^{t/τ_l} and integrate over t from $t_n - \Delta t/2$ to $t_n + \Delta t/2$. The strain is approximated as a constant $\epsilon(t) \approx \epsilon(t_n)$ over the integration interval.

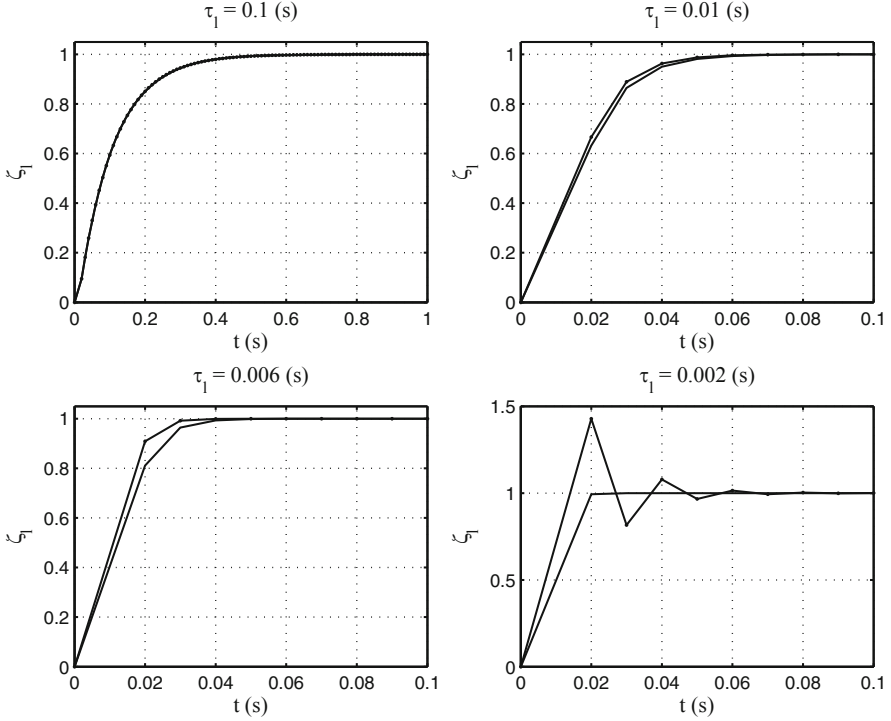


Fig. 2.1 Solution of the memory variable obtained using the exponential integrator method Eq. 2.21 (*solid lines*) and that obtained using the finite-difference method Eq. 2.22 (*dotted solid lines*) for different relaxation time τ_l . The initial condition is set as $\zeta_l(0) = 0$, the weight coefficient $w_l = 1$, the strain is set to a constant $\epsilon = 1$, the time-step length $\Delta t = 0.01$ s.

where Δt is the time-step length and $t_n = (n - 1)\Delta t$ is the time of the n^{th} time step. In general, both solutions are consistent with each other. If we approximate $\exp(-\Delta t/\tau_l)$ in Eq. 2.21 using its Padé approximant¹⁰

$$(1 - 0.5\Delta t/\tau_l) / (1 + 0.5\Delta t/\tau_l), \quad (2.23)$$

we arrive at Eq. 2.22. For very small τ_l (i.e., the ODE is stiff¹¹), the exponential integrator solution 2.21 can alleviate some of the stiffness because part of the ODE is integrated exactly (Cox and Matthews 2002). Comparisons of the two solutions for a fixed Δt and different τ_l are shown in Fig. 2.1. When τ_l is reduced to 0.002 s,

¹⁰ The Padé approximant provides an approximation of a function using rational functions and often gives better accuracy than a truncated Taylor series (Baker and Graves-Morris 1996).

¹¹ The solutions of some differential equations may contain both rapid-varying, transient modes and longer-period slow modes. The numerical algorithms for solving such equations need to use very small time steps to ensure numerical stability even after the rapid-varying, transient modes are no longer visible in the solutions. Such differential equations are said to be stiff (Heath 2005).

the finite-difference solution shows strong oscillation, while the exponential integrator solution increases monotonically to 1. The oscillation in the finite-difference solution can be reduced by decreasing Δt , at the expense of higher computational cost. In each time step, after the memory variables are updated using Eq. 2.21 or 2.22, the stress can then be updated using Eq. 2.19.

2.1.2.2 Physical Interpretation of the Approximate Modulus

The approximation in Eq. 2.13 can be understood in multiple ways. For example, the $2N$ values of τ_l and ω_l can be related to the poles and residues of the rational approximation for $\hat{M}(\omega)$ (Day and Minster 1984). A more physical interpretation is in terms of the Generalized Zener Body (GZB, i.e., the Standard Linear Solid model) or the Generalized Maxwell Body (GMB-EK) rheological models. The GZB and the GMB-EK models are equivalent to each other and one can be derived from the other (Moczo and Kristek 2005). These generalized rheological models are based on serial and parallel compositions of two simplest linear rheological models: the Hooke body (i.e., the stress is proportional to the strain), which has the stress-strain relation

$$\hat{\sigma}_H(\omega) = M_H \hat{\epsilon}_H(\omega) \quad (2.24)$$

and is often depicted as a spring, and the Stokes body (i.e., the stress is proportional to the strain rate), which in the frequency domain has the stress-strain relation

$$\hat{\sigma}_S(\omega) = i\omega\eta_S \hat{\epsilon}_S(\omega) \quad (2.25)$$

and is often depicted as a dash pot. Here η_S is the time-independent viscosity. A Maxwell body is a Hooke body connected in series with a Stokes body. The stress in the Maxwell body $\hat{\sigma}_M$ should be equal to that in its Hooke and Stokes components, i.e.,

$$\hat{\sigma}_M = \hat{\sigma}_H = \hat{\sigma}_S \quad (2.26)$$

and the strain $\hat{\epsilon}_M$ should be equal to the summation of the strain in its Hooke and Stokes components, i.e.,

$$\hat{\epsilon}_M = \hat{\epsilon}_H + \hat{\epsilon}_S = \frac{\hat{\sigma}_H}{M_H} + \frac{\hat{\sigma}_S}{i\omega\eta_S} = \frac{M_H + i\omega\eta_S}{i\omega\eta_S M_H} \hat{\sigma}_M \quad (2.27)$$

therefore the effective modulus of a single Maxwell body in the frequency domain \hat{M}_M can be expressed using the modulus of its Hooke component and the viscosity of its Stokes component as

$$\hat{M}_M(\omega) = \frac{i\omega\eta_S M_H}{M_H + i\omega\eta_S}. \quad (2.28)$$

A Generalized Maxwell Body (GMB) is N Maxwell bodies connected in parallel. A GMB-EK body is N Maxwell bodies and a single Hooke body connected in parallel

(Emmerich and Korn 1987). The stress in a GMB-EK body should be equal to the summation of the stresses in all of its parallel components and the strain should be equal to that in each of its individual parallel components. The effective modulus for such a GMB-EK body is therefore

$$\hat{M}(\omega) = M_H + \sum_{l=1}^N \frac{i\omega\eta_l M_l}{M_l + i\omega\eta_l} \quad (2.29)$$

where M_l and η_l are the modulus and the viscosity of the Hooke body and the Stokes body in the l^{th} Maxwell body and M_H is the modulus of the single Hooke body that is connected in parallel with the N Maxwell bodies. The unrelaxed and relaxed moduli for the GMB-EK body can therefore be derived using Eq. 2.29

$$M_U = \lim_{\omega \rightarrow \infty} \hat{M}(\omega) = M_H + \sum_{l=1}^N M_l, \quad (2.30)$$

$$M_R = \lim_{\omega \rightarrow 0} \hat{M}(\omega) = M_H. \quad (2.31)$$

If we introduce

$$\Delta M = M_U - M_R = \sum_{l=1}^N M_l \quad (2.32)$$

we can express the effective modulus of the GMB-EK body Eq. 2.29 using M_U and ΔM as

$$\begin{aligned} \hat{M}(\omega) &= M_U - \Delta M + \sum_{l=1}^N \frac{i\omega\eta_l M_l}{M_l + i\omega\eta_l} \\ &= M_U - \Delta M \left(\sum_{l=1}^N \frac{M_l}{\Delta M} - \sum_{l=1}^N \frac{M_l}{\Delta M} \frac{i\omega\eta_l}{M_l + i\omega\eta_l} \right) \\ &= M_U - \Delta M \sum_{l=1}^N \frac{M_l}{\Delta M} \left(1 - \frac{i\omega\eta_l}{M_l + i\omega\eta_l} \right) \\ &= M_U \left[1 - \frac{\Delta M}{M_U} \sum_{l=1}^N \frac{M_l}{\Delta M} \left(\frac{1}{1 + i\omega\eta_l/M_l} \right) \right] \\ &= M_U \left(1 - \sum_{l=1}^N \frac{M_l/M_U}{1 + i\omega\eta_l/M_l} \right). \end{aligned}$$

Compare with Eq. 2.13, the parameters w_l and τ_l in Eq. 2.13 can be understood as

$$w_l = M_l/M_U, \quad (2.33)$$

$$\tau_l = \eta_l / M_l, \quad (2.34)$$

and in the literature τ_l and w_l are often called the relaxation time and the weight coefficient, respectively. The inverse of the relaxation time

$$\omega_l = 1/\tau_l \quad (2.35)$$

is called the relaxation frequency. Another widely used notation for the weight coefficient is a_l , which is related to w_l through

$$w_l = a_l \frac{\Delta M}{M_U}. \quad (2.36)$$

Considering Eqs. 2.32 and 2.33, we have

$$\sum_{l=1}^N a_l = 1. \quad (2.37)$$

Bring Eq. 2.36 into the approximate modulus Eq. 2.13 and the ODEs for the memory variables Eq. 2.20, we obtain another set of equations also widely used in the literature

$$\hat{M}(\omega) = M_U - \Delta M \sum_{l=1}^N \frac{a_l \omega_l}{\omega_l + i\omega}, \quad (2.38)$$

$$\dot{\zeta}_l(t) + \omega_l \zeta_l(t) = \omega_l a_l \frac{\Delta M}{M_U} \epsilon(t). \quad (2.39)$$

Here I have used the relaxation frequency ω_l instead of the relaxation time τ_l in Eqs. 2.38 and 2.39. The approximate modulus can also be expressed in terms of the relaxed modulus M_R . Since $M_U = M_R + \Delta M$, considering Eq. 2.38, we have

$$\hat{M}(\omega) = M_R + \Delta M \left(1 - \sum_{l=1}^N \frac{a_l \omega_l}{\omega_l + i\omega} \right).$$

Considering Eq. 2.37, we obtain (Emmerich and Korn 1987)

$$\begin{aligned} \hat{M}(\omega) &= M_R + \Delta M \left(\sum_{l=1}^N a_l - \sum_{l=1}^N \frac{a_l \omega_l}{\omega_l + i\omega} \right) \\ &= M_R + \Delta M \sum_{l=1}^N \frac{i a_l \omega}{\omega_l + i\omega}. \end{aligned} \quad (2.40)$$

Based on this notation, the inverse of the quality factor can be expressed as (Emmerich and Korn 1987)

$$Q^{-1}(\omega) = \frac{\sum_{l=1}^N Y_l \frac{\omega/\omega_l}{1+(\omega/\omega_l)^2}}{1 + \sum_{l=1}^N Y_l \frac{(\omega/\omega_l)^2}{1+(\omega/\omega_l)^2}}, \quad (2.41)$$

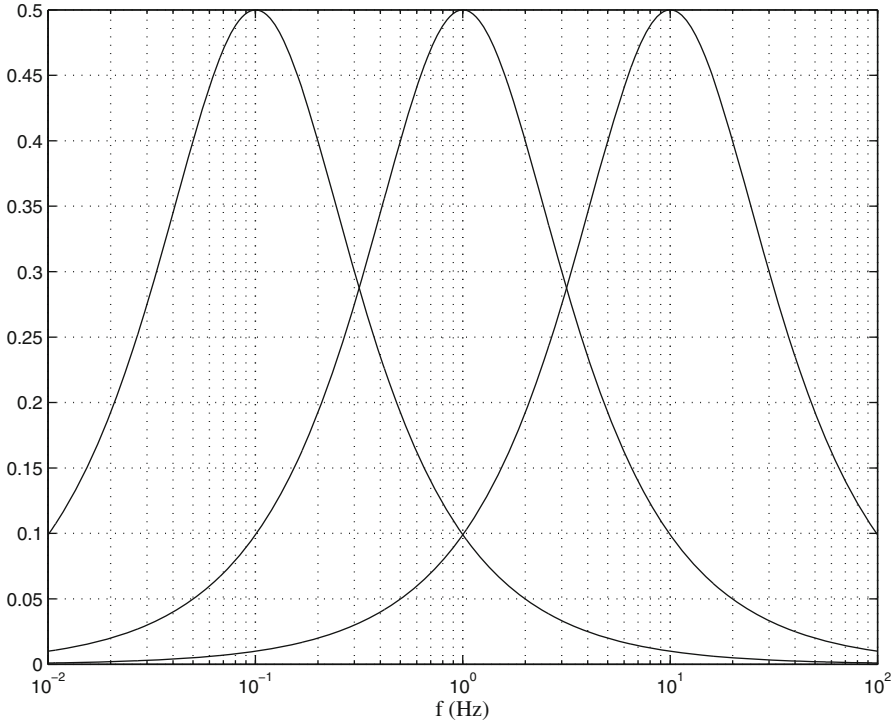


Fig. 2.2 Examples of the Debye function for $\omega_l = 0.2\pi, 2\pi, 20\pi$ and $Y_l = 1$.

where the weight coefficient Y_l is related to a_l through

$$Y_l = \frac{\Delta M}{M_R} a_l. \quad (2.42)$$

In the case that $\Delta M/M_R \ll 1$, we can approximate $Q^{-1}(\omega)$ as

$$Q^{-1}(\omega) \approx \sum_{l=1}^N Y_l \frac{\omega/\omega_l}{1 + (\omega/\omega_l)^2}, \quad (2.43)$$

which is a summation of Debye functions (Emmerich and Korn 1987). The Debye functions are symmetric on a logarithmic scale and centered at frequency ω_l . Examples of the Debye function are shown in Fig. 2.2.

A Zener body is composed of a Maxwell body and a Hooke body connected in parallel. The GZB consists of N Zener bodies connected in parallel. An analysis of Eq. 2.13 based on the GZB rheological model can also be carried out (Liu et al. 1976; Carcione 2007; Moczo et al. 2007).

2.1.2.3 Isotropic Viscoelastic Media

For isotropic media, the stress-strain relation Eq. 2.6 can be partitioned into two separate relations analogous to Eq. 2.19. Both the stress tensor and the strain tensor can be decomposed into a volumetric component and a deviatoric component. For the stress tensor, we have

$$\sigma_{ij} = \bar{\sigma}\delta_{ij} + \sigma'_{ij} \quad (2.44)$$

where the volumetric stress¹² is defined as

$$\bar{\sigma} = \sigma_{kk}/3 \quad (2.45)$$

and the deviatoric stress is defined as

$$\sigma'_{ij} = \sigma_{ij} - \delta_{ij}\sigma_{kk}/3. \quad (2.46)$$

The strain tensor can be decomposed in a similar way,

$$\epsilon_{ij} = \bar{\epsilon}\delta_{ij} + \epsilon'_{ij} \quad (2.47)$$

where the volumetric strain is

$$\bar{\epsilon} = \epsilon_{kk}/3 \quad (2.48)$$

and the deviatoric strain is

$$\epsilon'_{ij} = \epsilon_{ij} - \delta_{ij}\epsilon_{kk}/3. \quad (2.49)$$

Here repeated indices imply summation (i.e., Einstein convention). For elastic isotropic media, the volumetric stress $\bar{\sigma}$ and the volumetric strain $\bar{\epsilon}$ are related through the bulk modulus κ

$$\bar{\sigma} = 3\kappa\bar{\epsilon} = \kappa\epsilon_{kk} \quad (2.50)$$

and the deviatoric stress σ'_{ij} and the deviatoric strain ϵ'_{ij} are related through the shear modulus μ

$$\sigma'_{ij} = 2\mu\epsilon'_{ij}. \quad (2.51)$$

Analogous to Eq. 2.19, we can introduce two sets of memory variables $\bar{\zeta}_l$ and $l\zeta'_{ij}$,

$$\bar{\sigma}(t) = \kappa_U \left[\epsilon_{kk}(t) - \sum_{l=1}^N \bar{\zeta}_l(t) \right], \quad (2.52)$$

$$\sigma'_{ij}(t) = 2\mu_U \left[\epsilon'_{ij}(t) - \sum_{l=1}^N l\zeta'_{ij}(t) \right], \quad (2.53)$$

¹² The volumetric stress is also called the mean stress or the hydrostatic stress in the literature.

where κ_U and μ_U are the unrelaxed bulk and shear moduli. Analogous to Eq. 2.39, the ODEs that the memory variables must follow are

$$\dot{\zeta}_l + \omega_l \bar{\zeta}_l = \omega_l a_l \frac{\Delta\kappa}{\kappa_U} \epsilon_{kk}, \quad (2.54)$$

$$l\dot{\zeta}'_{ij} + \omega_l l\zeta'_{ij} = \omega_l a_l \frac{\Delta\mu}{\mu_U} \epsilon'_{ij}, \quad (2.55)$$

where $\Delta\kappa = \kappa_U - \kappa_R$, $\Delta\mu = \mu_U - \mu_R$ and κ_R and μ_R are the relaxed bulk and shear moduli respectively. Here the Einstein convention applies to the repeated index k , but does not apply to the repeated index l .

In principle, the relaxation frequency ω_l and the weight coefficient a_l in Eq. 2.54 do not need to be identical to those in Eq. 2.55, which means that the bulk and the shear moduli do not need to follow the same frequency-dependence. However, if we assume that they are identical (i.e., the bulk and the shear moduli follow the same frequency-dependence), we can simplify the equations (Day and Bradley 2001). Considering Eqs. 2.44, 2.52 and 2.53, we introduce a new set of memory variables

$$l\xi_{ij} = \kappa_U \bar{\zeta}_l + 2\mu_U l\zeta'_{ij}. \quad (2.56)$$

Multiplying Eq. 2.54 and 2.55 with $\kappa_U \delta_{ij}$ and $2\mu_U$, respectively, and adding them together, considering that the relaxation frequency ω_l and the weight coefficient a_l in Eqs. 2.54 and 2.55 are identical, we obtain the ODEs for the new memory variables $l\xi_{ij}$ as

$$l\dot{\xi}_{ij} + \omega_l l\xi_{ij} = \omega_l a_l \left(\Delta\kappa \epsilon_{kk} \delta_{ij} + 2\Delta\mu \epsilon'_{ij} \right). \quad (2.57)$$

Considering Eq. 2.49, we can re-write the ODEs in terms of the strain tensor ϵ_{ij} as

$$l\dot{\xi}_{ij} + \omega_l l\xi_{ij} = \omega_l a_l \left[\left(\Delta\kappa - \frac{2}{3} \Delta\mu \right) \epsilon_{kk} \delta_{ij} + 2\Delta\mu \epsilon_{ij} \right]. \quad (2.58)$$

Considering Eqs. 2.44, 2.52 and 2.53, the stress tensor σ_{ij} can be expressed in terms of the new memory variables as

$$\sigma_{ij} = \left(\kappa_U - \frac{2\mu_U}{3} \right) \epsilon_{kk} \delta_{ij} + 2\mu_U \epsilon_{ij} - \sum_{l=1}^N l\xi_{ij}. \quad (2.59)$$

Analogous to the elastic stress-strain relation Eq. 2.3, we can write the stress rate in terms of the strain rate as

$$\dot{\sigma}_{ij} = \lambda_U \dot{\epsilon}_{kk} \delta_{ij} + 2\mu_U \dot{\epsilon}_{ij} - \sum_{l=1}^N l\dot{\xi}_{ij}, \quad (2.60)$$

where $\lambda_U = \kappa_U - 2\mu_U/3$ and the memory variables $l\dot{\xi}_{ij} = l\dot{\zeta}'_{ij}$. Considering Eq. 2.58, the memory variables $l\zeta'_{ij}$ must follow the ODEs

$$l\dot{\zeta}'_{ij} + \omega_l l\zeta'_{ij} = \omega_l a_l \left[\Delta\lambda \dot{\epsilon}_{kk} \delta_{ij} + 2\Delta\mu \dot{\epsilon}_{ij} \right], \quad (2.61)$$

where $\Delta\lambda = \Delta\kappa - 2\Delta\mu/3$. Here the Einstein convention applies to the repeated index k , but does not apply to the repeated index l .

The formulation for incorporating viscoelasticity into anisotropic media is more involved and I refer to (Carcione and Cavallini 1994; Carcione 2007) for more detailed discussion. At the current stage, the AWP-ODC code does not account for anisotropy. It accounts for viscoelasticity in isotropic media using quality factors $Q_p(\omega)$ and $Q_s(\omega)$, which approximate the frequency-independent Q model over a broad frequency range.

2.1.2.4 Frequency-Independent Q

The relaxation time τ_l and the weight coefficient w_l in Eq. 2.15 or the relaxation frequency ω_l and the weight coefficient Y_l in Eq. 2.41 can be determined using numerical optimization algorithms. In such algorithms, one searches for an optimal set of τ_l and w_l or ω_l and Y_l by minimizing the differences between the $Q(\omega)$ model computed using Eq. 2.15 or 2.41 with a given target model $\tilde{Q}(\omega)$. For many problems of practical interest, \tilde{Q} is assumed to be frequency-independent or varying slowly with frequency, although the power-law frequency-dependence $\tilde{Q}(\omega) = Q_0\omega^\nu$ is sometimes of interest (e.g., Müller 1983).

Considering the Debye function approximation of Q^{-1} in Eq. 2.43, in order to obtain a Q^{-1} model that is fairly constant over a given frequency range $[\Omega_1, \Omega_2]$, a natural choice of the relaxation frequencies ω_l is a logarithmically uniform distribution over $[\Omega_1, \Omega_2]$ (Emmerich and Korn 1987). The distance between adjacent Debye functions is then given by ω_l/ω_{l-1} , which should be constant for a logarithmically uniform distribution of ω_l . The approximation can be improved by increasing the number of relaxation frequencies inside $[\Omega_1, \Omega_2]$, thereby reducing ω_l/ω_{l-1} .

Once we have fixed the values for ω_l , the weight coefficient Y_l in Eq. 2.41 can be determined by requiring $Q^{-1}(\tilde{\omega}_m) = \tilde{Q}^{-1}(\tilde{\omega}_m)$ at a set of different frequencies $\tilde{\omega}_m$, $m = 1, 2, \dots, K$. The choice of $\tilde{\omega}_m$ suggested in (Emmerich and Korn 1987) is $\tilde{\omega}_1 = \omega_1$ and $\tilde{\omega}_{m+1} = \tilde{\omega}_m (\omega_l/\omega_{l-1})^{1/2}$, then $K = 2N - 1$. Considering Eq. 2.41, the N weight coefficients Y_l are constrained by $2N - 1$ linear equations in the form (Emmerich and Korn 1987)

$$\sum_{l=1}^N \frac{\tilde{\omega}_m \left[\omega_l - \tilde{Q}^{-1}(\tilde{\omega}_m) \tilde{\omega}_m \right]}{\omega_l^2 + \tilde{\omega}_m^2} Y_l = \tilde{Q}^{-1}(\tilde{\omega}_m), \quad m = 1, 2, \dots, K. \quad (2.62)$$

This overdetermined linear system Eq. 2.62 can be solved using the least-squares algorithm.

In the AWP-ODC code, a simple, yet not optimal, scheme is adopted to approximate the frequency-independent Q model (Day and Bradley 2001). The distribution of the relaxation time τ_l is logarithmically uniform over the interval $[\tau_m, \tau_M]$,

$$\ln \tau_l = \ln \tau_m + \frac{2l - 1}{2N} (\ln \tau_M - \ln \tau_m), \quad (2.63)$$

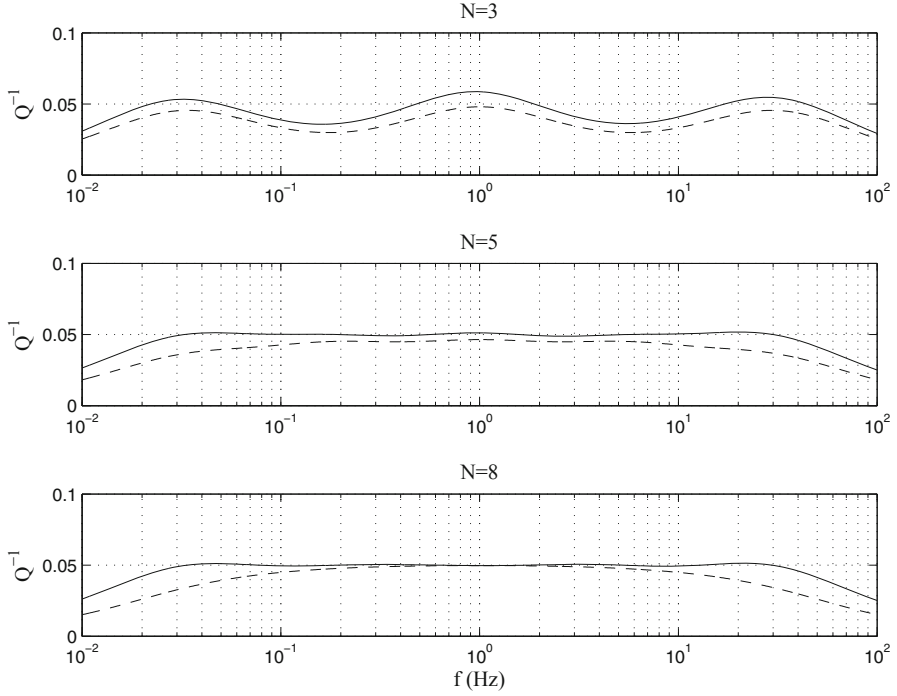


Fig. 2.3 Comparisons of the frequency-independent Q model obtained by solving Eq. 2.62 using a least-squares method (*solid lines*) and the more simplistic approach of (Day and Bradley 2001) (*dash lines*) for $N = 3, 5, 8$. The target Q model is set at $\tilde{Q}^{-1} = 0.05$.

where τ_m and τ_M are related to the Nyquist frequency $\omega_{Nyquist}$ and the total number of time steps in the calculation N_t ,

$$\tau_m = \omega_{Nyquist}^{-1}, \quad \tau_M = 5N_t\omega_{Nyquist}^{-1}. \quad (2.64)$$

The weight coefficient a_l is chosen to be the constant $1/N$. The resulting Q model is nearly constant over the entire usable frequency range of the simulation (Day and Bradley 2001). Figure 2.3 shows comparisons of the Q models obtained using this simplistic approach with those obtained using the least-squares optimization method for three different values of N . When $N = 3$, the separation between adjacent Debye functions is too large compared with their width (Fig. 2.2), which results in the oscillatory behavior of the Q models obtained using both methods. In the AWP-ODC code, N is fixed at 8, which provides a model sufficiently accurate for many applications.

In anelastic wave-propagation simulations, attenuation parameters are usually given in terms of Q_p and Q_s instead of the differences between the unrelaxed and relaxed moduli. Considering Eq. 2.15 or 2.41, once the Q model is determined, the differences between the unrelaxed and the relaxed moduli $\Delta\kappa$ and $\Delta\mu$ in the ODEs for the memory variables in Eq. 2.61 can be replaced with Q_p^{-1} and Q_s^{-1} . For the

frequency-independent Q model, the ODEs for the memory variables in Eq. 2.61 can be written in terms of the unrelaxed moduli κ_U and μ_U and Q_p^{-1} and Q_s^{-1} values at a given reference frequency ω_0 (Day and Bradley 2001)

$$i\dot{\zeta}_{ij} + \omega_l i\zeta_{ij} = \omega_l a_l \left\{ 2\mu_U A_S \dot{\epsilon}_{ij} + \left[\left(\kappa_U + \frac{4}{3}\mu_U \right) A_P - 2\mu_U A_S \right] \dot{\epsilon}_{kk} \delta_{ij} \right\}, \quad (2.65)$$

where the modulus relaxations A_P and A_S can be expressed as

$$A_P \equiv \frac{\Delta\kappa + 4\Delta\mu/3}{\kappa_U + 4\mu_U/3} = \frac{2}{\pi} Q_{p0}^{-1} \ln \frac{\tau_M}{\tau_m} \left[1 - \frac{2}{\pi} Q_{p0}^{-1} (\ln \omega_0 \tau_m) \right]^{-1}, \quad (2.66)$$

$$A_S \equiv \frac{\Delta\mu}{\mu_U} = \frac{2}{\pi} Q_{s0}^{-1} \ln \frac{\tau_M}{\tau_m} \left[1 - \frac{2}{\pi} Q_{s0}^{-1} (\ln \omega_0 \tau_m) \right]^{-1}. \quad (2.67)$$

Here Q_{p0} and Q_{s0} are the Q values at the reference frequency ω_0 . The values for τ_M and τ_m are given in Eq. 2.64.

The unrelaxed modulus can be determined from the phase velocity at a specified reference frequency and the density. Introducing c_0 , the phase velocity at the reference frequency ω_0 , the modulus at the reference frequency can be represented as

$$\hat{M}(\omega_0) = \rho c_0^2. \quad (2.68)$$

Considering Eq. 2.13, the unrelaxed modulus can then be expressed as (Liu and Archuleta 2006)

$$M_U = \rho c_0^2 \left(1 - \sum_{l=1}^N \frac{w_l}{1 + i\omega_0 \tau_l} \right)^{-1}. \quad (2.69)$$

For $c_0 = \beta_0$, where β_0 is the S-wave velocity at the reference frequency, we obtain

$$\mu_U = \rho \beta_0^2 \left(1 - \sum_{l=1}^N \frac{w_l}{1 + i\omega_0 \tau_l} \right)^{-1}, \quad (2.70)$$

and for $c_0 = \alpha_0$, where α_0 is the P-wave velocity at the reference frequency, we have

$$\kappa_U + 4\mu_U/3 = \rho \alpha_0^2 \left(1 - \sum_{l=1}^N \frac{w_l}{1 + i\omega_0 \tau_l} \right)^{-1}. \quad (2.71)$$

In general, the w_l and τ_l in Eq. 2.70 can be different from those in Eq. 2.71. In practice, we usually do not have information about the reference frequency at which the phase velocities are known. In such a case, we often assume $\omega_0 = 2\pi$.

2.1.3 Discretization

The AWP-ODC code uses a uniform staggered grid, which results in an algorithmically simple and computationally efficient discretization scheme. In a conventional uniform grid, different components of the field variables (i.e., the three-component particle velocity and the 6-component stress) are all located at the same grid points. In a staggered grid, different components of the particle velocity and the stress tensor can be located at different grid points. To achieve the same numerical accuracy, the staggered grid requires eight times less memory and 10–16 times less CPU time for 3D problems than schemes based on conventional grids (e.g., Moczo et al. 2007). The fourth-order staggered-grid finite-difference scheme introduced in (Virieux 1984, 1986) has been adopted by most finite-difference modelers and is also adopted in AWP-ODC.

A schematic diagram of a unit cell inside the staggered grid used in the AWP-ODC code is shown in Fig. 2.4. The edge length of a unit cell is $h/2$, however the grid spacing of the mesh is still h . For instance, the distance between two adjacent v_1 (black squares in Fig. 2.4) is h . Considering the scenario that v_1 is located at a grid point (I, J, K) , where I, J, K are all integers (Fig. 2.4), then v_2 is located at grid point $(I + 1/2, J + 1/2, K)$ and v_3 at $(I + 1/2, J, K + 1/2)$, the three diagonal components of the stress tensor $\sigma_{11}, \sigma_{22}, \sigma_{33}$ are located at $(I + 1/2, J, K)$ and the off-diagonal stress component σ_{12} at $(I, J + 1/2, K)$, σ_{13} at $(I, J, K + 1/2)$ and σ_{23} at $(I + 1/2, J + 1/2, K + 1/2)$.

In order to update σ_{11} located at $(I + 1/2, J, K)$ using Eq. 2.3, we need access to $v_{1,1}$, the first-order partial derivative of v_1 with respect to x_1 , at grid point $(I + 1/2, J, K)$. We can approximate this partial derivative through a differencing scheme. For a fourth-order central differencing scheme, we need values of v_1 at grid points (I, J, K) , $(I + 1, J, K)$, $(I - 1, J, K)$ and $(I + 2, J, K)$ (Fig. 2.4). If we denote the x_1 -coordinate of the grid point $(I + 1/2, J, K)$ as x_0 , the x_1 -coordinates of the four neighboring grid points are $x_0 - h/2$, $x_0 + h/2$, $x_0 - 3h/2$ and $x_0 + 3h/2$. The first-order partial derivative of v_1 with respect to x_1 at x_0 can be approximated as a linear combination of v_1 values at the four neighboring grid points as

$$v_{1,1}(x_0) \approx \frac{1}{h} \left\{ a \left[v_1 \left(x_0 + \frac{3}{2}h \right) - v_1 \left(x_0 - \frac{3}{2}h \right) \right] + b \left[v_1 \left(x_0 + \frac{1}{2}h \right) - v_1 \left(x_0 - \frac{1}{2}h \right) \right] \right\} + O(h^4). \quad (2.72)$$

Here I have dropped the dependence on x_2 and x_3 coordinates to reduce clutter. We can expand the v_1 at the four neighboring grid points in Taylor series at x_0 and truncate to fourth order as

$$v_1 \left(x_0 + \frac{3}{2}h \right) \approx v_1(x_0) + \frac{3h}{2} v_1'(x_0) + \frac{9h^2}{8} v_1''(x_0) + \frac{27h^3}{48} v_1'''(x_0) + O(h^4), \quad (2.73)$$

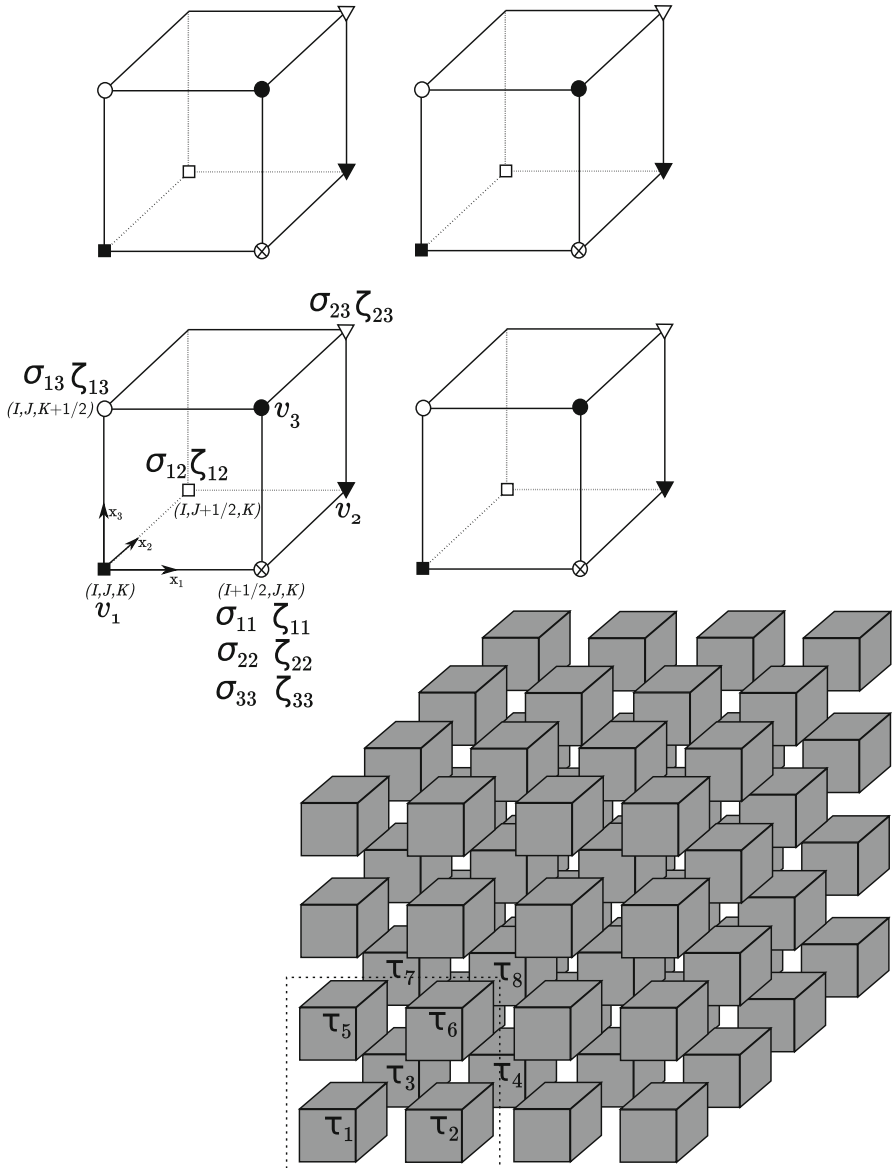


Fig. 2.4 Geometry of the staggered grid and the locations of the different components of the particle velocity, the stress tensor and the memory variables used in AWP-ODC. The *lower-right* plot shows a 3D array of unit cells. A unit cell is a cube with edge-length $h/2$, where h is the grid spacing. A blow-up view of the four unit cells inside the *dash-line box* is shown in the *upper-left plot*. The *three black arrows* on the *lower-left cell* indicate the coordinate system, in which the vertical axis x_3 points upward.

$$v_1 \left(x_0 - \frac{3}{2}h \right) \approx v_1(x_0) - \frac{3h}{2}v_1'(x_0) + \frac{9h^2}{8}v_1''(x_0) - \frac{27h^3}{48}v_1'''(x_0) + O(h^4), \quad (2.74)$$

$$v_1 \left(x_0 + \frac{1}{2}h \right) \approx v_1(x_0) + \frac{h}{2}v_1'(x_0) + \frac{h^2}{8}v_1''(x_0) + \frac{h^3}{48}v_1'''(x_0) + O(h^4), \quad (2.75)$$

$$v_1 \left(x_0 - \frac{1}{2}h \right) \approx v_1(x_0) - \frac{h}{2}v_1'(x_0) + \frac{h^2}{8}v_1''(x_0) - \frac{h^3}{48}v_1'''(x_0) + O(h^4). \quad (2.76)$$

Bring Eqs. 2.73–2.76 into Eq. 2.72, we obtain

$$v_{1,1}(x_0) \approx (3a + b)v_1'(x_0) + \frac{(27a + b)h^2}{24}v_1'''(x_0) + O(h^4). \quad (2.77)$$

For Eq. 2.72 to be fourth-order accurate, we require

$$\begin{cases} 3a + b = 1 \\ 27a + b = 0 \end{cases}, \quad (2.78)$$

which has the solution

$$\begin{cases} a = -\frac{1}{24} \\ b = \frac{9}{8} \end{cases}. \quad (2.79)$$

Partial derivatives with respect to other coordinates and for other particle-velocity and stress components are approximated in a similar way using the same fourth-order central differencing scheme.

The time derivatives are approximated using a second-order central differencing scheme in AWP-ODC. The particle-velocity fields at the $(n + 1/2)$ th time step, where n is an integer, are computed using their values at the $(n - 1/2)$ th time step and the stress fields at the n^{th} time step. And the stress fields at the $(n + 1)$ th time step are computed using their values at the n^{th} time step and the particle-velocity fields at the $(n + 1/2)$ th time step.

To incorporate viscoelasticity using Eqs. 2.60 and 2.61, we need to introduce N memory variables, as well as the associated relaxation frequencies and weight coefficients, for each stress component at all grid points. For a large number of relaxation mechanisms, for instance $N = 8$, the amount of computer memory required to store them and also the computational time can be significant. The coarse-grained

sampling of the memory variables, the relaxation frequencies and weight coefficients introduced in (Day 1998; Day and Bradley 2001) is adopted in AWP-ODC to significantly reduce the computational costs associated with viscoelasticity. In the coarse-grained sampling scheme, the distribution of the memory variables, the relaxation frequencies and the weight coefficients has a spatial period of $2h$. A schematic diagram of the spatial distribution of the relaxation times for $N = 8$ is shown in Fig. 2.4. The summation over all relaxation mechanisms at one grid point in Eq. 2.60 can be accounted for by taking the memory variables from neighboring positions with proper weighting. This coarse-grained spatial sampling scheme effectively reduces the computational cost for 8 different relaxation mechanisms to that of one relaxation mechanism. The numerical stability and accuracy of this sampling scheme were carefully analyzed in (Graves and Day 2003). The implementation of this scheme in the AWP-ODC code was tested and applied to derive empirical relations between S-wave velocity and Q_s and empirical relations between Q_s and Q_p in the Los Angeles Basin area in (Olsen et al. 2003).

2.1.4 Free-Surface Boundary

AWP-ODC only considers the planar free surface. Assuming the coordinate x_3 is in the vertical direction, the traction-free boundary condition for the free surface

$$\boldsymbol{\sigma} \cdot \hat{\mathbf{x}}_3 = \begin{bmatrix} \sigma_{13} \\ \sigma_{23} \\ \sigma_{33} \end{bmatrix} = \mathbf{0} \quad (2.80)$$

is modeled using the stress imaging formulation¹³ (Levander 1988; Graves 1996) in AWP-ODC. For the staggered grid adopted in AWP-ODC (Fig. 2.4), the particle velocity and stress components that lie on the free surface are v_3 , σ_{13} and σ_{23} , i.e., the components lying on the top of a unit cell. Suppose the free surface is the horizontal plane located at vertical grid index $K + 1/2$ (Fig. 2.4), the stress imaging formulation is implemented by setting

$$\sigma_{13}|_{K+1/2} = \sigma_{23}|_{K+1/2} = 0 \quad (2.81)$$

and imposing the antisymmetry condition for σ_{33} , which is located at vertical grid indices K and $K + 1$

$$\sigma_{33}|_{K+1} = -\sigma_{33}|_K. \quad (2.82)$$

Here I have dropped the dependence of the stress field on horizontal coordinates to reduce clutter. The stress components σ_{11} , σ_{22} and σ_{12} are not needed above the free surface.

¹³ The stress imaging formulation is sometimes called the zero-stress formulation in the literature.

The relations among the different components of the particle velocity in the vicinity of the free surface can be obtained based on the traction-free boundary condition in Eq. 2.80 and the stress-strain relation. Considering Eq. 2.3, without the moment source term, for $i = 1$ and $j = 3$

$$\dot{\sigma}_{13}|_{K+1/2} = \mu (v_{1,3} + v_{3,1})|_{K+1/2} = 0, \quad (2.83)$$

we have

$$v_{1,3}|_{K+1/2} = -v_{3,1}|_{K+1/2}, \quad (2.84)$$

where $v_{1,3} = \partial v_1 / \partial x_3$ and $v_{3,1} = \partial v_3 / \partial x_1$. The partial derivatives in Eq. 2.84 can be approximated using a second-order central differencing scheme. Considering the grid point $(I, J, K + 1/2)$ (Fig. 2.4), we have

$$\frac{v_1|_{(I,J,K+1)} - v_1|_{(I,J,K)}}{h} \approx -\frac{v_3|_{(I+1/2,J,K+1/2)} - v_3|_{(I-1/2,J,K+1/2)}}{h}$$

re-arranging the terms, we obtain

$$v_1|_{(I,J,K+1)} \approx v_1|_{(I,J,K)} - v_3|_{(I+1/2,J,K+1/2)} + v_3|_{(I-1/2,J,K+1/2)}. \quad (2.85)$$

Considering

$$\dot{\sigma}_{23}|_{K+1/2} = \mu (v_{2,3} + v_{3,2})|_{K+1/2} = 0 \quad (2.86)$$

at grid point $(I + 1/2, J + 1/2, K + 1/2)$ we have

$$v_2|_{(I+1/2,J+1/2,K+1)} \approx v_2|_{(I+1/2,J+1/2,K)} - v_3|_{(I+1/2,J+1,K+1/2)} + v_3|_{(I+1/2,J,K+1/2)}. \quad (2.87)$$

Considering

$$\dot{\sigma}_{33}|_{K+1/2} = (\lambda + 2\mu) v_{3,3}|_{K+1/2} + \lambda (v_{1,1} + v_{2,2})|_{K+1/2} = 0, \quad (2.88)$$

at grid point $(I + 1/2, J, K + 1/2)$, we have

$$\begin{aligned} \frac{v_3|_{(I+1/2,J,K+3/2)} - v_3|_{(I+1/2,J,K-1/2)}}{2h} &\approx -\frac{\lambda}{\lambda + 2\mu}. \\ \left[\frac{1}{2} \left(\frac{v_1|_{(I+1,J,K)} - v_1|_{(I,J,K)}}{h} + \frac{v_1|_{(I+1,J,K+1)} - v_1|_{(I,J,K+1)}}{h} \right) \right. \\ &+ \frac{1}{2} \left(\frac{v_2|_{(I+1/2,J+1/2,K)} - v_2|_{(I+1/2,J-1/2,K)}}{h} \right. \\ &\left. \left. + \frac{v_2|_{(I+1/2,J+1/2,K+1)} - v_2|_{(I+1/2,J-1/2,K+1)}}{h} \right) \right] \end{aligned}$$

and re-arrange terms we obtain

$$v_3|_{(I+1/2,J,K+3/2)} \approx v_3|_{(I+1/2,J,K-1/2)} \quad (2.89)$$

$$\begin{aligned}
& -\frac{\lambda}{\lambda + 2\mu} (v_1|_{(I+1,J,K)} - v_1|_{(I,J,K)} + v_1|_{(I+1,J,K+1)} - v_1|_{(I,J,K+1)}) \\
& \quad + v_2|_{(I+1/2,J+1/2,K)} - v_2|_{(I+1/2,J-1/2,K)} \\
& \quad + v_2|_{(I+1/2,J+1/2,K+1)} - v_2|_{(I+1/2,J-1/2,K+1)}.
\end{aligned}$$

Equations 2.85, 2.87 and 2.89 together with Eqs. 2.81 and 2.82 allow us to compute particle velocity components and stress components at grid points above the free surface using the values below and on the free surface.

The differencing scheme adopted for the free-surface boundary condition is second-order accurate. For the fourth-order scheme adopted for the interior, it is usually sufficient to have 4–5 grid points per minimum wavelength to ensure sufficient numerical accuracy. For the second-order scheme on the free-surface, we may need 8–10 grid points per minimum wavelength. To accurately model surface waves, which are particularly sensitive to the free-surface boundary, I usually use 9 grid points per minimum wavelength in simulations.

2.1.5 PML Boundary

The AWP-ODC code implements the perfectly-matched-layer (PML) for the 4 sides and the bottom of the modeling domain. The PML implementation is based on the split-field formulation, which was introduced for the electromagnetic wave equation in (Berenger 1994, 1996) and later adopted for acoustic and elastodynamic equations in (e.g., Chew and Liu 1996 Hastings et al. 1996 Liu and Tao 1997 Collino and Tsogka 2001 Marcinkovich and Olsen 2003). The details of the implementation used in AWP-ODC, as well as careful tests on numerical stability and performance comparisons with other types of absorbing boundary implementations, are documented in (Marcinkovich and Olsen 2003).

The split-field formulation can be understood in terms of a generalization of the real coordinates into complex space (e.g., Chew et al. 1997 Collino and Monk 1998 Teixeira and Chew 1999 Collino and Tsogka 2001). Suppose the PML lies at $0 < x_1 \leq n_b$ and the interior domain is located at $x_1 \leq 0$ (Fig. 2.5), we can define a damping profile $d(x_1)$, which is zero in the interior domain and increases monotonically to its maximum value d_0 at $x_1 = n_b$. The thickness of the PML n_b , the maximum damping value d_0 and the exact shape of the damping profile can be chosen based on the application. In (Marcinkovich and Olsen 2003), the formula for determining the maximum damping value d_0 is

$$d_0 = \frac{\tau\beta}{h} \left(c_1 + c_2 n_b + c_3 n_b^2 \right), \quad c_1 = \frac{8}{15}, c_2 = \frac{-3}{100}, c_3 = \frac{1}{1500}, \quad (2.90)$$

where h is the grid spacing, β is a representative S-wave velocity value and τ is a tuning coefficient that usually has a value between 3 and 4. We then introduce a

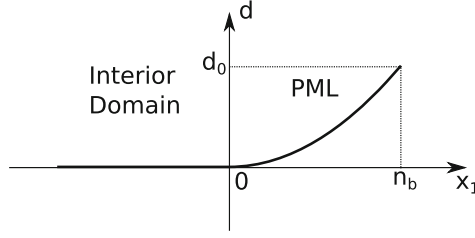


Fig. 2.5 The PML damping profile as a function of the distance from the Interior domain-PML interface located at $x_1 = 0$. The width of the PML is n_b and the maximum damping value d_0 is reached at $x_1 = n_b$. The damping profile shown here was generated using the equation $d_I = d_0 (I/n_b)^p$ (Marcinkovich and Olsen 2003), where I is the number of grid points counted from the Interior domain-PML interface, $p = 2$ and the PML width is $n_b = 12$.

complex coordinate \tilde{x}_1 defined in terms of the damping profile, the real coordinate x_1 and the angular frequency ω

$$\tilde{x}_1 = x_1 - \frac{i}{\omega} \int_0^{x_1} d(l) dl. \quad (2.91)$$

Based on this definition the partial derivative with respect to \tilde{x}_1 can be related to the partial derivative with respect to x_1 using the chain rule

$$\tilde{\partial}_1 = \frac{1}{s} \partial_1 \quad (2.92)$$

where the complex function s has the expression

$$s = 1 + \frac{d(x_1)}{i\omega}. \quad (2.93)$$

The next step is to generalize the elastodynamic equation using the complex coordinate \tilde{x}_1 . In order to do that, we need to split the gradient operator ∇ into two components, one parallel to x_1 and the other perpendicular to x_1 ,

$$\nabla = \hat{\mathbf{x}}_1 \partial_1 + \nabla^{\parallel}, \quad (2.94)$$

where

$$\nabla^{\parallel} = (\mathbf{I} - \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_1) \cdot \nabla \quad (2.95)$$

is the projection of the gradient operator onto the plane perpendicular to the $\hat{\mathbf{x}}_1$ direction. Here \mathbf{I} is the identity operator and $\hat{\mathbf{x}}_1$ is the normal direction perpendicular to the PML interface (Fig. 2.5). In the frequency domain, the elastodynamic equation in the velocity-stress form Eqs. 2.2 and 2.3 can be written using the gradient operator in its split form Eq. 2.94 as

$$i\omega\rho\hat{\mathbf{v}} = \hat{\mathbf{x}}_1 \partial_1 \cdot \hat{\boldsymbol{\sigma}} + \nabla^{\parallel} \cdot \hat{\boldsymbol{\sigma}}, \quad (2.96)$$

$$i\omega\hat{\boldsymbol{\sigma}} = \mathbf{c}:\hat{\mathbf{x}}_1\tilde{\partial}_1\hat{\mathbf{v}} + \mathbf{c}:\nabla^{\parallel}\hat{\mathbf{v}}, \quad (2.97)$$

where $\hat{\mathbf{v}}$ and $\hat{\boldsymbol{\sigma}}$ are Fourier transforms of the particle velocity field and the stress field, \mathbf{c} is the elastic tensor, the operator “:” represents double dot product¹⁴ and I have dropped the source fields to reduce clutter. In a homogeneous medium, Eqs. 2.96 and 2.97 admit the plane wave solution of the form

$$\mathbf{A} \exp[-i(k_1x_1 + k_2x_2 + k_3x_3 - \omega t)], \quad (2.98)$$

where the vector $\mathbf{k} = [k_1, k_2, k_3]$ is the wave vector¹⁵. If we generalize the ∂_1 operator in Eqs. 2.96 and 2.97 using the $\tilde{\partial}_1$ operator defined in Eq. 2.92, we obtain

$$i\omega\rho\hat{\mathbf{v}} = \hat{\mathbf{x}}_1 \cdot \tilde{\partial}_1\hat{\boldsymbol{\sigma}} + \nabla^{\parallel} \cdot \hat{\boldsymbol{\sigma}}, \quad (2.99)$$

$$i\omega\hat{\boldsymbol{\sigma}} = \mathbf{c}:\hat{\mathbf{x}}_1\tilde{\partial}_1\hat{\mathbf{v}} + \mathbf{c}:\nabla^{\parallel}\hat{\mathbf{v}}, \quad (2.100)$$

which are identical to Eqs. 2.96 and 2.97 in the interior domain, since $d = 0$ and $s = 1$ in the interior domain. In the PML, Eqs. 2.99 and 2.100 admit the plane-wave solution of the form

$$\begin{aligned} \mathbf{A} \exp[-i(k_1\tilde{x}_1 + k_2x_2 + k_3x_3 - \omega t)] &= \mathbf{A} \exp\left[-\frac{k_1}{\omega} \int_0^{x_1} d(l)dl\right] \\ &\exp[-i(\mathbf{k} \cdot \mathbf{x} - \omega t)], \end{aligned} \quad (2.101)$$

which decays exponentially in the x_1 direction. At the PML boundary $x_1 = 0$, the plane-wave solution in the form of Eq. 2.101 in the PML is identical to that in the form of Eq. 2.98 in the interior domain for all incident angles and frequencies, therefore no artificial reflections are generated at the PML boundary. The next step is to bring Eq. 2.92 into Eqs. 2.99 and 2.100 and also split the particle velocity and stress fields into two components

$$\hat{\mathbf{v}} = \hat{\mathbf{v}}_1 + \hat{\mathbf{v}}_2, \quad \hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}_1 + \hat{\boldsymbol{\sigma}}_2 \quad (2.102)$$

such that they satisfy

$$i\omega\rho\hat{\mathbf{v}}_1 = \frac{1}{s}\hat{\mathbf{x}}_1 \cdot \partial_1\hat{\boldsymbol{\sigma}}, \quad i\omega\rho\hat{\mathbf{v}}_2 = \nabla^{\parallel} \cdot \hat{\boldsymbol{\sigma}}, \quad (2.103)$$

¹⁴ $\mathbf{A}:\mathbf{B} = \sum_i \sum_j A_{ij} \bar{B}_{ij}$, where complex matrices \mathbf{A} and \mathbf{B} have the same dimension, \bar{B}_{ij} is the complex conjugate of B_{ij} . For the fourth-order elastic tensor \mathbf{c} , its double dot product with a real second-order tensor $\boldsymbol{\phi}$ can be calculated as $\mathbf{c}:\boldsymbol{\phi} = \sum_k \sum_l c_{ijkl} \phi_{kl}$ and the result is a second-order tensor.

¹⁵ If the medium is isotropic, for the P wave we have $\mathbf{A} \times \mathbf{k} = \mathbf{0}$ and $|\mathbf{k}| = \omega/\alpha$, and for the S wave we have $\mathbf{A} \cdot \mathbf{k} = \mathbf{0}$ and $|\mathbf{k}| = \omega/\beta$, where α and β are the P- and S-wave velocities.

$$i\omega\hat{\sigma}_1 = \frac{1}{s}\mathbf{c}:\hat{\mathbf{x}}_1\partial_1\hat{\mathbf{v}}, \quad i\omega\hat{\sigma}_2 = \mathbf{c}:\nabla^{\parallel}\hat{\mathbf{v}}. \quad (2.104)$$

Transforming Eqs. 2.103 and 2.104 back into the time domain, we obtain the equations that can be solved using the finite-difference method

$$(\partial_t + d)\rho\mathbf{v}_1 = \hat{\mathbf{x}}_1 \cdot \partial_1\boldsymbol{\sigma}, \quad \rho\dot{\mathbf{v}}_2 = \nabla^{\parallel} \cdot \boldsymbol{\sigma}, \quad (2.105)$$

$$(\partial_t + d)\boldsymbol{\sigma}_1 = \mathbf{c}:\hat{\mathbf{x}}_1\partial_1\mathbf{v}, \quad \dot{\boldsymbol{\sigma}}_2 = \mathbf{c}:\nabla^{\parallel}\mathbf{v}. \quad (2.106)$$

The same analysis can also be applied to other boundaries that are perpendicular to the x_2 and x_3 axes.

Equation 2.101 shows that the attenuation is strongest for waves traveling perpendicular to the PML interface and much weaker for waves traveling at grazing incidence angles to the PML interface (i.e., $k_1 \approx 0$ in Eq. 2.101). One drawback of the split-field PML is that its absorption performance can become very poor when many waves are traveling at grazing incidence, which is often the case if the simulation domain is elongated in one direction and/or the source is located very close to the boundary of the simulation domain. The convolutional PML (C-PML) or complex frequency shifted PML (CFS-PML) approach solves this problem effectively and also removes the requirement for splitting the fields (Roden and Gedney 2000; Bérenger 2002; Wrenger 2002; Drossaert and Giannopoulos 2007; Martin et al. 2008a,b; Martin and Komatitsch 2009). In C-PML, the complex function s in Eq. 2.93 is generalized to the form

$$s = c + \frac{d(x_1)}{e + i\omega} \quad (2.107)$$

by including two more variables $c \geq 1$ and $e \geq 0$. For the special case $c = 1$ and $e = 0$, the generalized s function becomes identical to that in Eq. 2.93. Spatial derivatives with respect to generalized coordinates inside the PML can be expressed as

$$\tilde{\partial}_1 = \frac{1}{c}\partial_1 + \psi, \quad (2.108)$$

where the memory variable ψ at each time step is updated through a recursive convolution process. For waves traveling perpendicular to the PML interface, the C-PML has similar performance as the split-field PML. For waves traveling at grazing incidence angles, the C-PML can also strongly attenuate their energy. However, the C-PML implementation has not yet been included into the AWP-ODC code at the current stage.

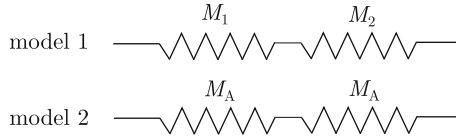


Fig. 2.6 The effective material model for two Hooke bodies connected in serial. M_1 and M_2 are the moduli of the two different Hooke bodies in *model 1*. We seek an effective material model, *model 2*, which is composed of two identical Hooke bodies with modulus M_A connected in serial and has the same stress-strain relation as *model 1*. The stress and strain in the two Hooke bodies in model 1 are σ_1 and ϵ_1 and σ_2 and ϵ_2 , respectively. The total stress and strain both in *model 1* and *model 2* are σ and ϵ .

2.1.6 Internal Welded Material Interfaces

The formulation discussed so far is applicable to smoothly heterogeneous media¹⁶. In realistic structural models of the Earth's interior, the media may contain material discontinuities, where density and/or elastic moduli may change discontinuously. One approach for dealing with material discontinuities is to apply the usual finite-difference scheme at grid points outside the discontinuities and then account for wavefields at or near the discontinuities using proper boundary conditions (e.g., the continuity of stress and displacement across internal welded material interfaces) (e.g., Alterman and Karal 1968 Ilan et al. 1975 Ilan and Loewenthal 1976). This type of approaches are often called homogeneous and are applicable when the geometry of the material discontinuities is simple. For models with complex curved material interfaces, this type of homogeneous approaches are highly cumbersome, especially for finite-difference schemes based on uniform grids.

An alternative approach is to replace the structural model with material discontinuities with a smoothly varying effective material model such that the application of the usual finite-difference scheme on this smooth effective material model produces the same wavefields as those obtained through applying the homogeneous approach on the original model with material discontinuities. This type of approaches are often called heterogeneous (e.g., Kelly et al. 1976 Muir et al. 1992 Zahradník and Priolo 1995 Graves 1996 Ohminato and Chouet 1997 Moczo et al. 2002).

The details about how to derive such an effective material model, as well as careful validations of the resulting finite-difference scheme based on the effective material model, is documented in (e.g., Moczo et al. 2002). Consider two different homogeneous elastic bodies separated by a welded material interface, a simple physical model analogous to this scenario is a model in which two different Hooke bodies connected in serial (Fig. 2.6). For the serial connection, the stresses inside the two Hooke bodies should be equal to each other and to the total stress (i.e., the

¹⁶ The characteristic length scale of the variations in the material properties is much larger than the grid spacing.

stress continuity boundary condition at the welded interface)

$$\sigma_1 = \sigma_2 = \sigma \quad (2.109)$$

and the total strain should be equal to the sum of the strain in each Hooke body

$$\epsilon_1 + \epsilon_2 = \epsilon. \quad (2.110)$$

For model 1 in Fig. 2.6, the total stress and the total strain follow

$$\epsilon = \frac{\sigma_1}{M_1} + \frac{\sigma_2}{M_2} = \sigma \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \quad (2.111)$$

and for the effective material model, model 2, we have

$$\epsilon = \sigma \frac{2}{M_A}. \quad (2.112)$$

In order for model 1 and model 2 to have identical stress-strain relations, we must have

$$\begin{aligned} \frac{2}{M_A} &= \frac{1}{M_1} + \frac{1}{M_2}, \\ M_A &= \frac{2}{\frac{1}{M_1} + \frac{1}{M_2}}. \end{aligned} \quad (2.113)$$

Equation 2.113 indicates that the modulus of the effective material should be the harmonic average of the moduli of the two different Hooke bodies M_1 and M_2 .

To obtain the effective density model, we consider two particles with mass m_1 and m_2 moving together with the same acceleration \mathbf{a} (i.e., continuity of displacement boundary condition at the welded interface). The forces acting on the two particles are \mathbf{F}_1 and \mathbf{F}_2 , respectively. Therefore we have

$$\mathbf{F}_1 = m_1 \mathbf{a}, \quad \mathbf{F}_2 = m_2 \mathbf{a}. \quad (2.114)$$

The total force on the system is then

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 = (m_1 + m_2) \mathbf{a}. \quad (2.115)$$

If we replace this system with two particles with the same mass m_A ,

$$\mathbf{F} = 2m_A \mathbf{a}, \quad (2.116)$$

in order to keep the total force \mathbf{F} and the acceleration \mathbf{a} unchanged, we must have

$$m_A = \frac{m_1 + m_2}{2}. \quad (2.117)$$

Equation 2.117 indicates that the density of the effective material should be the arithmetic average of the densities of the two different components.

For three-dimensional isotropic media, similar analysis can be carried out for the volumetric stress-strain relation Eq. 2.50 and the deviatoric stress-strain relation Eq. 2.51 and the effective bulk and shear moduli at a grid point are volume harmonic averages of the bulk and shear moduli at neighboring grid points. The effective density at a grid point is the volume arithmetic average of the densities at neighboring grid points. Consider the effective bulk modulus $\kappa_{I+1/2,J,K}^A$ at the grid point $(I + 1/2, J, K)$ (Fig. 2.4), we can define the volume integral with the integration volume $V_{I+1/2,J,K}^K$ centered at this grid point

$$\iiint_{V_{I+1/2,J,K}^K} dV = \int_{x_1^I}^{x_1^{I+1}} \int_{x_2^{J-1/2}}^{x_2^{J+1/2}} \int_{x_3^{K-1/2}}^{x_3^{K+1/2}} dx_1 dx_2 dx_3, \quad (2.118)$$

where the integration bounds on the right-hand-side are x_1 , x_2 and x_3 coordinates of boundaries of the integration volume. The effective bulk modulus can then be computed as the volume harmonic average

$$\kappa_{I+1/2,J,K}^A = \left[\frac{1}{h^3} \iiint_{V_{I+1/2,J,K}^K} \frac{1}{\kappa(\mathbf{x})} dV \right]^{-1}, \quad (2.119)$$

where h is the grid spacing. In practice, the volume integral is approximated through summation. The effective shear modulus can be evaluated in a similar way. Consider the effective density $\rho_{I,J,K}^A$ at grid point (I, J, K) (Fig. 2.4), the volume integral can be defined as

$$\iiint_{V_{I,J,K}^\rho} dV = \int_{x_1^{I-1/2}}^{x_1^{I+1/2}} \int_{x_2^{J-1/2}}^{x_2^{J+1/2}} \int_{x_3^{K-1/2}}^{x_3^{K+1/2}} dx_1 dx_2 dx_3, \quad (2.120)$$

where $V_{I,J,K}^\rho$ is the integration volume. The effective density can then be computed as the volume arithmetic average

$$\rho_{I,J,K}^A = \frac{1}{h^3} \iiint_{V_{I,J,K}^\rho} \rho(\mathbf{x}) dV. \quad (2.121)$$

Numerical tests on the stability and the accuracy of this heterogeneous scheme were documented in (e.g., Moczo et al. 2002). This scheme is also adopted in the AWP-ODC code. In AWP-ODC, the Lamé parameters are harmonically averaged and the density is arithmetically averaged (Olsen et al. 2003).

2.1.7 Source Representation

Seismic sources can be included into the finite-difference scheme through either adding body-force components to the corresponding velocity components in Eq. 2.2

(e.g., Yomogida and Etgen 1993 Graves 1996) or subtracting moment tensor components from the corresponding stress components in Eq. 2.3 (e.g., Coutant et al. 1995 Olsen et al. 1995). In AWP-ODC, a body-force source is automatically multiplied with the buoyancy (i.e., the inverse of the density) at the source location before adding to the corresponding velocity component at the source location. For a moment-rate source, the AWP-ODC code automatically normalizes it using the cell volume h^3 , where h is the grid spacing, to obtain the moment-rate-density tensor before subtracting it from the stress tensor at the source location.

2.2 Software

The software that is explained in this book, named F3DWI, can be compiled and run on any typical Beowulf clusters, SMP clusters and MPP systems. The minimum requirement is a system with at least one node that runs a certain distribution of the Linux operating system and has the MPI library correctly installed. The system used for running the half-space example in the following is a Linux box with one AMD FX-8350 8-core processor running the Ubuntu 14.04 LTS operating system, the MPI library is the MVAPICH2¹⁷ developed and maintained at the Ohio State University. The Fortran 77/90 and C/C++ compilers are from the GNU Compiler Collection version 4.8¹⁸.

For problems of larger sizes, systems with more nodes are needed. We carried out most of the computations for the full-3D tomography of the crustal structure in Southern California on the IBM Blue Gene/P system, named Intrepid, at the Argonne Leadership Computing Facility (ALCF), Argonne National Laboratory. One simulation used 4096 CPU cores for about 15 minutes of wall time and the full inversion involved tens of thousands of simulations. The Blue Gene/P system has now been replaced with the more advanced Blue Gene/Q system, named Mira, at ALCF. The computational details of our inversion for southern California have been documented in (Lee et al. 2014b).

The F3DWI software package is organized into three directories, `src`, `bin` and `py`. The `src` directory contains a modified version of the AWP-ODC source codes and also other Fortran 77/90, C/C++ source codes related to solving the seismic source and structural inverse problems. The source codes need to be compiled using the GNU `make` utility¹⁹ and examples of `Makefiles` for different systems we have tested the codes on are also provided in the `src` directory. Once the source codes are correctly compiled, the binary executables generated from those source codes are automatically moved to the `bin` directory. All the binary executables require input files and generate output files. In the `py` directory, there is a collection

¹⁷ <http://mvapich.cse.ohio-state.edu>

¹⁸ <http://gcc.gnu.org/>

¹⁹ <http://www.gnu.org/software/make/>

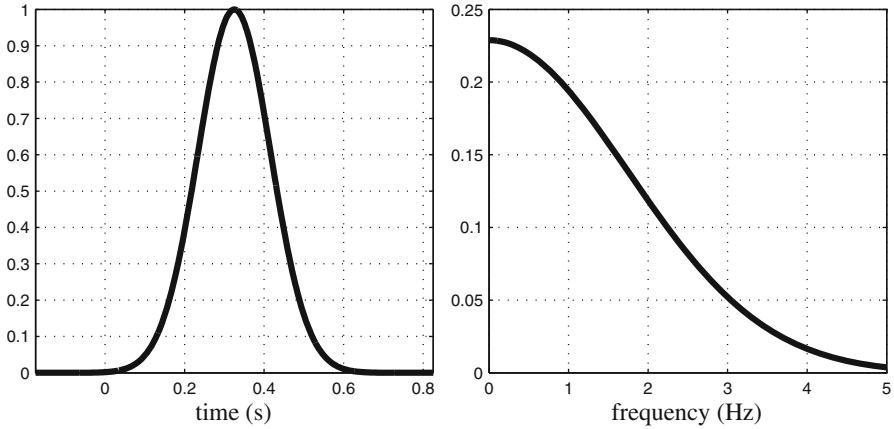


Fig. 2.7 The source-time-function described by Eq. 2.122 in the time domain (*left*) and its amplitude spectrum in the frequency domain (*right*).

of `python`²⁰ scripts, which can help users to generate the input files and process the output files of the binary executables.

In this chapter, I will only discuss codes that are related to the forward wave-propagation simulations. Codes for solving seismic inverse problems will be explained in later chapters after the theoretical backgrounds of those codes have been discussed.

I will use the simplest structural model, a uniform half-space with constant density, P- and S-wave velocities and a free surface, to demonstrate the basic steps for running forward wave-propagation simulations. The material properties are density $\rho = 3000 \text{ kg/m}^3$, P-wave velocity $\alpha = 6500 \text{ m/s}$ and S-wave velocity $\beta = 3500 \text{ m/s}$. I will use an explosive point source with the source-time-function given by

$$s(t) = \exp \left[-c_1 \left(t - \frac{c_2}{2} \right)^2 \right], \quad (2.122)$$

where $c_1 = 60$ and $c_2 = 0.65$. This source-time-function and its amplitude spectrum in the frequency domain are shown in Fig. 2.7. Both the source and the receiver are buried at 24-km depth from the free surface and the source-receiver distance is 32.2 km. This half-space configuration was used in (Zhao et al. 2005) to benchmark full-3D Fréchet kernels and will also be used in later chapters in this book.

The forward wave-propagation simulation is carried out by the binary executable `awp` in the `bin` directory. The `awp` code requires a set of input files related to the

²⁰ Python is a high-level programming language and is often used as a scripting language. Python interpreters are available on almost every platform. Details about the Python language can be found at <https://www.python.org/> and http://en.wikipedia.org/wiki/Python_programming_language.

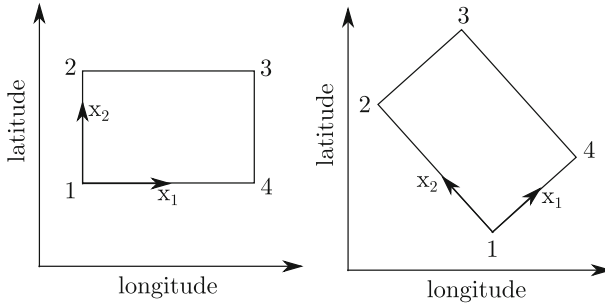


Fig. 2.8 The *top view* of the rectangular modeling region and the box coordinate system used in the forward wave-propagation simulations. *Left*: the two vertices on the bottom lie at the same longitude; *right*: an arbitrarily oriented rectangular modeling region. The directions of the x_1 and x_2 axes are marked with *arrows*. The x_3 axis points out of the page towards the viewer and forms a right-hand system with x_1 and x_2 .

mesh, the seismic structural model and the seismic source model. These input files can be generated using other binary executables in the `bin` directory and also scripts in the `py` directory. I will start with the mesh generation process and go through all the necessary steps for preparing the input files in the following.

2.2.1 Mesh Generation

To correctly generate the rectangular uniform mesh required by `awp`, we first need to understand the coordinate system used in the code and how it is mapped to a geographic region. We assume the modeling region is sufficiently small, such that the curvature of the Earth's surface can be safely ignored. For a rectangular modeling region on the surface of the Earth, we can number its four vertices clockwise (Fig. 2.8). For the situation shown in Fig. 2.8 (left), the southwest vertex is numbered as 1. For an arbitrarily oriented rectangle as shown in Fig. 2.8 (right), the southernmost vertex is numbered as 1. The generic coordinate system used inside the `awp` code, which I call the box coordinate, is defined using the four vertices. If the modeling volume has a vertical thickness of H meters, the origin of the coordinate system lies at H meters below vertex 1 inside the Earth. The x_1 axis points from the origin towards the point lying H meters below vertex 4, the x_2 axis points from the origin towards the point lying H meters below vertex 2, the x_3 axis points upwards from the origin to the free surface and forms a right-hand Cartesian coordinate system with x_1 and x_2 .

Locations defined in longitude/latitude are first mapped to Cartesian coordinates in Easting/Northing using the Universal Transverse Mercator (UTM) coordinate

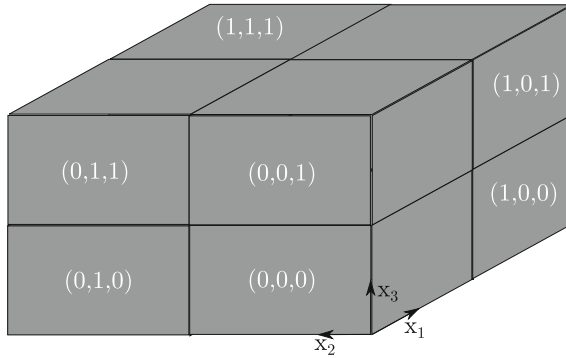


Fig. 2.9 The sub-meshes for the configuration with 2 MPI processes in each dimension. The x_1 , x_2 and x_3 axes of the box coordinate system are shown as three *black arrows*. The 3-number index of each sub-mesh is shown in the *parentheses in white*. The sub-mesh with the index (1,1,0) is located underneath the sub-mesh with the index (1,1,1) and is not visible from the viewing angle used in this figure.

system²¹ and then transformed to the box coordinate used in `awp`. In UTM, the globe is divided into 60 longitudinal zones starting at 180° W. Each longitudinal zone is 6° wide. Within each longitudinal zone, the transverse mercator projection is used to map locations in longitude/latitude to Cartesian coordinates defined in terms of Easting and Northing in meters. For the Easting, the origin lies at 500,000 m west of the central meridian of the longitudinal zone. The origin of the Northing is defined at the equator for the northern hemisphere and at 10,000,000 m south of the equator for the southern hemisphere. In `F3DWI` the conversion between longitude/latitude and Easting/Northing is handled by the subroutine in `utm_geo.f90` in the `src` directory. The comment lines at the beginning of the source code give more details about this code. The default behavior of this code is to keep the Northing origin in the southern hemisphere at the equator instead of setting it at 10,000,000 m south of equator, which gives negative Northing values for longitude/latitude locations in the southern hemisphere. This definition gives a uniform scale across the equator. To change this behavior, the `north` variable defined on line 34 of this code needs to be changed to 10000000 and all the source codes need to be re-compiled.

The `awp` code was parallelized using a domain-decomposition approach. The uniform mesh is divided into a number of sub-meshes of equal size and each MPI process is responsible for all the calculations on a single sub-mesh. The total number of MPI processes usually equals to the total number of CPU cores used in a simulation. Each sub-mesh is index using 3 integers, which indicate its relative position on the x_1 , x_2 and x_3 axes in the box coordinate (Fig. 2.9). If we have eight MPI processes, we can divide the mesh into two sub-meshes in each dimension, as shown

²¹ Details about UTM can be found at <http://www.dmap.co.uk/utmworld.htm> and also on Wikipedia.

Table 2.1 Format of the text input file for `FD_GRID_XYZ_rect`

Line number	Description
1	Full path to the mesh output file named XYZGRD
2	UTM longitudinal zone number of the modeling region
3	Longitude and latitude of vertex 1 of the rectangular modeling region
4	Longitude and latitude of vertex 2 of the rectangular modeling region
5	Distance between vertex 1 and vertex 4 in meter
6	Vertical thickness of the modeling volume in meter
7	Grid spacing in meter
8	Number of sub-meshes (MPI processes) in each spatial dimension

```

/home/pochen/wrk/F3DWI/HS/VM/
11
-118.005738 33.690067
-118.030457 33.82023
47500
35500
200
2 2 2

```

Listing 2.1 An example input file for the mesh generation code on my Linux box.

in Fig. 2.9, or we can divide it into two sub-meshes in the x_1 dimension and four sub-meshes in the x_2 dimension. There are different ways for evenly partitioning the mesh among the MPI processes. To minimize the inter-process MPI communication overhead, we often partition the mesh in a way such that the number of grid points in every dimension of the sub-mesh is about the same²².

The binary executable for mesh generation is `FD_GRID_XYZ_rect`, which is located in the `bin` directory. It requires a command-line text input file. The format of the input file is listed in Table 2.1 and an example input file, named `FD_GRID_XYZ.in` is shown in Listing 2.1.

²² Parallelization based on the domain-decomposition approach requires us to exchange stress and particle velocity fields on the boundaries of the sub-meshes between neighboring MPI processes. The communication overhead is therefore proportional to the number of grid points on the boundaries of each sub-mesh. Suppose the number of grid points in each dimension of a sub-mesh is NXT , NYT and NZT , the total number of grid points on the boundaries of each sub-mesh is therefore $S = 2(NXT \times NYT + NXT \times NZT + NYT \times NZT)$. The total number of grid points of the whole sub-mesh is $V = NXT \times NYT \times NZT$. For a given V , we need to minimize S in order to minimize the communication overhead. To minimize S , we can replace the NZT in the expression for S with $V/(NXT \times NYT)$ and set the first-order partial derivatives of S with respect to NXT and NYT to zeros to obtain two equations for NXT and NYT . The solution of these two equations gives $NXT = NYT$. The same analysis can also be applied to NXT and NZT and the solution gives $NXT = NZT$. Therefore the minimum number of boundary grid points of the sub-mesh for a given V is obtained when $NXT = NYT = NZT$ (i.e., a cube-shaped sub-mesh).

In this example input file, the first line specifies that the mesh output file XYZGRD will be stored under `/home/pochen/wrk/F3DWI/HS/VM/`. In the second line, I specify that my target modeling area is inside the UTM longitudinal zone 11. The next two lines are the longitude/latitude locations of vertex 1 and 2, which were picked using Google Maps. The distance between vertex 1 and 4 is 47,500 m and the vertical thickness of the modeling volume is 35,500 m. The grid spacing is 200 m. The last line specifies that the mesh is partitioned into two sub-meshes in each direction, therefore the total number of MPI processes needed for the simulation is eight.

On my Linux box, the binary executables are stored under the directory `/home/pochen/wrk/F3DWI/bin/`. Suppose the text input file for the mesh generation code `FD_GRID_XYZ.in` is under the current directory, I can run the code in the terminal using the command

```
/home/pochen/wrk/F3DWI/bin/FD_GRID_XYZ_rect FD_GRID_XYZ.in
```

If the `bin` directory is added to the search path²³, in the terminal I can omit the full path to the binary executable and use the command

```
FD_GRID_XYZ_rect FD_GRID_XYZ.in
```

Two text output files will be generated by `FD_GRID_XYZ_rect`, a file named `box.dat` under the current directory and the file `XYZGRD` under the directory specified on the first line of `FD_GRID_XYZ.in` (Listing 2.1). In `box.dat` (Listing 2.2), the longitudes (left column) and latitudes (right column) of vertex 1–4 of the rectangular modeling region are listed from line 1 to 4. The last line repeats the longitude/latitude of vertex 1. This file can be used for drawing the rectangular modeling region using the Generic Mapping Tools (GMT)²⁴ (Wessel et al. 2013). Figure 2.10, which was generated using GMT, shows the geographic area and the target rectangular modeling region for the half-space example.

²³ Under the `bash` shell, all directories listed in the `PATH` environment variable are automatically searched for executables. The `bin` directory can be added to `PATH` by adding the line `export PATH=/home/pochen/wrk/F3DWI/bin:$PATH` to the `.bashrc` or `.bash_profile` file under the home directory.

²⁴ The GMT is a set of command-line binary executables for generating and processing figures related to geographic data sets. More detailed information can be found at <http://gmt.soest.hawaii.edu/>.

-118.00573799999999	33.690066999999999
-118.03045700000000	33.820230000000002
-117.52090189901836	33.890036719992096
-117.49629718561751	33.756393997202352
-118.00573799999999	33.690066999999999

Listing 2.2 The content of the box.dat output file.

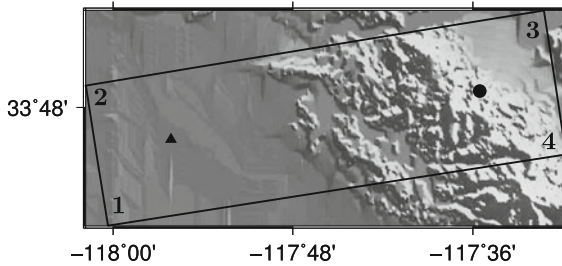


Fig. 2.10 The geographic location of the half-space example. The *black box* shows the rectangular modeling area and its four projection vertices are numbered. *Black circle*: epicenter of the seismic source; *black triangle*: surface projection of the receiver; *gray-scale* background: topography.

Snippets of the output file XYZGRD for the half-space example are shown in Listing 2.3. The first three lines specify the total number of grid points in each spatial dimension, the total number of sub-meshes (i.e., MPI processes) in each spatial dimension and the number of grid points in each spatial dimension of a sub-mesh. The next four lines specify the UTM Easting (left column) and Northing (right column) coordinates in meter for vertex 1–4 of the rectangular modeling region. The rest of XYZGRD lists the grid numbers in the vertical direction and their corresponding depths in meter. In Listing 2.3 the grid number 180 is located at the free surface with depth of 0 m and the grid number 1 is at 35,800 m below the free surface (i.e., the depth of the origin of the box coordinate). The vertical thickness of the modeling volume is slightly larger than the vertical thickness 35,500 m specified in FD_GRID_XYZ.in (Listing 2.1) because FD_GRID_XYZ_rect has adjusted the vertical thickness such that it is a multiple of the grid spacing specified in FD_GRID_XYZ.in and the total number of grid points in the vertical direction is also a multiple of the number of sub-meshes in the vertical direction. The distance between vertex 1 and 4 is also slightly extended from 47,500 m in FD_GRID_XYZ.in (Listing 2.1) to 47800 m in XYZGRD for the same reason.

2.2.2 Structural Model

After the mesh is generated, we need to assign material properties to every grid point in the mesh. For elastic media, we need three properties per grid point, the P- and

```

240   76   180   ! Global grid numbers NX, NY, NZ
    2    2    2   ! Chunks (#of proc.) in X, Y, Z
120   38   90   ! Grid numbers per chunk NXT, NYT, NZT
406782.38411618437      3728040.2806620458
404578.62796418497      3742868.9354490079
451832.60788530414      3749891.5717200460
454036.36403730355      3735062.9169330839
    180  -0.0000000000000000
    179  -200.00000000000000
    178  -400.00000000000000
      .
      .
      .
    3   -35400.000000000000
    2   -35600.000000000000
    1   -35800.000000000000

```

Listing 2.3 The first ten lines and the last three lines in the XYZGRD file.

S-wave velocities and the density. For viscoelastic media, we need two additional properties per grid point, the attenuation quality factors Q_p and Q_s . For demonstration purposes, I list the C++ code named `vm_hs_gen.cpp` (Listing 2.4) for assigning constant elastic material properties to the half-space mesh generated in Sect. 2.2.1.

This code can be compiled in the terminal using the command

```
g++ -o vm_hs_gen vm_hs_gen.cpp
```

which will generate the executable `vm_hs_gen` in the current directory. After executing it by typing its name in the terminal, it will generate 8 different binary files named as `HSGRD000000000`, etc., under `/home/pochen/wrk/F3DWI/HS/VM/`, which is the same directory for storing the mesh output file `XYZGRD` on my Linux box. Each binary output file contains the material properties for all grid points of a sub-mesh and the number of files equals the number of sub-meshes.

Detailed discussions about the C/C++ programming language are beyond the scope of this book. But we can understand the code without delving too much into the language. The basic parameters about the mesh are defined on line 6–10 based upon the information in `XYZGRD`. The total number of grid points in each spatial dimension is specified on line 6 and the number of sub-meshes in each spatial dimension is specified on line 7. The number of grid points in each dimension of a sub-mesh is then computed on line 8–10. The constant P- and S-wave velocities in m/s and the density in kg/m^3 are defined on line 11 as four-byte floating point numbers.

From line 13–15, we have a three-level nested loop, which allows us to loop through every sub-mesh. The three iterators, `ipx`, `ipy` and `ipz`, range from zero to the number of sub-meshes in each spatial dimension subtracting 1. Inside this

```

1  #include <iostream>
2  #include <fstream>
3  #include <algorithm>
4  using namespace std;
5  int main() {
6      const int NX=240, NY=76, NZ=180;
7      const int NPX=2, NPY=2, NPZ=2;
8      int NXT=NX/NPX;
9      int NYT=NY/NPY;
10     int NZT=NZ/NPZ;
11     const float vp=6500.0, vs=3500.0, rho=3000.0;
12     char mdl_out[200]={0};
13     for (int ipz=0; ipz<NPZ; ipz++) {
14         for (int ipy=0; ipy<NPY; ipy++) {
15             for (int ipx=0; ipx<NPX; ipx++) {
16                 sprintf(mdl_out, "/home/pochen/wrk/F3DWI/HS/VM/HSGRD%3.3d
17                     %3.3d%3.3d", ipx, ipy, ipz);
18                 FILE *bin_out;
19                 bin_out=fopen(mdl_out, "wb");
20                 for (int iz=-2; iz<=NZT+1; iz++) {
21                     for (int iy=-2; iy<=NYT+1; iy++) {
22                         for (int ix=-2; ix<=NXT+1; ix++) {
23                             int ixg=min(max(0, ipx*NXT+ix), NX-1);
24                             int iyg=min(max(0, ipy*NXT+iy), NY-1);
25                             int izg=min(max(0, ipz*NXT+iz), NZ-1);
26                             fwrite(&vp, sizeof(float), 1, bin_out);
27                             fwrite(&vs, sizeof(float), 1, bin_out);
28                             fwrite(&rho, sizeof(float), 1, bin_out);
29                         }}}}
30                 fclose(bin_out);
31             }}}}
32     cout << "constant_Vp:_" << vp << endl;
33     cout << "constant_Vs:_" << vs << endl;
34     cout << "constant_Rho:_" << rho << endl;
35     return 0;}

```

Listing 2.4 C++ source code for generating the half-space structural model. The line numbers at the beginning of each line in this and some other listings are not part of the file. They are added to the listings to make it easier to reference specific lines in the text.

nested loop, the three-tuple (ipx, ipy, ipz) is actually the index of a sub-mesh (Fig. 2.9). On line 16, the output file name `mdl_out` is constructed using this sub-mesh index. The string `%3.3d` means that each number in the index tuple is printed using three digits with zero-padding on the left. For example, the output file for sub-mesh $(0,1,1)$ will have a file name ending with the nine digits `000001001`. The full path of the output file and the prefix `HSGRD` are also printed into `mdl_out`. After the file name is generated, we can open the file for writing binary data on line 18.

From line 19–21, we have another three-level nested loop, which allows us to loop through every grid point of the sub-mesh (ipx , ipy , ipz). The range of the iterators ix , iy and iz is from -2 to the number of grid points in each dimension of the sub-mesh plus 1. The reason for having two extra grid points on each end of a dimension is that each MPI process needs some ghost layers on the boundaries of its sub-mesh to store the information obtained from its neighbors in order to carry out the finite-difference calculations on the interior grid points of its sub-mesh. For the fourth-order differencing scheme discussed in Sect. 2.1.3, the ghost layers are 2 grid points thick. On line 25–27, I write the P- and S-wave velocities and the density (must be written in this order) at the grid point (ix , iy , iz) in the sub-mesh (ipx , ipy , ipz) as four-byte-long²⁵ binary floating-point numbers into the output file.

This code can be modified and extended in many different ways to account for much more complex situations and to improve the efficiency. I mention a few in the following.

- One obvious modification is that instead of hard-wiring the mesh information into the code on line 6–10, we can read the mesh information directly from XYZGRD.
- For viscoelastic media, we need to add two more lines between line 27 and line 28 to write Q_p and Q_s (must be written in this order) also as four-byte-long binary floating-point numbers into the output file.
- The three integers, ixg , iyg and izg defined on line 22–24 are not used in the current code, but these three lines allow us to convert the index of an arbitrary grid point (ix , iy , iz) in the sub-mesh (ipx , ipy , ipz) to its index in the full mesh of the whole modeling volume. We call (ixg , iyg , izg) the global grid index of a grid point and (ix , iy , iz) its local grid index in the sub-mesh (ipx , ipy , ipz). The global index can be used, for instance, to locate and retrieve the correct material properties from a 3D structural model defined on 3D arrays.
- For very large meshes, this code needs to be parallelized. The parallelization can be based on domain-decomposition, i.e., each MPI process is responsible for one or a few sub-meshes.

After all grid points in the full mesh have been assigned appropriate material properties, the code should print out the minimum S-wave velocity and the maximum P-wave velocity in the full mesh, which will be used in the parameter file described in Sect. 2.2.3. For the half-space model, this task is easily completed on line 31–32. For a large 3D structural model processed in parallel, we can first obtain the minimum S-velocity and maximum P-velocity on each sub-mesh using conditional

²⁵ The `sizeof` function used on line 25–27 returns the number of bytes for a given data type. On most systems, the `float` data type is four-byte long. At the current stage the `awp` code only reads binary material property files written in four-byte-long floating-point numbers.

statements (e.g., `if...else...`) or operators (e.g., the `?:` operator in C/C++) and then obtain the minimum S-velocity and maximum P-velocity on the full mesh by using the MPI reduction operations²⁶.

2.2.3 Parameter File

The parameter file, a text file written in key-value pairs, stores the parameters that will be used throughout the entire inversion project. The names of the keys and short descriptions about their values are listed in Table 2.2 and the parameter file for the half-space example, named `param_HS.dat`, is shown in Listing 2.5. The order of the key-value pairs in the file can be arbitrary, but each key-value pair must occupy a single line in the file and follow the format “key = value”, using “=” to separate the key from the value. The key names cannot be changed and should be all in capital letters.

The values for the keys `BINDIR` and `SCRIPTDIR` are the full paths to the `bin` and the `py` directories of the `F3DWI` package on your system. A good practice is to create a new work directory for each simulation/inversion project and assign its full path to `WRKDIR`. For the half-space example, I have created a work directory named `HS` for storing all the files related to this project and its full path on my Linux box is assigned to `WRKDIR` on line 3 of Listing 2.5. I often use a sub-directory named `JOBS` under the work directory for storing some input/output files and also some short scripts for interacting with the job scheduling system²⁷. The full path to the `JOBS` directory is assigned to the key `JOBDIR`. The parameter file itself, `param_HS.dat` for the half-space example, is also stored under the `JOBS` directory.

The full path of the mesh file `XYZGRD` is assigned to the key `XYZGRD` on line 5 (Listing 2.5). The coordinates used in the mesh file are defined in terms of the UTM and the corresponding UTM longitudinal zone number is assigned to the key `UTMZONE` on line 6 (Listing 2.5).

Basic information about the structural model is registered into the parameter file on line 7–9 (Listing 2.5). The full path together with the prefix `HSGRD` of the material property files (line 16 in Listing 2.4, Sect. 2.2.2) is assigned to the key `MEDIA`. The minimum S- and maximum P-wave velocities in the structural

²⁶ The MPI functions `MPI_Reduce` and `MPI_Allreduce` can combine the values provided in the input buffer of each MPI process, using a specified operation, such as taking the minimum or maximum, and return the combined value to the output buffer of either the `root` process (in the case of `MPI_Reduce`) or all processes (in the case of `MPI_Allreduce`).

²⁷ On large computing systems, the computational tasks (i.e., batch jobs) are allocated among the available computing resources through a job scheduling software. One example is the Portable Batch System (PBS) software. To interact with the job scheduling system (e.g., submit jobs, monitor the progress of jobs, cancel previously submitted jobs), we often need to write short scripts.

Table 2.2 Format of the parameter file for storing control parameters used in F3DWI

Key name	Value description
BINDIR	Full path to directory for storing the binary executables
SCRIPTDIR	Full path to directory for storing the <code>python</code> scripts
WRKDIR	Full path to the project work directory
JOBDIR	Full path to the project job directory
XYZGRD	Full path of the XYZGRD file
UTMZONE	The UTM longitudinal zone number of the modeling region
MEDIA	Full path and the prefix of the material property files
MINIMUM VELOCITY	Minimum S-wave velocity in the structure model
MAXIMUM VELOCITY	Maximum P-wave velocity in the structure model
SOURCE RECEIVER LIST	Full path of a text file that stores a list of sources and receivers used in the project
TMAX	The maximum length of the synthetic seismograms in seconds
DT	The time-step length in seconds
CFL	The CFL stability condition
USEPML	1: use the PML boundary condition (Sect. 2.1.5); 0: use the absorbing boundary condition
PMLCOEF	The tuning coefficient τ for PML (Eq. 2.90)
BOUNDARY NODE NUMBER	The width of the PML n_b (Fig. 2.5)
VISCO	1: viscoelastic material; 0: elastic material
SOURCE STENCIL SIZE IN X	The number of grid points in the x_1 dimension of a small source-centered volume
SOURCE STENCIL SIZE IN Y	The number of grid points in the x_2 dimension of a small source-centered volume
SOURCE STENCIL SIZE IN Z	The number of grid points in the x_3 dimension of a small source-centered volume
NXSKP	The decimation rate in x_1
NYSKP	The decimation rate in x_2
NZSKP	The decimation rate in x_3
NTSKP	The decimation rate in time

model are assigned to keys `MINIMUM VELOCITY` and `MAXIMUM VELOCITY`, respectively.

On line 10 of Listing 2.5, the key `SOURCE RECEIVER LIST` is assigned the full path of the text file describing the sources and receivers used in the project. The details about this file are described in Sect. 2.2.4. This file can be generated after we finish editing the parameter file and the content of this file can be changed anytime if necessary.

```

1 BINDIR = /home/pochen/wrk/F3DWI/bin
2 SCRIPTDIR = /home/pochen/wrk/F3DWI/py
3 WRKDIR = /home/pochen/wrk/F3DWI/HS
4 JOBDIR = /home/pochen/wrk/F3DWI/HS/JOBS
5 XYZGRD = /home/pochen/wrk/F3DWI/HS/VM/XYZGRD
6 UTMZONE = 11
7 MEDIA = /home/pochen/wrk/F3DWI/HS/VM/HSGRD
8 MINIMUM VELOCITY = 3500.0
9 MAXIMUM VELOCITY = 6500.0
10 SOURCE RECEIVER LIST = /home/pochen/wrk/F3DWI/HS/JOBS/srinfo/
    srlst
11 TMAX = 15.0
12 DT = 0.015
13 CFL = 0.9792
14 USEPML = 1
15 PMLCOEF = 3.5
16 BOUNDARY NODE NUMBER = 10
17 VISCO = 0
18 SOURCE STENCIL SIZE IN X = 1
19 SOURCE STENCIL SIZE IN Y = 1
20 SOURCE STENCIL SIZE IN Z = 2
21 NXSKP = 4
22 NYSKP = 4
23 NZSKP = 4
24 NTSKP = 2

```

Listing 2.5 The parameter file for the half-space example.

On line 11, the value for the key `TMAX` determines the maximum length in seconds for the synthetic seismograms computed using `awp`. On line 12, the value for `DT` specifies the time-step length in seconds. On line 13, the key `CFL` specifies the Courant-Friedrichs-Lewy (CFL) stability condition²⁸, which should have a value smaller than 1. In `F3DWI`, the CFL condition is verified using the equation

$$\frac{\alpha_{max} \Delta t}{h} < 0.5, \quad (2.123)$$

where α_{max} is the maximum P-wave velocity assigned to key `MAXIMUM VELOCITY`, Δt is the time-step length assigned to key `DT`, h is the grid spacing of the mesh, which can be obtained from the `XYZGRD` file assigned to the key `XYZGRD`. If the value assigned to `DT` in the parameter file violates Eq. 2.123, a new value will be computed using the formula

$$\Delta t = \frac{0.5h}{\alpha_{max}} \text{CFL}, \quad (2.124)$$

²⁸ The CFL condition is a necessary condition for an explicit time-integration scheme to be stable (Courant et al. 1967). Roughly speaking, for a given mesh with grid spacing h , the time-step length Δt must be sufficiently small in order for the numerical scheme to be causal.

where CFL is the value assigned to the key CFL in the parameter file. Any script in the `py` directory that uses the parameter file as one of its command-line inputs will automatically modify the parameter file to replace an incorrect value assigned to DT with the value computed using Eq. 2.124. If the value assigned to DT satisfies Eq. 2.123, the value will be left unchanged.

Line 14–16 (Listing 2.5) are related to the use of the PML boundary condition in `awp`. If the key USEPML is set to 1, the PML boundary condition as described in Sect. 2.1.5 is used, otherwise, the absorbing boundary condition of Cerjan (Cerjan et al. 1985) is used instead. If USEPML is set to 0, the keys PMLCOEF and BOUNDARY NODE NUMBER assigned on line 16–17 have no effect on the wave-propagation simulations. If USEPML is set to 1, appropriate values of these two keys can be chosen based on the descriptions in Sect. 2.1.5. The key PMLCOEF usually has a value between 3 and 4 and the key BOUNDARY NODE NUMBER can have any integer between 8 and 20.

For the half-space example, I will only consider elastic media and on line 17 of Listing 2.5, I set the key VISCO to 0. For wave-propagation simulations in viscoelastic media, the key VISCO should be set to 1 and the material properties at every grid point should include P- and S-wave velocities, density and the attenuation quality factors Q_p and Q_s , written as four-byte-long floating point numbers in this order (e.g., line 25–27, Listing 2.4).

For seismic source parameter inversions (e.g., centroid moment tensor inversions), the wavefields on a small volume surrounding the source need to be stored. The three keys assigned on line 18–20 in Listing 2.5 define the dimensions of this source-centered volume in the x_1 , x_2 and x_3 directions of the box coordinate. For instance, if the key SOURCE STENCIL SIZE IN Z is assigned a value of 2, the source-centered volume will have 5 grid points in the x_3 direction with the point source located at the middle and 2 grid points both above and below. For situations involving multiple point sources, the same dimensions are applied to every source in the source-receiver list file. If we do not plan to invert for seismic source parameters, these three keys can be set to zeros. Seismic source parameter inversions using F3DWI are discussed in Sects. 3.4 and 4.3.8.

For computing sensitivity (Fréchet) kernels used in full-3D tomography, we often need to store the wavefield on a decimated mesh and at decimated time samples to reduce the disk storage amount. I will discuss the details about Fréchet kernel calculations in Chap. 4. On line 21–24 in Listing 2.5 I specify the decimation rate in the three spatial dimensions and also the time dimension. For example, on line 21–23 in Listing 2.5 I have assigned the value 4 to the keys NXSKP, NYSKP and NZSKP and the value 2 to the key NTSKP, which will request the `awp` code to write out the wavefield at grid points whose x_1 , x_2 and x_3 indices are multiples of 4 (i.e., 4, 8, 12,...) at those time steps that are multiples of 2 (i.e., 2, 4, 6,...) during the wave-propagation simulation. The values assigned to NXSKP, NYSKP and NZSKP do not need to be identical. The appropriate values for the spatial decimation rates can be determined based on the desired spatial resolution of the Fréchet kernels. The temporal decimation rate has effects on the accuracy of the kernel value at each grid

```

1 0 IN.RC01 -1 -117.93578183557393 33.770387530932254 24.0
2 1 100001 -1 -117.59233339866256 33.815181833226426 24.0 1 1 1 1
   0 0 0 /home/pochen/wrk/F3DWI/HS/JOBS/srinfo/src_hs.dat
3 2 IN.RC01.1 -1 -117.93578183557393 33.770387530932254 24.0 -2 1 0
   0 /home/pochen/wrk/F3DWI/HS/JOBS/srinfo/imp_hs.dat
4 2 IN.RC01.2 1 40 40 60 -2 0 1 0 /home/pochen/wrk/F3DWI/HS/JOBS/
   srinfo/imp_hs.dat
5 2 IN.RC01.3 1 40 40 60 -2 0 0 1 /home/pochen/wrk/F3DWI/HS/JOBS/
   srinfo/imp_hs.dat
6 2 100001.adj 2 /home/pochen/wrk/F3DWI/HS/JOBS/srinfo/100001.adj
7 ...

```

Listing 2.6 An example of the source-receiver file.

point and it is usually sufficient to have ten time samples per shortest period of the synthetic seismogram.

2.2.4 Sources and Receivers

Two types of sources are considered in F3DWI, body-force sources specified as force vectors and moment sources specified as moment tensors. Both types of sources can have both spatial- and temporal-dependence. For moment sources, the source-time function at a grid point specifies the moment-rate function $\partial m_{ij}/\partial t$ at that grid point (Eq. 2.3). In some situations (e.g., calculations involving reciprocity) we need to put sources at receiver locations and retrieve synthetic wavefields at source locations. The distinction between sources and receivers starts to blur under such circumstances, which is also reflected in the format of the source-receiver file.

2.2.4.1 File Format

A source-receiver list file for the half-space example is shown in Listing 2.6, which includes a simple point receiver, a point moment source, three point body-force sources and one finite source. Each source or receiver occupies a single line in the file. The different columns of a line can be separated by any amount of spaces. This source-receiver list file, named `sr.lst`, is stored under a sub-directory `srinfo` inside the project `JOBS` directory (Line 10, Listing 2.5).

The format for a simple point receiver is shown on line 1. The first column is a flag for indicating whether this line is for a source or a receiver. For a simple receiver line, this source-receiver flag must be 0. The second column is a unique receiver ID, which is composed of a seismic network ID, `IN`, and a station ID, `RC01`, separated by a dot. The network ID and the station ID can have any number of characters. The third column is a flag for indicating whether the receiver location specified in the following three columns is in terms of longitude, latitude and depth or in terms of

Table 2.3 The source-type flag on column 7 of the source-receiver list file, the description of the source type and the meaning of the following columns

Flag	Source description	Following columns
1	Moment tensor in A&R format	$M_{xx}, M_{yy}, M_{zz}, M_{xy}, M_{xz}, M_{yz}$
2	Focal mechanism in A&R format	strike, dip, rake (degrees)
3	Moment tensor in x -East, y -North, z -Up	$M_{xx}, M_{yy}, M_{zz}, M_{xy}, M_{xz}, M_{yz}$
4	Moment tensor in the box coordinate	$M_{xx}, M_{yy}, M_{zz}, M_{xy}, M_{xz}, M_{yz}$
-1	Body-force vector in x -East, y -North, z -Up	$\gamma_x, \gamma_y, \gamma_z$
-2	Body-force vector in the box coordinate	$\gamma_x, \gamma_y, \gamma_z$

global grid indices in the box coordinate. For the former, the third column must be -1 and for the latter, the third column must be 1 . For the example shown in Listing 2.6, the third column is -1 , therefore the next three columns specify the longitude, latitude²⁹ and the depth³⁰ of the point receiver.

Line 2 in Listing 2.6 shows an example of a simple point source. The source-receiver flag on the first column for a point source must be 1 . The second column is a unique source ID, which is 100001 in this example. Column 3–6 have the same format as in line 1. Column 7 is a flag for indicating the type of the source specified in the following columns. It can take 6 different values (Table 2.3), 1 for a moment tensor in the A&R format³¹, 2 for a focal mechanism in the A&R format³², 3 for a moment tensor specified in the x -East, y -North, z -Up coordinate, 4 for a moment tensor specified in the box coordinate, -1 for a body-force vector specified in the coordinate: x -East, y -North, z -Up, -2 for a body-force vector specified in the box coordinate. For body-force vectors, we need to specify the projections (i.e., the cosine of the angle between the vector and the coordinate axis) of the body-force vector on the three axes of the coordinate system, γ_x , γ_y and γ_z , in the following three columns. For the example shown on line 2, the source-type flag on column 7 is set to 1 , therefore column 8–13 specify the A&R moment tensor in the order: M_{xx} , M_{yy} , M_{zz} , M_{xy} , M_{xz} , M_{yz} . The values shown on line 2 indicate that this source is an explosive source (i.e., $M_{xx} = M_{yy} = M_{zz} = 1$, $M_{xy} = M_{xz} = M_{yz} = 0$). Column 14 is the full path of the text file for storing the source-time function. If the source-type flag on column 7 is set to 2 , column 8–10 specify the A&R focal

²⁹ In F3DWT, a geodetic location is specified using real-valued longitude and latitude in degrees, the western hemisphere has negative longitude and the southern hemisphere has negative latitude.

³⁰ The depth is specified in km and should always be larger than or equal to 0 . The free-surface is at 0 km depth.

³¹ The A&R moment tensor format is the format specified on page 112 in (Aki and Richards 2002). Note that the coordinate system for this format (x : North, y : East, z : Down) is different from the one used in F3DWT. Necessary coordinate transforms are done automatically in F3DWT.

³² The A&R focal mechanism format is the format specified on page 101 in (Aki and Richards 2002). The strike, dip and rake angles should be in degrees in the source-receiver file.

```

1      1.0100000e+02 ! total number of time samples
2      0.0000000e+00 ! beginning time
3      2.0000000e-02 ! time-sampling interval
4      1.7687185e+07 ! 1st sample
5      3.7669109e+07 ! 2nd sample
6      ...

```

Listing 2.7 The first 5 lines in the example source-time function file.

mechanism, in the order: strike, dip and rake, and column 11 should be the full path of the source-time-function file.

The source-time-function file specifies a uniformly sampled time series. Listing 2.7 shows the first five lines of the source-time-function file on line 2 of Listing 2.6. The graphs of this time series (normalized by dividing the scalar moment) in the time domain and its amplitude spectrum in the frequency domain are shown in Fig. 2.7. Line 1–3 in Listing 2.7 specify the total number of time samples, the beginning time of the first sample in seconds and the time-sampling interval in seconds. The amplitude of the first time sample is on line 4, followed by the amplitudes of the rest of the time samples. Note that the time-sampling interval specified on line 3 in Listing 2.7, 0.02 s, is different from the time-step length used in `awp`, which is stored in key `DT` (Table 2.2; Line 12, Listing 2.5). The script `proc_srlst.py` (Sect. 2.2.4.2) will re-sample it to the time-step length stored in `DT` by using the cubic spline interpolation algorithm and generates source files that can be read by `awp`.

Line 3–5 in Listing 2.6 specify three point body-force sources located at the same location as the receiver `IN.RC01`. For a source that is applied at a receiver location, the source-receiver flag on the first column must be 2. The source ID on the second column is composed of the receiver ID and a suffix indicating the direction of the body force. For example, on line 4, the suffix `.2` indicates that the body force points to the x_2 direction of the box coordinate.

On line 3 in Listing 2.6, column 3–6 are identical to those on line 1. On lines 4 and 5, column 3 is set to 1, therefore column 4–6 should be global grid indices in the box coordinate, which must be positive integers. `F3DWI` understands both formats. If you prefer to use grid indices, there is a script `geo2grd.py` under the `py` directory that allows you to convert a list of geodetic locations into grid indices. This script is essentially a wrapper around the binary executable `geo2grd` under the `bin` directory. Suppose the `py` directory has been added to the search path (Footnote 23), this script can be used by typing the following command in the terminal

```
geo2grd.py param_HS.dat geo2grd.in 2 1 3 geo2grd.ou
```

The first command-line input is the name of the parameter file (Sect. 2.2.3; Listing 2.5); the second input is the name of a text input file containing a list of the geodetic locations; the third to the fifth input are the column numbers inside the

33.815181833226426	-117.59233339866256	24
33.770387530932254	-117.93578183557393	24
33.770387530932254	-117.93578183557393	0
33.815181833226426	-117.59233339866256	0

Listing 2.8 An example input file for `geo2grd.py`.

201	40	60	33.815181833226426	-117.59233339866256	24
40	40	60	33.770387530932254	-117.93578183557393	24
40	40	180	33.770387530932254	-117.93578183557393	0
201	40	180	33.815181833226426	-117.59233339866256	0

Listing 2.9 The output file generated by `geo2grd.py`.

text input file, `geo2grd.in`, for longitude, latitude and depth (Footnote 29, 30), in this order; the sixth input is the name of the output file that will be generated by this script. For all the file names used in this command, you will need to include the full paths if they are not under the current directory. An example of the input file is shown in Listing 2.8 and its corresponding output file is shown in Listing 2.9. The global grid indices in the box coordinate for the four geodetic locations in the input file are shown on column 1–3 in the output file. The conversion was done by finding the grid point in the mesh that has the smallest distance from an input geodetic location, therefore the maximum possible mislocation error in each direction is $h/2$, where h is the grid spacing of the mesh.

The source-type flag on column 7 is set to -2 on line 3–5 in Listing 2.6, which indicates that column 8–10 on these 3 lines specify body-force vectors in the box coordinate. For example, on line 3, the vector defined on column 8–10, $(1, 0, 0)$, specify a unit-length body-force vector pointing to the x_1 direction and on lines 4 and 5, the vectors $(0, 1, 0)$ and $(0, 0, 1)$ specify unit-length body-forces pointing to the x_2 and x_3 directions, respectively. Column 11 on these three lines specifies the full path of a source-time-function file for these body-force sources. The format of this source-time-function file is identical to that in Listing 2.7. The graphs of this source-time function in the time domain and its amplitude spectrum in the frequency domain are shown in Fig. 2.11. This source-time function was generated by applying a sixth-order Butterworth low-pass filter with the corner at 1.5 Hz to a delta impulse signal. The wavefields generated by the three unit body-force sources applied at the receiver `IN.RC01` are examples of what we call receiver Green’s tensors (RGTs), whose utility in solving seismic inverse problems will be demonstrated in later chapters.

Line 6 in Listing 2.6 specifies a finite source that is composed of 2 point sources. In this particular example, this finite source actually represents the adjoint source field for the point moment source `100001` specified on line 2 of Listing 2.6. Details about the adjoint source field will be discussed in later chapters. In terms of the format, the point sources making up the adjoint source field are all applied at receiver locations, therefore the source-receiver flag on column 1 is set to 2. The source ID on column 2 is set to the source ID of the point moment-source `100001` with the

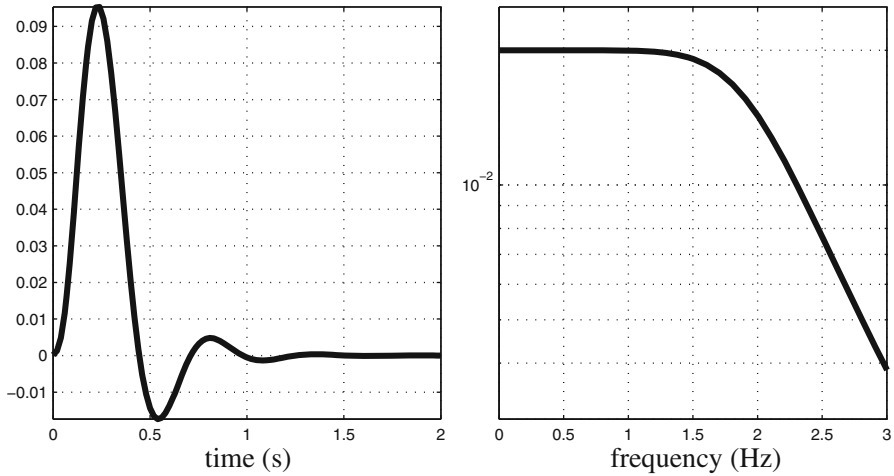


Fig. 2.11 The source-time function for the three body-force sources in the time domain (*left*) and its amplitude spectrum in the frequency domain (*right*).

```

1  2  IN.RC01.1  1  40  40  60  -2  1  0  0  /home/pochen/wrk/F3DWI/HS/SYN
    /100001/IN.RC01.1.adj
2  2  SF.RC02.1  1  40  40  180  -2  1  0  0  /home/pochen/wrk/F3DWI/HS/SYN
    /100001/SF.RC02.1.adj

```

Listing 2.10 An example of an adjoint-source-field file.

suffix `.adj`. The grid flag on column 3 is set to 2, which is the total number of point sources making up this finite source. Column 4 specifies the full path of the file that describes all the point sources in this adjoint source field. Listing 2.10 shows the content of this file. This file has the same format as that shown in Listing 2.6. It has two lines, the first line specifies a point body-force source applied at receiver `IN.RC01`, pointing to the x_1 direction and having a source-time-function file specified in the last column. The second line specifies another point body-force source applied at a receiver named `SF.RC02`, also pointing to the x_1 direction and having a different source-time-function file specified in the last column.

2.2.4.2 File Processing

The source-receiver file needs to be processed by the script `proc_srlst.py` in the `py` directory. This script is a wrapper around the binary executable `write_point_source` in the `bin` directory. It processes the information stored in the source-receiver file (e.g., the `srlst` file shown in Listing 2.6) and the associated source-time-function files. It generates a set of source files, which have file

1	0	IN.RC01	40	40	60
2	1	100001	201	40	60
3	2	IN.RC01.1	40	40	60
4	2	IN.RC01.2	40	40	60
5	2	IN.RC01.3	40	40	60
6	2	100001.adj_IN.RC01.1	40	40	60
7	2	100001.adj_SF.RC02.1	40	40	180

Listing 2.11 The output file `srlst.grd` generated by `proc_srlst.py`.

names with the suffix `.src` and will be read by `awp`, and also a text file with the suffix `.grd`, which will be read by the script `set_sr_grd.py` (Sect. 2.2.5).

Assuming the `py` directory has been added to the search path (Footnote 23) and the parameter file `param_HS.dat` is under the current directory³³, I can type the following command in the terminal,

```
proc_srlst.py param_HS.dat 1 6
```

The first command-line input is the parameter file (use the full path if it is not under the current directory). This script will read the source-receiver file associated with the key `SOURCE RECEIVER LIST` in the parameter file (Line 10, Listing 2.5). The second and third command-line inputs indicate the range of lines inside the source-receiver file that this script will process. For the command shown above, the script `proc_srlst.py` will process line 1–6 in the `srlst` file shown in Listing 2.6.

The output files are stored under the same directory as the `srlst` file, which is under the `srinfo` sub-directory inside the project `JOBS` directory (Line 10, Listing 2.5) on my Linux box. There is a text output file named `srlst.grd`, whose content is shown in Listing 2.11. For point sources and receivers (e.g., Line 1–5, Listing 2.6) the first two columns in `srlst.grd` are identical to those in the corresponding source-receiver file, `srlst`. The following three columns are global grid indices of the point source/receiver in the x_1 , x_2 and x_3 directions of the box coordinate. Geodetic locations specified in longitude, latitude and depth in `srlst` have been converted to grid indices. For the finite source shown on line 6 in Listing 2.6, information about the two point sources making up this finite source is listed on lines 6 and 7 in Listing 2.11. The first column on these two lines is identical to that in the corresponding adjoint-source-field file shown in Listing 2.10. The second column is a new source ID composed of the point-source ID in the adjoint-source-field file (Column 2, Listing 2.10) and a prefix given by the ID of the finite source,

³³ We can execute any of the scripts in `py` and any sequential binary executables in `bin` from inside the project `JOBS` directory, where the parameter file and other input/output files are stored. In such a case, the current directory is the project `JOBS` directory.

```

1      201      40      60      1
2      0.1768718E+08  0.1768718E+08  0.1768718E+08  0.0000000E+00
      0.0000000E+00  -0.3927344E-08
3      0.3115688E+08  0.3115688E+08  0.3115688E+08  0.0000000E+00
      0.0000000E+00  -0.6918216E-08
4      ...

```

Listing 2.12 The source file for the point moment source 100001.

100001.adj_ (Column 2, Line 6, Listing 2.6). The following three columns on line 6–7 in Listing 2.11 specify the grid indices of the two point sources associated with the finite source, which are identical to those shown in Listing 2.10.

The output source files, which will be read by `awp`, are under the same directory as the source-receiver file `srlst` (Line 10, Listing 2.5) and have file names composed of the source/receiver ID shown on the second column of `srlst` (Listing 2.6) and the suffix `.src`. For the point moment source 100001, the first three lines of the source file `100001.src` are shown in Listing 2.12. Line 1 shows the global grid indices of this point source (Column 3–5, Line 2, Listing 2.11) and the corresponding source-type flag (Column 7, Line 2, Listing 2.6). Starting from line 2 in Listing 2.12, column 1–6 show the time histories of the moment tensor components in the order: M_{xx} , M_{yy} , M_{zz} , M_{xy} , M_{xz} , M_{yz} . The moment tensor listed here has been transformed from the coordinate used in the A&R format to the box coordinate and the time histories have been obtained by re-sampling the corresponding source-time function (Listing 2.7) with a time-sampling interval given by the key `DT` in the parameter file (Line 12, Listing 2.5). If the source-type flag is 2 (i.e., focal mechanism specified using strike, dip and rake angles), the moment tensor used in the six columns in Listing 2.12 will then be the one computed from the focal mechanism using Equation (1) on page 112 in (Aki and Richards 2002) and then transformed to the box coordinate. If the source-type flag is either -1 or -2 , the first three columns will list the time histories of the three components of the body-force vector in the x_1 , x_2 and x_3 directions of the box coordinate and the next three columns will be set to 0 and be neglected by `awp` during the wave-propagation simulation. The source file of a finite source is a concatenation of the source files of the individual point sources making up the finite source. Snippets of the source file `100001.adj.src` for the adjoint source field `100001.adj` (Listing 2.10) are shown in Listing 2.13. The time histories of the body-force components for adjoint sources will be explained in later chapters. On lines 1 and 4 in Listing 2.13, the source-type flags for both point sources in this adjoint source field are -2 and the global grid indices are identical to those listed on lines 6 and 7 in Listing 2.11. The source file shown in Listing 2.13 is actually the concatenation of two other source files under the same directory, `100001.adj_IN.RC01.1.src` and `100001.adj_SF.RC02.1.src`, both of which were generated by `proc_srlst.py`.

```

1      40      40      60      -2
2      0.000000E+00  0.000000E+00  0.000000E+00  0.000000E+00
3      0.000000E+00  0.000000E+00
3 ... rest of time histories omitted
4      40      40      180      -2
5      0.000000E+00  0.000000E+00  0.000000E+00  0.000000E+00
6      0.000000E+00  0.000000E+00
6 ... rest of time histories omitted

```

Listing 2.13 Snippets of the source file for the adjoint source 100001.adj.

2.2.5 Recording Grid Points

The synthetic wavefields at selected grid points can be recorded and written onto the disk during wave-propagation simulations. In the `py` directory there are two scripts, `set_sr_grd.py` and `set_ker_grd.py`, that allow us to register those selected grid points with `awp`. These selected grid points are called the recording grid points in the following. The script `set_sr_grd.py` registers the recording grid points associated with the sources and receivers, which are often randomly or non-uniformly distributed in the mesh. The script `set_ker_grd.py` registers a set of recording grid points that are uniformly distributed throughout the mesh. Fréchet kernels for tomography are often calculated and stored on those uniformly distributed recording grid points.

The script `set_sr_grd.py` reads the source-receiver grid file, `srlst.grd` (Listing 2.11), generated by `proc_srlst.py`, adds more recording grid points according to the `SOURCE STENCIL SIZE` keys in the parameter file (Line 18–20, Listing 2.5) around the sources whose source-receiver flags equal to 1 (Column 1, Listing 2.11), removes any duplicate recording grid points and then converts the global grid indices into the corresponding local grid indices for each sub-mesh (Sect. 2.2.2). For the half-space example, I can execute this script by typing the following command in the terminal

```
set_sr_grd.py param_HS.dat
```

It takes the parameter file as its single command-line input. It generates a text output file named `srcbuffgrd.out` under the same directory as the source-receiver grid file `srlst.grd` (Line 10, Listing 2.5). For the half-space example, it also generates eight output files with the prefix `HSGRDRECRDM` followed by the nine-digit sub-mesh ID under the directory for storing the material property files (Line 7, Listing 2.5). The number of such files equals to the number of sub-meshes defined in `XYZGRD` (Sect. 2.2.1; Listing 2.3).

In the source-receiver grid file `srlst.grd`, only the point moment source 100001 has its source-receiver flag equal to 1 (Column 1, Line 2, Listing 2.11). The script `set_sr_grd.py` calls another script `srcbuffgrd.py` in the `py`

```

1 45          ! total number of grid point in this sub-mesh
2 81 2 60     ! local grid index of the 1st point
3 80 1 58     ! local grid index of the 2nd point
4 ...

```

Listing 2.14 Snippets of the grid file for recording the wavefields at source/receiver locations.

directory to generate the small source-centered volume surrounding the source grid point based on the values assigned to the `SOURCE STENCIL SIZE` keys in the parameter file (Line 18–20, Listing 2.5). For this example, the small source-centered volume has 5 grid points in the vertical direction (Line 20, Listing 2.5) (i.e., the source grid point and 2 grid points both above and below the source grid point) and 3 grid points in each horizontal direction (Line 18–19, Listing 2.5), therefore the total number of grid points in this volume is 45. The text output file `srcbuffgrd.out` is only a list of the global grid indices of these 45 points. The total number of points listed in `srcbuffgrd.out` should equal $S_x \times S_y \times S_z \times N_s$, where S_x , S_y and S_z are the number of grid points in each spatial dimension of the source-centered volume as specified by the `SOURCE STENCIL SIZE` keys and N_s is the total number of point sources with the source-receiver flags on the first column equal to 1 in the source-receiver grid file `srlst.grd` (Listing 2.11).

The script `set_sr_grd.py` combines the grid points in `srcbuffgrd.out` with the grid points of other sources and receivers in `srlst.grd` whose source-receiver flags on column 1 are either 0 or 2. The small source-centered volumes for different sources may have overlaps if the sources are close to each other. Some sources and receivers may have identical grid indices. In such cases, `set_sr_grd.py` removes any duplicate grid points from the combined recording grid point set such that every grid point in the final recording grid point set is unique. It then converts the global grid indices of those recording grid points into their corresponding local grid indices in the sub-meshes (Sect. 2.2.2) and writes out the local grid indices of the recording grid points in each sub-mesh into a separate file named using the sub-mesh index (Sect. 2.2.2). For example, the local grid indices in the sub-mesh (1, 1, 0) (Fig. 2.9) are written into the recording grid file `HSGRDRECRDM001001000` under the directory for storing material property files, which is specified by the key `MEDIA` in the parameter file (Line 7, Listing 2.5). The first three lines in this recording grid file are shown in Listing 2.14. The total number of recording grid points in this sub-mesh is shown on the first line, followed by the local grid indices of all those recording grid points. If a sub-mesh does not have any recording grid points in it, the number 0 will be written into its corresponding recording grid file.

The script `set_ker_grd.py` registers a set of uniformly distributed recording grid points with `awp`. Those recording grid points are obtained by decimating the mesh defined in `XYZGRD` (Sect. 2.2.1; Listing 2.3) according to the decimation rates stored in key `NXSKP`, `NYSKP` and `NZSKP` in the parameter file (Table 2.2; Line

21–23, Listing 2.5). It can be executed by typing the following command in the terminal

```
set_ker_grd.py param_HS.dat
```

It takes the parameter file as its single command-line input. For the half-space example, it generates eight output files with the prefix HSGRDREC followed by the sub-mesh ID under the directory for storing material property files (Line 7, Listing 2.5). The format of these output files is identical to the format of the recording grid files generated by `set_sr_grd.py`.

After executing `set_sr_grd.py` and `set_ker_grd.py`, there should be 25 files, including the mesh file `XYZGRD`, under the directory for storing material property files. There should be 8 files named as `HSGRDprocID`, which store the material properties on each sub-mesh (Sect. 2.2.2), 8 files named as `HSGRDRECRDMprocID`, which store the recording grid points associated with sources and receivers in each sub-mesh, and 8 files named as `HSGRDRECprocID`, which store the uniformly distributed recording grid points in each sub-mesh. Here `procID` in the file names is the nine-digit sub-mesh ID. Since each sub-mesh corresponds to a unique MPI process during the wave-propagation simulations, the sub-mesh ID can also be used as the unique process ID.

2.2.6 Wavefield Simulation

Wave-propagation simulations are carried out using `awp`, whose theoretical background has been summarized in Sect. 2.1. The `awp` code is also located in the `bin` directory. But unlike other sequential binary executables in the same directory, `awp` has been parallelized using MPI and the procedure for executing it is closely tied with the MPI parallel job launcher and the job scheduling system (Footnote 27), such as the Portable Batch System (PBS)³⁴, installed on your cluster. On my Linux box, I have installed a version of the MPI library, called `MVAPICH2` (Footnote 17), which comes with a parallel job launcher `mpiexec`³⁵ installed under the `~/tools/mpi/bin` directory. I do not have a job scheduling system installed on my Linux box, so I will be using `mpiexec` in the terminal. On clusters with job schedulers, it is often necessary to invoke the MPI job launcher inside a job script and then submit the job script to the scheduler for execution (Footnote 34).

³⁴ A good introduction to PBS can be found at <http://hpc.sissa.it/pbs/pbs.html>.

³⁵ On your system, you may have different parallel job launchers, such as `mpirun`, `mpirun_rsh`, `mpirun.lsf`, `aprun`, `runjob`. You may need to look into the user guide of your cluster or consult with your system administrator.

The script `setup_awp.py` in the `py` directory can help to set up the correct environment for running `awp`. For the half-space example, I can execute it in the terminal using the following command

```
setup_awp.py param_HS.dat FWF 2 5
```

The first command-line input is the parameter file (Table 2.2; Listing 2.5). The second command-line input `FWF` stands for Forward Wave Field. In `F3DWTI`, two different types of wave-propagation simulations are considered, the forward simulation and the adjoint simulation. The details about the adjoint simulations will be discussed in later chapters. To set up the environment for adjoint simulations, the `FWF` on the second command-line input needs to be replaced with `AWF`, which stands for Adjoint Wave Field. Lowercase letters, `fwf` and `awf`, also works. The two numbers on the third and fourth command-line inputs give the range of the line numbers inside the source-receiver list file `srlst` (Listing 2.6) for which the synthetic wavefields will be computed using `awp`. Any lines for point receivers (i.e., the source-receiver flag on column 1 of `srlst` equal to 0) within the line-number range will be automatically skipped. The command above will create a directory named `FWF` under the project work directory specified by key `WRKDIR` in the parameter file (Line 3, Listing 2.5) and 4 sub-directories named using the source ID in `srlst` (Column 2, Listing 2.6), `100001`, `IN.RC01.1`, `IN.RC01.2` and `IN.RC01.3`, under the `FWF` directory. Under each source sub-directory, for instance the `100001` sub-directory under `FWF`, there are two sub-directories named `output_ckp` and `output_vlm` for storing output files generated by `awp`. There are also a soft link that points to the `awp` binary executable in the `bin` directory, a text file named `IN3D`, which is an input file for `awp`, and a copy of the corresponding source file generated by `proc_srlst.py` previously (Sect. 2.2.4.2) and then copied from the source-receiver directory (Line 10, Listing 2.5) by `setup_awp.py`. The script `setup_awp.py` calls another script `write_IN3D.py` under the `py` directory to generate the `IN3D` file in each source sub-directory. The `IN3D` file summarizes information about the mesh, full paths to the source and the structural model files and a set of parameters that control the behavior of `awp`. In practice, I do not need to edit this file manually.

If the source sub-directories already exist (e.g., generated during previous `awp` runs), executing `setup_awp.py` for the same sources will clean up the content inside those source sub-directories generated from previous runs and create new environments based on the current parameter file and source-receiver configurations.

To run `awp` using the MPI parallel job launcher `mpiexec` installed on my Linux box, I can type the following command in the terminal under a source sub-directory, say `FWF/100001`,

1	0.000E+00	0.000E+00	0.000E+00
2	0.000E+00	0.000E+00	0.000E+00
3	0.000E+00	0.000E+00	0.000E+00
...			
999	0.508E-12	-0.393E-12	0.128E-12
1000	-0.121E-12	0.235E-12	0.327E-12
1001	0.521E-12	-0.421E-12	0.202E-13

Listing 2.15 Snippets of the CHKJ96PV file generated by `awp` during the example run.

```
~/tools/mpi/bin/mpirun -n 8 -f ~/tools/mpi/share/my
hosts./awp
```

The first option `-n` specifies the total number of MPI processes to use, the second option `-f` specifies the full path of a host file that lists the names of the machines to use and the last command-line input to `mpirun` is the full path of the parallel binary executable `awp`. The number of sub-meshes for the half-space example is 8 (Listing 2.3), therefore I set the total number of MPI processes for the `-n` option to 8. On my Linux box, the host file has a single line in it, `fresnel:8`, which is the name of the host, `fresnel`, followed by a colon and then the number of cores on that host. You can find the name of your machine by looking into the `/etc/hostname` file on your system or consulting with your system administrator. Since there is a soft link to the `awp` in the `bin` directory under the current directory, I can use this link for the last command-line input to `mpirun` instead. By default, `awp` will read the `IN3D` file under the current directory.

It takes about 2 min of wall time for the run to complete on my Linux box. After the run finishes successfully, there should be two text files named `CHKJ96PV` and `CHKP96PV` under the sub-directory `output_ckp`. These two files contain basic information for quickly evaluating results of the run. The file `CHKJ96PV` contains the time history of the particle velocity at the local grid index $(nd+2, nd+2, nd+2)$ of the sub-mesh $(0, 0, 0)$, where `nd` is the number of grid points used in the PML boundary (key `BOUNDARY NODE NUMBER`, Line 16, Listing 2.5). This file can be used to check the stability of the PML boundary condition and also the stability of the full finite-difference scheme during the run. The first three lines and the last three lines inside `CHKJ96PV` are shown in Listing 2.15. Column 1 shows the time step, column 2–4 show the particle velocity in the x_1 , x_2 and x_3 directions of the box coordinate. If a simulation is unstable, the particle velocity listed in this file may have unrealistically large amplitudes or become NaN (i.e., Not A Number).

The content of the output file `CHKP96PV` is shown in Listing 2.16. Line 1 shows the result of the CFL stability test (Equation 2.123). If the last number on this line is larger than 0.5, the simulation is unstable and you will have NaN inside the `CHKJ96PV` file. But this situation is unlikely to happen because any of the scripts in the `py` directory that I have discussed so far can automatically adjust the time

```

1 STABILITY CRITERIA .5 > CMAX*DT/DX = 0.4875000E+00
2 # OF X NODES PER PROC 120
3 # OF Y NODES PER PROC 38
4 # OF Z NODES PER PROC 90
5 # OF TIME STEPS 1001
6 DISCRETIZATION IN SPACE 200.0000
7 DISCRETIZATION IN TIME 0.0150
8 PML COEFFICIENT 3.500
9 HIGHEST P-VELOCITY ENCOUNTERED 6500.00
10 LOWEST P-VELOCITY ENCOUNTERED 6500.00
11 HIGHEST S-VELOCITY ENCOUNTERED 3500.00
12 LOWEST S-VELOCITY ENCOUNTERED 3500.00
13 HIGHEST DENSITY ENCOUNTERED 3000.00
14 LOWEST DENSITY ENCOUNTERED 3000.00
15 ABC CONDITION, PML=1 OR CERJAN=0: 1
16 FD SCHEME, VISCO=1 OR ELASTIC=0: 0
17 Q FREQUENCY BAND, LOW, CENTER, HIGH: 0.50000E-02 0.87500 1.7500

```

Listing 2.16 Content of the CHKP96PV file generated by awp during the example run.

step length stored in key `DT` (Line 12, Listing 2.5) using Eq. 2.124 to satisfy the CFL stability condition. Line 2 to 7 show the basic information about the mesh and the discretization in space and time. The PML tuning coefficient (τ in Eq. 2.90, key `PMLCOEF` on Line 15, Listing 2.5) is shown on line 8. Line 9–14 show the highest and lowest P- and S-wave velocities and density inside the structural model. The lowest S-wave velocity and the highest P-wave velocity should be similar to those specified by key `MINIMUM VELOCITY` and `MAXIMUM VELOCITY` (Line 8–9, Listing 2.5). Line 15 is simply a flag for indicating if the PML or the absorbing boundary condition is used, which should be identical to the value associated with key `USEPML` (Line 14, Listing 2.5). Line 16 is another flag for indicating if viscoelasticity is accounted for during the simulation and should have a value identical to that associated with key `VISCO` (Line 17, Listing 2.5). Line 17 lists the low, central and high frequencies of the frequency band for the frequency-independent Q approximation (Sect. 2.1.2.4). For elastic simulations (i.e., the flag on line 16 is 0), numbers listed on line 17 do not have any effects on the simulation.

In the sub-directory `output_vlm`, there are two groups of binary output files. The group of files named as `SX96PVRDMprocID` store the spatial-temporal wavefields at the recording grid points associated with sources and receivers registered by the script `set_sr_grd.py` and the group of files named as `SX96PVprocID` store the spatial-temporal wavefields at the uniformly-distributed recording grid points registered by the script `set_ker_grd.py`. Here `procID` is the 9-digit sub-mesh (process) ID. At each recording grid point, the 9-tuple that contains the six components of the symmetric strain tensor and the three components of the particle velocity in the box coordinate

$$(\epsilon_{11}, \epsilon_{22}, \epsilon_{33}, 2\epsilon_{12}, 2\epsilon_{13}, 2\epsilon_{23}, v_1, v_2, v_3) \quad (2.125)$$

is stored at the time steps decimated according to the temporal decimation rate specified by key `NTSKP` (Line 24, Listing 2.5). Each number is stored as a four-byte-long binary floating-point number. For the recording grid file `HSGRDRECRDM001001000` shown in Listing 2.14, which has 45 recording grid points, the corresponding binary wavefield output file `SX96PVRDM001001000`, which contains the wavefields on these 45 recording grid points at 500 time steps, should be 810,000-byte long³⁶.

The three off-diagonal components of the symmetric strain tensor defined in Eq. 2.5 are multiplied by 2 in Eq. 2.125. The three off-diagonal components $2\epsilon_{12}$, $2\epsilon_{13}$, $2\epsilon_{23}$ are sometimes called the *engineering shear strain* and they represent the changes in the angles of two originally orthogonal lines before and after the deformation. The strain and velocity fields, as ordered in Eq. 2.125, will be used in constructing the Fréchet kernels in Chap. 4.

For the half-space example, wave-propagation simulations for the source sub-directories `IN.RC01.1`, `IN.RC01.2` and `IN.RC01.3` under the `FWF` directory can be carried out using the same procedure as that for the source sub-directory `100001`. Synthetic wavefields generated by `awp` inside these three source directories are examples of receiver-side Green's tensors (RGTs) since the three point body-force sources are applied at the location of the receiver `IN.RC01` and the directions of the three body-force vectors are orthogonal to each other. The RGTs are low-pass filtered by the Butterworth filter (Fig. 2.11) that is used as the source-time function for these three orthogonal point body-force sources.

2.2.7 Synthetic Seismograms

Two scripts in the `py` directory, `fwf2syn.py` and `rgt2syn.py`, are provided for extracting synthetic seismograms from the binary output files generated by `awp`. The script `fwf2syn.py` is merely a wrapper around the binary executables `rdstat` and `rotate_syn` in the `bin` directory. It reads the binary wavefield output files generated by `awp` at specified receiver grid locations by calling `rdstat` and then rotates the two horizontal components of the particle-velocity vector into the East-North and Radial-Transverse directions by calling `rotate_syn`. For the half-space example, I can use the following command in the terminal

```
fwf2syn.py param_HS.dat 2 2 1 1
```

³⁶ 9 components \times 45 grid points \times 500 time steps \times 4 bytes = 810000 bytes. You can verify the exact byte size of the file by typing the command `ls -l SX96PVRDM001001000` in the terminal.

The first command-line input is the parameter file. The second and third command-line inputs specify the line-number range inside the source-receiver file `srlst` (Line 10, Listing 2.5; Listing 2.6) for sources that `awp` has generated synthetic wavefields for. The fourth and fifth command-line inputs specify the line-number range inside `srlst` for receivers at which we would like to extract synthetic seismograms. I would like to extract synthetic seismograms for the source 100001 (Line 2, Listing 2.6) at the receiver `IN.RC01` (Line 1, Listing 2.6), therefore the second and third command-line inputs are both 2 and the fourth and fifth command-line inputs are both 1. The script generates a directory `SYN` under the project work directory (Line 3, Listing 2.5) and a source sub-directory 100001 under `SYN`. Inside the source sub-directory 100001, there are four text files, `rdstat.in`, `IN.RC01.rtin`, `IN.RC01` and `IN.RC01.rot`. The text files `rdstat.in` and `IN.RC01.rtin` are input files for the binary executables `rdstat` and `rotate_syn`, respectively. The script `fwf2syn.py` has generated them automatically and used them when calling the two binary executables. The text file `IN.RC01` contains the time histories of the six components of the strain (column 1–6) and the three components of the particle velocity (column 7–9) in the box coordinate at the receiver `IN.RC01`. The ordering of the columns in this file follows that shown in Eq. 2.125 and the number of time samples (i.e., the number of lines in the file) is identical to the number of time steps decimated according the decimation rate specified in key `NTSKP` (Line 24, Listing 2.5). The text file `IN.RC01.rot` has 7 columns and contains the time histories of the particle velocity vector in the x_1 , x_2 and x_3 directions of the box coordinate (column 1–3), the East and North directions (column 4–5) and the Radial and Transverse directions (column 6–7). The number of time samples has been decimated in the same way as in the strain-particle-velocity output file `IN.RC01`. If the receiver and the point source are aligned in the vertical direction (e.g., the receiver is located at exactly the same longitude and latitude as the point source), the Radial and Transverse components are undefined and will not be computed. If the source has finite spatial dimensions (i.e., not a point source), the Radial and Transverse components are not computed either.

The script `rgt2syn.py` is a wrapper of three binary executables in `bin`, `rdstat`, `rgt2syn` and `rotate_syn`. It calculates synthetic seismograms from RGTs by applying the reciprocity principle, which will be discussed in detail in Sects. 3.3.2, 3.4.2 and 3.4.3. For the half-space example, I can run this script using the following command in the terminal

```
rgt2syn.py param_HS.dat 3 5 2 2
```

The first command-line input is the parameter file. The second and third command-line inputs specify the line-number range in the source-receiver file `srlst` (Line 10, Listing 2.5; Listing 2.6) for the unit body-force sources applied at receiver locations. The fourth and fifth command-line inputs specify the line-number range in `srlst` for the sources that I am computing synthetic

seismograms for. In this case, the three orthogonal unit body-force sources acting at the receiver `IN.RC01` are on line 3–5 in `srlst` and I am computing synthetic seismograms for the source 100001, which is on line 2 in `srlst`, therefore the second and third command-line inputs are three and five, respectively, and the fourth and fifth command-line inputs are both 2. The script will generate three source sub-directories under `SYN`, `IN.RC01.1`, `IN.RC01.2` and `IN.RC01.3` for the three body-force sources. Under each source sub-directory, there are two text input files, `rdstat.in` and `rgt2syn_100001.in`, which were generated by `rgt2syn.py` automatically and used as input files when `rgt2syn.py` calls the binary executables `rdstat` and `rgt2syn`, respectively. There are two text output files under each source sub-directory, `100001` and `100001.syn`. The text file `100001` contains the time histories of the strain and particle velocity components in the box coordinate at the grid location of the source 100001. The text files `100001.syn` in the source sub-directories `IN.RC01.1`, `IN.RC01.2` and `IN.RC01.3` contain the particle-velocity synthetic seismograms generated by the source 100001 and recorded at the receiver `IN.RC01` in the x_1 , x_2 and x_3 directions of the box coordinate, respectively. If the synthetic seismograms for the two horizontal components in the box coordinate (i.e., the `100001.syn` file in source sub-directories `IN.RC01.1` and `IN.RC01.2` under `SYN`) are available, such as in the example here, `rgt2syn.py` will attempt to rotate them into the East-North and Radial-Transverse directions. In such a case, `rgt2syn.py` will generate three more text files in the source sub-directory `100001` under `SYN`, `IN.RC01.rgt.box`, `IN.RC01.rgt.rtin` and `IN.RC01.rgt.rot`. The file `IN.RC01.rgt.box` has three columns, which are identical to the `100001.syn` file in the three source sub-directories `IN.RC01.1`, `IN.RC01.2` and `IN.RC01.3`, respectively. The `IN.RC01.3` sub-directory is allowed to be absent for the rotation operation, in which case `IN.RC01.rgt.box` will have only two columns corresponding to the two horizontal components of the particle-velocity in the box coordinate. The text file `IN.RC01.rgt.rtin` is used as the input to the binary executable `rgt2syn`. The text output file `IN.RC01.rgt.rot` has the same format as the output file `IN.RC01.rot` generated by `fwf2syn.py`. If the vertical component of the particle-velocity synthetic seismogram is missing (i.e., the source sub-directory `IN.RC01.3` does not exist), the file `IN.RC01.rgt.rot` will have 6 columns instead of 7, with the components in the x_1 and x_2 directions of the box coordinate (column 1–2) followed by those in the East and North directions (column 3–4) and those in the Radial and Transverse directions (column 5–6). If any of the two horizontal components of the particle velocity in the box coordinate is missing, `rgt2syn.py` will not attempt to rotate the horizontal-component synthetic seismograms.

The synthetic seismograms in the output file `IN.RC01.rgt.rot` and those in the output file `IN.RC01.rot` in the source sub-directory `100001` under `SYN` should satisfy reciprocity (Sect. 3.4.3). However, I note that the source-time function of the body-force sources (Last Column, Line 3–5, Listing 2.6; Fig. 2.11) used for computing the RGTs is different from the source-time function of the point moment source 100001 (Last Column, Line 2, Listing 2.6; Fig. 2.7), therefore the synthetics in `IN.RC01.rgt.rot`, which were computed from the RGTs by applying

the reciprocity principle, are different from those in `IN.RC01.rot`, which were extracted from the forward wavefield generated by the source `100001` and recorded at the receiver `IN.RC01`. If I convolve the synthetics in `IN.RC01.rot` with the source-time function of the body-force sources (Fig. 2.11), which is equivalent to filtering those synthetics using the low-pass Butterworth filter, and convolve the synthetics in `IN.RC01.rgt.rot` with the source-time function of the point moment source `100001` (Fig. 2.7), the two sets of synthetics after the convolution operations should be identical to each other, which is indeed the case, as shown in Fig. 2.12. The theoretical backgrounds of this numerical experiment will be discussed in Sects. 3.3.2, 3.4.2 and 3.4.3.

For the half-space example, both the source and the receiver lie at 24-km depth beneath the free-surface. Because the source-receiver path is parallel to the x_1 axis of the box coordinate and the Radial direction points from the source toward the receiver (Fig. 2.10), the Radial-component synthetic seismogram equals the x_1 -component synthetic seismogram multiplying -1 (Fig. 2.12). The Transverse direction is parallel to the x_2 axis of the box coordinate and is 90° clockwise from the Radial direction (Fig. 2.10), therefore the Transverse-component synthetic seismogram is identical to the x_2 -component synthetic seismogram (Fig. 2.12). Because the point moment source `100001` has an explosive source mechanism (Column 8–13, Line 2, Listing 2.6), the direct P-wave arriving at about 5 s on the Radial component has positive polarity (Fig. 2.12). The wave arriving at about 9 s on the Radial component is the pP-wave (Fig. 2.12), which is the P-wave reflected once at the free surface. The wave arriving at about 12.5 s on the Radial component is the pS-wave (Fig. 2.12), which is the P-to-S converted wave caused by the free surface. I will compute the sensitivity (Fréchet) kernels of these three waves on the x_1 component in Chap. 4.

2.2.8 Half-Space Example Summary

The complete sequence of the terminal commands I have used so far for the half-space example can be summarized into a shell script, `hs_syn.csh`, which is shown in Listing 2.17. This script assumes that the `bin` and the `py` directories under `F3DWI` have been added to the search path (Footnote 23). The two shell variables `MPIEXEC` and `HOSTS` defined on line 2–3 in Listing 2.17 store the full paths to the MPI parallel job launcher and the host file on my Linux box (Sect. 2.2.6). You may need to change these two shell variables based on your system configuration. The input file for the mesh generator `FD_GRID_XYZ.in` (Listing 2.1), the C++ source code for generating the structural model files `vm_hs_gen.cpp` (Listing 2.4) and the parameter file `param_HS.dat` (Listing 2.5) are all located under the project `JOBS` directory. The source-receiver file `srlst` (Listing 2.6), the two source-time-function files `src_hs.dat` (Fig. 2.7; Listing 2.7) and `imp_hs.dat` (Fig. 2.11) and the adjoint-source-field file `100001.adj` (Listing 2.10) used in `srlst` are all stored inside the `srinfo` sub-directory under the project `JOBS` directory (Line 10, Listing 2.5).

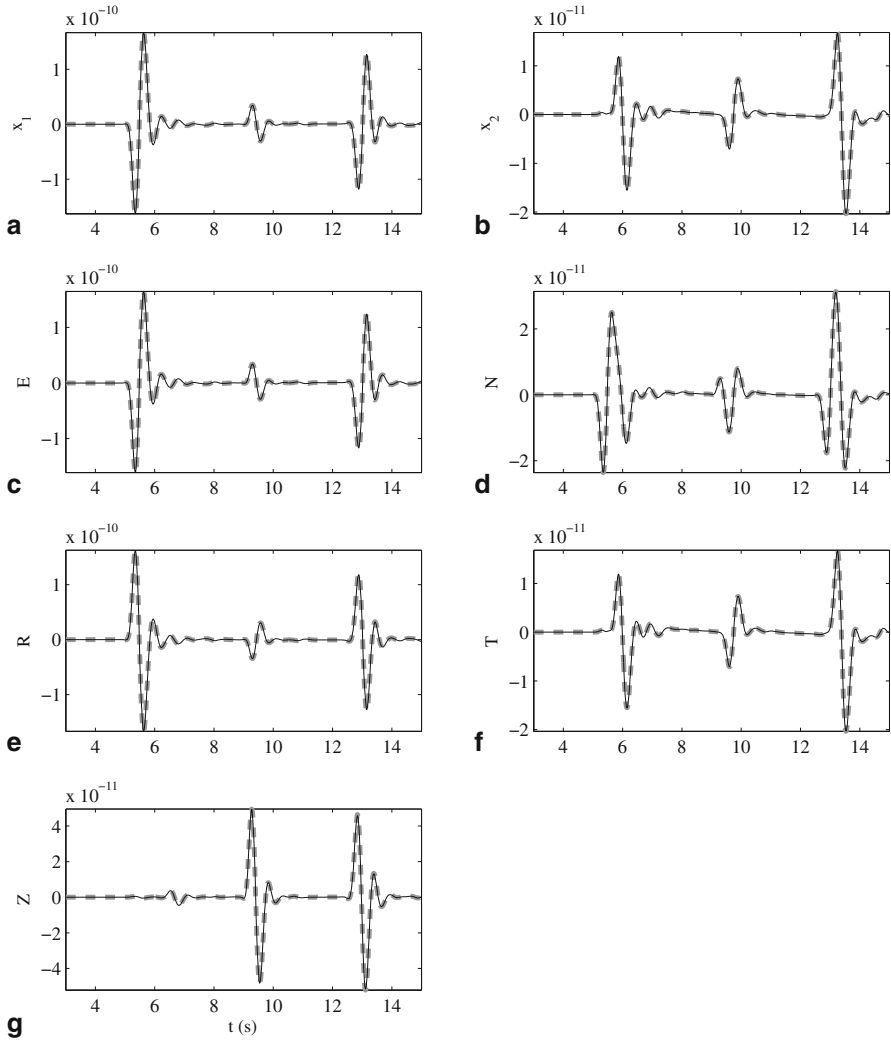


Fig. 2.12 Verifying the reciprocity principle for the half-space example. *Thin black solid lines:* synthetics in `IN.RC01.rgt.rot` after the convolution with the source-time function shown in Fig. 2.7; *thick gray dash-lines:* synthetics in `IN.RC01.rot` after the convolution with the source-time function shown in Fig. 2.11; *horizontal axis:* time in seconds; *vertical axis:* particle velocity in m/s. From left to right, top to bottom, the sub-plots (a)–(g) show synthetics in the x_1, x_2 directions of the box coordinate, the East and North directions, the Radial and Transverse directions and the vertical direction.

```

1  #!/bin/csh
2  set MPIEXEC=/home/pochen/tools/mpi/bin/mpixec
3  set HOSTS=/home/pochen/tools/mpi/share/myhosts
4
5  echo "Changing to the JOBS directory..."
6  cd /home/pochen/wrk/F3DWI/HS/JOBS
7
8  echo "Generating the mesh..."
9  mkdir ~/wrk/F3DWI/HS/VM
10 FD_GRID_XYZ_rect FD_GRID_XYZ.in
11
12 echo "Generating the structure model files..."
13 g++ -o vm_hs_gen vm_hs_gen.cpp
14 vm_hs_gen
15
16 echo "Processing the source-receiver file..."
17 proc_srlst.py param_HS.dat 1 6
18
19 echo "Register recording grid points..."
20 set_sr_grd.py param_HS.dat
21 set_ker_grd.py param_HS.dat
22
23 echo "Set up the running environment for awp..."
24 setup_awp.py param_HS.dat FWF 2 5
25
26 echo "Running awp for 100001, IN.RC01.1, IN.RC01.2 and IN.RC01.3"
27 cd ../FWF/100001
28 $MPIEXEC -n 8 -f $HOSTS ./awp
29
30 cd ../IN.RC01.1
31 $MPIEXEC -n 8 -f $HOSTS ./awp
32
33 cd ../IN.RC01.2
34 $MPIEXEC -n 8 -f $HOSTS ./awp
35
36 cd ../IN.RC01.3
37 $MPIEXEC -n 8 -f $HOSTS ./awp
38
39 cd ../../JOBS
40
41 echo "Getting synthetics from the wavefield of 100001 at IN.RC01"
42 fwf2syn.py param_HS.dat 2 2 1 1
43
44 echo "Computing synthetics from the RGTs of IN.RC01.? for 100001"
45 rgt2syn.py param_HS.dat 3 5 2 2

```

Listing 2.17 Content of the shell script that summarizes the half-space example.

On line 6 in Listing 2.17, I change to the project JOBS directory. Sequential binary executables and scripts in the following are executed from inside the JOBS directory. On line 9, I generate a new directory VM under the project work directory for storing the mesh output file XYZGRD, which will be generated by the binary executable FD_GRID_XYZ_rec on line 10 (Sect. 2.2.1), and also the structural model files, which will be generated by the binary executable vm_hs_gen on line 14 (Sect. 2.2.2). The full path of the VM directory should be consistent with that on the first line of FD_GRID_XYZ.in (Listing 2.1) and key XYZGRD and MEDIA in the parameter file (Lines 5 and 7, Listing 2.5). I process the first six lines in the source-receiver file srlst using the script proc_srlst.py on line 17 (Sect. 2.2.4.2), register the recording grid points associated with the sources and the receivers using the script set_sr_grd.py on line 20 and register the uniformly distributed recording grid points using the script set_ker_grd.py on line 21 (Sect. 2.2.5). I set up the correct awp running environment for the point moment source 100001 and the three orthogonal body-force sources at the receiver IN.RC01 using the script setup_awp.py on line 24 (Sect. 2.2.6). On line 27–37, I run awp from inside each source sub-directory under FWF (Sect. 2.2.6). On line 39, I go back to the project JOBS directory. Particle-velocity synthetic seismograms are extracted from the forward wavefield of the point moment source 100001 at the receiver IN.RC01 and rotated to the East-North and Radial-Transverse directions using the script fwf2syn.py on line 42 (Sect. 2.2.7). The synthetic seismograms are written into the text file IN.RC01.rot inside the source sub-directory 100001, which is located inside the SYN directory under the project work directory (Line 3, Listing 2.5). On line 45, I compute the same set of particle-velocity synthetic seismograms of the point moment source 100001 from the RGTs of the receiver IN.RC01 by applying the reciprocity principle using the script rgt2syn.py (Sect. 2.2.7). The synthetic seismograms are written into the text file IN.RC01.rgt.rot under the same directory as the text file IN.RC01.rot. Note that synthetics in IN.RC01.rot are still different from those in IN.RC01.rgt.rot due to differences in the source-time functions in the two files src_hs.dat and imp_hs.dat. After I convolve the synthetics in IN.RC01.rot with the source-time function in imp_hs.dat and convolve the synthetics in IN.RC01.rgt.rot with the source-time function in src_hs.dat, the two sets of synthetic seismograms are identical as shown in Fig. 2.12. The MATLAB script reciprocity_hs.m for computing the convolutions and generating Fig. 2.12 is included in the mfiles sub-directory in the F3DWI package. Theoretical background for reciprocity and the RGT calculations are explained in Sects. 3.3.2 and 3.4.2.

Because the inputs and outputs of one command in Listing 2.17 are closely tied with those of the preceding and the following commands, if you change the inputs to one command at some point, you may need to re-execute all the following commands in order for the changes to take effect through all the calculations.

2.3 Application

Southern California is one region where seismic hazard is high (e.g., Jackson 1996 Field et al. 2009) and the subsurface seismic structure is complex. Computer simulations based on realistic seismic structural models have shown that three-dimensional structural heterogeneities, in particular, sedimentary basins, can have significant effects on the subsurface propagation of seismic waves and the amplification of ground motions (e.g., Gao et al. 1996; Graves et al. 1998; Davis et al. 2000). The effects of realistic 3D seismic structural models can be incorporated into probabilistic seismic hazard estimations and ground-motion forecasts in an accurate and computationally efficient way using physics-based seismic hazard analysis tools, such as the CyberShake computational platform (Graves et al. 2011) provided by the Southern California Earthquake Center (SCEC). An important challenge now lies in improving the accuracy of the seismic structural models used in those physics-based seismic hazard analysis tools.

2.3.1 SCEC Community Velocity Models (CVMs)

In the past twenty years, researchers in SCEC created two 3D seismic structural models for Southern California based on geological and geophysical observations. These two models are named Community Velocity Model-SCEC (CVM-S) and Community Velocity Model-Harvard (CVM-H). CVM-S was constructed by embedding rule-based seismic velocity models of major basins in Southern California (Magistrale et al. 1996) within a background seismic velocity model determined from regional seismic travel-time tomography (Hauksson 2000; Magistrale et al. 2000). It was later improved by adding geotechnical layers on top of the basins, a laterally varying Moho depth determined from receiver function studies (Zhu and Kanamori 2000), an upper-mantle seismic velocity model (Kohler et al. 2003) and models for the Salton Trough region and the San Bernardino Valley region. The latest version of this model is referred to as CVM-S4.

CVM-H was also constructed by embedding high-resolution basin structural models within the regional seismic travel-time tomography model of (Hauksson 2000). In CVM-H, the basin models were mainly determined using a large number of sonic logs and seismic reflection profiles from the oil industry (Süss and Shaw 2003). It includes a geotechnical layer based on the Vs30 map of (Ely et al. 2010), the Moho interface of (Yan and Clayton 2007) and an upper-mantle structural model determined through finite-frequency teleseismic surface wave tomography (Prindle and Tanimoto 2006). Most notably, the crustal structure in CVM-H was improved through 16 iterations of full-3D tomography based on the adjoint-wavefield method (F3DT-AW) (Tape et al. 2009, 2010). The latest official release was in November 2011, named CVM-H11.9.

Studies in the Los Angeles Basin region (Chen et al. 2007) showed that synthetic seismograms computed using either CVM-S or CVM-H provided substantially better fit to observed seismograms than synthetics computed using either a laterally homogeneous 1D seismic velocity model (Hadley and Kanamori 1977) or a set of path-averaged 1D models. In the Los Angeles Basin region, CVM-H provided a slightly better fit to the observed P-waves than CVM-S, while CVM-S provided a slightly better fit to the observed S-waves than CVM-H. Full-3D tomography (F3DT) was applied to improve the crustal structure of CVM-S in the Los Angeles Basin region in (Chen et al. 2007). The improved model was named LAF3D. Synthetic seismograms computed using LAF3D provided substantially better fit to observed waveforms at frequencies up to 1.2 Hz than those computed using the 3D starting model CVM-S. The inversion extended the area of the Los Angeles Basin and increased the velocities inside the basin, which brought the basin structure more into agreement with that in CVM-H.

CVM-S4.26 is the latest official release of the CVM-S series. It includes improvements made through 26 iterations of our F3DT inversion procedure. Our F3DT used CVM-S4 as the starting model and successively improved the fit to data sets that eventually included more than 550,000 waveform misfit measurements at frequencies up to 0.2 Hz, obtained from more than 38,000 earthquake seismograms and 12,000 ambient-noise Green's functions. The size of our final data set is comparable to the largest used in global tomographic studies (e.g., Kustowski et al. 2008). Among the 26 iterations, 18 iterations were based on the adjoint-wavefield method (F3DT-AW) and 8 iterations were based on the scattering-integral method (F3DT-SI). The details of these two methods will be explained in later chapters. In each iteration, synthetic seismograms for the updated structural model were calculated using *awp* (Sect. 2.2) and misfit measurements were made between the observed waveforms and the synthetics, accounting for the nonlinearity of the inversion. High structural resolution was obtained by using misfit measurements that can capture subtle differences in the waveforms. The result is CVM-S4.26, a 3D seismic structural model with unprecedented resolution of crustal features throughout Southern California. Synthetic seismograms computed using CVM-S4.26 show substantially better fit to low-frequency (≤ 0.2 Hz) observed seismograms than those computed using our starting model CVM-S4 and those computed using CVM-H11.9. Quantitative results have shown that CVM-S4.26 is the most accurate model in predicting low-frequency (≤ 0.2 Hz) observed seismograms among the three SCEC CVMs. Details of CVM-S4.26 and our F3DT inversion procedure have been documented in (Lee et al. 2014a, b) and will be discussed in Chap. 6.

2.3.2 Earthquake Source Models

We compared the low-frequency seismograms observed in two recent earthquakes in the Los Angeles area, 17 Mar 2014 Encino ($M_w 4.4$) and 29 Mar 2014 La Habra ($M_w 5.1$), with the synthetic seismograms computed using CVM-S4.26, CVM-S4

Table 2.4 Our revised CMT solutions for the Encino earthquake and the La Habra earthquake. The centroid location is specified by the longitude and the latitude in degrees and the depth in km. The source mechanism is specified by the strike, dip and rake angles in degrees

Event	Centroid time	Centroid location	M_w	Strike/Dip/Rake
Encino	03/17/2014 13:25:36.85	-118.492(°), 34.139(°), 9.0(km)	4.42	101/60/27
La Habra	03/29/2014 04:09:42.97	-117.930(°), 33.922(°), 5.0(km)	5.12	134/55/155

and CVM-H11.9. Note that because the seismograms from these two recent earthquakes were not used to derive the three CVMs, they provide prospective tests of the three CVMs' capabilities in predicting low-frequency observed waveforms.

Both the Encino earthquake and the La Habra earthquake were well recorded by three-component broadband seismic stations of the California Integrated Seismic Network (CISN). We used more than 900 seismograms from both earthquakes for our waveform prediction test. CISN routinely determines focal mechanisms from first motions and the centroid-moment tensors (CMTs) using synthetic Green's functions computed in the 1D crustal velocity model of (Dreger and Helmberger 1993). The CISN source-mechanism inversions were discussed in (e.g., Clinton et al. 2006 Hutton et al. 2010).

We revised the CISN CMT solutions for the two earthquakes by applying our fast reciprocity-based inversion algorithm documented in (Lee et al. 2011). Table 2.4 shows our revised CMT solutions for the two earthquakes. Details about our CMT inversion algorithm will be discussed in later chapters. Our revised CMT solutions for both earthquakes improved the waveform fits for all three CVMs. For the Encino earthquake, our centroid location is about 0.9 km shallower than the CISN hypocenter and 1 km deeper than the CISN CMT location. Our centroid time is about the same as the CISN origin time. For the La Habra earthquake, our centroid location is at the same depth as the CISN CMT solution, which is about 2.5 km shallower than the CISN hypocenter. Our centroid time is nearly 1 s later than the CISN origin time, which might be due to the finite source duration of this larger event. Our revised A&R focal mechanisms for both earthquakes (Footnote 32) are generally consistent with those determined by CISN.

2.3.3 Validation of SCEC CVMs

We computed synthetic seismograms for the Encino earthquake and the La Habra earthquake using CVM-S4.26, CVM-S4 and CVM-H11.9. The procedure is similar to that used in our half-space example (Sect. 2.2). Our modeling volume (Fig. 2.13), which is about 768-km long, 496-km wide and 50-km thick, is discretized using a uniform mesh with 500-m grid spacing and about 152 million grid points. The full mesh is partitioned into 8192 sub-meshes with 32, 64 and 4 sub-meshes in the x_1 , x_2 and x_3 directions of the box coordinate, respectively.

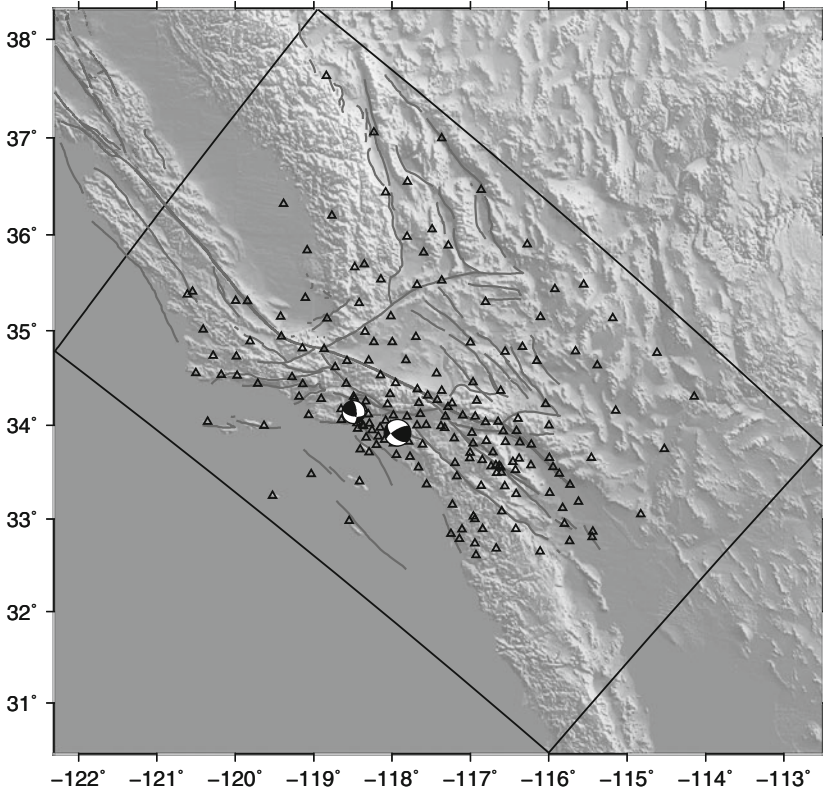


Fig. 2.13 *Black box*: our modeling domain; *black triangles*: three-component broadband stations used in our validation study; *beach balls*: source mechanisms for the Encino earthquake (*smaller beach ball*) and the La Habra earthquake (*larger beach ball*); *gray solid lines*: major faults; *gray-scale background*: topography.

Material properties were assigned to each grid point using the corresponding values in one of the CVMs. The minimum S-wave velocity is clamped to about 900 m/s in the three CVMs. Therefore our mesh gives about 9 grid points per minimum wavelength at our highest target frequency of 0.2 Hz. For the fourth-order scheme used in *awp*, 4–5 grid points per minimum wavelength are usually sufficient for modeling body waves accurately and 8–10 grid points per minimum wavelength are usually needed for modeling surface waves accurately (Sect. 2.1.4). The maximum P-wave velocity in our modeling volume is about 8200 m/s in the three CVMs. Considering the CFL stability condition (Eq. 2.123), we used a time-step length of 0.03 s in our simulations. The length of the synthetic seismogram is 180 s and the total number of time steps is 6000. The source models were our revised CMT solutions shown in Table 2.4. We used a sixth-order Butterworth band-pass filter with corners at 0.02 and 0.2 Hz as our filter for both earthquakes. On the IBM Blue Gene/Q system at the Argonne Leadership Computing Facility, each

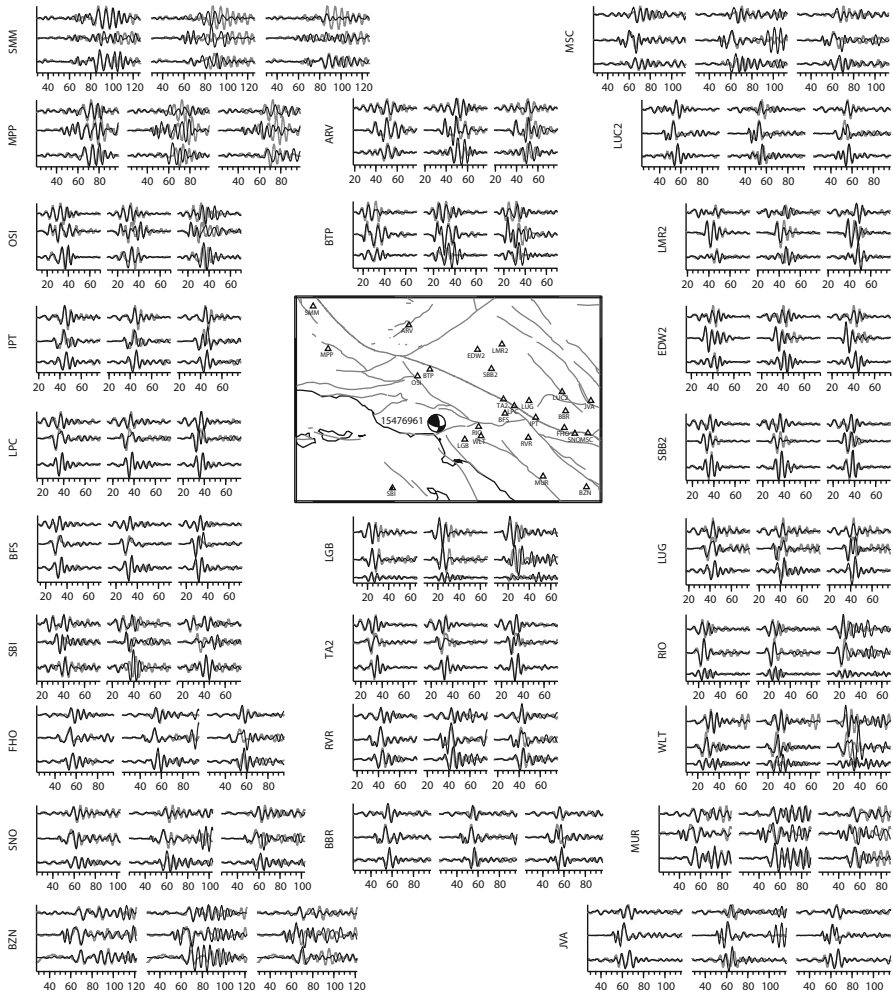


Fig. 2.14 Examples of observed and synthetic seismograms for the Encino earthquake. *Black solid lines*: synthetic seismograms; *gray solid lines*: observed seismograms. In the sub-plot for each station, *left column*: CVM-S4.26 synthetics; *center column*: CVM-S4 synthetics; *right column*: CVM-H11.9 synthetics; *top row*: radial components; *center row*: transverse components; *bottom row*: vertical components. In the map, *beach ball*: source mechanism of the earthquake at its epicenter; *black triangles*: stations whose seismograms are shown; *gray and black solid lines*: faults and coast lines.

wave-propagation simulation using `awp` took about 5 min of wall time on 8192 cores.

Observed velocity seismograms of the two earthquakes were downloaded from the Southern California Earthquake Data Center (SCEDC). Instrument responses were removed. Horizontal-component seismograms were rotated into the radial and

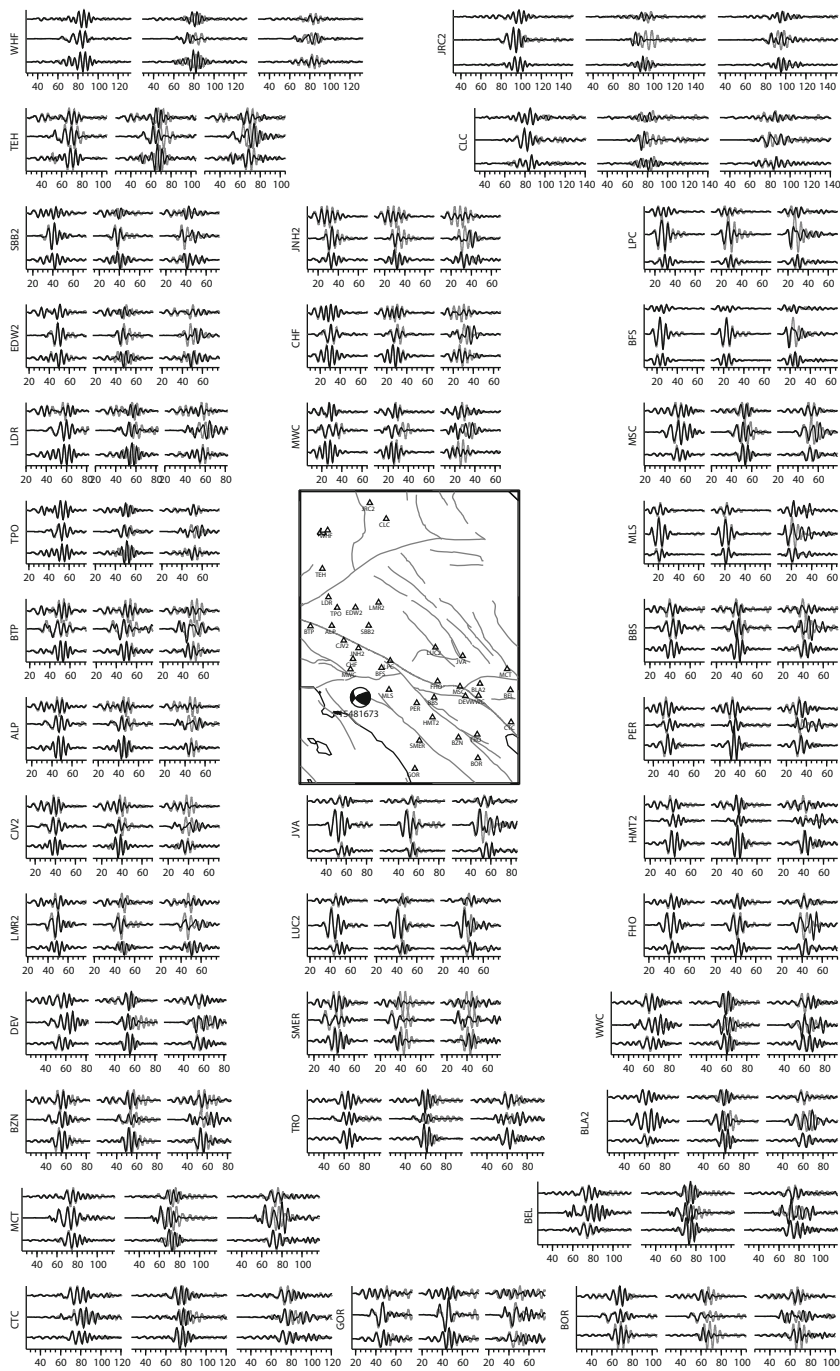


Fig. 2.15 Examples of observed and synthetic seismograms for the La Habra earthquake. The format is identical to that in Fig. 2.14.

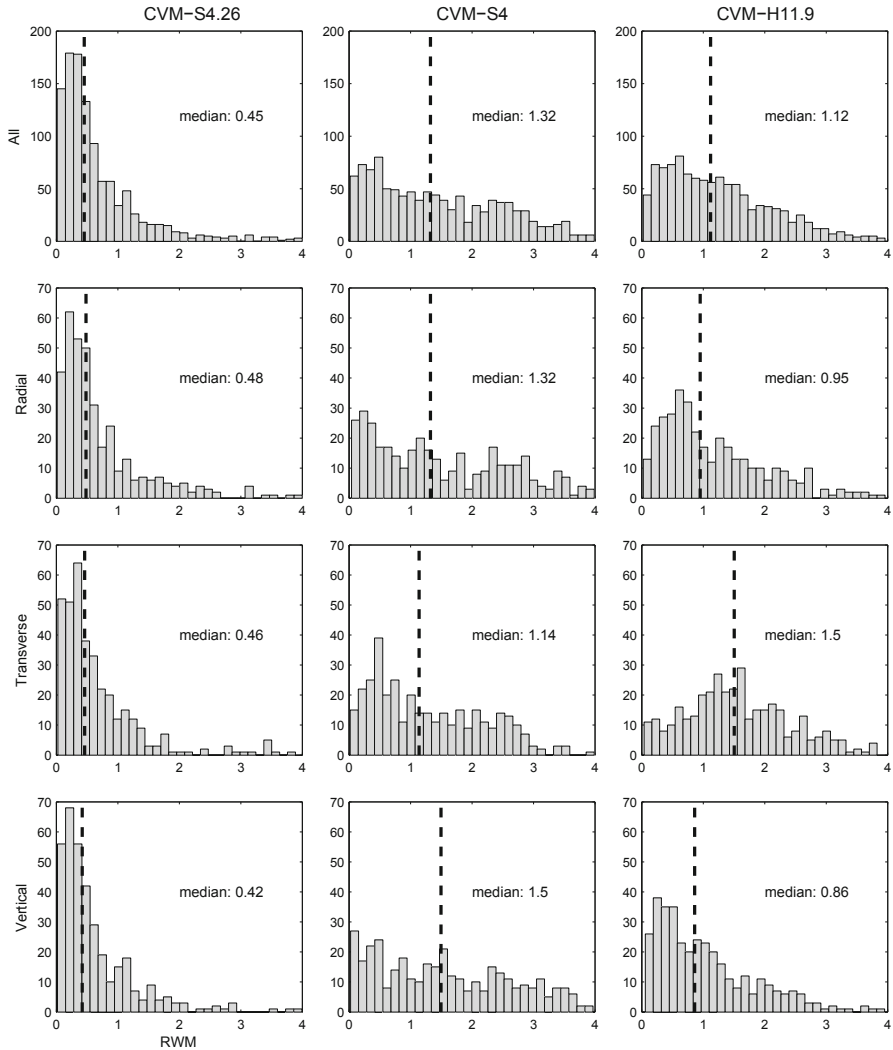


Fig. 2.16 Histograms of the RWMs for all synthetics (*top row*), radial-component synthetics (*second row from top*), transverse-component synthetics (*third row from top*) and vertical-component synthetics (*bottom row*) computed using CVM-S4.26 (*left column*), CVM-S4 (*center column*) and CVM-H11.9 (*right column*). Vertical black dash lines: locations of the RWM median values (*mRWMs*).

the transverse components. The mean and the linear trend of each seismogram were removed and a Hanning taper was then applied. We then filtered each observed seismogram using the same Butterworth filter that is used for the synthetics. Examples of the observed seismograms and the corresponding synthetic seismograms computed using CVM-S4.26, CVM-S4 and CVM-H11.9 for the Encino earthquake and

the La Habra earthquake are shown in Figs. 2.14 and 2.15, respectively. In general, synthetic seismograms computed using CVM-S4.26 provided substantially better fit to observed seismograms than those computed using CVM-S4 and CVM-H11.9.

We can quantify the differences between an observed seismogram $s(t)$ and its corresponding synthetic seismogram $v(t)$ using the relative waveform misfit (*RWM*) (e.g., Zhu and Helmberger 1996 Tape et al.2010),

$$RWM = \frac{\int_{t_0}^{t_1} [s(t) - v(t)]^2 dt}{\sqrt{\int_{t_0}^{t_1} [s(t)]^2 dt \int_{t_0}^{t_1} [v(t)]^2 dt}}, \quad (2.126)$$

where the integration time window $[t_0, t_1]$ includes the segment from the first arrival to the end of the main surface wave. Histograms of the *RWMs* computed for all the seismograms of the Encino and the La Habra earthquakes are shown in Fig. 2.16. If we use the median value of the *RWMs* (*mRWM*) as a measure of the overall fit to observed seismograms, CVM-S4.26 has much smaller *mRWMs* than CVM-S4 and CVM-H11.9. On average, CVM-H11.9 has a smaller *mRWM* than CVM-S4. However, the transverse-component *mRWM* of CVM-H11.9 is substantially larger than the *mRWMs* of the other two components, which means that the fits on transverse components are substantially worse than the fits on the other two components for CVM-H11.9. In (Tape et al. 2010), this observation was attributed to significant crustal anisotropy, which is not included in CVM-H11.9. Our study does not support this explanation, as CVM-S4.26, which is also isotropic, can fit all 3 components of the observed seismograms equally well.

We can compute a station misfit by averaging the *RWMs* over all components at that station and over both earthquakes. We can also compute a component-dependent station misfit by averaging the *RWMs* of each component at that station over both earthquakes. Spatial distributions of the station misfits of the three CVMs are shown in Fig. 2.17. Compared with CVM-S4 and CVM-H11.9, CVM-S4.26 has much better fits in the Mojave, southern Sierra Nevada and southern San Joaquin Valley, Basin and Ranges and Salton Trough. For CVM-H11.9, the fits on the transverse component are much worse than the fits on the other two components in the Mojave and Basin and Ranges.

A more detailed analysis of the seismograms from the Encino and the La Habra earthquakes and a more thorough discussion of the differences among the three CVMs, including careful comparisons with active-source refraction tomography models, have been documented in (Lee et al. 2014a).

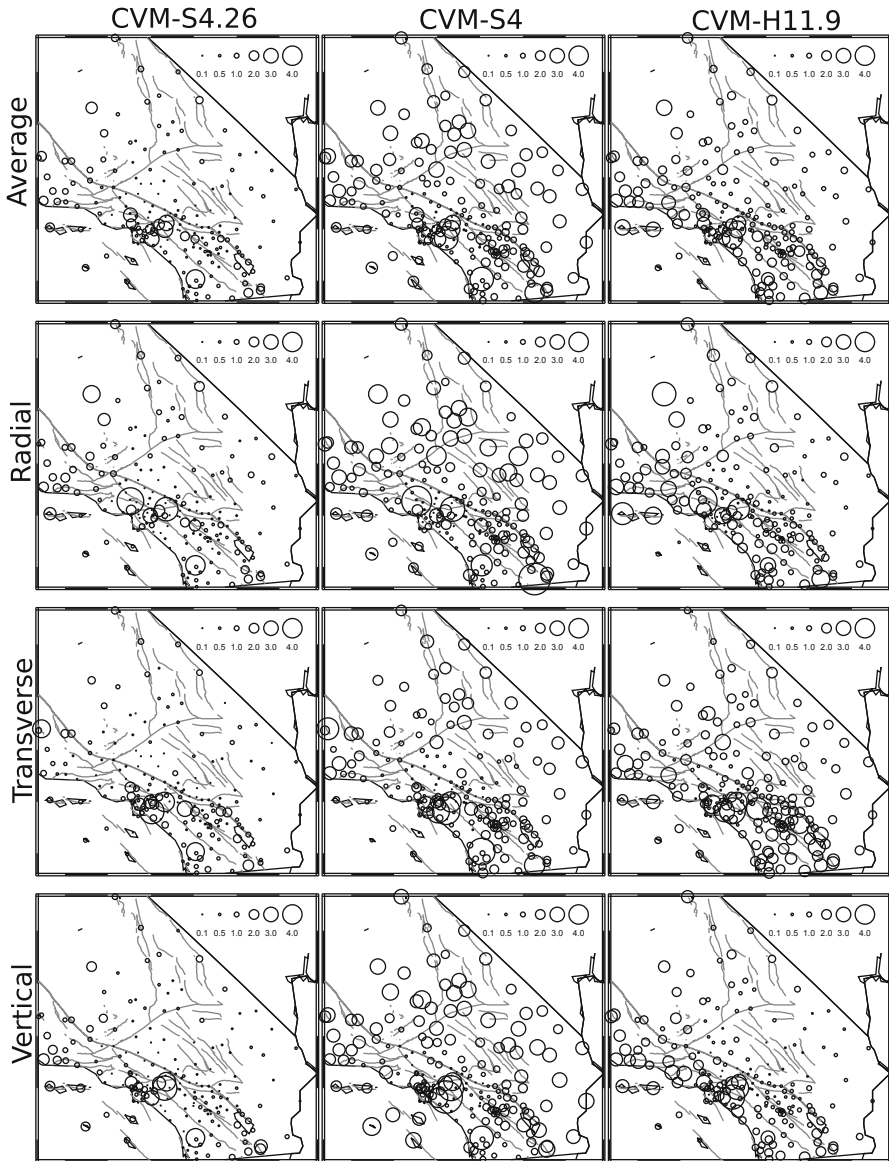


Fig. 2.17 Spatial distributions of station misfits for all synthetics (*top row*), radial-component synthetics (*second row from top*), transverse-component synthetics (*third row from top*) and vertical-component synthetics (*bottom row*) computed using CVM-S4.26 (*left column*), CVM-S4 (*center column*) and CVM-H11.9 (*right column*). Sizes of the *black circles* are proportional to the station misfits and the component-dependent station misfits and a scale is shown on the *top-right corner* of each sub-plot. The *circles* are centered at station locations.

References

- Akcelik, V., Bielak, J., Biros, G., Epanomeritakis, I., Fernandez, A., Ghattas, O., Kim, E. J., Lopez, J., O'Hallaron, D., & Tu, T., et al. (2003). *High resolution forward and inverse earthquake modeling on terascale computers*. In Supercomputing, 2003 ACM/IEEE Conference, pp. 52–52. IEEE.
- Aki, K., & Richards, P. G. (2002). *Quantitative seismology* (2nd ed.). Sausalito, California: University Science Books.
- Alterman, Z., & Karal, F. (1968). Propagation of elastic waves in layered media by finite difference methods. *Bulletin of the Seismological Society of America*, 58(1), 367–398.
- Baker, G., & Graves-Morris, P. (1996). *Padé Approximants. Encyclopedia of mathematics and its applications*. New York: Cambridge University Press.
- Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2), 185–200.
- Berenger, J.-P. (1996). Three-dimensional perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 127(2), 363–379.
- Bérenger, J.-P. (2002). Application of the CFS PML to the absorption of evanescent waves in waveguides. *IEEE Microwave and Wireless Components Letters*, 12(6), 218–220.
- Blanch, J. O., Robertsson, J. O., & Symes, W. W. (1995). Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysics*, 60(1), 176–184.
- Carcione, J. (2007). *Wave fields in real media: Wave propagation in anisotropic, anelastic, porous and electromagnetic media* (Handbook of geophysical exploration: seismic exploration). Oxford: Elsevier Science.
- Carcione, J. M., & Cavallini, F. (1994). A rheological model for anelastic anisotropic media with applications to seismic wave propagation. *Geophysical Journal International*, 119(1), 338–348.
- Cerjan, C., Kosloff, D., Kosloff, R., & Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4), 705–708.
- Certaine, J. (1960). The solution of ordinary differential equations with large time constants. In A. Ralston & H. S. Wilf (Eds.), *Mathematical Methods for Digital Computers*, 128–132. New York: Wiley.
- Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, 97(4), 1094–1120.
- Chew, W., & Liu, Q. (1996). Perfectly matched layers for elastodynamics: A new absorbing boundary condition. *Journal of Computational Acoustics*, 4(04), 341–359.
- Chew, W., Jin, J., & Michielssen, E. (1997). Complex coordinate stretching as a generalized absorbing boundary condition. *Microwave and Optical Technology Letters*, 15(6), 363–369.
- Christensen, R. (2003). *Theory of viscoelasticity* (Civil, mechanical and other engineering series). Mineola, New York: Dover Publications.
- Clinton, J. F., Hauksson, E., & Solanki, K. (2006). An evaluation of the SCSN moment tensor solutions: Robustness of the mw magnitude scale, style of faulting, and automation of the method. *Bulletin of the Seismological Society of America*, 96(5), 1689–1705.
- Collino, F., & Monk, P. B. (1998). Optimizing the perfectly matched layer. *Computer Methods in Applied Mechanics and Engineering*, 164(1), 157–171.
- Collino, F., & Tsogka, C. (2001). Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics*, 66(1), 294–307.
- Courant, R., Friedrichs, K., & Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2), 215–234.
- Coutant, O., Virieux, J., & Zollo, A. (1995). Numerical source implementation in a 2d finite difference scheme for wave propagation. *Bulletin of the Seismological Society of America*, 85(5), 1507–1512.

- Cox, S., & Matthews, P. (2002). Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2), 430–455.
- Cui, Y., Olsen, K., Chourasia, A., Moore, R., Maechling, P., & Jordan, T. (2009). The terashake computational platform for large-scale earthquake simulations. In *Advances in geocomputing* (pp. 229–277). New York: Springer.
- Davis, P. M., Rubinstein, J. L., Liu, K. H., Gao, S. S., & Knopoff, L. (2000). Northridge earthquake damage caused by geologic focusing of seismic waves. *Science*, 289(5485), 1746–1750.
- Day, S. M. (1998). Efficient simulation of constant Q using coarse-grained memory variables. *Bulletin of the Seismological Society of America*, 88(4), 1051–1062.
- Day, S. M., & Bradley, C. R. (2001). Memory-efficient simulation of anelastic wave propagation. *Bulletin of the Seismological Society of America*, 91(3), 520–531.
- Day, S. M., & Minster, J. B. (1984). Numerical simulation of attenuated wavefields using a Padé approximant method. *Geophysical Journal International*, 78(1), 105–118.
- Day, S., Bielak, J., Dreger, D., Larsen, S., Graves, R., Pitarka, A., & Olsen, K. (2001). *Test of 3d elastodynamic codes*. Final Report to Pacific Earthquake Engineering Research Center, Lifelines Program TASK 1A01, pp. 1–24.
- Day, S., Bielak, J., Dreger, D., Larsen, S., Graves, R., Pitarka, A., & Olsen, K. (2003). *Tests of 3d elastodynamic codes*. Final Report to Pacific Earthquake Engineering Research Center, Lifelines Program TASK 1A02, pp. 1–32.
- Day, S., Bielak, J., Dreger, D., Larsen, S., Graves, R., Pitarka, A., & Olsen, K. (2005). *Tests of 3d elastodynamic codes*. Final Report to Pacific Earthquake Engineering Research Center, Lifelines Program TASK 1A03, pp. 1–18.
- Dreger, D., & Helmberger, D. (1993). Determination of source parameters at regional distances with three-component sparse network data. *Journal of Geophysical Research*, 98(B5), 8107–8125.
- Drossaert, F. H., & Giannopoulos, A. (2007). Complex frequency shifted convolution PML for FDTD modelling of elastic waves. *Wave Motion*, 44(7), 593–604.
- Ely, G. P., Jordan, T., Small, P., & Maechling, P. J. (2010). A Vs30-derived near-surface seismic velocity model. In *Abstract S51A-1907, Fall Meeting*.
- Emmerich, H., & Korn, M. (1987). Incorporation of attenuation into time-domain computations of seismic wave fields. *Geophysics*, 52(9), 1252–1264.
- Evans, L. (2010). *Partial differential equations* (Graduate studies in mathematics). Washington, DC: American Mathematical Society.
- Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., & Weldon, R., et al. (2009). Uniform california earthquake rupture forecast, version 2 (ucerf 2). *Bulletin of the Seismological Society of America*, 99(4), 2053–2107.
- Gao, S., Liu, H., Davis, P., & Knopoff, L. (1996). Localized amplification of seismic waves and correlation with damage due to the Northridge earthquake: Evidence for focusing in Santa Monica. *Bulletin of the Seismological Society of America*, 86(1B), S209–S230.
- Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America*, 86(4), 1091–1106.
- Graves, R. W., & Day, S. M. (2003). Stability and accuracy analysis of coarse-grain viscoelastic simulations. *Bulletin of the Seismological Society of America*, 93(1), 283–300.
- Graves, R. W., Pitarka, A., & Somerville, P. G. (1998). Ground-motion amplification in the Santa Monica area: Effects of shallow basin-edge structure. *Bulletin of the Seismological Society of America*, 88(5), 1224–1242.
- Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., & Milner, K., et al. (2011). Cybershake: A physics-based seismic hazard model for Southern California. *Pure and Applied Geophysics*, 168(3–4), 367–381.
- Hadley, D., & Kanamori, H. (1977). Seismic structure of the transverse ranges, california. *Geological Society of America Bulletin*, 88(10), 1469–1478.

- Hastings, F. D., Schneider, J. B., & Broschat, S. L. (1996). Application of the perfectly matched layer (PML) absorbing boundary condition to elastic wave propagation. *The Journal of the Acoustical Society of America*, 100(5), 3061–3069.
- Hauksson, E. (2000). Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in Southern California. *Journal of Geophysical Research*, 105(B6), 13875–13903.
- Heath, M. (2005). *Scientific computing: An introductory survey*. New York: McGraw-Hill.
- Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake monitoring in Southern California for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, 100(2), 423–446.
- Ilan, A., & Loewenthal, D. (1976). Instability of finite difference schemes due to boundary conditions in elastic media. *Geophysical Prospecting*, 24(3), 431–453.
- Ilan, A., Ungar, A., & Alterman, Z. (1975). An improved representation of boundary conditions in finite difference schemes for seismological problems. *Geophysical Journal of the Royal Astronomical Society*, 43(3), 727–745.
- Jackson, D. D. (1996). Seismic hazards in Southern California: Probable earthquakes, 1994 to 2024. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstracts*, 33, 54A–54A. (Elsevier).
- Kelly, K., Ward, R., Treitel, S., & Alford, R. (1976). Synthetic seismograms: A finite-difference approach. *Geophysics*, 41(1), 2–27.
- Kohler, M., Magistrale, H., & Clayton, R. (2003). Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3. *Bulletin of the Seismological Society of America*, 93(2), 757–774.
- Kustowski, B., Ekström, G., & Dziewoński, A. (2008). Anisotropic shear-wave velocity structure of the earth's mantle: A global model. *Journal of Geophysical Research*, 113(B6), 1–23.
- Larsen, S., & Schultz, C. (1995). *Elas3d: 2d/3d elastic finite-difference wave propagation code. Lawrence Livermore National Laboratory, UCRLMA-121792.*
- Lee, E.-J., Chen, P., Jordan, T. H., & Wang, L. (2011). Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. *Geophysical Journal International*, 186(1), 311–330.
- Lee, E.-J., Chen, P., & Jordan, T. H. (2014a). Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes. *Seismological Research Letters*, 85(6), 311–330.
- Lee, E.-J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014b). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research*, 119(8), 6421–6451.
- Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11), 1425–1436.
- Liu, P., & Archuleta, R. J. (2006). Efficient modeling of Q for 3D numerical simulation of wave propagation. *Bulletin of the Seismological Society of America*, 96(4A), 1352–1358.
- Liu, Q.-H. & Tao, J. (1997). The perfectly matched layer for acoustic waves in absorptive media. *The Journal of the Acoustical Society of America*, 102(4), 2072–2082.
- Liu, H.-P., Anderson, D. L., & Kanamori, H. (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition. *Geophysical Journal of the Royal Astronomical Society*, 47(1), 41–58.
- Magistrale, H., McLaughlin, K., & Day, S. (1996). A geology-based 3D velocity model of the Los Angeles basin sediments. *Bulletin of the Seismological Society of America*, 86(4), 1161–1166.
- Magistrale, H., Day, S., Clayton, R. W., & Graves, R. (2000). The SCEC Southern California reference three-dimensional seismic velocity model version 2. *Bulletin of the Seismological Society of America*, 90(6B), S65–S76.
- Marcinkovich, C., & Olsen, K. (2003). On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme. *Journal of Geophysical Research*, 108(B5), ESE-18 1–16.

- Martin, R., & Komatitsch, D. (2009). An unsplit convolutional perfectly matched layer technique improved at grazing incidence for the viscoelastic wave equation. *Geophysical Journal International*, 179(1), 333–344.
- Martin, R., Komatitsch, D., & Ezziani, A. (2008a). An unsplit convolutional perfectly matched layer improved at grazing incidence for seismic wave equation in poroelastic media. *Geophysics*, 73(4), T51–T61.
- Martin, R., Komatitsch, D., & Gedney, S. D. (2008b). A variational formulation of a stabilized unsplit convolutional perfectly matched layer for the isotropic or anisotropic seismic wave equation. *Computer Modeling in Engineering and Sciences*, 37(3), 274–304.
- McDonal, F., Angona, F., Mills, R., Sengbush, R., Van Nostrand, R., & White, J. (1958). Attenuation of shear and compressional waves in pierre shale. *Geophysics*, 23(3), 421–439.
- Moczo, P. & Kristek, J. (2005). On the rheological models used for time-domain methods of seismic wave propagation. *Geophysical Research Letters*, 32(1), L01306 1–5.
- Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R., & Halada, L. (2002). 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities. *Bulletin of the Seismological Society of America*, 92, 3042–3066.
- Moczo, P., Kristek, J., Galis, M., Pazak, P., & Balazovjeh, M. (2007). The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. *Acta Physica Slovaca. Reviews and Tutorials*, 57(2), 177–406.
- Muir, F., Dellinger, J., Etgen, J., & Nichols, D. (1992). Modeling elastic fields across irregular boundaries. *Geophysics*, 57(9), 1189–1193.
- Müller, G. (1983). Rheological properties and velocity dispersion of a medium with power-law dependence of q on frequency. *Journal of Geophysics-Zeitschrift für Geophysik*, 54(1), 20–29.
- Murphy III, W. F. (1982). Effects of partial water saturation on attenuation in massilon sandstone and vycor porous glass. *The Journal of the Acoustical Society of America*, 71(6), 1458–1468.
- Ohminato, T., & Chouet, B. A. (1997). A free-surface boundary condition for including 3d topography in the finite-difference method. *Bulletin of the Seismological Society of America*, 87(2), 494–515.
- Olsen, K. B. (1994). *Simulation of three-dimensional wave propagation in the Salt Lake Basin*. PhD thesis, University of Utah, Salt Lake City, UT.
- Olsen, K. B., Archuleta, R. J., & Matarese, J. R. (1995). Three-dimensional simulation of a magnitude 7.75 earthquake. *Science*, 270, 8.
- Olsen, K., Day, S., & Bradley, C. (2003). Estimation of Q for long-period (> 2 sec) waves in the Los Angeles basin. *Bulletin of the Seismological Society of America*, 93(2), 627–638.
- Olsen, K., Day, S., Minster, J., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P., & Jordan, T. (2006). Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophysical Research Letters*, 33(7), L07305 1–4.
- Pitarka, A. (1999). 3D elastic finite-difference modeling of seismic motion using staggered grids with nonuniform spacing. *Bulletin of the Seismological Society of America*, 89(1), 54–68.
- Pope, D. A. (1963). An exponential method of numerical integration of ordinary differential equations. *Communications of the ACM*, 6(8), 491–493.
- Prindle, K., & Tanimoto, T. (2006). Teleseismic surface wave study for S-wave velocity structure under an array: Southern California. *Geophysical Journal International*, 166(2), 601–621.
- Roden, J. A., & Gedney, S. D. (2000). Convolutional PML (CPML): An efficient FDTD implementation of the CFS-PML for arbitrary media. *Microwave and Optical Technology Letters*, 27(5), 334–338.
- Spencer, J. W. (1981). Stress relaxations at low frequencies in fluid-saturated rocks: Attenuation and modulus dispersion. *Journal of Geophysical Research*, 86(B3), 1803–1812.
- Süss, M. P., & Shaw, J. H. (2003). P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California. *Journal of Geophysical Research*, 108(B3), ESE-13 1–18.

- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the Southern California crust. *Science*, 325(5943), 988–992.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the Southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180(1), 433–462.
- Teixeira, F. & Chew, W. (1999). On causality and dynamic stability of perfectly matched layers for FDTD simulations. *IEEE Transactions Microwave Theory and Techniques*, 47(6), 775–785.
- Virieux, J. (1984). SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 49, 1933–1942.
- Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51, 889–901.
- Wessel, P., Smith, W. H., Scharroo, R., Luis, J., & Wobbe, F. (2013). Generic mapping tools: Improved version released. *Eos, Transactions American Geophysical Union*, 94(45), 409–410.
- Wrenger, J.-P. (2002). Numerical reflection from FDTD-PMLs: A comparison of the split PML with the unsplit and CFS PMLs. *IEEE Transactions on Antennas Propagation*, 50(3), 258–265.
- Yan, Z., & Clayton, R. (2007). Regional mapping of the crustal structure in Southern California from receiver functions. *Journal of Geophysical Research*, 112(B5), B05311 1–20.
- Yomogida, K., & Etgen, J. T. (1993). 3-D wave propagation in the Los Angeles basin for the Whittier-Narrows earthquake. *Bulletin of the Seismological Society of America*, 83(5), 1325–1344.
- Zahradník, J., & Priolo, E. (1995). Heterogeneous formulations of elastodynamic equations and finite-difference schemes. *Geophysical Journal International*, 120(3), 663–676.
- Zhao, L., Jordan, T. H., Olsen, K. B., & Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, 95(6), 2066–2080.
- Zhou, J., Cui, Y., Poyraz, E., Choi, D. J., & Guest, C. C. (2013). Multi-GPU implementation of a 3D finite difference time domain earthquake code on heterogeneous supercomputers. *Proceedia Computer Science*, 18(0), 1255–1264. (2013 International Conference on Computational Science).
- Zhu, L., & Helmberger, D. (1996). Advancement in source estimation techniques using broadband regional seismograms. *Bulletin of the Seismological Society of America*, 86(5), 1634–1641.
- Zhu, L., & Kanamori, H. (2000). Moho depth variation in Southern California from teleseismic receiver functions. *Journal of Geophysical Research*, 105(B2), 2969–2980.

Chapter 3

Green's Functions

The theory of Green's functions is one of the most elegant pieces of machinery for solving both the forward problem, often defined in terms of differential equations coupled with initial and/or boundary conditions, and also the inverse problem, which often requires evaluations of functional derivatives. For the forward problem, the Green's function provides the inverse of the differential operator. In solving the inverse problem, the functional derivatives can often be expressed conveniently using the Green's function. In this book, I also use Green's functions extensively in developing a theory of F3DWI. In practice, approximations of Green's functions in three-dimensional earth structural models can be computed by solving the elastodynamic equation using various numerical algorithms, such as the finite-difference method discussed in Chap. 2.

3.1 Mathematical Preliminary

Before discussing the details of Green's functions, we will need some mathematical preparations. In the following, I will collect some of the important mathematical results that are useful to our discussion on F3DWI. In-depth studies on those mathematical results can be found in e.g., (Zeidler 1995; Axler 1997; Hassani 1999). I will not use the Einstein summation convention in equations of this chapter. All summations over indices are written out explicitly using \sum , which might make it easier to convert the equations to computer codes if some readers are interested in developing F3DWI codes.

3.1.1 Vector Spaces

Mathematical objects are often organized into *spaces*. A *vector space* \mathcal{V} is a collection of objects called *vectors*, denoted by $|a\rangle$, $|b\rangle$, $|x\rangle$, $|y\rangle$ ¹ and so on, that can be

¹ Here I use the *bra* $\langle |$ and *ket* $| \rangle$ notation invented by Dirac for vectors.

added together and *multiplied* by real or complex scalars. The sum of two vectors $|a\rangle$ and $|b\rangle$ in \mathcal{V} corresponds to a vector $|a\rangle + |b\rangle$, also in \mathcal{V} , and has the properties

$$|a\rangle + |b\rangle = |b\rangle + |a\rangle,$$

$$|a\rangle + (|b\rangle + |c\rangle) = (|a\rangle + |b\rangle) + |c\rangle.$$

There exists a unique *zero vector* $|0\rangle \in \mathcal{V}$ such that

$$|x\rangle + |0\rangle = |x\rangle \text{ for any } |x\rangle \in \mathcal{V}.$$

For any $|x\rangle \in \mathcal{V}$ there exists a unique vector $-|x\rangle \in \mathcal{V}$ such that

$$|x\rangle + (-|x\rangle) = |0\rangle.$$

For any complex number α , also called a *scalar*, and any vector $|x\rangle \in \mathcal{V}$, the scalar multiplication $\alpha|x\rangle \in \mathcal{V}$ and has the properties

$$\alpha(\beta|x\rangle) = (\alpha\beta)|x\rangle,$$

$$1|x\rangle = |x\rangle.$$

The multiplication with scalars is distributive for the vectors and also the scalars, that is,

$$\alpha(|x\rangle + |y\rangle) = \alpha|x\rangle + \alpha|y\rangle,$$

$$(\alpha + \beta)|x\rangle = \alpha|x\rangle + \beta|x\rangle.$$

A *subspace* \mathcal{W} of a vector space \mathcal{V} is a non-empty subset of \mathcal{V} and must also be a vector space in its own right, i.e., if $\forall |x\rangle, |y\rangle \in \mathcal{W}$, then $\forall \alpha, \beta \in \mathbb{C}, \alpha|x\rangle + \beta|y\rangle \in \mathcal{W}$, where \mathbb{C} is the space of complex numbers.

For example, it can be verified that \mathbb{C} is a vector space, but the space of real numbers \mathbb{R} is not a vector space according to the definition above. However, if we restrict the scalars in the multiplication to real numbers, i.e., $\forall \alpha, \beta \in \mathbb{R}$, then both \mathbb{R} and \mathbb{C} are vector spaces and \mathbb{R} is a subspace of \mathbb{C} . This example shows the importance of identifying the properties of the scalars used in the multiplication when defining vector spaces. If the scalars belong to \mathbb{C} , the vector space \mathcal{V} is called a *complex vector space* or a vector space *over* \mathbb{C} . If the scalars belong to \mathbb{R} , \mathcal{V} is called a *real vector space* or a vector space *over* \mathbb{R} . In the previous example, \mathbb{R} is a vector space over \mathbb{R} , but not a vector space over \mathbb{C} ; \mathbb{C} is a vector space over \mathbb{R} , and also a vector space over \mathbb{C} .

3.1.2 Inner Product

The inner product is a rule that associates a scalar with two vectors. We can denote this association as $f: \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$ and it has the properties

$$\begin{aligned}
f(|x\rangle, |y\rangle) &= f^*(|y\rangle, |x\rangle), \\
f(|x\rangle, \alpha|y\rangle + \beta|z\rangle) &= \alpha f(|x\rangle, |y\rangle) + \beta f(|x\rangle, |z\rangle), \forall \alpha, \beta \in \mathbb{C}, \\
f(|x\rangle, |x\rangle) &\geq 0, \text{ and } f(|x\rangle, |x\rangle) = 0 \text{ if and only if } |x\rangle = |0\rangle.
\end{aligned}$$

The “*” in the first property denotes complex conjugation. The second property indicates that the inner product f is linear in its second argument. The last property is called the *positive definite* property of the inner product and it is often used to define the *length* of a vector in \mathcal{V} .

To understand why the complex conjugation in the first property is needed, let's suppose that we can remove it and we have $f(|x\rangle, |y\rangle) = f(|y\rangle, |x\rangle)$, which indicates that the inner product f is symmetric. Since f is linear in its second argument, this symmetry indicates that f is also linear in its first argument. Consider a nonzero vector $|x\rangle$, we can compute the inner product of $i|x\rangle$ with itself, $f(i|x\rangle, i|x\rangle)$, where i is the imaginary unit. The linearity in both of its arguments indicates that $f(i|x\rangle, i|x\rangle) = i^2 f(|x\rangle, |x\rangle) = -f(|x\rangle, |x\rangle)$. Since $|x\rangle$ is nonzero, $i|x\rangle$ is nonzero too. The third property indicates that either the left-hand-side or the right-hand-side of this equation must be negative, which is inconsistent with the positive definite property. If we introduce complex conjugation in the first property, we then have $f(i|x\rangle, i|x\rangle) = i f(i|x\rangle, |x\rangle) = i f^*(|x\rangle, i|x\rangle) = -i^2 f^*(|x\rangle, |x\rangle) = f^*(|x\rangle, |x\rangle) = f(|x\rangle, |x\rangle)$. The inconsistency is then removed.

Using the Dirac notation, the inner product of two vectors $|x\rangle, |y\rangle \in \mathcal{V}$ is denoted as $\langle x | y \rangle \in \mathbb{C}$ and the three properties can be expressed as

$$\langle x | y \rangle = \langle y | x \rangle^*, \quad (3.1)$$

$$\langle x | \alpha y + \beta z \rangle = \alpha \langle x | y \rangle + \beta \langle x | z \rangle, \quad (3.2)$$

$$\langle x | x \rangle \geq 0 \text{ and } \langle x | x \rangle = 0 \text{ if and only if } |x\rangle = |0\rangle. \quad (3.3)$$

The inner product is linear in the second vector, i.e., the vector in the *ket*, as shown in Eq. 3.2, but not linear in the first vector, i.e., the vector in the *bra*. Consider the inner product $\langle \alpha x + \beta y | z \rangle$, using Eq. 3.1, we have

$$\langle \alpha x + \beta y | z \rangle = \langle z | \alpha x + \beta y \rangle^*.$$

Using Eq. 3.2 and considering Eq. 3.1, we have

$$\begin{aligned}
\langle \alpha x + \beta y | z \rangle &= \langle z | \alpha x + \beta y \rangle^* \\
&= \alpha^* \langle z | x \rangle^* + \beta^* \langle z | y \rangle^* \\
&= \alpha^* \langle x | z \rangle + \beta^* \langle y | z \rangle.
\end{aligned} \quad (3.4)$$

Because of the extra complex conjugation operation on the scalars $\alpha, \beta \in \mathbb{C}$ in Eq. 3.4, the inner product on a complex vector space is not linear in its first vector, but called *sesquilinear* or *antilinear*. A vector space equipped with an inner product is called an *inner product space*.

As an example of one of the most widely used inner products, let's consider two vectors $|x\rangle, |y\rangle \in \mathbb{C}^n$, where \mathbb{C}^n is the set of all complex n -tuples, with $|x\rangle = (x_1, x_2, \dots, x_n)$ and $|y\rangle = (y_1, y_2, \dots, y_n)$, we have an inner product defined on \mathbb{C}^n as

$$\langle x | y \rangle \equiv x_1^* y_1 + x_2^* y_2 + \dots + x_n^* y_n = \sum_{i=1}^n x_i^* y_i. \quad (3.5)$$

It can be verified that this inner product satisfies all the properties listed in Eq. 3.1–3.3. If we represent $|y\rangle \in \mathbb{C}^n$ as a column vector,

$$|y\rangle = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

and $\langle x|$ as a row vector,

$$\langle x| = (x_1^*, x_2^*, \dots, x_n^*),$$

then Eq. 3.5 can be written as the matrix multiplication

$$\langle x | y \rangle = (x_1^*, x_2^*, \dots, x_n^*) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^n x_i^* y_i. \quad (3.6)$$

The inner product defined in Eqs. 3.5 and 3.6 is sometimes called the *natural inner product* for \mathbb{C}^n .

The *norm* or *length* of a vector $|x\rangle$ in an inner product space is defined as

$$\|x\| = \sqrt{\langle x | x \rangle}. \quad (3.7)$$

It can be shown that the norm induced from the inner product, Eq. 3.7, has the following properties,

$$\|0\rangle = 0,$$

$$\|x\| \geq 0, \text{ and } \|x\| = 0 \text{ if and only if } |x\rangle = |0\rangle,$$

$$\|\alpha x\| = |\alpha| \|x\|, \quad \forall \alpha \in \mathbb{C},$$

$$\|x + y\| \leq \|x\| + \|y\|.$$

A vector space equipped with a norm is called a *normed linear space*. An inner product space is a normed space, but the converse may not be true.

A set of vectors $|a_1\rangle, |a_2\rangle, \dots, |a_n\rangle$ are *linearly independent* if for any $\alpha_i \in \mathbb{C}$, the relation $\sum_{i=1}^n \alpha_i |a_i\rangle = 0$ implies $\alpha_i = 0$ for all i . If every vector in \mathcal{V} can be

represented as a linear combination of vectors in a set S , we say S spans \mathcal{V} . A *basis* of a vector space \mathcal{V} is a set of linearly independent vectors that spans \mathcal{V} . The number of vectors n in a basis of the vector space \mathcal{V} is the dimension of \mathcal{V} . If n is finite, \mathcal{V} is called *finite-dimensional*; otherwise, it is called *infinite-dimensional*. Two vectors, $|x\rangle$ and $|y\rangle$, are *orthogonal* if their inner product is zero, i.e., $\langle x | y \rangle = 0$. If the length of a vector $|e\rangle$ is 1, i.e., $\langle e | e \rangle = 1$, $|e\rangle$ is called a *normal vector* or a *normalized vector*. An *orthonormal basis* $B = \{|e_i\rangle\}_{i=1}^n$ is a basis in which every vector is a normal vector and any two different vectors are orthogonal to each other, i.e.,

$$\langle e_i | e_j \rangle = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \quad (3.8)$$

where δ_{ij} is called the *Kronecker delta*. An example of orthonormal bases is the standard basis of \mathbb{C}^n (or \mathbb{R}^n) defined as

$$|e_1\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, |e_2\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, |e_n\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \quad (3.9)$$

It can be verified that this basis is orthonormal under the natural inner product of \mathbb{C}^n as defined in Eq. 3.5. We can express an arbitrary vector $|x\rangle \in \mathbb{C}^n$ in the standard basis as

$$|x\rangle = \sum_{i=1}^n x_i |e_i\rangle, \quad (3.10)$$

where the scalar coefficients x_i , $i = 1, 2, \dots, n$ can be evaluated using the inner product with each of the basis vectors $|e_j\rangle$, $j = 1, 2, \dots, n$,

$$\langle e_j | x \rangle = \left\langle e_j \left| \sum_{i=1}^n x_i |e_i\rangle \right. \right\rangle = \sum_{i=1}^n x_i \langle e_j | e_i \rangle = \sum_{i=1}^n x_i \delta_{ij} = x_j. \quad (3.11)$$

In Eq. 3.11 above, I have used the linearity of the inner product in its second vector (i.e., Eq. 3.2) at the second equality and used Eq. 3.8 at the third equality.

We can convert any basis into an orthonormal basis using a process called *Gram-Schmidt orthonormalization*. The Gram-Schmidt process is neatly illustrated through a 3D animation on its Wikipedia page², together with references.

A useful inequality for inner product spaces, known as the *Cauchy-Schwarz inequality*, states that

$$|\langle x | y \rangle|^2 \leq \|x\|^2 \|y\|^2. \quad (3.12)$$

² http://en.wikipedia.org/wiki/Gram-Schmidt_process.

The equality in Eq. 3.12 holds when $|x\rangle$ is proportional to $|y\rangle$ (i.e., $|x\rangle = \alpha |y\rangle$ and α is a scalar). In two- or three-dimensional Euclidean spaces, the inner product $\langle x | y \rangle = \|x\| \cdot \|y\| \cdot \cos \theta$, where θ is the angle between $|x\rangle$ and $|y\rangle$, and the Cauchy-Schwarz inequality is equivalent to the fact that $\cos \theta \leq 1$ and equality holds when $\theta = 0$ (i.e., $|x\rangle$ and $|y\rangle$ are parallel to each other). The Cauchy-Schwarz inequality, Eq. 3.12, is applicable to inner product spaces of both finite and infinite dimensions. The proof is short (e.g., Axler 1997) and I include it here to illustrate the connection with geometrical operations in Euclidean spaces. We assume that $|x\rangle$ and $|y\rangle$ are nonzero vectors (if one of them is $|0\rangle$, Eq. 3.12 holds trivially). We can construct a new vector

$$|z\rangle = |y\rangle - \frac{\langle x | y \rangle}{\langle x | x \rangle} |x\rangle, \quad (3.13)$$

and take the inner product with $|y\rangle$ on both sides to obtain

$$\begin{aligned} \langle y | z \rangle &= \langle y | y \rangle - \frac{\langle x | y \rangle \langle y | x \rangle}{\langle x | x \rangle} = \langle y | y \rangle - \frac{\langle x | y \rangle \langle x | y \rangle^*}{\langle x | x \rangle} \\ &= \langle y | y \rangle - \frac{|\langle x | y \rangle|^2}{\langle x | x \rangle}, \end{aligned} \quad (3.14)$$

where I have used Eq. 3.1 at the second equality. If we take the inner product with $|x\rangle$ on both sides of Eq. 3.13, we have

$$\langle x | z \rangle = 0, \quad (3.15)$$

which shows that $|x\rangle$ and $|z\rangle$ are orthogonal to each other. If we take the inner product with $|z\rangle$ on both sides of Eq. 3.13, considering Eq. 3.15, we have

$$\langle z | z \rangle = \langle z | y \rangle \Rightarrow \langle y | z \rangle = \langle z | z \rangle. \quad (3.16)$$

Bring the equality $\langle y | z \rangle = \langle z | z \rangle$ into Eq. 3.14, we have

$$\begin{aligned} \langle z | z \rangle &= \langle y | y \rangle - \frac{|\langle x | y \rangle|^2}{\langle x | x \rangle} \\ \Rightarrow 0 &\leq \|z\|^2 = \|y\|^2 - \frac{|\langle x | y \rangle|^2}{\|x\|^2}. \end{aligned} \quad (3.17)$$

Re-arranging Eq. 3.17 gives the Cauchy-Schwarz inequality, Eq. 3.12. Comparing with geometrical operations in Euclidean spaces, the second term on the right-hand-side (RHS) of Eq. 3.13 is actually the projection of vector $|y\rangle$ in the direction of vector $|x\rangle$. By subtracting this projection from $|y\rangle$ on the RHS of Eq. 3.13, we have obtained a vector, $|z\rangle$, that is on a plane perpendicular (i.e., orthogonal) to $|x\rangle$, therefore we have Eq. 3.15. Equation 3.13 can be considered as decomposing $|y\rangle$ into two perpendicular components, i.e.,

$$|y\rangle = |z\rangle + \frac{\langle x | y \rangle}{\langle x | x \rangle} |x\rangle. \quad (3.18)$$

In Euclidean geometry, the Pythagorean theorem states that

$$\|y\|^2 = \|z\|^2 + \left\| \frac{\langle x | y \rangle}{\langle x | x \rangle} |x\rangle \right\|^2 = \|z\|^2 + \frac{|\langle x | y \rangle|^2}{\|x\|^2}, \quad (3.19)$$

which is identical to Eq. 3.17 after re-arranging the terms. In fact, the Pythagorean theorem is applicable on general inner product spaces.

3.1.3 Linear Transformations

A *transformation* is a mapping that brings a vector in a vector space \mathcal{V} to a vector in another vector space \mathcal{W} . A *linear transformation* is a mapping, denoted as $\mathbf{T}: \mathcal{V} \rightarrow \mathcal{W}$, where \mathcal{V} and \mathcal{W} are both complex vector spaces, \mathcal{V} is the domain and \mathcal{W} is the codomain, such that,

$$\mathbf{T}(\alpha |x\rangle + \beta |y\rangle) = \alpha \mathbf{T}(|x\rangle) + \beta \mathbf{T}(|y\rangle), \quad \forall \alpha, \beta \in \mathbb{C}, \forall |x\rangle, |y\rangle \in \mathcal{V}. \quad (3.20)$$

The action of a linear transformation on a vector can be written without the parentheses, i.e., $\mathbf{T}(|x\rangle) \equiv \mathbf{T}|x\rangle$. The set of all linear transformations from \mathcal{V} to \mathcal{W} is often denoted as $\mathcal{L}(\mathcal{V}, \mathcal{W})$, which is also a vector space when equipped with the following addition and scalar multiplication

$$(\mathbf{T}_1 + \mathbf{T}_2)(|x\rangle) = \mathbf{T}_1|x\rangle + \mathbf{T}_2|x\rangle,$$

$$(\alpha \mathbf{T}_1)(|x\rangle) = \alpha(\mathbf{T}_1|x\rangle),$$

$$\forall \mathbf{T}_1, \mathbf{T}_2 \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \forall \alpha \in \mathbb{C}, \forall |x\rangle \in \mathcal{V}.$$

The set of vectors in \mathcal{V} that are mapped to the zero vector in \mathcal{W} by the transformation \mathbf{T} are called the *kernel* (not to be confused with the Fréchet kernel in Chap. 4) or *null space* of \mathbf{T} and is often denoted as $\ker \mathbf{T}$. It can be verified that $\ker \mathbf{T}$ is a subspace of \mathcal{V} . It can also be verified that the range of the linear transformation, $\mathbf{T}(\mathcal{V})$, forms a subspace of \mathcal{W} .

A special case of a linear transformation occurs when $\mathcal{W} = \mathbb{C}$, i.e., the codomain of the mapping happens to be the space of one-dimensional vectors (i.e., scalars). In such a case, the linear transformation is called a *linear functional* and the space of linear functionals is often denoted as $\mathcal{L}(\mathcal{V}, \mathbb{C})$. It can be verified that $\mathcal{L}(\mathcal{V}, \mathbb{C})$ is a vector space. We call $\mathcal{L}(\mathcal{V}, \mathbb{C})$ the *dual space* of \mathcal{V} and denote it as \mathcal{V}^* ³.

Given a basis in \mathcal{V} , $B = \{|e_i\rangle\}_{i=1}^n$, it is possible to construct a corresponding basis for the dual space \mathcal{V}^* , which we call the *dual basis* of B and denote it as $B^* = \{\mathbf{f}_i\}_{i=1}^n$. The linear functionals $\mathbf{f}_i \in \mathcal{V}^*$ satisfy the relations

$$\mathbf{f}_i |e_j\rangle = \delta_{ij}, \quad (3.21)$$

³ When applied on a vector space, the superscript “*” does not denote complex conjugation, but represents the corresponding dual space.

where δ_{ij} is the Kronecker delta. Equation 3.21 has a different meaning from Eq. 3.8. In Eq. 3.8, $|e_i\rangle, |e_j\rangle \in \mathcal{V}$, while in Eq. 3.21, $\mathbf{f}_i \in \mathcal{V}^*, |e_j\rangle \in \mathcal{V}$.

The inner product defined in Sect. 3.1.2 can be associated with linear functionals. If we consider the second vector (i.e., the vector in the *ket*) as the input to the inner product transformation defined using the vector in the *bra*, then this transformation maps the vector space \mathcal{V} to \mathbb{C} . We can verify that it is a linear functional by using the linearity in the second vector of the inner product, Eq. 3.2. In fact, for any vector $|x\rangle$ in \mathcal{V} , we can define such a linear functional, which we denote as $\langle x|$, and we can introduce the notation

$$(|x\rangle)^\dagger \equiv \langle x|, \quad (3.22)$$

where the superscript \dagger means the *dual* of. Considering that the inner product defined in Sect. 3.1.2 is sesquilinear in the first vector, we speculate that the duality operation is also sesquilinear. To verify this, let's consider the linear combination of two vectors

$$|z\rangle = \alpha |x\rangle + \beta |y\rangle, \forall |x\rangle, |y\rangle \in \mathcal{V}, \forall \alpha, \beta \in \mathbb{C}.$$

We can take the inner product of $|z\rangle$ with an arbitrary vector $|f\rangle \in \mathcal{V}$. Considering the linearity in the second vector, Eq. 3.2, we have

$$\langle f | z \rangle = \alpha \langle f | x \rangle + \beta \langle f | y \rangle.$$

We can complex conjugate both sides of this equation. Considering Eq. 3.1, we have

$$\text{LHS} = \langle f | z \rangle^* = \langle z | f \rangle,$$

$$\text{RHS} = \alpha^* \langle x | f \rangle + \beta^* \langle y | f \rangle = (\alpha^* \langle x| + \beta^* \langle y|) |f\rangle.$$

Since the LHS must equal to the RHS for any $|f\rangle \in \mathcal{V}$, we must have

$$|z\rangle^\dagger \equiv \langle z| = \alpha^* \langle x| + \beta^* \langle y|. \quad (3.23)$$

Because the complex scalars in the linear combination must be complex conjugated in Eq. 3.23, the duality operation is sesquilinear.

Let's use the natural inner product for \mathbb{C}^n , Eq. 3.6, as an example. For a vector $|x\rangle \in \mathbb{C}^n$, we can represent it as a column vector

$$|x\rangle = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \quad (3.24)$$

Considering Eq. 3.6, the inner product of $|x\rangle$ with an arbitrary vector

$$|y\rangle = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

can then be expressed as,

$$\langle y | x \rangle = \sum_{i=1}^n y_i^* x_i. \quad (3.25)$$

We now complex conjugate both sides of Eq. 3.25. Considering Eq. 3.1, we obtain

$$\text{LHS} = \langle y | x \rangle^* = \langle x | y \rangle, \quad (3.26)$$

$$\text{RHS} = \left(\sum_{i=1}^n y_i^* x_i \right)^* = \sum_{i=1}^n x_i^* y_i = (x_1^*, x_2^*, \dots, x_n^*) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad (3.27)$$

Since the LHS must equal to the RHS and vector $|y\rangle$ is arbitrary, Eqs. 3.26 and 3.27 suggest that

$$(|x\rangle)^\dagger \equiv \langle x| = (x_1^*, x_2^*, \dots, x_n^*). \quad (3.28)$$

Comparing with Eq. 3.24, Eq. 3.28 shows that the dual of a column vector in \mathbb{C}^n is a row vector with its elements complex conjugated.

For the standard basis of \mathbb{C}^n , $B = \{|e_i\rangle\}_{i=1}^n$, where the vectors $|e_i\rangle$ are represented as column vectors in Eq. 3.9, the corresponding standard dual basis is often denoted as $B^* = \{|e_j\rangle\}_{j=1}^n$, where $\langle e_j|$ are row vectors,

$$\langle e_1| = (1, 0, 0, \dots, 0), \langle e_2| = (0, 1, 0, \dots, 0), \dots, \langle e_n| = (0, 0, 0, \dots, 1). \quad (3.29)$$

Using a basis vector $|e_j\rangle$ and the corresponding dual basis vector $\langle e_j|$, we can construct a *projection operator*

$$\mathbf{P}_j = |e_j\rangle\langle e_j|. \quad (3.30)$$

To understand how a projection operator works, let's apply it to an arbitrary vector $|x\rangle \in \mathbb{C}^n$. Considering the expansion in Eq. 3.10, we have

$$\begin{aligned} \mathbf{P}_j |x\rangle &= \mathbf{P}_j \left(\sum_{i=1}^n x_i |e_i\rangle \right) = \sum_{i=1}^n x_i (\mathbf{P}_j |e_i\rangle) \\ &= \sum_{i=1}^n x_i |e_j\rangle \langle e_j | e_i \rangle = \sum_{i=1}^n x_i |e_j\rangle \delta_{ij} = x_j |e_j\rangle. \end{aligned} \quad (3.31)$$

If we introduce an identity transform $\mathbf{1}$, such that

$$\mathbf{1} |x\rangle = |x\rangle, \forall |x\rangle \in \mathbb{C}^n, \quad (3.32)$$

then we have

$$\mathbf{1}|x\rangle = \sum_{j=1}^n x_j |e_j\rangle = \sum_{j=1}^n \mathbf{P}_j |x\rangle = \left(\sum_{j=1}^n \mathbf{P}_j \right) |x\rangle, \quad (3.33)$$

where I have used Eq. 3.10 at the first equality, Eq. 3.31 at the second equality. Since $|x\rangle$ is arbitrary, we must have

$$\sum_{j=1}^n \mathbf{P}_j = \sum_{j=1}^n |e_j\rangle\langle e_j| = \mathbf{1}. \quad (3.34)$$

Equation 3.34 is known as the *completeness relation*. If we choose the first $l < n$ basis vectors in B and construct the projection operator $\mathbf{P}^{(l)} = \sum_{j=1}^l |e_j\rangle\langle e_j|$, then $\mathbf{P}^{(l)}$ projects an arbitrary vector into the subspace spanned by the first l basis vectors in B .

The linear transformation $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ also has its dual, often denoted as $\mathbf{T}^* : \mathcal{W}^* \rightarrow \mathcal{V}^*$ and defined as

$$(\mathbf{T}^* \mathbf{g}) |x\rangle = \mathbf{g} (\mathbf{T} |x\rangle), \quad \forall |x\rangle \in \mathcal{V}, \mathbf{g} \in \mathcal{W}^*. \quad (3.35)$$

In the literature, the dual \mathbf{T}^* is also called the *transpose* (Trèves 1967) of \mathbf{T} . The linear functional $\mathbf{T}^* \mathbf{g} \in \mathcal{V}^*$ is called the *pullback* of \mathbf{g} along \mathbf{T} . If the vector spaces \mathcal{V} and \mathcal{W} are both equipped with inner products as defined in Sect. 3.1.2, the transpose is characterized by the identity

$$\langle f | (\mathbf{T} |x\rangle) \rangle = (\langle f | \mathbf{T}^*) |x\rangle, \quad \forall |x\rangle \in \mathcal{V}, \forall \langle f | \in \mathcal{W}^*. \quad (3.36)$$

Note that the inner product on the LHS of Eq. 3.36 is for \mathcal{W} , while the inner product on the RHS of Eq. 3.36 is for \mathcal{V} .

As an example, let's consider $\mathcal{V} = \mathbb{C}^n$ and $\mathcal{W} = \mathbb{C}^m$ and the linear transformation \mathbf{T} is represented by the real $m \times n$ matrix \mathbf{A} . For a column vector $|x\rangle \in \mathbb{C}^n$, $\mathbf{T} |x\rangle$ can be represented as the matrix-vector multiplication

$$y_i = \sum_{j=1}^n A_{ij} x_j,$$

where A_{ij} and x_j are elements of the matrix \mathbf{A} and the vector $|x\rangle$, respectively. The result is a column vector $|y\rangle \in \mathbb{C}^m$ with its elements given by y_i . For a row-vector $\langle f | \in (\mathbb{C}^m)^*$, the LHS of Eq. 3.36 gives

$$\text{LHS} = \langle f | y \rangle = \sum_{i=1}^m f_i^* y_i = \sum_{i=1}^m \sum_{j=1}^n f_i^* A_{ij} x_j, \quad (3.37)$$

where $f_i, i = 1, 2, \dots, m$ are elements of $\langle f |$ and $*$ represents complex conjugation. The transpose \mathbf{T}^* can then be represented as \mathbf{A}^T , the transpose of the matrix \mathbf{A} ,

with its elements given by A_{ji} ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$). Alternatively, we can also represent \mathbf{T}^* using the same matrix \mathbf{A} but acting on the right of row vectors, i.e., $\langle f | \mathbf{T}^*$, rather than acting on the left of column vectors like \mathbf{T} does, i.e., $\mathbf{T} | x$. The result of $\langle f | \mathbf{T}^*$ can be represented as

$$z_j = \sum_{i=1}^m f_i A_{ij},$$

where $z_j, j = 1, 2, \dots, n$ are elements of $\langle z | \in (\mathbb{C}^n)^*$. Considering our assumption that \mathbf{A} is real, the RHS of Eq. 3.36 can be expressed as

$$\text{RHS} = \langle z | x \rangle = \sum_{j=1}^n z_j^* x_j = \sum_{j=1}^n \sum_{i=1}^m f_i^* A_{ij} x_j. \quad (3.38)$$

Comparing with Eq. 3.37, the LHS is identical to the RHS, therefore we have verified the identity 3.36. Note that in the second equality in Eq. 3.37, the inner product is for \mathbb{C}^m , therefore the summation index i is from 1 to m ; while in the second equality in Eq. 3.38, the inner product is for \mathbb{C}^n , therefore the summation index j is from 1 to n .

An extension of the example above is to allow the matrix \mathbf{A} to be complex. In such a case, the transpose \mathbf{T}^* can no longer be represented as \mathbf{A}^T for the identity 3.36 to hold. In addition to taking the transpose of \mathbf{A} , we also need to take the complex conjugate of every element in \mathbf{A} . The resulting matrix, often denoted as \mathbf{A}^\dagger , is often called the *conjugate transpose*, or the *Hermitian transpose*, or the *Hermitian conjugate*, or the *adjoint matrix* of \mathbf{A} . When \mathbf{T}^* is represented by \mathbf{A}^\dagger , the identity 3.36 holds for complex \mathbf{A} .

3.1.4 Hilbert Spaces

The mathematical results we have discussed so far are mainly demonstrated using examples of finite-dimensional vector spaces. Many important applications, such as those involving functions $f: \mathbb{R} \rightarrow \mathbb{C}$, require us to extend those results to infinite-dimensional vector spaces. Such an extension is not trivial and requires some additional machinery. To understand the differences between finite-dimensional and infinite-dimensional vector spaces, let's use the inner product defined in Eq. 3.5 as an example. For finite-dimensional vector spaces, the sum involves only a finite number of terms; while for infinite-dimensional vector spaces the sum involves an infinite number of terms and we have to investigate the convergence of such an infinite sum before we can extend the definition of the inner product to infinite-dimensional vector spaces.

One useful apparatus for studying convergence is the concept of *Cauchy sequences*, which are sequences whose members become arbitrarily close to each

other once we go far enough into the sequences. Formally, if an infinite sequence of vectors $\{|a_i\rangle\}_{i=1}^{\infty}$ in a normed vector space satisfies

$$\lim_{i \rightarrow \infty, j \rightarrow \infty} \|a_i - a_j\| = 0,$$

then this sequence is a Cauchy sequence. The usefulness of the Cauchy sequence lies in the fact that it allows us to test convergence without knowing what the limit of the sequence might be.

For a normed vector space \mathcal{V} , if every Cauchy sequence in \mathcal{V} converges to a vector in \mathcal{V} , we say \mathcal{V} is a *complete vector space*. For example, every Cauchy sequence in \mathbb{R} converges to a limit in \mathbb{R} with respect to the absolute value norm $\|a\| = |a|$, therefore \mathbb{R} is complete (for a proof, see e.g., Schröder 2008). It can be shown that every Cauchy sequence in \mathbb{C} converges to a limit in \mathbb{C} with respect to the norm $\|\alpha\| = \sqrt{(\Re\alpha)^2 + (\Im\alpha)^2}$, where $\Re\alpha$ and $\Im\alpha$ are the real and imaginary parts of $\alpha \in \mathbb{C}$, therefore \mathbb{C} is complete. The space of rational numbers \mathbb{Q} is not complete. Consider the sequence defined by $a_0 = 1$, $a_{n+1} = (a_n + 2/a_n)/2$, which is a Cauchy sequence and is inside \mathbb{Q} , the limit of this sequence is the irrational number $\sqrt{2} \notin \mathbb{Q}$.

It can be shown that every finite-dimensional inner product space is complete with respect to the norm induced by its inner product (i.e., Eq. 3.7). Suppose \mathcal{V} is a finite-dimensional vector space with dimension n and $\{|a_i\rangle\}_{i=1}^{\infty}$ is a Cauchy sequence with $|a_i\rangle \in \mathcal{V}$. We can express $|a_i\rangle$ and $|a_j\rangle$ using an orthonormal basis $B = \{|e_k\rangle\}_{k=1}^n$,

$$|a_i\rangle = \sum_{k=1}^n \alpha_k^i |e_k\rangle, \quad |a_j\rangle = \sum_{k=1}^n \alpha_k^j |e_k\rangle,$$

where the superscripts i and j on the scalar coefficients are vector indices, not exponents and $\alpha_k^i, \alpha_k^j \in \mathbb{C}$. The norm of the difference $|a_i\rangle - |a_j\rangle$ can be expressed as

$$\begin{aligned} \|a_i - a_j\|^2 &= \langle (a_i - a_j) | (a_i - a_j) \rangle \\ &= \left\| \sum_{k=1}^n (\alpha_k^i - \alpha_k^j) |e_k\rangle \right\|^2 = \sum_{k=1}^n |\alpha_k^i - \alpha_k^j|^2, \end{aligned} \quad (3.39)$$

where I have used the orthonormal property of $|e_k\rangle$, Eq. 3.8, in the last equality. Because $\{|a_i\rangle\}_{i=1}^{\infty}$ is assumed to be a Cauchy sequence, the LHS of Eq. 3.39 goes to 0 as i and j approach infinity. The summation on the RHS of Eq. 3.39 involve non-negative terms $|\alpha_k^i - \alpha_k^j|^2$, therefore we must have $|\alpha_k^i - \alpha_k^j| \rightarrow 0$ as $i, j \rightarrow \infty$ for all k , which shows that the sequence $\{\alpha_k^i\}_{i=1}^{\infty}$ is a Cauchy sequence. Because \mathbb{C} is complete, the limit of this Cauchy sequence $\alpha_k \equiv \lim_{i \rightarrow \infty} \alpha_k^i$ must be inside \mathbb{C} . If we construct a vector using the limits $\alpha_k, k = 1, 2, \dots, n$,

$$|a\rangle = \sum_{k=1}^n \alpha_k |e_k\rangle, \quad (3.40)$$

we have $|a\rangle \in \mathcal{V}$, because \mathcal{V} is a vector space and $|a\rangle$ is expressed as a linear combination of vectors inside \mathcal{V} . We can also show that $|a\rangle$ constructed in Eq. 3.40 is the limit of $\{|a_i\rangle\}_{i=1}^{\infty}$,

$$\lim_{i \rightarrow \infty} \|a - a_i\|^2 = \lim_{i \rightarrow \infty} \sum_{k=1}^n |\alpha_k - \alpha_k^i|^2 = \sum_{k=1}^n \lim_{i \rightarrow \infty} |\alpha_k - \alpha_k^i|^2 = 0.$$

We have shown that an arbitrary Cauchy sequence $\{|a_i\rangle\}_{i=1}^{\infty}$ in \mathcal{V} converges to a vector $|a\rangle$ that also lies inside \mathcal{V} , therefore an arbitrary finite-dimensional vector space is complete.

The proof of completeness for finite-dimensional vector spaces cannot be directly extended to infinite-dimensional vector spaces. The problem lies in Eq. 3.40. The definition of vector spaces, including both finite- and infinite-dimensional cases, only guarantees that a linear combination of a *finite* number of vectors in \mathcal{V} still lies in \mathcal{V} . When n goes to infinity, not all vector spaces contain the limits of infinite sums. The infinite-dimensional vector spaces that do contain those limits are therefore of special interest. Formally, if an inner product space \mathcal{H} is complete with respect to the norm induced by the inner product, we call \mathcal{H} a *Hilbert space*. In principle, all finite-dimensional vector spaces are Hilbert spaces according to this definition. However, when we talk about Hilbert spaces, we often imply that the vector spaces are infinite-dimensional.

If we confine ourselves to a Hilbert space \mathcal{H} , we can re-visit Eq. 3.40 with n replaced by ∞ . We now consider the set of infinite sequence $B = \{|e_k\rangle\}_{k=1}^{\infty}$ of orthonormal vectors with $|e_k\rangle \in \mathcal{H}$ for all k . Whether B forms a basis of \mathcal{H} requires some additional considerations, as we will discuss later. We can take a vector of finite length $|a\rangle \in \mathcal{H}$ and compute its inner product with the orthonormal vectors $|e_k\rangle$,

$$\alpha_k = \langle e_k | a \rangle. \quad (3.41)$$

Using these scalars α_k , we can construct a sequence of vectors $\{|a_n\rangle\}_{n=1}^{\infty}$ with

$$|a_n\rangle = \sum_{k=1}^n \alpha_k |e_k\rangle, \quad (3.42)$$

and we can show that this sequence converges to $|a\rangle \in \mathcal{H}$. To show that this sequence converges, we take the inner product with $|a\rangle$ on both sides of Eq. 3.42 and obtain

$$\langle a | a_n \rangle = \sum_{k=1}^n \alpha_k \langle a | e_k \rangle = \sum_{k=1}^n \alpha_k \langle e_k | a \rangle^* = \sum_{k=1}^n \alpha_k \alpha_k^* = \sum_{k=1}^n |\alpha_k|^2, \quad (3.43)$$

where I have used Eq. 3.1 at the second equality and Eq. 3.41 at the third equality. The Cauchy-Schwarz inequality, Eq. 3.12, gives

$$|\langle a | a_n \rangle|^2 \leq \|a\|^2 \|a_n\|^2. \quad (3.44)$$

Because vectors $|e_k\rangle$ are orthonormal, considering Eq. 3.42, we have

$$\|a_n\|^2 = \sum_{k=1}^n |\alpha_k|^2. \quad (3.45)$$

Bring Eqs. 3.43 and 3.45 into the Cauchy-Schwarz inequality, Eq. 3.44, we obtain

$$\|a_n\|^2 = \sum_{k=1}^n |\alpha_k|^2 \leq \|a\|. \quad (3.46)$$

Since $|a\rangle$ has finite length by assumption, the length of $|a_n\rangle$ is finite for $n \rightarrow \infty$, therefore the vector sequence $\{|a_n\rangle\}_{n=1}^{\infty}$ converges.

In order to show that $\{|a_n\rangle\}_{n=1}^{\infty}$ converges to $|a\rangle$, we need to impose an additional condition on the set of orthonormal vectors $B = \{|e_k\rangle\}_{k=1}^{\infty}$. We require that the only vector in \mathcal{H} that is orthogonal to all the $|e_k\rangle$ in B is the zero vector. A set of orthonormal vectors satisfying this condition is called *complete* and forms a *basis* of \mathcal{H} . To show that $\{|a_n\rangle\}_{n=1}^{\infty}$ converges to $|a\rangle$, we construct the vector

$$|f\rangle \equiv |a\rangle - |a_{\infty}\rangle = |a\rangle - \sum_{k=1}^{\infty} \alpha_k |e_k\rangle \quad (3.47)$$

and take the inner product with $|e_i\rangle$ on both sides for $i = 1, 2, \dots, \infty$. Considering Eq. 3.41, we have

$$\langle e_i | f \rangle = \langle e_i | a \rangle - \sum_{k=1}^{\infty} \langle e_i | e_k \rangle \alpha_k = \alpha_i - \sum_{k=1}^{\infty} \delta_{ik} \alpha_k = \alpha_i - \alpha_i = 0. \quad (3.48)$$

Since we have Eq. 3.48 for $i = 1, 2, \dots, \infty$, vector $|f\rangle$ is therefore orthogonal to all vectors in B . If B is complete, we have $|f\rangle = |0\rangle$, therefore

$$|a\rangle = \sum_{k=1}^{\infty} \alpha_k |e_k\rangle \quad (3.49)$$

and the sequence $\{|a_n\rangle\}_{n=1}^{\infty}$ converges to $|a\rangle \in \mathcal{H}$.

Within the confinement of Hilbert spaces, the results and techniques we have developed in Sects. 3.1.1–3.1.3 can be safely applied. Several identities are useful in the following discussions. Bring Eq. 3.41 into Eq. 3.49 and use the identity transform defined in Eq. 3.32, we have

$$\mathbf{1} |a\rangle = |a\rangle = \sum_{k=1}^{\infty} |e_k\rangle \langle e_k | a \rangle = \left(\sum_{k=1}^{\infty} |e_k\rangle \langle e_k | \right) |a\rangle. \quad (3.50)$$

Since Eq. 3.50 holds for all $|a\rangle \in \mathcal{H}$, we must have

$$\mathbf{1} = \sum_{k=1}^{\infty} |e_k\rangle \langle e_k |, \quad (3.51)$$

which is analogous to Eq. 3.34 and is the *completeness relation* in \mathcal{H} . For two arbitrary vectors $|f\rangle, |g\rangle \in \mathcal{H}$, we have their inner product

$$\langle g|f\rangle = \langle g|\mathbf{1}|f\rangle = \langle g|\left(\sum_{k=1}^{\infty}|e_k\rangle\langle e_k|\right)|f\rangle = \sum_{k=1}^{\infty}\langle g|e_k\rangle\langle e_k|f\rangle, \quad (3.52)$$

where I have used Eq. 3.51 at the second equality. Equation 3.52 is comparable to the finite-dimensional case in Eq. 3.5, if we take into account Eqs. 3.11 and 3.1. If we replace $|g\rangle$ with $|f\rangle$ in Eq. 3.52, we obtain an identity for the norm of $|f\rangle$ in \mathcal{H} .

3.1.5 Functions As Vectors

Once we allow the number of dimensions of a vector space to be infinite, most functions we encounter in realistic applications can be considered as vectors in suitably chosen Hilbert spaces. To establish a direct connection between vectors and functions, we need to establish the concept of a *counting set*. For a finite-dimensional vector space with dimension n , the components f_i of a vector $|f\rangle$ with respect to a basis, say $B = \{|e_i\rangle\}_{i=1}^n$, can be considered as values of a function f with its domain given by the finite set $\{1, 2, \dots, n\}$. The function values can be denoted as $f(i)$ with $i = 1, 2, \dots, n$ and we have $f(i) \equiv f_i = \langle e_i|f\rangle$ (Eq. 3.11). If we allow the counting set to be a continuum, i.e., \mathbb{R} , and change the counting index from $i \in \mathbb{N}$ (\mathbb{N} is the set of natural numbers) to $x \in \mathbb{R}$, for a vector $|f\rangle \in \mathcal{H}$, its components with respect to the infinite basis $B = \{|e_x\rangle\}_{x \in \mathbb{R}}$ can then be denoted as $f(x)$ and we can express them as

$$f(x) = \langle e_x|f\rangle. \quad (3.53)$$

The values of a function $f(x), x \in \mathbb{R}$ can be considered as components of the vector $|f\rangle \in \mathcal{H}$ with respect to the basis $B = \{|e_x\rangle\}_{x \in \mathbb{R}}$, therefore we have a correspondence between functions and vectors. We will see in Sect. 3.1.6 that the basis vectors $|e_x\rangle \in \mathcal{H}$ that can give rise to Eq. 3.53 have connections with the *Dirac delta function*.

Consider $\forall |f\rangle, |g\rangle \in \mathcal{H}$, their inner product is given by Eq. 3.52. If we replace the index $k \in \mathbb{N}$ with the continuous index $x \in \mathbb{R}$ in Eq. 3.52, the summation over k needs to be replaced with the integration over x and we obtain

$$\langle g|f\rangle = \int_{-\infty}^{\infty}\langle g|e_x\rangle\langle e_x|f\rangle dx = \int_{-\infty}^{\infty}g^*(x)f(x)dx, \quad (3.54)$$

where I have used Eqs. 3.53 and 3.1 at the second equality. Equation 3.54 gives a definition for the inner product between functions. In certain applications, we may introduce a strictly positive, real-valued, continuous *weight function* $w(x)$ and define the inner product as

$$\langle g|f\rangle = \int_{-\infty}^{\infty}\langle g|e_x\rangle\langle e_x|f\rangle w(x)dx = \int_{-\infty}^{\infty}g^*(x)f(x)w(x)dx. \quad (3.55)$$

If we fix $w(x) = 1$ in Eq. 3.55, we recover Eq. 3.54. It can be verified that the inner product defined in Eq. 3.55 satisfies all the properties of a general inner product described in Eqs. 3.1–3.3.

An important class of functions, which are used widely in many applications, is the space of square-integrable functions defined on a finite interval, $[a, b]$. These functions satisfy

$$\|f\|^2 = \langle f|f \rangle = \int_a^b f^*(x)f(x)w(x)dx = \int_a^b |f(x)|^2w(x)dx < \infty, \quad (3.56)$$

and the space of such functions is often denoted as $\mathcal{L}_w^2(a, b)$. The Riesz-Fischer theorem states that $\mathcal{L}_w^2(a, b)$ is complete, therefore a Hilbert space (for a proof, see e.g., MacCluer 2008). We will focus on this function space in our following discussions⁴.

For arbitrary $f, g \in \mathcal{L}_w^2(a, b)$, Eq. 3.55 can also be written as

$$\langle g|f \rangle = \langle g| \left(\int_a^b |e_x\rangle w(x) \langle e_x| dx \right) |f \rangle,$$

which suggests the identity

$$\mathbf{1} = \int_a^b |e_x\rangle w(x) \langle e_x| dx. \quad (3.57)$$

Equation 3.57 is the extension of the completeness relation in Eq. 3.51 to the continuous index case. Using Eq. 3.57, we can write

$$\begin{aligned} |f\rangle &= \mathbf{1}|f\rangle = \left(\int_a^b |e_x\rangle w(x) \langle e_x| dx \right) |f\rangle = \int_a^b |e_x\rangle w(x) \langle e_x|f\rangle dx \\ &= \int_a^b f(x)w(x) |e_x\rangle dx, \end{aligned} \quad (3.58)$$

where we have used Eq. 3.53 in the last equality. Equation 3.58 is the extension of the expansion formula, Eq. 3.49, to the continuous index case.

3.1.6 Dirac Delta Function

If we take the inner product with $|e_{x'}\rangle$, $a \leq x' \leq b$ on both sides of Eq. 3.58, we obtain

$$\langle e_{x'}|f\rangle = f(x') = \int_a^b f(x)w(x) \langle e_{x'}|e_x\rangle dx, \quad (3.59)$$

⁴ It can be shown that all infinite-dimensional complete inner product spaces are isomorphic to $\mathcal{L}_w^2(a, b)$ (e.g., Friedman 1970), therefore all Hilbert spaces are “alike”.

where I have used Eq. 3.53 at the first equality. We can break the integral on the RHS of Eq. 3.59 into three pieces,

$$\begin{aligned} f(x') &= \int_a^{x'-\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx + \int_{x'-\epsilon}^{x'+\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx \\ &\quad + \int_{x'+\epsilon}^b f(x)w(x) \langle e_{x'}|e_x \rangle dx. \end{aligned} \quad (3.60)$$

If we let $\epsilon \rightarrow 0$, the second term on the RHS can be approximated as

$$\int_{x'-\epsilon}^{x'+\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx \approx f(x') \int_{x'-\epsilon}^{x'+\epsilon} w(x) \langle e_{x'}|e_x \rangle dx. \quad (3.61)$$

The function $f(x)$ can be arbitrary. Suppose we pick a function such that $f(x') = 0$, considering Eq. 3.60–3.61, we have

$$\begin{aligned} f(x') &= \lim_{\epsilon \rightarrow 0} \left[\int_a^{x'-\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx \right] \\ &\quad + \lim_{\epsilon \rightarrow 0} \left[\int_{x'-\epsilon}^{x'+\epsilon} w(x) \langle e_{x'}|e_x \rangle dx \right] f(x') \\ &\quad + \lim_{\epsilon \rightarrow 0} \left[\int_{x'+\epsilon}^b f(x)w(x) \langle e_{x'}|e_x \rangle dx \right] \end{aligned} \quad (3.62)$$

$$\begin{aligned} \Rightarrow 0 &= \lim_{\epsilon \rightarrow 0} \left[\int_a^{x'-\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx \right] + 0 \\ &\quad + \lim_{\epsilon \rightarrow 0} \left[\int_{x'+\epsilon}^b f(x)w(x) \langle e_{x'}|e_x \rangle dx \right]. \end{aligned} \quad (3.63)$$

Since $f(x)$ can still be arbitrary on $[a, x' - \epsilon) \cup (x' + \epsilon, b]$, to satisfy Eq. 3.63 we must have

$$w(x) \langle e_{x'}|e_x \rangle = 0, \forall x \neq x', x \in [a, b]. \quad (3.64)$$

We now consider functions satisfying $f(x') \neq 0$. Equation 3.64 allows us to re-write Eq. 3.60 as

$$\begin{aligned} f(x') &= \lim_{\epsilon \rightarrow 0} \left[\int_{x'-\epsilon}^{x'+\epsilon} f(x)w(x) \langle e_{x'}|e_x \rangle dx \right] = f(x') \lim_{\epsilon \rightarrow 0} \left[\int_{x'-\epsilon}^{x'+\epsilon} w(x) \langle e_{x'}|e_x \rangle dx \right] \\ \Rightarrow 1 &= \lim_{\epsilon \rightarrow 0} \left[\int_{x'-\epsilon}^{x'+\epsilon} w(x) \langle e_{x'}|e_x \rangle dx \right] = \int_a^b w(x) \langle e_{x'}|e_x \rangle dx. \end{aligned} \quad (3.65)$$

In the last equality in Eq. 3.65, I have used Eq. 3.64 again. Considering Eq. 3.64, the only point that contributes to the integral on the RHS of Eq. 3.65 is at $x = x'$. We can break the integral on the RHS of Eq. 3.65 into 3 pieces $[a, x' - \epsilon) \cup$

$[x' - \epsilon, x' + \epsilon] \cup (x' + \epsilon, b]$ and let $\epsilon \rightarrow 0$. The intervals $[a, x' - \epsilon) \cup (x' + \epsilon, b]$ do not contribute to the integral, therefore we have

$$\begin{aligned} 1 &= \lim_{\epsilon \rightarrow 0} \left[\int_{x'-\epsilon}^{x'+\epsilon} w(x) \langle e_{x'} | e_x \rangle dx \right] = w(x') \langle e_{x'} | e_{x'} \rangle \lim_{\epsilon \rightarrow 0} 2\epsilon \\ &\Rightarrow w(x') \langle e_{x'} | e_{x'} \rangle = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} = \infty. \end{aligned} \quad (3.66)$$

If we define the Dirac delta function as

$$\delta(x - x') = w(x) \langle e_{x'} | e_x \rangle, \quad (3.67)$$

then we have the following properties

$$\int_a^b f(x) \delta(x - x') dx = f(x'), \quad \forall x, x' \in [a, b], \quad (3.68)$$

$$\delta(x - x') = \begin{cases} \infty, & x = x' \\ 0, & x \neq x' \end{cases}, \quad \forall x, x' \in [a, b], \quad (3.69)$$

$$\int_a^b \delta(x - x') dx = 1, \quad \forall x, x' \in [a, b]. \quad (3.70)$$

Equation 3.68 is also known as the *sifting property* of the Dirac delta function. Equation 3.69 shows that the Dirac delta function is not a function in the traditional sense. We can re-write the definition in Eq. 3.67 as

$$\langle e_{x'} | e_x \rangle = \frac{\delta(x - x')}{w(x)}, \quad (3.71)$$

which is the extension of the orthonormal condition, Eq. 3.8, to the continuous index case.

The argument of the Dirac delta function can be generalized to a function $g(x)$, which we assume to have continuous derivatives $g'(x)$ on $[a, b]$. Suppose the equation $g(x) = 0$ has a root at $x = x_0 \in [a, b]$, make the change of variable $y = g(x)$, we can write the integral as

$$\int_a^b f(x) \delta(g(x)) dx = \int_a^b f(x) \delta(y) \frac{dy}{|g'(x)|} = \frac{f(x_0)}{|g'(x_0)|}, \quad \text{if } g'(x_0) \neq 0. \quad (3.72)$$

If $g(x)$ has n roots $x_i \in [a, b], i = 1, 2, \dots, n$, using the same derivation for each root we can generalize Eq. 3.72 to

$$\int_a^b f(x) \delta(g(x)) dx = \sum_{i=1}^n \frac{f(x_i)}{|g'(x_i)|}, \quad \text{if } g'(x_i) \neq 0. \quad (3.73)$$

Equation 3.73 can be used to define

$$\delta(g(x)) = \sum_{i=1}^n \frac{\delta(x - x_i)}{|g'(x_i)|}, \quad g'(x_i) \neq 0. \quad (3.74)$$

A special case is when $g(x) = cx$, where c is a scalar, and we have

$$\delta(cx) = \frac{\delta(x)}{|c|}. \quad (3.75)$$

If we fix $c = -1$, Eq. 3.75 states

$$\delta(-x) = \delta(x), \quad (3.76)$$

which shows that the Dirac delta function is an even function.

The derivatives of the Dirac delta function are not functions in the traditional sense either. The first-order derivative can be computed as

$$\begin{aligned} \int_a^b f(x)\delta'(x - x')dx &= f(x)\delta(x - x')\Big|_{x=a}^{x=b} - \int_a^b f'(x)\delta(x - x')dx \\ &= -f'(x'), \quad \forall x' \in (a, b), \end{aligned} \quad (3.77)$$

where I have applied integration-by-parts at the first equality. Higher-order derivatives of the Dirac delta function can be evaluated by repeated application of integration-by-parts,

$$\int_a^b f(x)\delta^{(n)}(x - x')dx = (-1)^n f^{(n)}(x'), \quad \forall x' \in (a, b), \quad (3.78)$$

where the superscript (n) denotes the order of the derivative while the superscript n is an exponent. In fact, in most applications, these unconventional functions appear inside an integral, which allows us to describe their behavior indirectly using their actions inside the integral on well-behaved conventional functions. A *distribution* or *generalized function* ϕ is a continuous linear functional that maps compactly supported, infinitely differentiable *test functions* f to a scalar defined as

$$\phi[f] = \int_{-\infty}^{\infty} \phi(x)f(x)dx. \quad (3.79)$$

Another notation commonly used in place of $\phi[f]$ is $\langle \phi, f \rangle$. Note that this notation is different from the Dirac *bra-ket* notation for vectors and inner products. An example of a compactly supported, infinitely differentiable test function is

$$f(x) = \begin{cases} e^{-\frac{1}{1-x^2}}, & |x| < 1 \\ 0, & \text{otherwise} \end{cases}. \quad (3.80)$$

It can be verified that this function has continuous derivatives of all orders at all $x \in \mathbb{R}$. It is nonzero on the finite interval $(-1, 1)$, therefore is also compactly supported. The space of test functions is often denoted as $\mathcal{C}_c^\infty(\mathbb{R})$.

Because the distribution is linear with respect to the test functions, we have

$$\langle \phi, \alpha f + \beta g \rangle = \alpha \langle \phi, f \rangle + \beta \langle \phi, g \rangle, \forall \alpha, \beta \in \mathbb{C}, \forall f, g \in \mathcal{C}_c^\infty(\mathbb{R}). \quad (3.81)$$

The set of all such linear functionals also forms a vector space over \mathbb{R} , therefore if ϕ_1 and ϕ_2 are two distributions, we have

$$\langle \alpha \phi_1 + \beta \phi_2, f \rangle = \alpha \langle \phi_1, f \rangle + \beta \langle \phi_2, f \rangle, \forall \alpha, \beta \in \mathbb{R}, \forall f \in \mathcal{C}_c^\infty(\mathbb{R}). \quad (3.82)$$

The *derivative of a distribution* ϕ is another distribution, denoted as ϕ' , defined by

$$\langle \phi', f \rangle = -\langle \phi, f' \rangle, \forall f \in \mathcal{C}_c^\infty(\mathbb{R}). \quad (3.83)$$

Higher-order derivatives of a distribution can be derived from Eq. 3.83. Because the derivatives on the test function are well defined for all orders, the derivatives of a distribution are also well defined for all orders.

For the Dirac delta function, the corresponding distribution is

$$\delta[f] = \int_{-\infty}^{\infty} \delta(x)f(x)dx = f(0) \quad (3.84)$$

for every test function $f(x)$. For the Dirac delta function $\delta(x - x')$, the corresponding distribution is defined as

$$\delta_{x'}[f] = \int_{-\infty}^{\infty} \delta(x - x')f(x)dx = f(x'). \quad (3.85)$$

Similarly, for $\delta'(x - x')$, the corresponding distribution is

$$\begin{aligned} \delta'_{x'}[f] &= \int_{-\infty}^{\infty} \delta'(x - x')f(x)dx \\ &= f(x)\delta(x - x')\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)\delta(x - x')dx \\ &= -f'(x'). \end{aligned} \quad (3.86)$$

Note that because the test function $f(x)$ is compact (i.e., nonzero on a finite set, therefore $f(\pm\infty) = 0$), the term $f(x)\delta(x - x')\Big|_{-\infty}^{\infty}$ is always 0 even when the Dirac delta function $\delta(x - x')$ is replaced with other functions that are nonzero at infinity. We can also obtain the same result by applying Eq. 3.83 directly.

3.1.7 Fourier Transforms

As an example, we apply the machinery we have developed in Sects. 3.1.4–3.1.6 to gain more insight into the Fourier transform pair defined as

$$\hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\xi x} f(x) dx, \quad (3.87)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi x} \hat{f}(\xi) d\xi. \quad (3.88)$$

Equation 3.88 can be written in vector form as

$$\begin{aligned} \langle e_x | f \rangle &= f(x) = \int_{-\infty}^{\infty} \langle e_x | \hat{e}_\xi \rangle \langle \hat{e}_\xi | \hat{f} \rangle d\xi \\ &= \langle e_x | \left(\int_{-\infty}^{\infty} |\hat{e}_\xi\rangle \langle \hat{e}_\xi| d\xi \right) | \hat{f} \rangle, \end{aligned} \quad (3.89)$$

where we have introduced the definition

$$\langle e_x | \hat{e}_\xi \rangle = \frac{e^{i\xi x}}{\sqrt{2\pi}}, \quad (3.90)$$

$$\langle \hat{e}_\xi | \hat{f} \rangle = \langle \hat{e}_\xi | f \rangle = \hat{f}(\xi) = \int_{-\infty}^{\infty} \frac{e^{-i\xi x}}{\sqrt{2\pi}} f(x) dx. \quad (3.91)$$

The first equality in Eq. 3.91 implies the identity

$$|\hat{f}\rangle = |f\rangle, \quad (3.92)$$

which is counter-intuitive at first glance because $\hat{f}(\xi)$ and $f(x)$ are clearly different functions. However, this identity is nevertheless correct in the context of vector representations of functions, because both $|f\rangle$ and $|\hat{f}\rangle$ represent the same vector and the two different functions $f(x)$ and $\hat{f}(\xi)$ are the coefficients of the same vector with respect to two different orthonormal bases $\{|e_x\rangle\}_{x \in \mathbb{R}}$ and $\{|\hat{e}_\xi\rangle\}_{\xi \in \mathbb{R}}$. To claim $\{|\hat{e}_\xi\rangle\}_{\xi \in \mathbb{R}}$ is an orthonormal basis, it must be complete. Considering the fact that $|f\rangle$ is arbitrary, Eq. 3.89 implies

$$\mathbf{1} = \int_{-\infty}^{\infty} |\hat{e}_\xi\rangle \langle \hat{e}_\xi| d\xi, \quad (3.93)$$

which shows that $\{|\hat{e}_\xi\rangle\}_{\xi \in \mathbb{R}}$ satisfies the continuous version of the completeness relation in \mathcal{H} , Eq. 3.57, with the weighting function $w(x) = 1$. To show that $\{|\hat{e}_\xi\rangle\}_{\xi \in \mathbb{R}}$ satisfies the continuous version of the orthonormal condition, Eq. 3.71, we compute the inner product

$$\begin{aligned}
\langle \hat{e}_\xi | \hat{e}_{\xi'} \rangle &= \langle \hat{e}_\xi | \mathbf{1} | \hat{e}_{\xi'} \rangle = \langle \hat{e}_\xi | \left(\int_{-\infty}^{\infty} |e_x\rangle \langle e_x| dx \right) | \hat{e}_{\xi'} \rangle = \int_{-\infty}^{\infty} \langle \hat{e}_\xi | e_x \rangle \langle e_x | \hat{e}_{\xi'} \rangle dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix(\xi' - \xi)} dx = \delta(\xi - \xi'), \tag{3.94}
\end{aligned}$$

where I have used the completeness of $\{|e_x\rangle_{x \in \mathbb{R}}\}$, Eq. 3.57, at the second equality and Eq. 3.90 at the fourth equality.

To prove the last equality in Eq. 3.94, consider the limit

$$\lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T e^{ix(\xi' - \xi)} dx = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \frac{e^{ix(\xi' - \xi)}}{i(\xi' - \xi)} \Bigg|_{-T}^T = \lim_{T \rightarrow \infty} \frac{T}{\pi} \left[\frac{\sin T(\xi' - \xi)}{T(\xi' - \xi)} \right].$$

The function inside the square bracket is known as the sinc function (Prudnikov 1986), which has the limit of 1 as $\xi' \rightarrow \xi$. Therefore we have

$$\lim_{\xi' \rightarrow \xi} \lim_{T \rightarrow \infty} \frac{T}{\pi} \left[\frac{\sin T(\xi' - \xi)}{T(\xi' - \xi)} \right] = \lim_{T \rightarrow \infty} \frac{T}{\pi} = \infty.$$

The width of the sinc function is given roughly by the distance between the points at which the sinc function drops from 1 to 0, i.e., $T(\xi' - \xi) = \pm\pi$, or

$$\xi' - \xi = \frac{2\pi}{T},$$

which approaches 0 as $T \rightarrow \infty$. Using methods from complex analysis, it can be shown that

$$\int_{-\infty}^{\infty} \frac{T}{\pi} \left[\frac{\sin T(\xi' - \xi)}{T(\xi' - \xi)} \right] d\xi' = 1.$$

These properties, when compared with Eqs. 3.68–3.70, suggest that

$$\delta(\xi' - \xi) = \lim_{T \rightarrow \infty} \frac{T}{\pi} \left[\frac{\sin T(\xi' - \xi)}{T(\xi' - \xi)} \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix(\xi' - \xi)} dx. \tag{3.95}$$

In fact, the Dirac delta function can be defined as the limits of not only the sinc function but also a range of other functions, such as the Gaussian and the box-car functions (e.g., Hassani 2008).

The analysis above shows that an arbitrary abstract vector $|f\rangle$ can be represented using its coefficients with respect to $\{|e_x\rangle\}_{x \in \mathbb{R}}$, $f(x) = \langle e_x | f \rangle$, or its coefficients with respect to $\{|\hat{e}_\xi\rangle\}_{\xi \in \mathbb{R}}$, $\hat{f}(\xi) = \langle \hat{e}_\xi | f \rangle$. Both representations are completely equivalent and we can choose the representation that is most convenient for a given application.

We can apply the same analysis to the n -dimensional Fourier transform pair

$$\hat{f}(\boldsymbol{\xi}) = \frac{1}{(2\pi)^{n/2}} \int e^{-i\boldsymbol{\xi} \cdot \mathbf{x}} f(\mathbf{x}) d^n x, \tag{3.96}$$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \int e^{i\boldsymbol{\xi} \cdot \mathbf{x}} \hat{f}(\boldsymbol{\xi}) d^n \boldsymbol{\xi}, \tag{3.97}$$

with the inner product

$$\langle e_{\mathbf{x}} | \hat{e}_{\boldsymbol{\xi}} \rangle = \frac{e^{i\boldsymbol{\xi} \cdot \mathbf{x}}}{(2\pi)^{n/2}}. \quad (3.98)$$

Here $\mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^n$ and the dot product $\boldsymbol{\xi} \cdot \mathbf{x} = \sum_{i=1}^n \xi_i x_i$ has the usual Euclidean geometric definition.

Consider two functions $f(t)$ and $g(t)$, which can be expressed using Eq. 3.88 as

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi t} \hat{f}(\xi) d\xi, \quad (3.99)$$

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi' t} \hat{g}(\xi') d\xi'. \quad (3.100)$$

The inner product of $f(t)$ and $g(t)$ can be expressed as

$$\begin{aligned} \int_{-\infty}^{\infty} g^*(t) f(t) dt &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \left[\int_{-\infty}^{\infty} e^{-i\xi' t} \hat{g}^*(\xi') d\xi' \right] \left[\int_{-\infty}^{\infty} e^{i\xi t} \hat{f}(\xi) d\xi \right] \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi' \int_{-\infty}^{\infty} d\xi \hat{g}^*(\xi') \hat{f}(\xi) \left[\int_{-\infty}^{\infty} e^{i(\xi - \xi') t} dt \right] \\ &= \int_{-\infty}^{\infty} d\xi' \int_{-\infty}^{\infty} d\xi \hat{g}^*(\xi') \hat{f}(\xi) \delta(\xi - \xi') \\ &= \int_{-\infty}^{\infty} \hat{g}^*(\xi) \hat{f}(\xi) d\xi, \end{aligned} \quad (3.101)$$

where I have used Eq. 3.95 at the third equality and Eq. 3.68 at the fourth equality. Equation 3.101 is known as the *Parseval theorem*.

3.1.8 Operator Algebra

A vector space equipped with a *vector-vector multiplication* is called an *algebra*. Note that the vector-vector multiplication we are discussing here is different from the scalar-vector multiplication used in defining vector spaces (Sect. 3.1.1). In Sect. 3.1.3, it is shown that the set of linear transformations $\mathcal{L}(\mathcal{V}, \mathcal{W})$ forms a vector space and linear transformations are vectors in this space. We can define a multiplication on this vector space through the composition of linear transformations. Suppose we have two linear transformations, $\mathbf{T}_1 : \mathcal{U} \rightarrow \mathcal{V}$ and $\mathbf{T}_2 : \mathcal{V} \rightarrow \mathcal{W}$, then the multiplication $\mathbf{T}_2 \mathbf{T}_1 : \mathcal{U} \rightarrow \mathcal{W}$ can be defined as

$$\mathbf{T}_2 \mathbf{T}_1 |x\rangle = \mathbf{T}_2 (\mathbf{T}_1 |x\rangle), \quad \forall |x\rangle \in \mathcal{U}. \quad (3.102)$$

This definition involves three vector spaces, $\mathbf{T}_1 \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, $\mathbf{T}_2 \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $\mathbf{T}_2 \mathbf{T}_1 \in \mathcal{L}(\mathcal{U}, \mathcal{W})$. If we let $\mathcal{U} = \mathcal{V} = \mathcal{W}$, the three spaces all become $\mathcal{L}(\mathcal{V}, \mathcal{V})$, which we abbreviate as $\mathcal{L}(\mathcal{V})$. A linear transformation $\mathbf{T} \in \mathcal{L}(\mathcal{V})$ is called a *linear*

operator and the space $\mathcal{L}(\mathcal{V})$ is the algebra of linear operators. Note that in general operator multiplications are not commutative, i.e., $\mathbf{T}_1\mathbf{T}_2 \neq \mathbf{T}_2\mathbf{T}_1, \forall \mathbf{T}_1, \mathbf{T}_2 \in \mathcal{L}(\mathcal{V})$.

Once we have a multiplication defined, we can discuss the inverse of the operator \mathbf{T} , which is denoted as \mathbf{T}^{-1} and satisfies the condition

$$\mathbf{T}^{-1}\mathbf{T} = \mathbf{T}\mathbf{T}^{-1} = \mathbf{1}. \quad (3.103)$$

Not all linear operators have inverses. If \mathbf{T}_1 and \mathbf{T}_2 are both invertible, then $\mathbf{T}_1\mathbf{T}_2$ is also invertible and has the inverse

$$(\mathbf{T}_1\mathbf{T}_2)^{-1} = \mathbf{T}_2^{-1}\mathbf{T}_1^{-1}. \quad (3.104)$$

The powers of an operator \mathbf{T} can be defined through induction

$$\mathbf{T}^n = \mathbf{T}^{n-1}\mathbf{T} = \mathbf{T}\mathbf{T}^{n-1}, \forall n \in \mathbb{N} \text{ and } \mathbf{T}^0 = \mathbf{1}. \quad (3.105)$$

Negative powers of \mathbf{T} can be defined in terms of \mathbf{T}^{-1} ,

$$\mathbf{T}^{-n} = \left(\mathbf{T}^{-1}\right)^n, \forall n \in \mathbb{N}, \quad (3.106)$$

as long as \mathbf{T} is invertible. For any two integers m and n , we have

$$\mathbf{T}^m\mathbf{T}^n = \mathbf{T}^{m+n}, \quad (3.107)$$

$$\left(\mathbf{T}^m\right)^n = \mathbf{T}^{mn}. \quad (3.108)$$

When $m = -n$, Eq. 3.107 and $\mathbf{T}^0 = \mathbf{1}$ imply that the inverse of \mathbf{T}^n is \mathbf{T}^{-n} . The powers of \mathbf{T} can be generalized to include all real numbers using the spectral theory of operators (e.g., Kubrusly 2012).

In Sect. 3.1.3, some concepts about linear transformations were demonstrated using matrices. The procedure can be formalized for finite-dimensional vector spaces. Suppose we are trying to solve an abstract operator equation

$$\mathbf{T}|g\rangle = |f\rangle \quad (3.109)$$

for $|g\rangle$ in the n -dimensional vector space \mathcal{V} . We can choose an orthonormal basis $B = \{|e_j\rangle\}_{j=1}^n$ for \mathcal{V} . Considering the completeness relation, Eq. 3.34, the LHS of Eq. 3.109 can be written as

$$\mathbf{T}|g\rangle = \mathbf{T}\mathbf{1}|g\rangle = \mathbf{T}\sum_{j=1}^n |e_j\rangle \langle e_j|g\rangle = \sum_{j=1}^n \mathbf{T}|e_j\rangle \langle e_j|g\rangle. \quad (3.110)$$

Bring Eq. 3.110 back into Eq. 3.109 and take the inner product with $|e_i\rangle$, $i = 1, 2, \dots, n$ on both sides of Eq. 3.109, we obtain

$$\sum_{j=1}^n \langle e_i|\mathbf{T}|e_j\rangle \langle e_j|g\rangle = \langle e_i|f\rangle, i = 1, 2, \dots, n. \quad (3.111)$$

If we define the elements of the matrix as

$$A_{ij} \equiv \langle e_i | \mathbf{T} | e_j \rangle, \quad (3.112)$$

and the elements of vectors

$$g_j \equiv \langle e_j | g \rangle, f_i \equiv \langle e_i | f \rangle, \quad (3.113)$$

we obtain the matrix-vector representation of the abstract operator equation

$$\sum_{j=1}^n A_{ij} g_j = f_i, \quad (3.114)$$

which can be solved for the coefficients of vector $|g\rangle$ with respect to the basis B . In finite-dimensional vector spaces, we can establish a one-to-one correspondence between operators and matrices, therefore operator algebra is essentially matrix algebra.

The same procedure can also be applied to the infinite-dimensional, continuous index case using the basis $B = \{|e_x\rangle\}_{x \in \mathbb{R}}$. Different vectors from this basis can be distinguished using different subscript, such as $|e_x\rangle, |e_y\rangle, x, y \in \mathbb{R}$. Using the continuous version of the completeness relation, Eq. 3.57, the operator Eq. 3.109 can be written as

$$\begin{aligned} \int_a^b \langle e_x | \mathbf{T} | e_y \rangle w(y) \langle e_y | g \rangle dy &= \langle e_x | f \rangle \\ \Rightarrow \int_a^b A(x, y) w(y) g(y) dy &= f(x), \end{aligned} \quad (3.115)$$

which is an *integral equation* with its *integration kernel* given by

$$A(x, y) \equiv \langle e_x | \mathbf{T} | e_y \rangle. \quad (3.116)$$

When we generalized finite-dimensional vector spaces to infinite-dimensional vector spaces in Sect. 3.1.4, we were concerned about the convergence, limits and completeness in infinite dimensions. Even though the analogy between matrices and integration kernels is obvious, we have similar concerns when we try to replace the matrix A_{ij} with the integration kernel $A(x, y)$.

A class of operators that are highly useful in practical applications are *bounded operators*. Suppose \mathcal{H}_1 and \mathcal{H}_2 are two Hilbert spaces equipped with the norm $\| \cdot \|_1$ and $\| \cdot \|_2$, respectively, and $\mathbf{T} \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$, the norm of \mathbf{T} can be defined as

$$\|\mathbf{T}\| = \max \left\{ \frac{\|\mathbf{T}x\|_2}{\|x\|_1}, |x\rangle \in \mathcal{H}_1 \text{ and } |x\rangle \neq |0\rangle \right\}. \quad (3.117)$$

The denominator $\|x\|_1$ in the definition makes the norm of \mathbf{T} independent of the vector length. Alternatively, the norm can equivalently be defined as

$$\|\mathbf{T}\| = \max \{ \|\mathbf{T}x\|_2, \|x\|_1 \leq 1 \}. \quad (3.118)$$

If $\|\mathbf{T}\| < \infty$, we call \mathbf{T} a *bounded linear transformation*. A bounded linear transformation from a Hilbert space to itself is called a bounded operator. A direct consequence of this definition is

$$\|\mathbf{T}x\|_2 \leq \|\mathbf{T}\| \|x\|_1, \forall |x\rangle \in \mathcal{H}_1. \quad (3.119)$$

The space of bounded operators is often denoted as $\mathcal{B}(\mathcal{H})$. It can be verified that $\mathcal{B}(\mathcal{H})$ is a subspace of $\mathcal{L}(\mathcal{H})$. In Eq. 3.119, if we replace $|x\rangle$ with $|x\rangle - |y\rangle$, we have

$$\|\mathbf{T}x - \mathbf{T}y\|_2 \leq \|\mathbf{T}\| \|x - y\|_1. \quad (3.120)$$

Since $\|\mathbf{T}\|$ is finite, if $|x\rangle$ and $|y\rangle$ are arbitrarily close to each other in \mathcal{H}_1 , $\mathbf{T}|x\rangle$ and $\mathbf{T}|y\rangle$ are arbitrarily close to each other in \mathcal{H}_2 , which means that all bounded operators are *continuous*.

We can apply the definition of the operator norm, Eq. 3.117, to the product of two operators $\forall \mathbf{T}_1, \mathbf{T}_2 \in \mathcal{L}(\mathcal{H})$. The norm of the product $\mathbf{T}_2\mathbf{T}_1$ can be written as

$$\begin{aligned} \|\mathbf{T}_2\mathbf{T}_1\| &= \max \left\{ \frac{\|\mathbf{T}_2\mathbf{T}_1x\|}{\|x\|}, |x\rangle \in \mathcal{H}, |x\rangle \neq |0\rangle \right\} \\ &= \max \left\{ \frac{\|\mathbf{T}_2\mathbf{T}_1x\|}{\|\mathbf{T}_1x\|} \frac{\|\mathbf{T}_1x\|}{\|x\|}, |x\rangle \in \mathcal{H}, |x\rangle \neq |0\rangle \neq \mathbf{T}_1|x\rangle \right\} \\ &\leq \max \left\{ \frac{\|\mathbf{T}_2(\mathbf{T}_1|x\rangle)\|}{\|\mathbf{T}_1|x\rangle\|}, \mathbf{T}_1|x\rangle \neq |0\rangle \right\} \max \left\{ \frac{\|\mathbf{T}_1x\|}{\|x\|}, |x\rangle \neq |0\rangle \right\} \\ &= \max \left\{ \frac{\|\mathbf{T}_2(\mathbf{T}_1|x\rangle)\|}{\|\mathbf{T}_1|x\rangle\|}, \mathbf{T}_1|x\rangle \neq |0\rangle \right\} \|\mathbf{T}_1\|. \end{aligned}$$

Note that in the first term of the RHS, the vector $\mathbf{T}_1|x\rangle$ is in the range of operator \mathbf{T}_1 , which is a subspace of \mathcal{H} . If we include all vectors in \mathcal{H} into our scan for maximality, we may find a larger number. Therefore we have

$$\begin{aligned} &\max \left\{ \frac{\|\mathbf{T}_2(\mathbf{T}_1|x\rangle)\|}{\|\mathbf{T}_1|x\rangle\|}, \mathbf{T}_1|x\rangle \neq |0\rangle \right\} \\ &\leq \max \left\{ \frac{\|\mathbf{T}_2y\|}{\|y\|}, |y\rangle \in \mathcal{H}, |y\rangle \neq |0\rangle \right\} = \|\mathbf{T}_2\|. \end{aligned}$$

Therefore we have shown that

$$\|\mathbf{T}_2\mathbf{T}_1\| \leq \|\mathbf{T}_2\| \|\mathbf{T}_1\|. \quad (3.121)$$

A direct consequence of Eq. 3.121 is

$$\|\mathbf{T}^n\| \leq \|\mathbf{T}\|^n. \quad (3.122)$$

Using Eq. 3.122, we can show that the series $\sum_{n=0}^{\infty} \mathbf{T}^n$ converges if $\|\mathbf{T}\| \leq 1$. The norm of this series can be expressed as

$$\left\| \sum_{n=0}^{\infty} \mathbf{T}^n \right\| \leq \sum_{n=0}^{\infty} \|\mathbf{T}^n\| \leq \sum_{n=0}^{\infty} \|\mathbf{T}\|^n = \frac{1}{1 - \|\mathbf{T}\|}, \quad (3.123)$$

which is finite for $\mathbf{T} \in \mathcal{B}(\mathcal{H})$, $\|\mathbf{T}\| \leq 1$. In the last equality in Eq. 3.123, we used the property of a geometric series because $\|\mathbf{T}\|$ is merely a scalar. In fact, we can show that this series converges to the inverse of the operator $\mathbf{1} - \mathbf{T}$. Let's consider the operator product

$$\begin{aligned} (\mathbf{1} - \mathbf{T}) \sum_{n=0}^{\infty} \mathbf{T}^n &= (\mathbf{1} - \mathbf{T}) \left(\lim_{l \rightarrow \infty} \sum_{n=0}^l \mathbf{T}^n \right) = \lim_{l \rightarrow \infty} (\mathbf{1} - \mathbf{T}) \sum_{n=0}^l \mathbf{T}^n \\ &= \lim_{l \rightarrow \infty} \left(\sum_{n=0}^l \mathbf{T}^n - \sum_{n=1}^{l+1} \mathbf{T}^n \right) \\ &= \lim_{l \rightarrow \infty} (\mathbf{1} - \mathbf{T}^{l+1}) = \mathbf{1}. \end{aligned}$$

The last equality holds because $\lim_{l \rightarrow \infty} \|\mathbf{T}^{l+1}\| \leq \lim_{l \rightarrow \infty} \|\mathbf{T}\|^{l+1} = 0$ if $\|\mathbf{T}\| < 1$ and the operator with 0 norm is the zero operator. Considering the definition of the operator inverse, Eq. 3.103, we therefore have

$$(\mathbf{1} - \mathbf{T})^{-1} = \sum_{n=0}^{\infty} \mathbf{T}^n, \text{ if } \mathbf{T} \in \mathcal{B} \text{ and } \|\mathbf{T}\| < 1. \quad (3.124)$$

Equation 3.124 can be generalized to the operator \mathbf{T}/α , $\alpha \in \mathbb{C}$ and we have

$$(\alpha \mathbf{1} - \mathbf{T})^{-1} = \alpha^{-1} \sum_{n=0}^{\infty} \left(\frac{\mathbf{T}}{\alpha} \right)^n, \text{ if } \mathbf{T} \in \mathcal{B} \text{ and } \|\mathbf{T}\| < |\alpha|. \quad (3.125)$$

Consider two operators $\mathbf{T}_1, \mathbf{T}_2 \in \mathcal{B}(\mathcal{H})$, if \mathbf{T}_1 is invertible and \mathbf{T}_2 satisfies the relation

$$\|\mathbf{T}_1 - \mathbf{T}_2\| < \frac{1}{\|\mathbf{T}_1^{-1}\|} \quad (3.126)$$

then \mathbf{T}_2 is also invertible. To show this, consider $\mathbf{P} \equiv \mathbf{T}_1^{-1} (\mathbf{T}_1 - \mathbf{T}_2)$, its norm is

$$\|\mathbf{P}\| \leq \|\mathbf{T}_1^{-1}\| \|\mathbf{T}_1 - \mathbf{T}_2\| < \|\mathbf{T}_1^{-1}\| \frac{1}{\|\mathbf{T}_1^{-1}\|} = 1,$$

therefore the operator $\mathbf{1} - \mathbf{P}$ is invertible considering Eq. 3.124. The operator $\mathbf{1} - \mathbf{P}$ can be expressed as

$$\mathbf{1} - \mathbf{P} = \mathbf{1} - \mathbf{T}_1^{-1} (\mathbf{T}_1 - \mathbf{T}_2) = \mathbf{1} - \mathbf{1} + \mathbf{T}_1^{-1} \mathbf{T}_2 = \mathbf{T}_1^{-1} \mathbf{T}_2,$$

therefore the operator $\mathbf{Q} \equiv \mathbf{T}_1^{-1} \mathbf{T}_2$ is invertible. Since we have $\mathbf{T}_2 = \mathbf{T}_1 \mathbf{Q}$ and both \mathbf{T}_1 and \mathbf{Q} are invertible, considering Eq. 3.104, \mathbf{T}_2 must also be invertible. The LHS in Eq. 3.126 is actually a measure of the “distance” between two operators. The fact that \mathbf{T}_2 is invertible means that operators that are sufficiently “close” to an invertible operator are also invertible.

In Sect. 3.1.3, the dual of a linear transformation and the concept of the adjoint matrix in finite-dimensional vector spaces were introduced. For a linear operator $\mathbf{T} \in \mathcal{B}(\mathcal{H})$, its *adjoint* or *Hermitian conjugate*, often denoted as \mathbf{T}^\dagger , is defined by the relation

$$\langle f | \mathbf{T} | g \rangle^* = \langle g | \mathbf{T}^\dagger | f \rangle, \forall |f\rangle, |g\rangle \in \mathcal{H}, \forall \mathbf{T} \in \mathcal{B}(\mathcal{H}), \quad (3.127)$$

which can also be written equivalently as

$$\langle \mathbf{T} g | f \rangle = \langle g | \mathbf{T}^\dagger f \rangle, \forall |f\rangle, |g\rangle \in \mathcal{H}, \forall \mathbf{T} \in \mathcal{B}(\mathcal{H}). \quad (3.128)$$

It can be verified that for arbitrary $\mathbf{S}, \mathbf{T} \in \mathcal{B}(\mathcal{H})$ and $\alpha \in \mathbb{C}$, the adjoint satisfies the following properties

$$(\alpha \mathbf{T})^\dagger = \alpha^* \mathbf{T}^\dagger, \quad (3.129)$$

$$(\mathbf{S} + \mathbf{T})^\dagger = \mathbf{S}^\dagger + \mathbf{T}^\dagger, \quad (3.130)$$

$$(\mathbf{S}\mathbf{T})^\dagger = \mathbf{T}^\dagger \mathbf{S}^\dagger, \quad (3.131)$$

$$\left(\mathbf{T}^\dagger\right)^\dagger = \mathbf{T}. \quad (3.132)$$

An operator \mathbf{T} is called *self-adjoint* or *Hermitian* if

$$\mathbf{T}^\dagger = \mathbf{T}, \quad (3.133)$$

and called *anti-self-adjoint* or *anti-Hermitian* if

$$\mathbf{T}^\dagger = -\mathbf{T}. \quad (3.134)$$

It can be verified that any operator \mathbf{T} can be decomposed into a self-adjoint operator

$$\mathbf{H} = \frac{1}{2} (\mathbf{T} + \mathbf{T}^\dagger) \quad (3.135)$$

plus an anti-self-adjoint operator

$$\mathbf{A} = \frac{1}{2} (\mathbf{T} - \mathbf{T}^\dagger) \quad (3.136)$$

and we have the decomposition

$$\mathbf{T} = \mathbf{H} + \mathbf{A}. \quad (3.137)$$

If \mathbf{T} is self-adjoint, we can show that $\langle f | \mathbf{T} f \rangle$ is real for any $|f\rangle \in \mathcal{H}$. In fact the complex conjugate of $\langle f | \mathbf{T} f \rangle$ can be written as

$$\langle f | \mathbf{T} f \rangle^* = \langle \mathbf{T} f | f \rangle = \langle f | \mathbf{T}^\dagger f \rangle = \langle f | \mathbf{T} f \rangle, \quad (3.138)$$

where I have used Eq. 3.1 at the first equality, Eq. 3.128 at the second equality and Eq. 3.133 at the last equality. Conversely, we can show that if $\langle f|\mathbf{T}f\rangle$ is real for any $f \in \mathcal{H}$, then \mathbf{T} is self-adjoint. If $\langle f|\mathbf{T}f\rangle$ is real, then we have

$$\begin{aligned}\langle f|\mathbf{T}f\rangle &= \langle f|\mathbf{T}f\rangle^* = \langle \mathbf{T}f|f\rangle = \langle f|\mathbf{T}^\dagger f\rangle \\ \Rightarrow \langle f|(\mathbf{T} - \mathbf{T}^\dagger)f\rangle &= 0, \forall f \in \mathcal{H} \Rightarrow \mathbf{T} = \mathbf{T}^\dagger.\end{aligned}$$

If \mathbf{T} is self-adjoint and $\langle f|\mathbf{T}f\rangle \geq 0, \forall |f\rangle \in \mathcal{H}$, we call \mathbf{T} a *positive operator*. If a positive operator \mathbf{T} satisfies the extra condition that $\langle f|\mathbf{T}f\rangle = 0$ implies $|f\rangle = |0\rangle$, then \mathbf{T} is called a *positive definite operator*.

The norm of the adjoint operator \mathbf{T}^\dagger satisfies the identity

$$\|\mathbf{T}^\dagger\| = \|\mathbf{T}\|. \quad (3.139)$$

To verify this identity, consider the inner product $\langle \mathbf{T}f|\mathbf{T}f\rangle$ for $|f\rangle \in \mathcal{H}$ and $\|f\| \leq 1$,

$$\begin{aligned}\|\mathbf{T}f\|^2 &= \langle \mathbf{T}f|\mathbf{T}f\rangle = \langle f|\mathbf{T}^\dagger\mathbf{T}f\rangle \leq \|f\| \|\mathbf{T}^\dagger\mathbf{T}f\| \leq \|f\|^2 \|\mathbf{T}^\dagger\mathbf{T}\| \\ &\leq \|\mathbf{T}^\dagger\mathbf{T}\| \leq \|\mathbf{T}^\dagger\| \|\mathbf{T}\|.\end{aligned}$$

Here I used the Cauchy-Schwarz inequality (Eq. 3.12) at the first inequality in the equation above, used Eq. 3.119 at the second inequality and Eq. 3.121 at the last inequality. Considering the alternative definition of the operator norm, Eq. 3.118, we have $\|\mathbf{T}\|^2 \leq \|\mathbf{T}^\dagger\| \|\mathbf{T}\| \Rightarrow \|\mathbf{T}\| \leq \|\mathbf{T}^\dagger\|$. If we replace \mathbf{T} with \mathbf{T}^\dagger in the derivation above and use Eq. 3.132, we can arrive at $\|\mathbf{T}^\dagger\| \leq \|\mathbf{T}\|$. Therefore the identity in Eq. 3.139 must hold.

In this very brief discussion about linear operators, I have neglected many important topics, such as the spectral theory of operators, which is concerned with the “diagonalization” of operators and generalization of eigenvalues and eigenvectors. For readers who are interested in this subject and its applications, I refer to (DeVito 1990; Sunder 1997; Kubrusly 2012).

3.1.9 Fréchet Derivative

In calculus, the definition of the derivative for real valued functions, $f: \mathbb{R} \rightarrow \mathbb{R}$ is

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}, \quad (3.140)$$

if the limit exists. In seismic inverse problems, the independent “variable” is usually not $x \in \mathbb{R}$, but some function $m(\mathbf{x})$ that might belong to a certain Hilbert space. We therefore need to extend the notion of differentiation to general normed vector

spaces. The reason for requiring a norm is that it is indispensable in discussing continuity. The derivative defined on general normed vector spaces is often called the *Fréchet derivative*.

If the limit in Eq. 3.140 exists, we can define a function of h as

$$\phi(h) = \begin{cases} \frac{f(x+h)-f(x)}{h} - f'(x), & h \neq 0 \\ 0 & h = 0 \end{cases}. \quad (3.141)$$

It can be verified that $\phi(h)$ is a continuous function of h everywhere, including the point $h = 0$. We therefore have the equation

$$f(x+h) = f(x) + f'(x)h + h\phi(h). \quad (3.142)$$

We can now generalize Eq. 3.142 to normed vector spaces. Suppose \mathcal{V} and \mathcal{W} are vector spaces equipped with the norm $\|\cdot\|_{\mathcal{V}}$ and $\|\cdot\|_{\mathcal{W}}$ and $f: \mathcal{V} \supset \mathcal{U} \rightarrow \mathcal{W}$, we say f is *differentiable* at $|x\rangle \in \mathcal{U}$ if there exists a linear transformation, often denoted as $\mathbf{D}f(|x\rangle) \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, and also a continuous function $\phi(|h\rangle), |h\rangle \in \mathcal{V}$ with $\phi(|0\rangle) = 0$, such that the equation

$$f(|x\rangle + |h\rangle) = f(|x\rangle) + [\mathbf{D}f(|x\rangle)]|h\rangle + \|h\|_{\mathcal{V}}\phi(|h\rangle) \quad (3.143)$$

holds for all $|h\rangle$ in an open neighborhood of $|0\rangle \in \mathcal{V}$. The linear transformation $\mathbf{D}f(|x\rangle) \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is called the Fréchet derivative of f at $|x\rangle$. An equivalent definition that resembles Eq. 3.140 is

$$\lim_{\|h\|_{\mathcal{V}} \rightarrow 0} \frac{\|f(|x\rangle + |h\rangle) - f(|x\rangle) - [\mathbf{D}f(|x\rangle)]|h\rangle\|_{\mathcal{W}}}{\|h\|_{\mathcal{V}}} = 0, |x\rangle \in \mathcal{U}. \quad (3.144)$$

If f is differentiable at $|x\rangle$, the action of its Fréchet derivative on an arbitrary vector $|h\rangle$ can be computed as

$$[\mathbf{D}f(|x\rangle)]|h\rangle = \lim_{\alpha \rightarrow 0} \frac{f(|x\rangle + \alpha|h\rangle) - f(|x\rangle)}{\alpha}, \alpha \in \mathbb{R}. \quad (3.145)$$

To show that Eq. 3.145 is correct, let's assume $\alpha > 0$ is sufficiently small so that $\alpha|h\rangle$ is in the neighborhood of $|0\rangle$. By Eq. 3.143, we have

$$\begin{aligned} f(|x\rangle + \alpha|h\rangle) &= f(|x\rangle) + [\mathbf{D}f(|x\rangle)](\alpha|h\rangle) + \|\alpha h\| \phi(\alpha|h\rangle) \\ &= f(|x\rangle) + \alpha [\mathbf{D}f(|x\rangle)]|h\rangle + \alpha \|h\| \phi(\alpha|h\rangle) \\ \Rightarrow [\mathbf{D}f(|x\rangle)]|h\rangle &= \frac{f(|x\rangle + \alpha|h\rangle) - f(|x\rangle)}{\alpha} - \|h\| \phi(\alpha|h\rangle), \end{aligned}$$

where I have used the linearity of $\mathbf{D}f(|x\rangle)$ at the second equality. If we let $\alpha \rightarrow 0$ on both sides, the LHS does not depend on α , therefore can be treated as a constant. The term $\phi(\alpha|h\rangle)$ on the RHS goes to 0 because $\phi(|0\rangle) = 0$, therefore Eq. 3.145 holds for $\alpha > 0$. The argument for $\alpha < 0$ is similar.

The reader may recognize that the RHS in Eq. 3.145 resembles the directional derivative in calculus. In fact, the derivative defined using the RHS of Eq. 3.145 is an extension of the directional derivative to general vector spaces and is often called the *Gâteaux derivative*. Our argument above has shown that if the Fréchet derivative exists at $|x\rangle$, the Gâteaux derivative also exists at $|x\rangle$ and the two derivatives coincide. However, the converse is not true in general. Because the Gâteaux derivative is allowed to be nonlinear while the Fréchet derivative must be linear, it is possible to have a function that has a nonlinear Gâteaux derivative at a point and does not have a linear Fréchet derivative at the same point.

If we express $|x\rangle$ using a basis $B = \{|e_i\rangle\}_{i=1}^n$, where n can be ∞ for a Hilbert space, $|x\rangle = \sum_{i=1}^n x_i |e_i\rangle$, the function f can then be written as $f(x_1, x_2, \dots, x_n)$. Equation 3.145 allows us to obtain a representation of the linear transformation $\mathbf{D}f(|x\rangle)$ with respect to this basis. In Eq. 3.145, we can choose $|h\rangle = |e_j\rangle$, $j = 1, 2, \dots, n$ and we have

$$\begin{aligned} [\mathbf{D}f(|x\rangle)]|h\rangle &= [\mathbf{D}f(|x\rangle)]|e_j\rangle \\ &= \lim_{\alpha \rightarrow 0} \frac{f(x_1, x_2, \dots, x_j + \alpha, \dots, x_n) - f(x_1, x_2, \dots, x_j, \dots, x_n)}{\alpha} \\ &= \frac{\partial f}{\partial x_j}, j = 1, 2, \dots, n, \end{aligned} \quad (3.146)$$

which is identical to the partial derivative of f with respect to x_j if the function is $f: \mathbb{R}^n \supset \mathcal{U} \rightarrow \mathbb{R}$ and the basis is the standard basis on \mathbb{R}^n given by Eq. 3.9. If we express $\forall |h\rangle \in \mathbb{R}^n$ as column vectors, then $\mathbf{D}f(|x\rangle) \in (\mathbb{R}^n)^*$ can be expressed as row vectors with its elements given by $\partial f / \partial x_j$, i.e.,

$$\mathbf{D}f(|x\rangle) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right). \quad (3.147)$$

The reader may identify the RHS of Eq. 3.147 with the definition of the *gradient* ∇f in calculus. In fact, the action of the linear transformation $\mathbf{D}f(|x\rangle)$ on a vector $|h\rangle$ is identical to taking the inner product between ∇f and $|h\rangle$,

$$[\mathbf{D}f(|x\rangle)]|h\rangle = \langle \nabla f | h \rangle = \sum_{i=1}^n \frac{\partial f}{\partial x_i} h_i. \quad (3.148)$$

For the continuous-index case, the summation over i in Eq. 3.148 needs to be replaced with an integral in the form of Eq. 3.55.

If $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, we can write $f = (f_1, f_2, \dots, f_m)$ with respect to the standard basis in \mathbb{R}^m and apply the same process to each component f_i , $i = 1, 2, \dots, m$. With respect to the standard basis, the linear transformation $\mathbf{D}f(|x\rangle)$ is now

represented by an $m \times n$ matrix with its elements given by $\partial f_i / \partial x_j$, i.e.,

$$\mathbf{D}f(|x\rangle) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}. \quad (3.149)$$

The reader may identify the RHS of Eq. 3.149 with the *Jacobian matrix* in calculus. If $|x\rangle$ belongs to an infinite-dimensional Hilbert space, e.g., the space of square-integrable functions $\mathcal{L}_w^2(a, b)$ (Eq. 3.56), the same process can be applied to obtain a representation of the Fréchet derivative with respect to a basis for this Hilbert space, e.g., $B = \{|e_x\rangle\}_{x \in \mathbb{R}}$ for the continuous-index case.

Suppose we have two functions $f, g: \mathcal{V} \supset \mathcal{U} \rightarrow \mathcal{W}$ and both are differentiable at $|x\rangle \in \mathcal{U}$, it can be verified that the function $\alpha f + \beta g: \mathcal{V} \supset \mathcal{U} \rightarrow \mathcal{W}$, $\forall \alpha, \beta \in \mathbb{C}$ is also differentiable at $|x\rangle$ and its Fréchet derivative is given by

$$\mathbf{D}(\alpha f + \beta g)(|x\rangle) = \alpha \mathbf{D}f(|x\rangle) + \beta \mathbf{D}g(|x\rangle). \quad (3.150)$$

If $\forall f, g$ are continuous functions, Eq. 3.150 indicates that the set of all differentiable functions forms a subspace of all continuous functions. If the point-wise multiplication is defined at $|x\rangle$, for example, if $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$, the point-wise multiplication $f(|x\rangle) \cdot g(|x\rangle)$ is defined for $\mathbb{R} \times \mathbb{R}$, then we can define the function $f \cdot g: \mathcal{V} \supset \mathcal{U} \rightarrow \mathcal{W}$ and the Fréchet derivative of $f \cdot g$ is

$$\mathbf{D}(f \cdot g)(|x\rangle) = g(|x\rangle) \mathbf{D}f(|x\rangle) + f(|x\rangle) \mathbf{D}g(|x\rangle). \quad (3.151)$$

In calculus, there exists a similar relation known as the *product rule* or the *Leibniz rule*. The *chain rule* in calculus can also be generalized to Fréchet derivatives. Suppose $f: \mathcal{U} \rightarrow \mathcal{V}$ is differentiable at $|x\rangle \in \mathcal{U}$ and $g: \mathcal{V} \rightarrow \mathcal{W}$ is differentiable at $f(|x\rangle) \in \mathcal{V}$, then the composition of the two functions, often denoted as $g \circ f: \mathcal{U} \rightarrow \mathcal{W}$, is differentiable at $|x\rangle$ and its Fréchet derivative is given by

$$\mathbf{D}(g \circ f)(|x\rangle) = \mathbf{D}g(f(|x\rangle)) \circ \mathbf{D}f(|x\rangle). \quad (3.152)$$

Higher-order Fréchet derivatives are multi-linear transformations. I illustrate the concept with an example. For a function $f: \mathbb{R}^n \supset \mathcal{U} \rightarrow \mathbb{R}$, its first-order Fréchet derivative $\mathbf{D}f$ is a linear transformation applied on $|h\rangle \in \mathbb{R}^n$ as shown in Eqs. 3.147 and 3.148. Its second-order Fréchet derivative, often denoted as $\mathbf{D}^2 f$, is a bi-linear transform applied on the 2-tuple $(|h\rangle)^{(2)} = (|h\rangle, |h\rangle)$, and with respect to the standard basis on \mathbb{R}^n has the matrix representation

$$\mathbf{D}^2 f = \mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}, \quad (3.153)$$

which is called the *Hessian matrix* of f at $|x\rangle$. If we represent $|h\rangle$ using a column vector \mathbf{h} , the action of the second-order Fréchet derivative on the 2-tuple $(|h\rangle)^{(2)}$, denoted as $[\mathbf{D}^2 f(|x\rangle)](|h\rangle, |h\rangle)$, is then represented as the linear algebra operation $\mathbf{h}^T \mathbf{H} \mathbf{h}$, where \mathbf{h}^T is the transpose of the column vector \mathbf{h} . Written in index form, we have $\mathbf{h}^T \mathbf{H} \mathbf{h} = \sum_{i=1}^n \sum_{j=1}^n h_i \frac{\partial^2 f}{\partial x_i \partial x_j} h_j$. The third-order Fréchet derivative is a tri-linear transformation and when applied to the 3-tuple $(|h\rangle)^{(3)} = (|h\rangle, |h\rangle, |h\rangle)$ produces the summation $\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k$. The Fréchet derivatives of higher-orders can be constructed in a similar way.

Using Fréchet derivatives we can generalize the *Taylor expansion* to vector spaces. Suppose the function $f : \mathcal{V} \supset \mathcal{U} \rightarrow \mathcal{W}$ has continuous Fréchet derivatives of all orders, denote by $(|h\rangle)^{(q)}$ the q -tuple $(|h\rangle, |h\rangle, \dots, |h\rangle)$, then we have

$$f(|x\rangle + |h\rangle) = \frac{f(|x\rangle)}{0!} + \frac{[\mathbf{D}f(|x\rangle)]|h\rangle}{1!} + \frac{[\mathbf{D}^2 f(|x\rangle)](|h\rangle)^{(2)}}{2!} + \dots + \frac{[\mathbf{D}^q f(|x\rangle)](|h\rangle)^{(q)}}{q!} + \dots \tag{3.154}$$

For $f : \mathbb{R}^n \supset \mathcal{U} \rightarrow \mathbb{R}$, with respect to the standard basis of \mathbb{R}^n Eq. 3.154 can be written as

$$f(|x\rangle + |h\rangle) = f(|x\rangle) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} h_i + \frac{1}{2!} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} h_i h_j + \frac{1}{3!} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k + \dots \tag{3.155}$$

In calculus, we can find the stationary points of a function (i.e., minima, maxima and saddle points) by looking where the first-order derivative is zero and the type of the stationary point is indicated by the second-order derivative at those stationary points. With Fréchet derivatives we can now extend a similar procedure to general normed vector spaces.

3.2 The Adjoint of a Differential Operator

For solving seismic inverse problems, we are interested in the description of mechanical waves propagating inside a volume with a bounded surface. Depending upon the type of the material through which the wave propagates, the fundamental laws of physics of mechanical wave propagation are described by different types of wave equations, which are partial differential equations (PDEs).

Differential equations are usually accompanied by initial and/or boundary conditions, which restrict the functions that the differential operator can act upon. The domain of a linear differential operator \mathbf{L}_x , often denoted as $\mathcal{D}(\mathbf{L}_x)$, provides a

natural setting for encoding not only the initial/boundary conditions but also other restrictions, such as the smoothness requirements, of the functions on which the differential operator acts.

Suppose we have two linear differential operators \mathbf{L}_x and \mathbf{L}_x^\dagger , which for two arbitrary functions $u, v \in \mathcal{D}(\mathbf{L}_x) \cap \mathcal{D}(\mathbf{L}_x^\dagger)$ satisfy the relation

$$v^* \mathbf{L}_x [u] - u \left(\mathbf{L}_x^\dagger [v] \right)^* = \nabla \cdot \mathbf{Q} [u, v^*], \quad (3.156)$$

then \mathbf{L}_x^\dagger is called the *formal adjoint* of \mathbf{L}_x . The reason for using “formal” here is because we have not specified the domain of \mathbf{L}_x^\dagger yet. If $\mathbf{L}_x^\dagger = \mathbf{L}_x$, then we say that \mathbf{L}_x is *formally self-adjoint*. If $\mathbf{L}_x^\dagger = \mathbf{L}_x$ and the domains of \mathbf{L}_x and \mathbf{L}_x^\dagger satisfy the relation $\mathcal{D}(\mathbf{L}_x^\dagger) \supset \mathcal{D}(\mathbf{L}_x)$, we say that \mathbf{L}_x is *symmetric* or *Hermitian*. If in addition to $\mathbf{L}_x^\dagger = \mathbf{L}_x$ we also have $\mathcal{D}(\mathbf{L}_x^\dagger) = \mathcal{D}(\mathbf{L}_x)$, then \mathbf{L}_x is said to be *self-adjoint*. The RHS of Eq. 3.156 is sometimes called the *surface term* and $\mathbf{Q} [u, v^*]$ is called the *conjunct* of the functions u and v^* . The RHS of Eq. 3.156 is the *divergence* of the conjunct and in n -dimensions we have

$$\nabla \cdot \mathbf{Q} \equiv \sum_{i=1}^n \frac{\partial Q_i}{\partial x_i} [u, v^*]. \quad (3.157)$$

If the physical problem is defined over a closed manifold $\Omega \subset \mathbb{R}^n$ with its boundary denoted as $\partial\Omega$, we can integrate both sides of Eq. 3.156 over Ω , using the *divergence theorem* on the RHS, we obtain

$$\int_{\Omega} \left\{ v^* \mathbf{L}_x [u] - u \left(\mathbf{L}_x^\dagger [v] \right)^* \right\} dV(\mathbf{x}) = \int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da, \quad (3.158)$$

where $dV(\mathbf{x})$ is a volume measure of the n -dimensional volume Ω , $\hat{\mathbf{e}}_n$ is the n -dimensional unit vector normal to $\partial\Omega$ and da is a surface measure of the n -dimensional surface $\partial\Omega$. The LHS of Eq. 3.158 can be expressed in terms of the inner product in Eq. 3.55 with the weighting function $w(\mathbf{x}) = 1$,

$$\langle v | \mathbf{L}_x u \rangle - \langle \mathbf{L}_x^\dagger v | u \rangle = \int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da, \quad (3.159)$$

which can also be written equivalently as

$$\langle v | \mathbf{L} | u \rangle - \langle u | \mathbf{L}^\dagger | v \rangle^* = \int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da. \quad (3.160)$$

Equations 3.156, 3.158 and 3.160 are equivalent forms of the *generalized Green's identity*. If the surface term on the RHS of Eq. 3.160 is zero, we recover the definition of the adjoint operator in Eq. 3.127. From the definitions above we can see that the adjoint of a differential operator depends on the definition of the inner product as well as the differential operator itself.

The evaluation of the adjoint of a differential operator and the domain of the adjoint can be separated into two steps. In the first step, we compute the formal adjoint and the surface term through integration-by-parts. In the second step, we require the surface term to be zero, which gives us a set of adjoint initial/boundary conditions coupled with the formal adjoint operator. In Sects. 3.2.1–3.2.5, I will construct the formal adjoint operators and the surface terms for ordinary differential operators, partial differential operators, scalar wave equations, the gradient, divergence and curl operators commonly used in vector calculus and the anisotropic viscoelastic wave equation. In Sect. 3.2.6 I will examine the adjoint initial/boundary conditions.

3.2.1 Ordinary Differential Operators

As an example, let's consider the differential operator $\mathbf{L}_x = \frac{d}{dx}$ applied on functions defined on the interval $[0, 1]$. The inner product $\langle v | \mathbf{L}u \rangle$ on the LHS of Eq. 3.159 can be expressed as

$$\langle v | \mathbf{L}u \rangle = \int_0^1 v^*(x) \frac{du(x)}{dx} dx.$$

Integration by parts yields

$$\langle v | \mathbf{L}u \rangle = \int_0^1 v^*(x) \frac{du(x)}{dx} dx = v^*(x)u(x) \Big|_0^1 - \int_0^1 \frac{dv^*(x)}{dx} u(x) dx.$$

The last term on the RHS is actually an inner product. If we define the formal adjoint of \mathbf{L}_x as

$$\mathbf{L}_x^\dagger = -\frac{d}{dx} \quad (3.161)$$

we then have

$$\langle v | \mathbf{L}u \rangle - \langle \mathbf{L}^\dagger v | u \rangle = v^*(x)u(x) \Big|_{x=0}^{x=1}.$$

Comparing with Eq. 3.159, for this one-dimensional problem the surface term is then given by

$$\int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da = Q[u, v^*] \Big|_{x=1} - Q[u, v^*] \Big|_{x=0} = v^*(1)u(1) - v^*(0)u(0).$$

Comparing with Eq. 3.134, since we have $\mathbf{L}_x^\dagger = -\mathbf{L}_x$, we can say that the differential operator $\mathbf{L}_x = \frac{d}{dx}$ is formally anti-self-adjoint, or formally anti-Hermitian, or formally *skew-adjoint*.

As a second example, let's consider the differential operator $\mathbf{L}_x = \frac{d^2}{dx^2}$ applied on functions defined on the interval $[0, 1]$. The inner product $\langle v | \mathbf{L}u \rangle$ on the LHS of Eq. 3.159 can be expressed as

$$\langle v | \mathbf{L}u \rangle = \int_0^1 v^*(x) \frac{d^2u(x)}{dx^2} dx.$$

Integration by parts once yields

$$\langle v|\mathbf{L}u\rangle = \int_0^1 v^*(x) \frac{d^2u(x)}{dx^2} dx = v^*(x) \frac{du(x)}{dx} \Big|_{x=0}^{x=1} - \int_0^1 \frac{dv^*(x)}{dx} \frac{du(x)}{dx} dx.$$

Integration by parts once more on the last term of the RHS yields

$$\langle v|\mathbf{L}u\rangle = v^*(x) \frac{du(x)}{dx} \Big|_{x=0}^{x=1} - \frac{dv^*(x)}{dx} u(x) \Big|_{x=0}^{x=1} + \int_0^1 \frac{d^2v^*(x)}{dx^2} u(x) dx.$$

If we define the formal adjoint of $\mathbf{L}_x = \frac{d^2}{dx^2}$ as

$$\mathbf{L}_x^\dagger = \frac{d^2}{dx^2} \quad (3.162)$$

we then have

$$\langle v|\mathbf{L}u\rangle - \langle \mathbf{L}^\dagger v|u\rangle = v^*(x) \frac{du(x)}{dx} \Big|_{x=0}^{x=1} - \frac{dv^*(x)}{dx} u(x) \Big|_{x=0}^{x=1}$$

and the surface term is given by

$$\begin{aligned} \int_{\partial V} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da &= \mathcal{Q}[u, v^*] \Big|_{x=0}^{x=1} = \left[v^*(x) \frac{du(x)}{dx} - \frac{dv^*(x)}{dx} u(x) \right] \Big|_{x=0}^{x=1} \\ &= v^*(x) \frac{du(x)}{dx} \Big|_{x=0}^{x=1} - \frac{dv^*(x)}{dx} u(x) \Big|_{x=0}^{x=1}. \end{aligned} \quad (3.163)$$

Since we have $\mathbf{L}_x^\dagger = \mathbf{L}_x$, we can say that the different operator $\mathbf{L}_x = \frac{d^2}{dx^2}$ is formally self-adjoint.

We now consider a general second-order ordinary differential operator of the form

$$\mathbf{L}_x = a(x) \frac{d^2}{dx^2} + b(x) \frac{d}{dx} + c(x). \quad (3.164)$$

The inner product $\langle v|\mathbf{L}u\rangle$ can be expressed as

$$\langle v|\mathbf{L}u\rangle = \int_0^1 v^*(x) \left[a(x) \frac{d^2u(x)}{dx^2} + b(x) \frac{du(x)}{dx} + c(x)u(x) \right] dx.$$

Integration by parts once yields

$$\begin{aligned} \langle v|\mathbf{L}u\rangle &= \left[a(x)v^*(x) \frac{du(x)}{dx} + b(x)v^*(x)u(x) \right] \Big|_{x=0}^{x=1} \\ &+ \int_0^1 \left\{ -\frac{d[a(x)v^*(x)]}{dx} \frac{du(x)}{dx} - \frac{d[b(x)v^*(x)]}{dx} u(x) + c(x)v^*(x)u(x) \right\} dx. \end{aligned}$$

Integration by parts once more yields

$$\begin{aligned} \langle v | \mathbf{L}u \rangle &= \left[a(x)v^*(x) \frac{du(x)}{dx} + b(x)v^*(x)u(x) - \frac{d[a(x)v^*(x)]}{dx} u(x) \right] \Big|_{x=0}^{x=1} \\ &+ \int_0^1 \left\{ \frac{d^2[a^*(x)v(x)]}{dx^2} - \frac{d[b^*(x)v(x)]}{dx} + c^*(x)v(x) \right\}^* u(x) dx. \end{aligned}$$

If we define the formal adjoint of \mathbf{L}_x as

$$\mathbf{L}_x^\dagger [v(x)] = \frac{d^2[a^*(x)v(x)]}{dx^2} - \frac{d[b^*(x)v(x)]}{dx} + c^*(x)v(x), \quad (3.165)$$

then we have

$$\langle v | \mathbf{L}u \rangle - \langle \mathbf{L}^\dagger v | u \rangle = \left[av^* \frac{du}{dx} + bv^*u - \frac{d[av^*]}{dx} u \right] \Big|_{x=0}^{x=1}. \quad (3.166)$$

Suppose we would like to have $\mathbf{L}_x^\dagger = \mathbf{L}_x$, apply both \mathbf{L}_x^\dagger and \mathbf{L}_x on an arbitrary $u(x)$, we obtain

$$a \frac{d^2u}{dx^2} + b \frac{du}{dx} + c = a^* \frac{d^2u}{dx^2} + \left(2 \frac{da^*}{dx} - b^* \right) \frac{du}{dx} + \left(\frac{d^2a^*}{dx^2} - \frac{db^*}{dx} + c^* \right) u. \quad (3.167)$$

Since $u(x)$ is arbitrary, we therefore must have

$$a = a^*, b = 2 \frac{da^*}{dx} - b^*, c = \frac{d^2a^*}{dx^2} - \frac{db^*}{dx} + c^* \quad (3.168)$$

in order for \mathbf{L}_x to be formally self-adjoint. Equation 3.168 indicates that the coefficient $a(x)$ must be real for \mathbf{L}_x to be formally self-adjoint. If we assume that both $b(x)$ and $c(x)$ are also real, Eq. 3.168 indicates that $b = da/dx$. If we introduce two real functions $p(x)$ and $q(x)$ so that $a = p$, $b = dp/dx$ and $c = q$, then the second-order ordinary differential operator in Eq. 3.164 can be written as

$$\mathbf{L}_x [u(x)] = \frac{d}{dx} \left[p(x) \frac{du(x)}{dx} \right] + q(x)u(x), \quad (3.169)$$

which is known as the *Sturm-Liouville operator*. Therefore all Sturm-Liouville operators are formally self-adjoint.

3.2.2 Partial Differential Operators

A general linear second-order partial differential operator in n -dimension can be written as

$$\mathbf{L}_x [u(\mathbf{x})] = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i \frac{\partial u}{\partial x_i} + cu, \quad (3.170)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \Omega \subset \mathbb{R}^n$ and the coefficients $\{a_{ij}\}, \{b_i\}, c$ are all functions of \mathbf{x} . For the general partial differential operator, it is more straightforward to work with Eq. 3.156 directly to obtain its adjoint.

The adjoint of a typical first-order term in Eq. 3.170 can be computed by multiplying with v^* and applying the product rule for derivatives, i.e.,

$$\begin{aligned} v^* b_i \frac{\partial u}{\partial x_i} &= \frac{\partial (v^* b_i u)}{\partial x_i} - \frac{\partial (b_i v^*)}{\partial x_i} u \\ \Rightarrow v^* b_i \frac{\partial u}{\partial x_i} - \left[-\frac{\partial (b_i^* v)}{\partial x_i} \right]^* u &= \frac{\partial (v^* b_i u)}{\partial x_i}. \end{aligned} \quad (3.171)$$

Therefore the contribution of a typical first-order term to the formal adjoint is given by

$$-\frac{\partial (b_i^* v)}{\partial x_i} \quad (3.172)$$

and its contribution to the surface term is given by

$$\frac{\partial (v^* b_i u)}{\partial x_i}. \quad (3.173)$$

The same procedure can also be applied to a typical second-order term in Eq. 3.170 and we obtain

$$v^* a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = v^* a_{ij} \frac{\partial}{\partial x_i} \left(\frac{\partial u}{\partial x_j} \right) = \frac{\partial}{\partial x_i} \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) - \frac{\partial (a_{ij} v^*)}{\partial x_i} \frac{\partial u}{\partial x_j}. \quad (3.174)$$

Applying the product rule on the last term in Eq. 3.174, we obtain

$$-\frac{\partial (a_{ij} v^*)}{\partial x_i} \frac{\partial u}{\partial x_j} = -\frac{\partial}{\partial x_j} \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right] + \frac{\partial^2 (a_{ij} v^*)}{\partial x_i \partial x_j} u. \quad (3.175)$$

Bring Eq. 3.175 back into equation 3.174, we obtain

$$v^* a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) - \frac{\partial}{\partial x_j} \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right] + \frac{\partial^2 (a_{ij} v^*)}{\partial x_i \partial x_j} u, \quad (3.176)$$

which can be re-arranged to give

$$\begin{aligned} v^* a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} &- \left[\frac{\partial^2 (a_{ij}^* v)}{\partial x_i \partial x_j} \right]^* u \\ &= \frac{\partial}{\partial x_i} \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) - \frac{\partial}{\partial x_j} \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right]. \end{aligned} \quad (3.177)$$

Therefore the contribution of a typical second-order term to the formal adjoint is given by

$$\frac{\partial^2 (a_{ij}^* v)}{\partial x_i \partial x_j}, \quad (3.178)$$

and its contribution to the surface term is given by

$$\frac{\partial}{\partial x_i} \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) - \frac{\partial}{\partial x_j} \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right]. \quad (3.179)$$

Combining Eqs. 3.172 and 3.178, the formal adjoint of \mathbf{L}_x in Eq. 3.170 can be expressed as

$$\mathbf{L}_x^\dagger [v(x)] = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 (a_{ij}^* v)}{\partial x_i \partial x_j} - \sum_{i=1}^n \frac{\partial (b_i^* v)}{\partial x_i} + c^* v, \quad (3.180)$$

which can be viewed as the generalization of Eq. 3.165 to n -dimension. Combining Eqs. 3.173 and 3.179, the surface term can be written as

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{\partial}{\partial x_i} \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) - \frac{\partial}{\partial x_j} \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right] \right\} + \sum_{i=1}^n \frac{\partial (v^* b_i u)}{\partial x_i} \\ &= \sum_{i=1}^n \frac{\partial}{\partial x_i} \left[\sum_{j=1}^n \left(v^* a_{ij} \frac{\partial u}{\partial x_j} \right) \right] - \sum_{j=1}^n \frac{\partial}{\partial x_j} \sum_{i=1}^n \left[\frac{\partial (a_{ij} v^*)}{\partial x_i} u \right] + \sum_{i=1}^n \frac{\partial (v^* b_i u)}{\partial x_i} \\ &= \nabla \cdot \mathbf{Q}, \end{aligned} \quad (3.181)$$

where explicit expressions of \mathbf{Q} can be obtained by comparing Eq. 3.181 with Eq. 3.157.

3.2.3 Scalar Wave Equations

Consider the scalar wave-equation operator

$$\mathbf{L}_{x,t} [u(x, t)] = \left[\frac{\partial^2}{\partial x^2} - \frac{1}{\alpha^2(x)} \frac{\partial^2}{\partial t^2} \right] u(x, t), \quad (3.182)$$

where $\alpha(x)$ is a positive real function representing the speed of the scalar wave. It is a partial differential operator with two dimensions, a spatial dimension x and a time dimension t . If we set the spatial dimension as the first dimension, i.e., $x_1 = x$, and the time dimension as the second dimension, $x_2 = t$, comparing with Eq. 3.170, we have $a_{11} = 1$, $a_{22} = -1/\alpha^2(x)$, $a_{12} = a_{21} = 0$, $b_1 = b_2 = 0$ and $c = 0$. Bring

these values into Eq. 3.180, we obtain the formal adjoint of the scalar wave equation operator

$$\begin{aligned} \mathbf{L}_{\mathbf{x},t}^\dagger [v(x,t)] &= \frac{\partial^2 (1 \cdot v)}{\partial x_1^2} + \frac{\partial^2 \left[-\frac{1}{\alpha^2(x_1)} v \right]}{\partial x_2^2} \\ &= \left[\frac{\partial^2}{\partial x_1^2} - \frac{1}{\alpha^2(x_1)} \frac{\partial^2}{\partial x_2^2} \right] v(x_1, x_2) \\ &= \left[\frac{\partial^2}{\partial x^2} - \frac{1}{\alpha^2(x)} \frac{\partial^2}{\partial t^2} \right] v(x, t). \end{aligned} \quad (3.183)$$

Comparing with Eq. 3.182, we have $\mathbf{L}_{\mathbf{x},t}^\dagger = \mathbf{L}_{\mathbf{x},t}$, therefore the scalar wave-equation operator is formally self-adjoint. The surface term can be obtained from Eq. 3.181 and we have

$$\nabla \cdot \mathbf{Q} = \frac{\partial}{\partial x} \left(v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \right) - \frac{1}{\alpha^2(x)} \frac{\partial}{\partial t} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right). \quad (3.184)$$

Comparing with Eq. 3.157, the conjunct \mathbf{Q} can be expressed as a column vector with 2 elements

$$\mathbf{Q} = \begin{bmatrix} v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \\ -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \end{bmatrix}. \quad (3.185)$$

We now consider the scalar wave-equation operator in 3 spatial dimensions

$$\mathbf{L}_{\mathbf{x},t} [u(\mathbf{x}, t)] = \left[\nabla^2 - \frac{1}{\alpha^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \right] u(\mathbf{x}, t), \quad (3.186)$$

where $\mathbf{x} = (x_1, x_2, x_3) \in V \subset \mathbb{R}^3$ and $\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$. If we identify the time dimension $t = x_4$, then we have

$$a_{ij} = \begin{cases} 1, & i = j = 1, 2, 3 \\ -\frac{1}{\alpha^2(\mathbf{x})}, & i = j = 4 \\ 0, & i \neq j \end{cases}, \quad b_i = 0 \ (i = 1, 2, 3, 4), \quad c = 0. \quad (3.187)$$

Bring Eq. 3.187 into Eq. 3.180, we obtain the formal adjoint of the scalar wave-equation operator in 3 spatial dimensions

$$\begin{aligned} \mathbf{L}_{\mathbf{x}}^\dagger [v(\mathbf{x}, t)] &= \left[\sum_{i=1}^3 \frac{\partial^2}{\partial x_i^2} - \frac{1}{\alpha^2(\mathbf{x})} \frac{\partial^2}{\partial x_4^2} \right] v(\mathbf{x}, x_4) \\ &= \left[\nabla^2 - \frac{1}{\alpha^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \right] v(\mathbf{x}, t). \end{aligned} \quad (3.188)$$

Comparing with Eq. 3.186, we have $\mathbf{L}_{\mathbf{x},t}^\dagger = \mathbf{L}_{\mathbf{x},t}$, therefore the scalar wave-equation operator in 3 spatial dimensions is also formally self-adjoint. The corresponding surface term can be obtained by bringing Eq. 3.187 into Eq. 3.181 and we have

$$\nabla \cdot \mathbf{Q} = \sum_{i=1}^3 \frac{\partial}{\partial x_i} \left(v^* \frac{\partial u}{\partial x_i} - \frac{\partial v^*}{\partial x_i} u \right) - \frac{1}{\alpha^2(\mathbf{x})} \frac{\partial}{\partial t} \left(v^* \frac{\partial u}{\partial t} - \frac{\partial v^*}{\partial t} u \right). \quad (3.189)$$

Comparing with Eq. 3.157, the components of the conjunct \mathbf{Q} can be expressed as

$$Q_i = \begin{cases} v^* \frac{\partial u}{\partial x_i} - \frac{\partial v^*}{\partial x_i} u, & i = 1, 2, 3 \\ -\frac{1}{\alpha^2(\mathbf{x})} \left(v^* \frac{\partial u}{\partial t} - \frac{\partial v^*}{\partial t} u \right), & i = 4 \end{cases}. \quad (3.190)$$

Note that if we use the integral form of the generalized Green's identity in Eq. 3.158, the integration volume Ω is 4-dimensional with 3 spatial dimensions and one time dimension and when applying the divergence theorem on the RHS, the boundary $\partial\Omega$ is a 4-dimensional hypersurface.

3.2.4 Gradient, Divergence and Curl

Let's now consider the gradient ∇ , the divergence $\nabla \cdot$ and the curl $\nabla \times$ operators and their formal adjoint operators in three spatial dimensions.

For the gradient operator ∇ , consider a scalar-valued function $u(\mathbf{x})$ and a vector-valued function $\mathbf{v}(\mathbf{x}) = [v_1(\mathbf{x}), v_2(\mathbf{x}), v_3(\mathbf{x})]^T$ with $\mathbf{x} = (x_1, x_2, x_3) \in V \subset \mathbb{R}^3$. We can compute the Euclidean dot product between the two column vectors \mathbf{v}^* and ∇u as

$$\begin{aligned} \mathbf{v}^* \cdot \nabla u &= v_1^* \frac{\partial u}{\partial x_1} + v_2^* \frac{\partial u}{\partial x_2} + v_3^* \frac{\partial u}{\partial x_3} \\ &= \frac{\partial (v_1^* u)}{\partial x_1} - \frac{\partial v_1^*}{\partial x_1} u + \frac{\partial (v_2^* u)}{\partial x_2} - \frac{\partial v_2^*}{\partial x_2} u + \frac{\partial (v_3^* u)}{\partial x_3} - \frac{\partial v_3^*}{\partial x_3} u \\ &= - \left(\frac{\partial v_1^*}{\partial x_1} + \frac{\partial v_2^*}{\partial x_2} + \frac{\partial v_3^*}{\partial x_3} \right) u + \left[\frac{\partial (v_1^* u)}{\partial x_1} + \frac{\partial (v_2^* u)}{\partial x_2} + \frac{\partial (v_3^* u)}{\partial x_3} \right] \\ &= - (\nabla \cdot \mathbf{v}^*) u + \nabla \cdot \mathbf{Q}, \end{aligned} \quad (3.191)$$

where the conjunct is defined as

$$\mathbf{Q} = (v_1^* u, v_2^* u, v_3^* u)^T. \quad (3.192)$$

Re-arranging Eq. 3.191, we obtain

$$\mathbf{v}^* \cdot \nabla u - [(-\nabla \cdot \mathbf{v}^*) u] = \nabla \cdot \mathbf{Q}. \quad (3.193)$$

Therefore the formal adjoint of the gradient operator ∇ is $-\nabla \cdot$, i.e., the minus divergence operator.

For the divergence operator $\nabla \cdot$, we consider a vector-valued function $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), u_2(\mathbf{x}), u_3(\mathbf{x})]^T$ and a scalar-valued function $v(\mathbf{x})$. We can compute

$$\begin{aligned} v^* \nabla \cdot \mathbf{u} &= v^* \left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} \right) = v^* \frac{\partial u_1}{\partial x_1} + v^* \frac{\partial u_2}{\partial x_2} + v^* \frac{\partial u_3}{\partial x_3} \\ &= \frac{\partial (v^* u_1)}{\partial x_1} - \frac{\partial v^*}{\partial x_1} u_1 + \frac{\partial (v^* u_2)}{\partial x_2} - \frac{\partial v^*}{\partial x_2} u_2 + \frac{\partial (v^* u_3)}{\partial x_3} - \frac{\partial v^*}{\partial x_3} u_3 \\ &= (-\nabla v^*) \cdot \mathbf{u} + \nabla \cdot \mathbf{Q}, \end{aligned} \quad (3.194)$$

where the conjunct is defined as

$$\mathbf{Q} = (v^* u_1, v^* u_2, v^* u_3)^T. \quad (3.195)$$

Re-arranging Eq. 3.194, we obtain

$$v^* \nabla \cdot \mathbf{u} - (-\nabla v^*) \cdot \mathbf{u} = \nabla \cdot \mathbf{Q}. \quad (3.196)$$

Therefore the formal adjoint of the divergence operator $\nabla \cdot$ is $-\nabla$, i.e., the minus gradient operator.

For the curl operator $\nabla \times$, we consider two vector-valued functions $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}), u_2(\mathbf{x}), u_3(\mathbf{x})]^T$ and $\mathbf{v}(\mathbf{x}) = [v_1(\mathbf{x}), v_2(\mathbf{x}), v_3(\mathbf{x})]^T$. We can compute the Euclidean dot product

$$\begin{aligned} \mathbf{v}^* \cdot (\nabla \times \mathbf{u}) &= v_1^* \left(\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3} \right) + v_2^* \left(\frac{\partial u_1}{\partial x_3} - \frac{\partial u_3}{\partial x_1} \right) + v_3^* \left(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} \right) \\ &= \left(\frac{\partial v_3^*}{\partial x_2} - \frac{\partial v_2^*}{\partial x_3} \right) u_1 + \left(\frac{\partial v_1^*}{\partial x_3} - \frac{\partial v_3^*}{\partial x_1} \right) u_2 + \left(\frac{\partial v_2^*}{\partial x_1} - \frac{\partial v_1^*}{\partial x_2} \right) u_3 \\ &\quad + \frac{\partial (v_3^* u_2 - v_2^* u_3)}{\partial x_1} + \frac{\partial (v_1^* u_3 - v_3^* u_1)}{\partial x_2} + \frac{\partial (v_2^* u_1 - v_1^* u_2)}{\partial x_3} \\ &= (\nabla \times \mathbf{v}^*) \cdot \mathbf{u} + \nabla \cdot \mathbf{Q}, \end{aligned} \quad (3.197)$$

where the conjunct is defined as

$$\mathbf{Q} = (v_3^* u_2 - v_2^* u_3, v_1^* u_3 - v_3^* u_1, v_2^* u_1 - v_1^* u_2)^T. \quad (3.198)$$

Re-arranging Eq. 3.197, we obtain

$$\mathbf{v}^* \cdot (\nabla \times \mathbf{u}) - (\nabla \times \mathbf{v}^*) \cdot \mathbf{u} = \nabla \cdot \mathbf{Q}. \quad (3.199)$$

Therefore the formal adjoint of the curl operator $\nabla \times$ is still $\nabla \times$, i.e., the curl operator is formally self-adjoint.

3.2.5 Viscoelastic Equation of Motion

The viscoelastic media considered in Sect. 2.1.2 is isotropic. We can extend our analysis to anisotropic viscoelastic media with the stress-strain relation defined as

$$\sigma_{ij}(\mathbf{x}, t) = \sum_{k=1}^3 \sum_{l=1}^3 \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \epsilon_{kl}(\mathbf{x}, \tau), \quad (3.200)$$

where Ψ_{ijkl} is the *rate-of-relaxation function*, $\mathbf{x} \in V \subset \mathbb{R}^3$, V is the region occupied by the material, and $t, \tau \in \mathbb{R}$. The rate-of-relaxation function is assumed to have the symmetries

$$\Psi_{ijkl} = \Psi_{ijlk} = \Psi_{jikl} = \Psi_{klij}, \quad (3.201)$$

and is causal, i.e.,

$$\Psi_{ijkl}(t) = 0 \quad \text{for } t < 0. \quad (3.202)$$

The anisotropic viscoelastic wave-equation operator in displacement $\mathbf{u}(\mathbf{x}, t)$ can then be written as

$$\begin{aligned} \mathbf{L}_{\mathbf{x},t}[\mathbf{u}(\mathbf{x}, t)] &= \rho(\mathbf{x}) \frac{\partial^2 \mathbf{u}(\mathbf{x}, t)}{\partial t^2} \\ &- \nabla \cdot \int_{-\infty}^{+\infty} d\tau \Psi(\mathbf{x}, t - \tau) : \nabla \mathbf{u}(\mathbf{x}, \tau), \end{aligned} \quad (3.203)$$

which can also be written in the index form as

$$\begin{aligned} \mathbf{L}_{\mathbf{x},t}[\mathbf{u}(\mathbf{x}, t)]_i &= \rho(\mathbf{x}) \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right]. \end{aligned} \quad (3.204)$$

The dual is then defined by a vector-valued function $\mathbf{v}(\mathbf{x}, t)$ of space and time⁵. The inner product involves the Euclidean dot product between two vectors, the spatial integration over V and the temporal integration over a certain duration, say $[0, T]$, i.e.,

$$\langle \mathbf{v} | \mathbf{L}\mathbf{u} \rangle = \int_0^T dt \int_V dV(\mathbf{x}) \sum_{i=1}^3 v_i^*(\mathbf{x}, t) \mathbf{L}_{\mathbf{x},t}[\mathbf{u}(\mathbf{x}, t)]_i, \quad (3.205)$$

where $\mathbf{L}_{\mathbf{x},t}[\mathbf{u}(\mathbf{x}, t)]_i$ is given by Eq. 3.204. We now can apply the same techniques we have been using in Sect. 3.2.1–3.2.4 to construct the adjoint of $\mathbf{L}_{\mathbf{x},t}$, together with the surface term.

⁵ Here \mathbf{v} does not represent the particle velocity field.

Let's consider the first term on the RHS of Eq. 3.204. Integration by parts twice with respect to time yields

$$\begin{aligned}
 & \int_0^T dt \int_V dV \sum_{i=1}^3 v_i^*(\mathbf{x}, t) \rho(\mathbf{x}) \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} \\
 &= \int_V dV \rho(\mathbf{x}) \sum_{i=1}^3 \left[\int_0^T v_i^*(\mathbf{x}, t) \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} dt \right] \\
 &= \int_V dV \rho(\mathbf{x}) \sum_{i=1}^3 \left\{ \left[v_i^* \frac{\partial u_i}{\partial t} - \frac{\partial v_i^*}{\partial t} u_i \right] \Big|_0^T + \int_0^T \frac{\partial^2 v_i^*}{\partial t^2} u_i dt \right\} \\
 &= \int_V dV \rho(\mathbf{x}) \sum_{i=1}^3 \left[v_i^* \frac{\partial u_i}{\partial t} - \frac{\partial v_i^*}{\partial t} u_i \right] \Big|_0^T + \int_0^T dt \int_V dV \sum_{i=1}^3 \rho(\mathbf{x}) \frac{\partial^2 v_i^*}{\partial t^2} u_i,
 \end{aligned}$$

which can be re-written as

$$\begin{aligned}
 & \int_0^T dt \int_V dV \sum_{i=1}^3 v_i^* \rho(\mathbf{x}) \frac{\partial^2 u_i}{\partial t^2} - \int_0^T dt \int_V dV \sum_{i=1}^3 \rho(\mathbf{x}) \frac{\partial^2 v_i^*}{\partial t^2} u_i \\
 &= \int_V dV \rho(\mathbf{x}) \sum_{i=1}^3 \left[v_i^* \frac{\partial u_i}{\partial t} - \frac{\partial v_i^*}{\partial t} u_i \right] \Big|_0^T. \tag{3.206}
 \end{aligned}$$

Therefore the contribution of the first term on the RHS of Eq. 3.204 to the formal adjoint is

$$\rho(\mathbf{x}) \frac{\partial^2 v_i^*}{\partial t^2} \tag{3.207}$$

and its contribution to the surface term is

$$\int_V dV \rho(\mathbf{x}) \sum_{i=1}^3 \left[v_i^* \frac{\partial u_i}{\partial t} - \frac{\partial v_i^*}{\partial t} u_i \right] \Big|_0^T. \tag{3.208}$$

We now consider the second term on the RHS of Eq. 3.204. Take the inner product with $\mathbf{v}(\mathbf{x}, t)$, we obtain

$$\begin{aligned}
 & - \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 v_i^*(\mathbf{x}, t) \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right] \\
 &= - \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[v_i^*(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right] \\
 &+ \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right].
 \end{aligned}$$

The first term on the RHS is a contribution to the surface term, let's denote it as Q_1 ,

$$Q_1 = - \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[v_i^*(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right]. \quad (3.209)$$

By splitting the integral $\int_{-\infty}^{+\infty} d\tau$ into 3 pieces, the second term on the RHS, denoted as A in the following, can be re-written as

$$\begin{aligned} A &= \int_V dV \sum_{i,j,k,l=1}^3 \int_{-\infty}^{+\infty} d\tau \int_0^T dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau) \\ &= \int_V dV \sum_{i,j,k,l=1}^3 \left(\int_{-\infty}^0 d\tau + \int_0^T d\tau + \int_T^{+\infty} d\tau \right) \int_0^T dt \\ &\quad \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau). \end{aligned}$$

Considering the causality condition of the rate-of-relaxation function, Eq. 3.202, the contribution from the integral $\int_T^{+\infty} d\tau$ is zero. Therefore we can express A as

$$\begin{aligned} A &= \int_V dV \sum_{i,j,k,l=1}^3 \int_{-\infty}^0 d\tau \int_0^T dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau) \\ &+ \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \int_0^T dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau). \end{aligned}$$

The first term on the RHS is a contribution to the surface term and we denote it as Q_2 ,

$$Q_2 = \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \int_{-\infty}^0 d\tau \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \Psi_{ijkl}(\mathbf{x}, t - \tau). \quad (3.210)$$

By re-writing the integral $\int_0^T dt$, the second term on the RHS of the expression for A , denoted as B in the following, can be re-written as

$$B = \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \int_0^T dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau)$$

$$\begin{aligned}
&= \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \left(\int_{-\infty}^{+\infty} dt - \int_{-\infty}^0 dt - \int_T^{+\infty} dt \right) \\
&\quad \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau).
\end{aligned}$$

Considering the causality condition, Eq. 3.202, the contribution from the integral $\int_{-\infty}^0 dt$ is zero. Therefore we can write B as

$$\begin{aligned}
B &= \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \int_{-\infty}^{+\infty} dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau) \\
&\quad - \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \int_T^{+\infty} dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau).
\end{aligned}$$

The second term on the RHS is a contribution to the surface term and we denote it as Q_3 ,

$$Q_3 = - \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \int_T^{+\infty} dt \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \Psi_{ijkl}(\mathbf{x}, t - \tau).$$

If we introduce an anti-causal rate-of-relaxation function

$$\Psi'_{ijkl}(\mathbf{x}, \tau - t) \equiv \Psi_{ijkl}(\mathbf{x}, t - \tau), \quad (3.211)$$

exchange the role of τ and t and use the symmetries in Eq. 3.201, we can re-write Q_3 as

$$\begin{aligned}
Q_3 &= - \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \\
&\quad \frac{\partial u_i(\mathbf{x}, t)}{\partial x_j} \int_T^{+\infty} d\tau \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \Psi'_{ijkl}(\mathbf{x}, t - \tau). \quad (3.212)
\end{aligned}$$

The first term on the RHS of the expression for B , denoted as C in the following, can be expressed as

$$\begin{aligned}
C &= \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \int_{-\infty}^{+\infty} dt \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \\
&= \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \frac{\partial}{\partial x_l} \left[u_k(\mathbf{x}, \tau) \int_{-\infty}^{+\infty} dt \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \right] \\
&\quad - \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau u_k(\mathbf{x}, \tau) \frac{\partial}{\partial x_l} \left[\int_{-\infty}^{+\infty} dt \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \right].
\end{aligned}$$

Considering the symmetry conditions of the rate-of-relaxation function in Eq. 3.201, we can re-write C as

$$C = \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T d\tau \frac{\partial}{\partial x_j} \left[u_i(\mathbf{x}, \tau) \int_{-\infty}^{+\infty} dt \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, t)}{\partial x_l} \right] \\ - \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T dt u_i(\mathbf{x}, t) \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right].$$

Using the anti-causal rate-of-relaxation function in Eq. 3.211 and exchanging τ and t , we can re-write C as

$$C = \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T dt \frac{\partial}{\partial x_j} \left[u_i(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right] \\ - \int_V dV \sum_{i,j,k,l=1}^3 \int_0^T dt u_i(\mathbf{x}, t) \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right].$$

The first term on the RHS is a contribution to the surface term and we denote it as Q_4 ,

$$Q_4 = \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial}{\partial x_j} \\ \left[u_i(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right]. \quad (3.213)$$

The second term on the RHS of the expression for C is a contribution to the formal adjoint. Therefore the contribution of the second term on the RHS of Eq. 3.204 to the formal adjoint is

$$- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right], \quad (3.214)$$

and its contribution to the surface term is given by $Q_1 + Q_2 + Q_3 + Q_4$.

Combining Eqs. 3.207 and 3.214, the formal adjoint of the anisotropic viscoelastic wave-equation operator is given by

$$\mathbf{L}_{\mathbf{x},t}^\dagger [\mathbf{v}(\mathbf{x}, t)]_i = \rho(\mathbf{x}) \frac{\partial^2 v_i}{\partial t^2} \\ - \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k(\mathbf{x}, \tau)}{\partial x_l} \right]. \quad (3.215)$$

The formal adjoint is almost identical to the anisotropic viscoelastic wave-equation operator in Eq. 3.204, except that the causal rate-of-relaxation function Ψ_{ijkl} in Eq. 3.204 has been replaced with the anti-causal rate-of-relaxation function Ψ'_{ijkl} defined in Eq. 3.211.

The surface term can be obtained by combining Eq. 3.208 and $Q_1 + Q_2 + Q_3 + Q_4$. In order to highlight the similarities among the different terms, the order of the summation in Eq. 3.216 in the following is actually $Q_4 + Q_1 + Q_2 + Q_3$. In Q_1 (Eq. 3.209) and Q_4 (Eq. 3.213), the expression $\sum_{j=1}^3 \frac{\partial}{\partial x_j}$ is a divergence operator and we can apply the divergence theorem to replace the volume integral $\int_V dV$ with the surface integral $\int_{\partial V} da$, where da is a measure of the boundary ∂V . If we denote the normal direction on ∂V as $\hat{\mathbf{n}}(\mathbf{x})$ or as $\hat{n}_j(\mathbf{x})$ in the index form, the surface term can be written as

$$\begin{aligned}
 & \int_V dV(\mathbf{x}) \rho(\mathbf{x}) \sum_{i=1}^3 \left[v_i^* \frac{\partial u_i}{\partial t} - \frac{\partial v_i^*}{\partial t} u_i \right] \Big|_{t=0}^{t=T} \tag{3.216} \\
 & + \int_0^T dt \int_{\partial V} da \sum_{i,j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \left[u_i(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \right] \\
 & - \int_0^T dt \int_{\partial V} da \sum_{i,j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \left[v_i^*(\mathbf{x}, t) \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right] \\
 & + \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial v_i^*(\mathbf{x}, t)}{\partial x_j} \int_{-\infty}^0 d\tau \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \Psi_{ijkl}(\mathbf{x}, t - \tau) \\
 & - \int_0^T dt \int_V dV \sum_{i,j,k,l=1}^3 \frac{\partial u_i(\mathbf{x}, t)}{\partial x_j} \int_T^{+\infty} d\tau \frac{\partial v_k^*(\mathbf{x}, \tau)}{\partial x_l} \Psi'_{ijkl}(\mathbf{x}, t - \tau).
 \end{aligned}$$

Readers can verify that the formal adjoint in Eq. 3.215 and the surface term in Eq. 3.216 derived here for the anisotropic viscoelastic wave-equation operator are identical to those documented in e.g., (Tarantola 1988).

3.2.6 Adjoint Initial and Boundary Conditions

After we have obtained the formal adjoint of a differential operator, the second step is to derive the adjoint initial/boundary conditions by requiring the surface term to be zero. I still start with ordinary differential operators, as they are simpler and also useful for illustrating many of the basic techniques.

3.2.6.1 ODE

For ordinary differential operators applied on functions defined on a finite interval, say $[0, 1]$, some common types of boundary conditions are

$$\left\{ \begin{array}{ll} u(0) = 0, u(1) = 0; & \text{Dirichlet} \\ \frac{du}{dx}(0) = 0, \frac{du}{dx}(1) = 0; & \text{Neumann} \\ u(0) = u(1), \frac{du}{dx}(0) = \frac{du}{dx}(1); & \text{periodic} \\ \alpha_0 u(0) + \beta_0 \frac{du}{dx}(0) = 0, \alpha_1 u(1) + \beta_1 \frac{du}{dx}(1) = 0; & \text{general unmixed,} \end{array} \right. \quad (3.217)$$

where $\alpha_0, \alpha_1, \beta_0$ and β_1 in the general unmixed boundary condition are scalars. Note that boundary conditions usually involve one condition imposed on the solution at each endpoint of the finite interval. In initial conditions, we impose two conditions at one of the endpoints, e.g.,

$$\left\{ \begin{array}{ll} u(0) = 0, \frac{du}{dx}(0) = 0; & \text{initial conditions,} \\ u(1) = 0, \frac{du}{dx}(1) = 0; & \text{final conditions.} \end{array} \right. \quad (3.218)$$

If we consider the independent variable $x \in [0, 1]$ as time, the conditions at $x = 0$ can be called initial conditions and the conditions imposed at $x = 1$ can be called *final conditions*.

In both Eqs. 3.217 and 3.218, the right-hand-sides of the conditions are all zeros. Such conditions are called *homogeneous* conditions. In practical applications, the right-hand-sides of the initial/boundary conditions are not always zeros, in which case we say that the initial/boundary conditions are *inhomogeneous*. As we will see in Sect. 3.3, the solutions of differential equations coupled with inhomogeneous initial/boundary conditions can be expressed using Green's functions that satisfy the same differential equations coupled with the corresponding homogeneous initial/boundary conditions. Therefore in the following I will only consider homogeneous conditions.

Consider the second-order differential operator $\mathbf{L}_x = \frac{d^2}{dx^2}$, whose formal adjoint and surface term have been derived in Eqs. 3.162 and 3.163 in Sect. 3.2.1. Suppose $u(x)$ satisfies the general unmixed boundary condition in Eq. 3.217, therefore we can express $\frac{du}{dx}(0)$ with $u(0)$ and $\frac{du}{dx}(1)$ with $u(1)$ as

$$\frac{du}{dx}(0) = -\frac{\alpha_0}{\beta_0} u(0), \quad \frac{du}{dx}(1) = -\frac{\alpha_1}{\beta_1} u(1). \quad (3.219)$$

Bring Eq. 3.219 into the surface term in Eq. 3.163, set the surface term to zero and re-arrange the terms we obtain

$$\left[\frac{\alpha_0}{\beta_0} v^*(0) + \frac{dv^*}{dx}(0) \right] u(0) - \left[\frac{\alpha_1}{\beta_1} v^*(1) + \frac{dv^*}{dx}(1) \right] u(1) = 0, \quad (3.220)$$

which holds for arbitrary $u(x)$ if and only if

$$\frac{\alpha_0}{\beta_0}v^*(0) + \frac{dv^*}{dx}(0) = 0 \quad \text{and} \quad \frac{\alpha_1}{\beta_1}v^*(1) + \frac{dv^*}{dx}(1) = 0. \quad (3.221)$$

Equation 3.221 can be re-written as

$$\alpha_0v^*(0) + \beta_0\frac{dv^*}{dx}(0) = 0 \quad \text{and} \quad \alpha_1v^*(1) + \beta_1\frac{dv^*}{dx}(1) = 0. \quad (3.222)$$

The boundary conditions on $v^*(x)$ in Eq. 3.222, which is the adjoint boundary condition, is identical to the general unmixed boundary condition on $u(x)$ in Eq. 3.217, therefore we have $\mathcal{D}(\mathbf{L}_x^\dagger) = \mathcal{D}(\mathbf{L}_x)$. Since $\mathbf{L}_x = \frac{d^2}{dx^2}$ is formally self-adjoint, as shown in Sect. 3.2.1, and now we also have $\mathcal{D}(\mathbf{L}_x^\dagger) = \mathcal{D}(\mathbf{L}_x)$, we can conclude that $\mathbf{L}_x = \frac{d^2}{dx^2}$ coupled with the general unmixed boundary condition in Eq. 3.217 is self-adjoint. It can be verified that $\mathbf{L}_x = \frac{d^2}{dx^2}$ coupled with the Dirichlet, Neumann or the periodic boundary conditions in Eq. 3.217 is also self-adjoint.

Suppose we replace the general unmixed boundary condition in Eq. 3.217 with the mixed boundary condition

$$\alpha_0u(0) + \beta_0\frac{du}{dx}(0) = 0 \quad \text{and} \quad \alpha_1u(1) + \beta_1\frac{du}{dx}(1) = 0; \quad (3.223)$$

apply the same procedure and we obtain the corresponding adjoint boundary condition

$$\alpha_0v^*(1) - \beta_0\frac{dv^*}{dx}(1) = 0 \quad \text{and} \quad \alpha_1v^*(0) - \beta_1\frac{dv^*}{dx}(0) = 0, \quad (3.224)$$

which is different from the mixed boundary condition in Eq. 3.223. Therefore the operator $\mathbf{L}_x = \frac{d^2}{dx^2}$ coupled with the mixed boundary condition in Eq. 3.223 is not self-adjoint.

For the initial conditions in Eq. 3.218, the surface term in Eq. 3.163 can be expressed as

$$v^*(1)\frac{du}{dx}(1) - \frac{dv^*}{dx}(1)u(1) = 0, \quad (3.225)$$

which holds for arbitrary $u(x)$ if and only if

$$v^*(1) = 0 \quad \text{and} \quad \frac{dv^*}{dx}(1) = 0. \quad (3.226)$$

So the final conditions in Eq. 3.218 are the adjoint of the initial conditions. Since the final conditions are different from the initial conditions, the operator $\mathbf{L}_x = \frac{d^2}{dx^2}$ coupled with the initial conditions in Eq. 3.218 is not self-adjoint.

We now consider the general second-order ordinary differential operator given by Eq. 3.164 coupled with boundary conditions of the form

$$\begin{aligned}\mathbf{R}_1[u] &\equiv \alpha_{11}u(0) + \alpha_{12}\frac{du}{dx}(0) + \beta_{11}u(1) + \beta_{12}\frac{du}{dx}(1) = 0, \\ \mathbf{R}_2[u] &\equiv \alpha_{21}u(0) + \alpha_{22}\frac{du}{dx}(0) + \beta_{21}u(1) + \beta_{22}\frac{du}{dx}(1) = 0,\end{aligned}\quad (3.227)$$

where we assume that the vectors $(\alpha_{11}, \alpha_{12}, \beta_{11}, \beta_{12})$ and $(\alpha_{21}, \alpha_{22}, \beta_{21}, \beta_{22})$ are linearly independent. The most general form of the conjunct for a second-order ordinary differential operator is

$$\begin{aligned}Q[u, v^*] &= q_{11}(x)u(x)v^*(x) + q_{12}(x)u(x)\frac{dv^*}{dx}(x) \\ &\quad + q_{21}(x)\frac{du}{dx}(x)v^*(x) + q_{22}(x)\frac{du}{dx}(x)\frac{dv^*}{dx}(x).\end{aligned}\quad (3.228)$$

For the surface term given in Eq. 3.166, we have $q_{11}(x) = b(x) - \frac{da}{dx}(x)$, $q_{12}(x) = -a(x)$, $q_{21}(x) = a(x)$ and $q_{22}(x) = 0$. The boundary conditions in Eq. 3.227 can be written in matrix form as

$$\begin{pmatrix} \mathbf{R}_1[u] \\ \mathbf{R}_2[u] \end{pmatrix} = (\mathbf{A} \ \mathbf{B}) \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \end{pmatrix} = 0,\quad (3.229)$$

where the sub-matrices are

$$\begin{aligned}\mathbf{A} &\equiv \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}, \quad \mathbf{B} \equiv \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}, \\ \mathbf{u}_0 &\equiv \begin{pmatrix} u(0) \\ \frac{du}{dx}(0) \end{pmatrix}, \quad \mathbf{u}_1 \equiv \begin{pmatrix} u(1) \\ \frac{du}{dx}(1) \end{pmatrix}.\end{aligned}\quad (3.230)$$

Since the two vectors $(\alpha_{11}, \alpha_{12}, \beta_{11}, \beta_{12})$ and $(\alpha_{21}, \alpha_{22}, \beta_{21}, \beta_{22})$ are linearly independent, the 2×4 matrix $(\mathbf{A} \ \mathbf{B})$ has a rank of 2, which means that the matrix $(\mathbf{A} \ \mathbf{B})$ has a 2×2 sub-matrix that is invertible. Let's assume that the sub-matrix \mathbf{B} is invertible. If it is not, we can always re-arrange the terms in Eq. 3.227 to make it invertible. We can also re-write the conjunct in Eq. 3.228 in the matrix form

$$Q[u, v^*](x) = \mathbf{u}_x^T \mathbf{Q}_x \mathbf{v}_x^*,\quad (3.231)$$

where the matrices are defined as

$$\mathbf{Q}_x \equiv \begin{pmatrix} q_{11}(x) & q_{12}(x) \\ q_{21}(x) & q_{22}(x) \end{pmatrix}, \quad \mathbf{u}_x \equiv \begin{pmatrix} u(x) \\ \frac{du}{dx}(x) \end{pmatrix}, \quad \mathbf{v}_x^* \equiv \begin{pmatrix} v^*(x) \\ \frac{dv^*}{dx}(x) \end{pmatrix}.\quad (3.232)$$

To make the surface term become zero, we must have

$$\mathbf{u}_1^T \mathbf{Q}_1 \mathbf{v}_1^* - \mathbf{u}_0^T \mathbf{Q}_0 \mathbf{v}_0^* = 0.\quad (3.233)$$

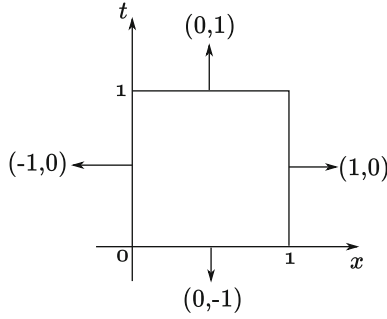


Fig. 3.1 The boundaries of the space-time domain for functions satisfying the scalar wave-equation defined in Eq. 3.182 and the normal directions on each boundary of the space-time domain.

Since matrix \mathbf{B} is assumed to be invertible, we can express \mathbf{u}_1 in terms of \mathbf{u}_0 by solving Eq. 3.229,

$$\mathbf{u}_1 = -\mathbf{B}^{-1}\mathbf{A}\mathbf{u}_0 \Rightarrow \mathbf{u}_1^T = -\mathbf{u}_0^T\mathbf{A}^T\left(\mathbf{B}^T\right)^{-1}. \quad (3.234)$$

Bring Eq. 3.234 into Eq. 3.233, we obtain

$$\begin{aligned} -\mathbf{u}_0^T\mathbf{A}^T\left(\mathbf{B}^T\right)^{-1}\mathbf{Q}_1\mathbf{v}_1^* &= \mathbf{u}_0^T\mathbf{Q}_0\mathbf{v}_0^* \\ \Rightarrow \mathbf{u}_0^T\left[\mathbf{Q}_0\mathbf{v}_0^* + \mathbf{A}^T\left(\mathbf{B}^T\right)^{-1}\mathbf{Q}_1\mathbf{v}_1^*\right] &= 0, \end{aligned} \quad (3.235)$$

which holds for arbitrary \mathbf{u}_0^T if and only if

$$\mathbf{Q}_0\mathbf{v}_0^* + \mathbf{A}^T\left(\mathbf{B}^T\right)^{-1}\mathbf{Q}_1\mathbf{v}_1^* = 0. \quad (3.236)$$

Equation 3.236 does not depend upon $u(x)$ or its derivative and is the adjoint boundary condition for the general boundary condition in Eq. 3.227.

3.2.6.2 Scalar Wave Equations

We now consider the scalar wave-equation operator defined by Eq. 3.182 acting on functions defined on $x \in [0, 1]$ and $t \in [0, 1]$. The conjunct for this operator is given in Eq. 3.185. The boundaries of this two-dimensional, space-time domain is shown in Fig. 3.1, in which the horizontal axis is the space dimension and the vertical axis is the time dimension. The normal directions on the four boundaries of the space-time area are marked with arrows. Considering Eq. 3.185, at the boundary $t = 0$, the Euclidean dot product between the conjunct and the normal direction is

$$\mathbf{Q} \cdot \hat{\mathbf{e}}_n = \begin{bmatrix} v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \\ -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=0}$$

and at the boundary $t = 1$, the dot product is

$$\mathbf{Q} \cdot \hat{\mathbf{e}}_n = \begin{bmatrix} v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \\ -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=1}.$$

At the boundary $x = 0$, the dot product is

$$\mathbf{Q} \cdot \hat{\mathbf{e}}_n = \begin{bmatrix} v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \\ -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \left(-v^* \frac{\partial u}{\partial x} + u \frac{\partial v^*}{\partial x} \right) \Big|_{x=0}$$

and at the boundary $x = 1$, the dot product is

$$\mathbf{Q} \cdot \hat{\mathbf{e}}_n = \begin{bmatrix} v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \\ -\frac{1}{\alpha^2(x)} \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \left(v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \right) \Big|_{x=1}.$$

The surface term can therefore be expressed as

$$\begin{aligned} \int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da &= \int_0^1 dx \frac{1}{\alpha^2(x)} \left[\left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=0} - \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=1} \right] \\ &+ \int_0^1 dt \left[\left(-v^* \frac{\partial u}{\partial x} + u \frac{\partial v^*}{\partial x} \right) \Big|_{x=0} + \left(v^* \frac{\partial u}{\partial x} - u \frac{\partial v^*}{\partial x} \right) \Big|_{x=1} \right]. \end{aligned} \quad (3.237)$$

Considering the *quiescent past* initial conditions for $u(x, t)$, i.e.,

$$u(x, t) = 0, \frac{\partial u}{\partial t}(x, t) = 0, \forall t \leq 0 \quad (3.238)$$

and the Dirichlet boundary condition for $u(x, t)$, i.e.,

$$u(0, t) = 0, u(1, t) = 0, \quad (3.239)$$

the surface term in Eq. 3.237 becomes zero for arbitrary $u(x, t)$ if and only if $v^*(x, t)$ satisfies the *quiescent future* final conditions

$$v^*(x, t) = 0, \frac{\partial v^*}{\partial t}(x, t) = 0, \forall t \geq 1 \quad (3.240)$$

and also the Dirichlet boundary condition

$$v^*(0, t) = 0, v^*(1, t) = 0. \quad (3.241)$$

If instead of the Dirichlet boundary condition, $u(x, t)$ satisfies the Neumann boundary condition

$$\frac{\partial u}{\partial x}(0, t) = 0, \frac{\partial u}{\partial x}(1, t) = 0, \quad (3.242)$$

then the surface term is zero for arbitrary $u(x, t)$ if and only if $v^*(x, t)$ also satisfies the Neumann boundary condition

$$\frac{\partial v^*}{\partial x}(0, t) = 0, \quad \frac{\partial v^*}{\partial x}(1, t) = 0. \quad (3.243)$$

A similar analysis can also be applied to the scalar wave-equation operator defined in Eq. 3.186, whose formal adjoint and conjunct are given in Eqs. 3.188 and 3.190. Suppose the spatial volume of the physical domain is given by $V \subset \mathbb{R}^3$ and the temporal duration is given by $[0, T]$. We denote the normal direction on the boundary of the spatial volume ∂V as $\mathbf{n}(\mathbf{x})$. Considering that the first 3 elements in the conjunct vector in Eq. 3.190 can be expressed as $v^* \nabla u - u \nabla v^*$, where ∇ is the spatial gradient operator in \mathbb{R}^3 , we can express the surface term as

$$\begin{aligned} \int_{\partial\Omega} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da &= \int_V dV \frac{1}{\alpha^2(\mathbf{x})} \left[\left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=0} - \left(v^* \frac{\partial u}{\partial t} - u \frac{\partial v^*}{\partial t} \right) \Big|_{t=T} \right] \\ &+ \int_0^T dt \int_{\partial V} da(\mathbf{x}) (v^* \nabla u - u \nabla v^*) \cdot \mathbf{n}(\mathbf{x}). \end{aligned} \quad (3.244)$$

The second term on the RHS of Eq. 3.244 can be expressed as

$$\begin{aligned} &\int_0^T dt \int_{\partial V} da(\mathbf{x}) (v^* \nabla u - u \nabla v^*) \cdot \mathbf{n}(\mathbf{x}) \\ &= \int_0^T dt \int_{\partial V} da(\mathbf{x}) \left(v^* \frac{\partial u}{\partial \mathbf{n}} - \frac{\partial v^*}{\partial \mathbf{n}} u \right), \end{aligned} \quad (3.245)$$

where $\partial/\partial \mathbf{n}$ is the directional derivative along $\mathbf{n}(\mathbf{x})$, the normal direction of ∂V . Using Eq. 3.244, we can verify that if $u(\mathbf{x}, t)$ satisfies the quiescent past initial conditions, i.e.,

$$u(\mathbf{x}, t) = 0, \quad \frac{\partial u}{\partial t}(\mathbf{x}, t) = 0, \quad \forall t \leq 0 \quad (3.246)$$

$v^*(\mathbf{x}, t)$ satisfies the quiescent future final conditions, i.e.,

$$v^*(\mathbf{x}, t) = 0, \quad \frac{\partial v^*}{\partial t}(\mathbf{x}, t) = 0, \quad \forall t \geq T. \quad (3.247)$$

Using Eq. 3.245, we can verify that if $u(\mathbf{x}, t)$ satisfies the Dirichlet boundary condition, i.e.,

$$u(\mathbf{x}, t) = 0, \quad \forall \mathbf{x} \in \partial V, \quad (3.248)$$

$v^*(\mathbf{x}, t)$ also satisfies the Dirichlet boundary condition, i.e.,

$$v^*(\mathbf{x}, t) = 0, \quad \forall \mathbf{x} \in \partial V. \quad (3.249)$$

If $u(\mathbf{x}, t)$ satisfies the Neumann boundary condition, i.e.,

$$\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x}, t) = 0, \quad \forall \mathbf{x} \in \partial V, \quad (3.250)$$

$v^*(\mathbf{x}, t)$ also satisfies the Neumann boundary condition, i.e.,

$$\frac{\partial v^*}{\partial \mathbf{n}}(\mathbf{x}, t) = 0, \quad \forall \mathbf{x} \in \partial V. \quad (3.251)$$

3.2.6.3 Anisotropic Viscoelastic Wave Equation

We now consider the anisotropic viscoelastic wave-equation operator defined in Eq. 3.203, whose formal adjoint and surface term are given in Eqs. 3.215 and 3.216. Suppose $u_i(\mathbf{x}, t)$ satisfies the quiescent past initial conditions, i.e.,

$$u_i(\mathbf{x}, t) = 0, \frac{\partial u_i}{\partial t}(\mathbf{x}, t) = 0, \forall t \leq 0, \quad (3.252)$$

and also the free-surface boundary condition, i.e.,

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k}{\partial x_l}(\mathbf{x}, \tau) = 0, \forall \mathbf{x} \in \partial V. \quad (3.253)$$

Considering the quiescent past initial condition for $u_i(\mathbf{x}, t)$ (Eq. 3.252), the first term in Eq. 3.216 is zero for arbitrary $u_i(\mathbf{x}, t)$ if and only if $v_i^*(\mathbf{x}, t)$ satisfies the quiescent future final conditions, i.e.,

$$v_i^*(\mathbf{x}, t) = 0, \frac{\partial v_i^*}{\partial t}(\mathbf{x}, t) = 0, \forall t \geq T. \quad (3.254)$$

A direct consequence of the quiescent past initial conditions and the quiescent future final conditions is that the last two terms in Eq. 3.216 become zeros. The free-surface boundary condition for $u_i(\mathbf{x}, t)$ in Eq. 3.253 makes the third term in Eq. 3.216 zero. Therefore the surface term is zero for arbitrary $u_i(\mathbf{x}, t)$ if and only if $v_i^*(\mathbf{x}, t)$ satisfies the free-surface condition defined using the anti-causal rate-of-relaxation function $\Psi'_{ijkl}(\mathbf{x}, t)$,

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial v_k^*}{\partial x_l}(\mathbf{x}, \tau) = 0, \forall \mathbf{x} \in \partial V. \quad (3.255)$$

If we replace the free-surface boundary condition for $u_i(\mathbf{x}, t)$ with the rigidity condition, i.e.,

$$u_i(\mathbf{x}, t) = 0, \forall \mathbf{x} \in \partial V, \quad (3.256)$$

we can verify that the corresponding adjoint boundary condition is

$$v_i^*(\mathbf{x}, t) = 0, \forall \mathbf{x} \in \partial V. \quad (3.257)$$

After we have obtained both the formal adjoint of the differential operator and also the adjoint initial/boundary conditions, the full adjoint problem is then completely specified.

At the beginning of Sect. 3.2, I mentioned that whether a differential operator is self-adjoint or not depends not only on the formal adjoint operator $\mathbf{L}_\mathbf{x}^\dagger$ but also on its domain $\mathcal{D}(\mathbf{L}_\mathbf{x}^\dagger)$. If in addition to $\mathbf{L}_\mathbf{x} = \mathbf{L}_\mathbf{x}^\dagger$, the adjoint initial/boundary conditions for $\mathbf{L}_\mathbf{x}^\dagger$ are identical to the initial/boundary conditions for $\mathbf{L}_\mathbf{x}$, then we have $\mathcal{D}(\mathbf{L}_\mathbf{x}^\dagger) = \mathcal{D}(\mathbf{L}_\mathbf{x})$ and we say that $\mathbf{L}_\mathbf{x}$ is self-adjoint.

3.3 Properties of Green's Functions

Consider a linear differential equation expressed as

$$\mathbf{L}_x [u] = f(\mathbf{x}), \quad (3.258)$$

where \mathbf{L}_x is a linear differential operator and $\mathbf{x} \in \Omega \subset \mathbb{R}^n$. In the abstract Dirac notation, we can write Eq. 3.258 as

$$\mathbf{L} |u\rangle = |f\rangle. \quad (3.259)$$

Suppose the linear operator \mathbf{L} has an inverse, \mathbf{L}^{-1} , we can then define the operator

$$\mathbf{G} \equiv \mathbf{L}^{-1}, \quad (3.260)$$

and we have

$$|u\rangle = \mathbf{L}^{-1} |f\rangle = \mathbf{G} |f\rangle. \quad (3.261)$$

For the continuous-index case, we can choose the basis $B = \{|e_{\mathbf{x}}\rangle\}_{\mathbf{x} \in \Omega \subset \mathbb{R}^n}$ (Sect. 3.1.5). Take the inner product with $|e_{\mathbf{x}}\rangle$ on both sides of Eq. 3.261 and insert the completeness relation $\mathbf{1} = \int_{\Omega} |e_{\mathbf{y}}\rangle w(\mathbf{y}) \langle e_{\mathbf{y}}| dV(\mathbf{y})$ between \mathbf{G} and $|f\rangle$, where dV is a measure of the manifold $\Omega \subset \mathbb{R}^n$, we obtain

$$u(\mathbf{x}) = \int_{\Omega} G(\mathbf{x}, \mathbf{y}) w(\mathbf{y}) f(\mathbf{y}) dV(\mathbf{y}). \quad (3.262)$$

Therefore the inverse of a differential operator can be expressed as an integral operator with its integration kernel given by the *Green's function* of the differential operator $G(\mathbf{x}, \mathbf{y})$.

Considering Eqs. 3.260 and 3.71, we have

$$\mathbf{L}\mathbf{G} = \mathbf{1} \quad (3.263)$$

$$\Rightarrow \langle e_{\mathbf{x}} | \mathbf{L}\mathbf{G} | e_{\mathbf{y}} \rangle = \langle e_{\mathbf{x}} | e_{\mathbf{y}} \rangle = \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})} \quad (3.264)$$

Replacing $\mathbf{1}$ with the completeness relation $\int |e_{\mathbf{x}'}\rangle w(\mathbf{x}') \langle e_{\mathbf{x}'}| d^n x'$ in Eq. 3.264, we obtain

$$\begin{aligned} \int \langle e_{\mathbf{x}} | \mathbf{L} | e_{\mathbf{x}'} \rangle w(\mathbf{x}') \langle e_{\mathbf{x}'} | \mathbf{G} | e_{\mathbf{y}} \rangle d^n x' &= \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})}, \\ \Rightarrow \int L(\mathbf{x}, \mathbf{x}') w(\mathbf{x}') G(\mathbf{x}', \mathbf{y}) d^n x' &= \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})}. \end{aligned} \quad (3.265)$$

Using Eqs. 3.115 and 3.258, we can verify that for linear differential operators, we have

$$L(\mathbf{x}, \mathbf{x}') = \frac{\delta(\mathbf{x} - \mathbf{x}')}{w(\mathbf{x}')} \mathbf{L}_x. \quad (3.266)$$

Bring Eq. 3.266 into Eq. 3.265, we obtain the differential equation satisfied by the Green's function

$$\mathbf{L}_x G(\mathbf{x}, \mathbf{y}) = \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})}. \quad (3.267)$$

Comparing with Eq. 3.258, the Green's function is the solution of the differential equation with the RHS replaced by the Dirac delta function, if we assume $w(\mathbf{x}) = 1$. Differential operators are usually accompanied by initial/boundary conditions, therefore we also require the Green's function to satisfy those initial/boundary conditions, i.e., $G(\mathbf{x}, \mathbf{y}) \in \mathcal{D}(\mathbf{L}_x)$. Since the Dirac delta function is a generalized function or a distribution (Sect. 3.1.6), the Green's function can also be understood in the distributional sense.

If we bring the representation of $u(\mathbf{x})$ in Eq. 3.262 into the LHS of Eq. 3.258, considering Eq. 3.267, we have

$$\begin{aligned} \mathbf{L}_x u(\mathbf{x}) &= \int [\mathbf{L}_x G(\mathbf{x}, \mathbf{y})] w(\mathbf{y}) f(\mathbf{y}) d^n y \\ &= \int \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})} w(\mathbf{y}) f(\mathbf{y}) d^n y = f(\mathbf{x}), \end{aligned} \quad (3.268)$$

which shows that Eq. 3.262 indeed gives a solution to the differential Eq. 3.258.

In general, linear differential operators are not bounded operators, such as those discussed in Sect. 3.1.8. However, the inverse of a linear differential operator, which can be expressed as an integral operator with its integration kernel given by the Green's function, is usually bounded. We can therefore study properties of unbounded linear differential operators by using their bounded inverses.

3.3.1 Adjoint Green's Functions

We can also study the Green's function for the adjoint differential operator \mathbf{L}_x^\dagger coupled with the adjoint initial/boundary conditions. If we denote this Green's function as $G^\dagger(\mathbf{x}, \mathbf{y}) \in \mathcal{D}(\mathbf{L}_x^\dagger)$, we have

$$\mathbf{L}_x^\dagger G^\dagger(\mathbf{x}, \mathbf{y}) = \frac{\delta(\mathbf{x} - \mathbf{y})}{w(\mathbf{x})}, \quad (3.269)$$

and we call $G^\dagger(\mathbf{x}, \mathbf{y})$ the *adjoint Green's function*. In the generalized Green's identity, Eq. 3.158, we can replace $v(\mathbf{x}, \mathbf{y})$ with $G^\dagger(\mathbf{x}, \mathbf{y})$. Since the surface term is zero by the construction of the adjoint system, we have

$$\begin{aligned} &\int_{\Omega} w(\mathbf{x}) [G^\dagger(\mathbf{x}, \mathbf{y})]^* \mathbf{L}_x [u] dV(\mathbf{x}) \\ &= \int_{\Omega} w(\mathbf{x}) u(\mathbf{x}) \left(\mathbf{L}_x^\dagger [G^\dagger] \right)^* dV(\mathbf{x}), \end{aligned} \quad (3.270)$$

where we have generalized the inner product by including the weighting function $w(\mathbf{x})$. Considering Eqs. 3.258 and 3.269, Eq. 3.270 can be re-written as

$$\int_{\Omega} w(\mathbf{x}) \left[G^{\dagger}(\mathbf{x}, \mathbf{y}) \right]^* f(\mathbf{x}) dV(\mathbf{x}) = \int_{\Omega} u(\mathbf{x}) \delta(\mathbf{x} - \mathbf{y}) dV(\mathbf{x}) = u(\mathbf{y}). \quad (3.271)$$

If we exchange \mathbf{x} and \mathbf{y} in Eq. 3.271, we obtain

$$u(\mathbf{x}) = \int_{\Omega} \left[G^{\dagger}(\mathbf{y}, \mathbf{x}) \right]^* w(\mathbf{y}) f(\mathbf{y}) dV(\mathbf{y}). \quad (3.272)$$

Suppose we have

$$\mathbf{L}_{\mathbf{x}}^{\dagger}[v] = h(\mathbf{x}). \quad (3.273)$$

If we replace $u(\mathbf{x})$ with $G(\mathbf{x}, \mathbf{y})$ in the Green's identity, Eq. 3.158, we obtain

$$\int_{\Omega} w(\mathbf{x}) v^*(\mathbf{x}) \mathbf{L}_{\mathbf{x}}[G] dV(\mathbf{x}) = \int_{\Omega} w(\mathbf{x}) G(\mathbf{x}, \mathbf{y}) \left(\mathbf{L}_{\mathbf{x}}^{\dagger}[v] \right)^* dV(\mathbf{x}),$$

considering Eqs. 3.267 and 3.273, we have

$$\begin{aligned} \int_{\Omega} v^*(\mathbf{x}) \delta(\mathbf{x} - \mathbf{y}) dV(\mathbf{x}) &= \int_{\Omega} w(\mathbf{x}) G(\mathbf{x}, \mathbf{y}) h^*(\mathbf{x}) dV(\mathbf{x}) \\ \Rightarrow v^*(\mathbf{y}) &= \int_{\Omega} w(\mathbf{x}) G(\mathbf{x}, \mathbf{y}) h^*(\mathbf{x}) dV(\mathbf{x}) \\ \Rightarrow v(\mathbf{y}) &= \int_{\Omega} G^*(\mathbf{x}, \mathbf{y}) w(\mathbf{x}) h(\mathbf{x}) dV(\mathbf{x}) \\ \Rightarrow v(\mathbf{x}) &= \int_{\Omega} G^*(\mathbf{y}, \mathbf{x}) w(\mathbf{y}) h(\mathbf{y}) dV(\mathbf{y}). \end{aligned} \quad (3.274)$$

Equations 3.272 and 3.274 may appear counter-intuitive at first glance, because unlike Eq. 3.262, which relates solutions of $\mathbf{L}_{\mathbf{x}}[u] = f(\mathbf{x})$ to the Green's function $G(\mathbf{x}, \mathbf{y})$, these two equations relate solutions of $\mathbf{L}_{\mathbf{x}}[u] = f(\mathbf{x})$ to the adjoint Green's function $G^{\dagger}(\mathbf{x}, \mathbf{y})$ and solutions of $\mathbf{L}_{\mathbf{x}}^{\dagger}[v] = h(\mathbf{x})$ to the Green's function $G(\mathbf{x}, \mathbf{y})$. To understand Eqs. 3.272 and 3.274, we substitute $G(\mathbf{z}, \mathbf{y})$ for $u(\mathbf{z})$ and $G^{\dagger}(\mathbf{z}, \mathbf{x})$ for $v(\mathbf{z})$ in Eq. 3.158 and we obtain

$$\begin{aligned} &\int_{\Omega} w(\mathbf{z}) \left[G^{\dagger}(\mathbf{z}, \mathbf{x}) \right]^* \mathbf{L}_{\mathbf{z}}[G(\mathbf{z}, \mathbf{y})] dV(\mathbf{z}) \\ &= \int_{\Omega} w(\mathbf{z}) G(\mathbf{z}, \mathbf{y}) \left(\mathbf{L}_{\mathbf{z}}^{\dagger} \left[G^{\dagger}(\mathbf{z}, \mathbf{x}) \right] \right)^* dV(\mathbf{z}). \end{aligned}$$

Considering Eqs. 3.267 and 3.269, we obtain

$$G(\mathbf{x}, \mathbf{y}) = \left[G^{\dagger}(\mathbf{y}, \mathbf{x}) \right]^* \Leftrightarrow G^{\dagger}(\mathbf{x}, \mathbf{y}) = G^*(\mathbf{y}, \mathbf{x}). \quad (3.275)$$

Bring Eq. 3.275 into Eq. 3.272, we recover Eq. 3.262. Bring Eq. 3.275 into Eq. 3.274, we obtain

$$v(\mathbf{x}) = \int_{\Omega} G^{\dagger}(\mathbf{x}, \mathbf{y}) w(\mathbf{y}) h(\mathbf{y}) dV(\mathbf{y}). \quad (3.276)$$

Note that Eq. 3.275 is not the *reciprocity relation* as derived in e.g., (Aki and Richards 2002; Dahlen and Tromp 1998). Equation 3.275 is a relation between G and G^\dagger while the reciprocity relation is a symmetry condition on G alone.

3.3.2 Reciprocity

Suppose \mathbf{L}_x is self-adjoint, i.e., $\mathbf{L}_x = \mathbf{L}_x^\dagger$, $\mathcal{D}(\mathbf{L}_x) = \mathcal{D}(\mathbf{L}_x^\dagger)$, then we have

$$G(\mathbf{y}, \mathbf{x}) = G^\dagger(\mathbf{y}, \mathbf{x}). \quad (3.277)$$

Bring Eq. 3.277 into Eq. 3.275, we obtain

$$G(\mathbf{x}, \mathbf{y}) = \left[G^\dagger(\mathbf{y}, \mathbf{x}) \right]^* = G^*(\mathbf{y}, \mathbf{x}). \quad (3.278)$$

Equation 3.278 is often called the reciprocity relation. As shown above, the reciprocity relation is a consequence of the self-adjointness of the differential operator. If the coefficients of \mathbf{L}_x are real, then G is a real function and we have $G^* = G$ and Eq. 3.278 can be expressed as

$$G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x}). \quad (3.279)$$

The reciprocity relations in Eqs. 3.278 and 3.279 are often used in problems involving scalar wave equations in exploration seismology.

3.3.2.1 Scalar Wave-Equation Operator

Consider the scalar wave-equation operator $\mathbf{L}_{y,\tau}$ given in Eq. 3.186 and its formal adjoint $\mathbf{L}_{y,\tau}^\dagger$ given in Eq. 3.188. The Green's function $G(\mathbf{y}, \tau; \mathbf{x}', t')$ and the adjoint Green's function $G^\dagger(\mathbf{y}, \tau; \mathbf{x}', t')$ satisfy the equations

$$\mathbf{L}_{y,\tau} [G(\mathbf{y}, \tau; \mathbf{x}', t')] = \delta(\mathbf{y} - \mathbf{x}')\delta(\tau - t'), \quad (3.280)$$

$$\mathbf{L}_{y,\tau}^\dagger [G^\dagger(\mathbf{y}, \tau; \mathbf{x}, t)] = \delta(\mathbf{y} - \mathbf{x})\delta(\tau - t), \quad (3.281)$$

and the homogeneous initial/boundary conditions (Eqs. 3.246 and 3.248) and the homogeneous adjoint initial/boundary conditions (Eqs. 3.247 and 3.249). Note that here I am using “;” instead of “,” to separate the spatial-temporal coordinates of the source from those of the wavefield inside the Green's functions. Bring the Green's function and the adjoint Green's function into the generalized Green's identity (Eq. 3.159) with the surface term set to zero, we obtain

$$\begin{aligned}
& \int_0^T d\tau \int_V dV(\mathbf{y}) \left[G^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right]^* \mathbf{L}_{\mathbf{y}, \tau} \left[G(\mathbf{y}, \tau; \mathbf{x}', t') \right] \\
&= \int_0^T d\tau \int_V dV(\mathbf{y}) G(\mathbf{y}, \tau; \mathbf{x}', t') \mathbf{L}_{\mathbf{y}, \tau}^\dagger \left[G^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right] \\
&\Rightarrow \int_0^T d\tau \int_V dV(\mathbf{y}) \left[G^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right]^* \delta(\mathbf{y} - \mathbf{x}') \delta(\tau - t') \\
&= \int_0^T d\tau \int_V dV(\mathbf{y}) G(\mathbf{y}, \tau; \mathbf{x}', t') \delta(\mathbf{y} - \mathbf{x}) \delta(\tau - t) \\
&\Rightarrow \left[G^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* = G(\mathbf{x}, t; \mathbf{x}', t'). \tag{3.282}
\end{aligned}$$

Note that Eq. 3.282 is not the reciprocity relation. If the coefficients of the scalar wave equation are real, both G^\dagger and G are real functions and we have

$$G^\dagger(\mathbf{x}', t'; \mathbf{x}, t) = G(\mathbf{x}, t; \mathbf{x}', t'). \tag{3.283}$$

The Green's function and the adjoint Green's function are translation-invariant in time since the material property $\alpha(\mathbf{x})$ does not vary with time. Therefore we can re-write Eq. 3.283 as

$$G^\dagger(\mathbf{x}', t' - t; \mathbf{x}, 0) = G(\mathbf{x}, t - t'; \mathbf{x}', 0). \tag{3.284}$$

If we replace $t - t'$ with τ , we have

$$G^\dagger(\mathbf{x}', -\tau; \mathbf{x}, 0) = G(\mathbf{x}, \tau; \mathbf{x}', 0). \tag{3.285}$$

If we apply Fourier transform on both sides of Eq. 3.285 to convert the time dimension τ to the frequency dimension Ω , we obtain

$$\left[\hat{G}^\dagger(\mathbf{x}', \omega; \mathbf{x}) \right]^* = \hat{G}(\mathbf{x}, \omega; \mathbf{x}'), \tag{3.286}$$

where \hat{G}^\dagger and \hat{G} are Fourier transforms of G^\dagger and G , respectively. The conjugation * on \hat{G}^\dagger on the LHS of Eq. 3.286 is due to the minus sign in front of τ in the LHS of Eq. 3.285.

Although the scalar wave-equation operator is formally self-adjoint, i.e., $\mathbf{L}_{\mathbf{y}, \tau} = \mathbf{L}_{\mathbf{y}, \tau}^\dagger$, the initial conditions (Eq. 3.246) and the corresponding adjoint initial conditions (Eq. 3.247) are different, therefore the scalar wave-equation operator is not self-adjoint. However, if we reverse the direction of the time axis in the adjoint system by introducing

$$\tilde{t} = T - t, \tag{3.287}$$

and substitute t for $T - \tilde{t}$, the adjoint system becomes identical to the forward system. The formal adjoint operator in Eq. 3.188 and the adjoint boundary condition

in Eq. 3.249 are both invariant under time-reversal. The adjoint initial (final) conditions in Eq. 3.247 become identical to the initial conditions in Eq. 3.246. To verify this point, we replace t with $T - \tilde{t}$ in Eq. 3.247 and obtain

$$v^*(\mathbf{x}, T - \tilde{t}) = 0, \frac{\partial v^*}{\partial \tilde{t}}(\mathbf{x}, T - \tilde{t}) = 0, \forall \tilde{t} \leq 0. \quad (3.288)$$

If we introduce the function

$$\tilde{v}(\mathbf{x}, \tilde{t}) \equiv v(\mathbf{x}, T - \tilde{t}), \quad (3.289)$$

Eq. 3.288 then becomes

$$\tilde{v}^*(\mathbf{x}, \tilde{t}) = 0, \frac{\partial \tilde{v}^*}{\partial \tilde{t}}(\mathbf{x}, \tilde{t}) = 0, \forall \tilde{t} \leq 0, \quad (3.290)$$

which is identical to the initial conditions in Eq. 3.246. A direct consequence of this self-adjointness after time-reversal is that we have

$$G^\dagger(\mathbf{x}', t'; \mathbf{x}, t) = G(\mathbf{x}', T - t'; \mathbf{x}, T - t). \quad (3.291)$$

Bring Eq. 3.291 into Eq. 3.283, we obtain

$$G(\mathbf{x}', T - t'; \mathbf{x}, T - t) = G(\mathbf{x}, t; \mathbf{x}', t'). \quad (3.292)$$

Considering that the Green's function is translation-invariant in time, we can rewrite Eq. 3.292 as

$$G(\mathbf{x}', t - t'; \mathbf{x}, 0) = G(\mathbf{x}, t - t'; \mathbf{x}', 0), \quad (3.293)$$

which is the reciprocity relation of the scalar wave equation in the time domain.

3.3.2.2 Anisotropic Viscoelastic Wave-Equation Operator

We now consider the Green's function for the anisotropic viscoelastic wave-equation operator defined in Eq. 3.204. The Green's function of the wave-equation operator $G_{pi}(\mathbf{y}, \tau; \mathbf{x}', t')$ and the adjoint Green's function of the adjoint wave-equation operator $G_{pj}^\dagger(\mathbf{y}, \tau; \mathbf{x}, t)$ satisfy the equations

$$\mathbf{L}_{\mathbf{y}, \tau} [\mathbf{G}(\mathbf{y}, \tau; \mathbf{x}', t')]_{pi} = \delta_{pi} \delta(\mathbf{y} - \mathbf{x}') \delta(\tau - t'), \quad (3.294)$$

$$\mathbf{L}_{\mathbf{y}, \tau}^\dagger [\mathbf{G}^\dagger(\mathbf{y}, \tau; \mathbf{x}, t)]_{pj} = \delta_{pj} \delta(\mathbf{y} - \mathbf{x}) \delta(\tau - t), \quad (3.295)$$

where $\mathbf{L}_{\mathbf{y}, \tau}$ and $\mathbf{L}_{\mathbf{y}, \tau}^\dagger$ are the anisotropic viscoelastic wave-equation operator defined in Eq. 3.204 and its formal adjoint given in Eq. 3.215, δ_{pi} and δ_{pj} are the Kronecker delta and $\delta(\cdot)$ represents the Dirac delta function. Bring the Green's function and the

adjoint Green's function into the generalized Green's identity with the surface term set to zero, we obtain

$$\begin{aligned} & \int_0^T d\tau \int_V dV(\mathbf{y}) \sum_{p=1}^3 \left[G_{pj}^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right]^* \mathbf{L}_{\mathbf{y}, \tau} [\mathbf{G}(\mathbf{y}, \tau; \mathbf{x}', t')]_{pj} \\ &= \int_0^T d\tau \int_V dV(\mathbf{y}) \sum_{p=1}^3 G_{pi}(\mathbf{y}, \tau; \mathbf{x}', t') \mathbf{L}_{\mathbf{y}, \tau}^\dagger \left[\mathbf{G}^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right]_{pj}, \end{aligned}$$

considering Eqs. 3.294 and 3.295, we have

$$\begin{aligned} & \int_0^T d\tau \int_V dV(\mathbf{y}) \sum_{p=1}^3 \left[G_{pj}^\dagger(\mathbf{y}, \tau; \mathbf{x}, t) \right]^* \delta_{pi} \delta(\mathbf{y} - \mathbf{x}') \delta(\tau - t') \\ &= \int_0^T d\tau \int_V dV(\mathbf{y}) \sum_{p=1}^3 G_{pi}(\mathbf{y}, \tau; \mathbf{x}', t') \delta_{pj} \delta(\mathbf{y} - \mathbf{x}) \delta(\tau - t) \\ &\Rightarrow \left[G_{ij}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* = G_{ji}(\mathbf{x}, t; \mathbf{x}', t'). \end{aligned} \quad (3.296)$$

Since the coefficients of the wave-equation operator and its formal adjoint are real in most seismic applications, \mathbf{G} and \mathbf{G}^\dagger are real functions, therefore we have

$$G_{ij}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) = G_{ji}(\mathbf{x}, t; \mathbf{x}', t'). \quad (3.297)$$

Equation 3.296 is analogous to Eq. 3.275. Equations 3.296 and 3.297 are not the reciprocity relation.

The anisotropic viscoelastic wave-equation operator is not self-adjoint. However, if we reverse the direction of the time axis by introducing

$$\tilde{t} = T - t, \quad (3.298)$$

and replace t with $T - \tilde{t}$ in the formal adjoint (Eq. 3.215) and the adjoint initial/boundary conditions (Eqs. 3.254, 3.255 or 3.257), the adjoint problem becomes identical to the forward problem. Let's first consider the adjoint initial conditions (i.e., the final conditions). Replacing t with $T - \tilde{t}$ in Eq. 3.255, we obtain

$$v_i^*(\mathbf{x}, T - \tilde{t}) = 0, \quad \frac{\partial v_i^*}{\partial \tilde{t}}(\mathbf{x}, T - \tilde{t}) = 0, \quad \forall \tilde{t} \leq 0. \quad (3.299)$$

If we introduce a new function

$$\tilde{v}_i(\mathbf{x}, \tilde{t}) = v_i(\mathbf{x}, T - \tilde{t}), \quad (3.300)$$

we then have

$$\tilde{v}_i^*(\mathbf{x}, \tilde{t}) = 0, \quad \frac{\partial \tilde{v}_i^*}{\partial \tilde{t}}(\mathbf{x}, \tilde{t}) = 0, \quad \forall \tilde{t} \leq 0 \quad (3.301)$$

which are identical to the quiescent past initial conditions in Eq. 3.252. Let's now consider the adjoint free-surface boundary condition in Eq. 3.255. Replacing t with $T - \tilde{t}$ in Eq. 3.255, we obtain

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, T - \tilde{t} - \tau) \frac{\partial v_k^*}{\partial x_l}(\mathbf{x}, \tau) = 0, \forall \mathbf{x} \in \partial V.$$

Replacing the anti-causal rate-of-relaxation function Ψ'_{ijkl} with the causal rate-of-relaxation function Ψ_{ijkl} using Eq. 3.211, we have

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, \tilde{t} - (T - \tau)) \frac{\partial v_k^*}{\partial x_l}(\mathbf{x}, \tau) = 0, \forall \mathbf{x} \in \partial V.$$

Introducing a new variable $\tilde{\tau} = T - \tau$, we then have

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tilde{\tau} \Psi_{ijkl}(\mathbf{x}, \tilde{t} - \tilde{\tau}) \frac{\partial v_k^*}{\partial x_l}(\mathbf{x}, T - \tilde{\tau}) = 0, \forall \mathbf{x} \in \partial V.$$

Using the definition of \tilde{v}_i in Eq. 3.300, we obtain

$$\sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tilde{\tau} \Psi_{ijkl}(\mathbf{x}, \tilde{t} - \tilde{\tau}) \frac{\partial \tilde{v}_k^*}{\partial x_l}(\mathbf{x}, \tilde{\tau}) = 0, \forall \mathbf{x} \in \partial V, \quad (3.302)$$

which is identical to the free-surface boundary condition in Eq. 3.253. Reversing the direction of the time axis for the adjoint boundary condition in Eq. 3.257 is straightforward.

Let's now consider the formal adjoint in Eq. 3.215. Replacing t with $T - \tilde{t}$, we obtain

$$\begin{aligned} \mathbf{L}_{\mathbf{x}, \tilde{t}}^\dagger [\mathbf{v}(\mathbf{x}, T - \tilde{t})]_i &= \rho(\mathbf{x}) \frac{\partial^2 v_i(\mathbf{x}, T - \tilde{t})}{\partial \tilde{t}^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, T - \tilde{t} - \tau) \frac{\partial v_k(\mathbf{x}, \tau)}{\partial x_l} \right]. \end{aligned}$$

Replacing Ψ'_{ijkl} with Ψ_{ijkl} in the second term on the RHS and substituting $T - \tau$ with $\tilde{\tau}$, we obtain

$$\begin{aligned} \mathbf{L}_{\mathbf{x}, \tilde{t}}^\dagger [\mathbf{v}(\mathbf{x}, T - \tilde{t})]_i &= \rho(\mathbf{x}) \frac{\partial^2 v_i(\mathbf{x}, T - \tilde{t})}{\partial \tilde{t}^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tilde{\tau} \Psi_{ijkl}(\mathbf{x}, \tilde{t} - \tilde{\tau}) \frac{\partial v_k(\mathbf{x}, T - \tilde{\tau})}{\partial x_l} \right]. \end{aligned}$$

Using the definition in Eq. 3.300, we obtain

$$\begin{aligned} \mathbf{L}_{\mathbf{x}, \tilde{t}}^\dagger [\tilde{\mathbf{v}}(\mathbf{x}, \tilde{t})]_i &= \rho(\mathbf{x}) \frac{\partial^2 \tilde{v}_i(\mathbf{x}, \tilde{t})}{\partial \tilde{t}^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tilde{\tau} \Psi_{ijkl}(\mathbf{x}, \tilde{t} - \tilde{\tau}) \frac{\partial \tilde{v}_k(\mathbf{x}, \tilde{\tau})}{\partial x_l} \right], \end{aligned} \quad (3.303)$$

which is identical to Eq. 3.204.

The analysis above shows that the adjoint anisotropic viscoelastic wave-equation operator becomes identical to the forward problem after reversing the direction of the time axis. A direct consequence is that we have

$$G_{ij}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) = G_{ij}(\mathbf{x}', T - t'; \mathbf{x}, T - t). \quad (3.304)$$

It can be verified that the anisotropic viscoelastic wave-equation operator in Eq. 3.204 is translation-invariant in time. We therefore can re-write the RHS of Eq. 3.304 as

$$G_{ij}(\mathbf{x}', T - t'; \mathbf{x}, T - t) = G_{ij}(\mathbf{x}', t - t'; \mathbf{x}, 0). \quad (3.305)$$

Considering Eq. 3.297, we obtain

$$\begin{aligned} G_{ij}(\mathbf{x}', t - t'; \mathbf{x}, 0) &= G_{ji}(\mathbf{x}, t; \mathbf{x}', t') \\ \Rightarrow G_{ij}(\mathbf{x}', t - t'; \mathbf{x}, 0) &= G_{ji}(\mathbf{x}, t - t'; \mathbf{x}', 0), \end{aligned} \quad (3.306)$$

which is the reciprocity relation for the Green's function of the anisotropic viscoelastic wave-equation operator (e.g., Aki and Richards 2002; Dahlen and Tromp 1998).

The derivation above shows that the requirements for obtaining the reciprocity relations (Eqs. 3.278 and 3.306) are more stringent than those for obtaining Eqs. 3.275 and 3.296. To obtain the reciprocity relation in Eq. 3.278, the differential operator must be self-adjoint, i.e., $\mathbf{L}_{\mathbf{x}}^\dagger = \mathbf{L}_{\mathbf{x}}$ and $\mathcal{D}(\mathbf{L}_{\mathbf{x}}^\dagger) = \mathcal{D}(\mathbf{L}_{\mathbf{x}})$. To obtain the reciprocity relation in Eq. 3.306, the formal adjoint operator and the adjoint initial/boundary conditions must be identical to the forward operator and its accompanying initial/boundary conditions after reversing the direction of the time axis and the forward problem must be translation-invariant in time. On the other hand, Eqs. 3.275 and 3.296 hold automatically because of the way in which the Green's functions and the corresponding adjoint Green's functions are constructed. In some situations where the reciprocity relations may not hold, it may still be possible to resort to Eqs. 3.275 and 3.296.

3.3.3 Inhomogeneous Initial/Boundary Conditions

By construction, the Green's function and the adjoint Green's function satisfy homogeneous initial/boundary conditions and homogeneous adjoint initial/boundary conditions, respectively. However, solutions of differential equations coupled

with inhomogeneous initial/boundary conditions can still be represented using those Green's functions obtained from homogeneous initial/boundary conditions. The foundation of this representation is the generalized Green's identity (Eqs. 3.156, 3.158 and 3.160).

3.3.3.1 General 2nd-Order ODE

Consider the general second-order ordinary differential equation \mathbf{L}_x given by 3.164 and the general homogeneous boundary conditions in Eq. 3.227 or 3.229. Suppose that the solution $u(x)$ satisfies the differential equation

$$\mathbf{L}_x [u(x)] = f(x), \quad (3.307)$$

and also the inhomogeneous boundary condition given by

$$\begin{pmatrix} \mathbf{R}_1 [u] \\ \mathbf{R}_2 [u] \end{pmatrix} = (\mathbf{A} \ \mathbf{B}) \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \end{pmatrix} = \mathbf{w}, \text{ and } \mathbf{w} \neq \mathbf{0}. \quad (3.308)$$

In the generalized Green's identity (Eq. 3.158), we can substitute $v(x) = G^\dagger(x, y)$ and obtain

$$\begin{aligned} \int_0^1 w(x) [G^\dagger(x, y)]^* \mathbf{L}_x [u(x)] dx - \int_0^1 w(x) \left\{ \mathbf{L}_x^\dagger [G^\dagger(x, y)] \right\}^* u(x) dx \\ = \mathcal{Q} \left[u, (G^\dagger)^* \right] \Big|_{x=0}^{x=1}, \end{aligned} \quad (3.309)$$

where we have generalized the inner product by including the weighting function $w(x)$. Because $u(x)$ does not satisfy the homogeneous boundary condition, the surface term on the RHS of Eq. 3.309 is therefore nonzero in general. Considering Eqs. 3.307 and 3.269, we can re-write Eq. 3.309 as

$$u(y) = \int_0^1 w(x) [G^\dagger(x, y)]^* f(x) dx - \mathcal{Q} \left[u, (G^\dagger)^* \right] \Big|_{x=0}^{x=1}. \quad (3.310)$$

Considering Eq. 3.275, we have

$$u(y) = \int_0^1 w(x) G(y, x) f(x) dx - \mathcal{Q} \left[u, (G^\dagger)^* \right] \Big|_{x=0}^{x=1}. \quad (3.311)$$

Considering the general form of the conjunct given in Eq. 3.231, the surface term on the RHS of Eq. 3.311 can be expressed as

$$\mathcal{Q} \left[u, (G^\dagger)^* \right] \Big|_{x=0}^{x=1} = \mathbf{u}_1^T \mathbf{Q}_1 \mathbf{g}_1^* - \mathbf{u}_0^T \mathbf{Q}_0 \mathbf{g}_0^*, \quad (3.312)$$

where the vector $\mathbf{g}(x, y)$ is given by

$$\mathbf{g}(x, y) \equiv \begin{bmatrix} G^\dagger(x, y) \\ \frac{dG^\dagger}{dx}(x, y) \end{bmatrix}, \quad (3.313)$$

and $\mathbf{g}_1 = \mathbf{g}(1, y)$, $\mathbf{g}_0 = \mathbf{g}(0, y)$. By solving Eq. 3.308 for \mathbf{u}_1 , we obtain

$$\mathbf{u}_1^T = \mathbf{w}^T (\mathbf{B}^T)^{-1} - \mathbf{u}_0^T \mathbf{A}^T (\mathbf{B}^T)^{-1}. \quad (3.314)$$

Bring Eq. 3.314 into Eq. 3.312, we obtain

$$\begin{aligned} Q \left[u, (G^\dagger)^* \right] \Big|_{x=0}^{x=1} &= \mathbf{w}^T (\mathbf{B}^T)^{-1} \mathbf{Q}_1 \mathbf{g}_1^* \\ &\quad - \mathbf{u}_0^T \left[\mathbf{A}^T (\mathbf{B}^T)^{-1} \mathbf{Q}_1 \mathbf{g}_1^* + \mathbf{Q}_0 \mathbf{g}_0^* \right]. \end{aligned} \quad (3.315)$$

Because G^\dagger satisfies the homogeneous adjoint boundary condition in Eq. 3.236, the term in the square bracket on the RHS of Eq. 3.315 is therefore zero. Bring Eq. 3.315 back into Eq. 3.311, we obtain

$$\begin{aligned} u(y) &= \int_0^1 w(x)G(y, x)f(x)dx - \mathbf{w}^T (\mathbf{B}^T)^{-1} \mathbf{Q}_1 \mathbf{g}_1^* \\ &= \int_0^1 w(x)G(y, x)f(x)dx - \mathbf{w}^T (\mathbf{B}^T)^{-1} \mathbf{Q}_1 \begin{bmatrix} G^\dagger(1, y) \\ \frac{dG^\dagger}{dx}(1, y) \end{bmatrix}^* \\ &= \int_0^1 w(x)G(y, x)f(x)dx - \mathbf{w}^T (\mathbf{B}^T)^{-1} \mathbf{Q}_1 \begin{bmatrix} G(y, 1) \\ \frac{dG}{dx}(y, x)|_{x=1} \end{bmatrix}, \end{aligned} \quad (3.316)$$

where I have used Eq. 3.275 in the last equality. Equation 3.316 shows that the solution of the general second-order ordinary differential equation coupled with the inhomogeneous boundary conditions (Eq. 3.308) can be represented entirely using the Green's function and its derivative with respect to the second argument.

3.3.3.2 Representation Theorem

A similar analysis can be applied to the anisotropic viscoelastic wave-equation operator $\mathbf{L}_{\mathbf{x},t}$ given in Eq. 3.204. Suppose that the space-time wavefield $u_i(\mathbf{x}, t)$ satisfies the equation

$$\mathbf{L}_{\mathbf{x},t} [\mathbf{u}(\mathbf{x}, t)]_i = f_i(\mathbf{x}, t) - \sum_{j=1}^3 \frac{\partial m_{ij}(\mathbf{x}, t)}{\partial x_j}, \quad (3.317)$$

where $f_i(\mathbf{x}, t)$ is the body-force source and $m_{ij}(\mathbf{x}, t)$ is the moment density tensor of any moment sources existing inside the spatial volume V and we assume that it

is zero on the boundary of V , i.e., $m_{ij}(\mathbf{x}, t) = 0, \forall \mathbf{x} \in \partial V$. Note that the minus sign in front of the second term on the RHS of Eq. 3.317 is consistent with Eq. 2.3. On the boundary ∂V , $u_i(\mathbf{x}, t)$ satisfies the inhomogeneous boundary condition

$$\begin{aligned} \sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right] \\ = T_i(\mathbf{x}, t), \forall \mathbf{x} \in \Sigma \subseteq \partial V, \end{aligned} \quad (3.318)$$

where $\hat{n}_j(\mathbf{x})$ is the normal direction on ∂V . The initial conditions for $u_i(\mathbf{x}, t)$ are also inhomogeneous, i.e.,

$$u_i(\mathbf{x}, t) = a_i(\mathbf{x}, t), \quad \frac{\partial u_i}{\partial t}(\mathbf{x}, t) = b_i(\mathbf{x}, t), \quad \forall t \leq 0. \quad (3.319)$$

In the generalized Green's identity (Eq. 3.159), we choose

$$v_j(\mathbf{x}', t') = G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t), \quad (3.320)$$

which satisfies the equation

$$\mathbf{L}_{\mathbf{x}', t'}^{\dagger} \left[\mathbf{G}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]_{ji} = \delta_{ji} \delta(\mathbf{x}' - \mathbf{x}) \delta(t' - t), \quad (3.321)$$

and the homogeneous adjoint initial/boundary conditions (Eqs. 3.254, 3.255). The inner product $\langle \mathbf{L}^{\dagger} v | u \rangle$ in the generalized Green's identity (Eq. 3.159) can then be expressed as

$$\begin{aligned} \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left\{ \mathbf{L}_{\mathbf{x}', t'}^{\dagger} \left[\mathbf{G}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]_{ji} \right\}^* u_j(\mathbf{x}', t') \\ = \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \delta_{ji} \delta(\mathbf{x}' - \mathbf{x}) \delta(t' - t) u_j(\mathbf{x}', t') = u_i(\mathbf{x}, t). \end{aligned} \quad (3.322)$$

Considering Eq. 3.317, the inner product $\langle v | \mathbf{L}u \rangle$ in Eq. 3.159 can be expressed as

$$\begin{aligned} \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]^* \mathbf{L}_{\mathbf{x}', t'} [\mathbf{u}(\mathbf{x}', t')]_j \\ = \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]^* f_j(\mathbf{x}', t') \\ - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]^* \sum_{k=1}^3 \frac{\partial m_{jk}(\mathbf{x}', t')}{\partial x'_k}. \end{aligned} \quad (3.323)$$

The second term on the RHS of Eq. 3.323 can be expressed as

$$\begin{aligned}
 & - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* \sum_{k=1}^3 \frac{\partial m_{jk}(\mathbf{x}', t')}{\partial x'_k} \\
 &= - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{k=1}^3 \frac{\partial}{\partial x'_k} \left\{ \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* m_{jk}(\mathbf{x}', t') \right\} \\
 & \quad + \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^*}{\partial x'_k} m_{jk}(\mathbf{x}', t').
 \end{aligned}$$

Apply the divergence theorem on the first term of the RHS, we obtain

$$\begin{aligned}
 & - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{k=1}^3 \frac{\partial}{\partial x'_k} \left\{ \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* m_{jk}(\mathbf{x}', t') \right\} \\
 &= - \int_0^T dt' \int_{\partial V} da(\mathbf{x}') \sum_{j,k=1}^3 \hat{n}_k(\mathbf{x}') \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* m_{jk}(\mathbf{x}', t') = 0,
 \end{aligned}$$

where I have used the assumption that m_{jk} is zero on the boundary ∂V . Therefore the inner product $\langle v | \mathbf{L}u \rangle$ can be expressed as

$$\begin{aligned}
 & \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* \mathbf{L}_{\mathbf{x}', t'} [\mathbf{u}(\mathbf{x}', t')]_j \\
 &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^* f_j(\mathbf{x}', t') \\
 & \quad + \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial \left[G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) \right]^*}{\partial x'_k} m_{jk}(\mathbf{x}', t'). \tag{3.324}
 \end{aligned}$$

Because $u_i(\mathbf{x}, t)$ satisfies inhomogeneous initial/boundary conditions, the surface term in Eq. 3.216 is not zero. However, because G_{ji}^\dagger satisfies the homogeneous adjoint initial/boundary conditions, the second and the fifth terms in the surface term (Eq. 3.216) are zeros and the first term is also zero at $t = T$. After removing terms that are zeros, the surface term is

$$\begin{aligned}
 & - \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^\dagger(\mathbf{x}', 0; \mathbf{x}, t) \right]^* b_j(\mathbf{x}', 0) \\
 & \quad + \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 \frac{\partial \left(G_{ji}^\dagger \right)^*}{\partial t'} (\mathbf{x}', 0; \mathbf{x}, t) a_j(\mathbf{x}', 0)
 \end{aligned}$$

$$\begin{aligned}
& - \int_0^T dt' \int_{\Sigma} da(\mathbf{x}') \sum_{j=1}^3 \left[G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]^* T_j(\mathbf{x}', t') \\
& + \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k,l,m=1}^3 \left\{ \frac{\partial \left[G_{ji}^{\dagger}(\mathbf{x}', t'; \mathbf{x}, t) \right]^*}{\partial x'_k} \right. \\
& \quad \left. \int_{-\infty}^0 d\tau \frac{\partial a_l(\mathbf{x}', \tau)}{\partial x_m} \Psi_{jklm}(\mathbf{x}', t' - \tau) \right\}, \tag{3.325}
\end{aligned}$$

where I have used the inhomogeneous initial conditions (Eq. 3.319) in the first term, the second term and the last term, used the inhomogeneous boundary condition (Eq. 3.318) in the third term.

Bring the surface term (Eq. 3.325) and expressions for the two inner products (Eqs. 3.322 and 3.324) into the generalized Green's identity (Eq. 3.159), considering Eq. 3.296, we obtain an expression for $u_i(\mathbf{x}, t)$ in terms of the Green's function $G_{ij}(\mathbf{x}, t; \mathbf{x}', t')$,

$$\begin{aligned}
u_i(\mathbf{x}, t) &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 G_{ij}(\mathbf{x}, t; \mathbf{x}', t') f_j(\mathbf{x}', t') \\
&+ \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ij}(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} m_{jk}(\mathbf{x}', t') \\
&+ \int_0^T dt' \int_{\Sigma} da(\mathbf{x}') \sum_{j=1}^3 G_{ij}(\mathbf{x}, t; \mathbf{x}', t') T_j(\mathbf{x}', t') \\
&+ \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 \left[G_{ij}(\mathbf{x}, t; \mathbf{x}', t') b_j(\mathbf{x}', 0) - \frac{\partial G_{ij}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial t'} a_j(\mathbf{x}', 0) \right] \\
&- \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k,l,m=1}^3 \left\{ \frac{\partial G_{ij}(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} \right. \\
& \quad \left. \int_{-\infty}^0 d\tau \frac{\partial a_l(\mathbf{x}', \tau)}{\partial x_m} \Psi_{jklm}(\mathbf{x}', t' - \tau) \right\}. \tag{3.326}
\end{aligned}$$

Equation 3.326 is also known as the *representation theorem* (e.g., Tarantola 1988; Dahlen and Tromp 1998; Aki and Richards 2002). As noted in (Tarantola 1988), the last term on the RHS of Eq. 3.326 represents the effect of the stress caused by the deformation before $t = 0$. If $u_i(\mathbf{x}, t)$ satisfies the quiescent past homogeneous initial conditions, the last two terms in Eq. 3.326 disappear. If $u_i(\mathbf{x}, t)$ satisfies the free-surface boundary condition (i.e., $T_j(\mathbf{x}', t') = 0, \forall \mathbf{x}' \in \partial V$), the third term in Eq. 3.326 disappears. The first term in Eq. 3.326 gives the displacement generated by body-force sources $f_j(\mathbf{x}', t')$. The second term in Eq. 3.326 gives the displacement generated by moment (stress) sources $m_{jk}(\mathbf{x}', t')$ inside the volume and is

often used for representing displacement caused by earthquake sources e.g., slip on a buried fault, stress glut (Backus and Mulcahy 1976a, b).

3.3.3.3 Adjoint Representation Theorem

The same analysis can also be applied to the adjoint system. Suppose the space-time adjoint wavefield $u_i^\dagger(\mathbf{x}, t)$ satisfies the adjoint anisotropic viscoelastic wave equation

$$\mathbf{L}_{\mathbf{x},t}^\dagger [\mathbf{u}^\dagger(\mathbf{x}, t)]_i = f_i^\dagger(\mathbf{x}, t) - \sum_{j=1}^3 \frac{\partial m_{ij}^\dagger(\mathbf{x}, t)}{\partial x_j}, \quad (3.327)$$

where the adjoint operator $\mathbf{L}_{\mathbf{x},t}^\dagger$ is given by Eq. 3.215, $f_i^\dagger(\mathbf{x}, t)$ is the adjoint body-force source field and $m_{ij}^\dagger(\mathbf{x}, t)$ is the adjoint moment density source field. And we assume that $u_i^\dagger(\mathbf{x}, t)$ satisfies the inhomogeneous adjoint initial/boundary conditions (Eqs. 3.254 and 3.255),

$$\begin{aligned} \sum_{j,k,l=1}^3 \hat{n}_j(\mathbf{x}) \int_{-\infty}^{+\infty} d\tau \Psi'_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k^\dagger(\mathbf{x}, \tau)}{\partial x_l} \\ = T_i^\dagger(\mathbf{x}, t), \forall \mathbf{x} \in \Sigma \subseteq \partial V, \end{aligned} \quad (3.328)$$

$$u_i^\dagger(\mathbf{x}, t) = a_i^\dagger(\mathbf{x}, t), \quad \frac{\partial u_i^\dagger(\mathbf{x}, t)}{\partial t} = b_i^\dagger(\mathbf{x}, t), \quad \forall t \geq T. \quad (3.329)$$

Here we have assumed that all functions are real and dropped the complex conjugation $*$ on the adjoint wavefield. We only consider the inhomogeneous adjoint free-surface boundary condition in Eq. 3.255 here. In the generalized Green's identity, Eq. 3.159, we choose

$$u_j(\mathbf{x}', t') = G_{ji}(\mathbf{x}', t'; \mathbf{x}, t), \quad (3.330)$$

which satisfies the equation

$$\mathbf{L}_{\mathbf{x}',t'} [\mathbf{G}(\mathbf{x}', t'; \mathbf{x}, t)]_{ji} = \delta_{ji} \delta(\mathbf{x}' - \mathbf{x}) \delta(t' - t), \quad (3.331)$$

and the homogeneous initial/boundary conditions (Eqs. 3.252, 3.253). The inner product $\langle v | \mathbf{L}u \rangle$ in the generalized Green's identity (Eq. 3.159) can be expressed as

$$\begin{aligned} \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 u_j^\dagger(\mathbf{x}', t') \left\{ \mathbf{L}_{\mathbf{x}',t'} [\mathbf{G}(\mathbf{x}', t'; \mathbf{x}, t)]_{ji} \right\} \\ = \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 u_j^\dagger(\mathbf{x}', t') \delta_{ji} \delta(\mathbf{x}' - \mathbf{x}) \delta(t' - t) = u_i^\dagger(\mathbf{x}, t). \end{aligned} \quad (3.332)$$

Considering Eq. 3.327, the inner product $\langle \mathbf{L}^\dagger v | u \rangle$ in the generalized Green's identity (Eq. 3.159) can be expressed as

$$\begin{aligned} & \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \mathbf{L}_{\mathbf{x}',t'}^\dagger [\mathbf{u}^\dagger(\mathbf{x}', t')]_j G_{ji}(\mathbf{x}', t'; \mathbf{x}, t) \\ &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 f_j^\dagger(\mathbf{x}', t') G_{ji}(\mathbf{x}', t'; \mathbf{x}, t) \\ & - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \left[\sum_{k=1}^3 \frac{\partial m_{jk}^\dagger(\mathbf{x}', t')}{\partial x'_k} \right] G_{ji}(\mathbf{x}', t'; \mathbf{x}, t), \end{aligned}$$

which, after applying the divergence theorem, can be re-written as

$$\begin{aligned} \langle \mathbf{L}^\dagger v | u \rangle &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 f_j^\dagger(\mathbf{x}', t') G_{ji}(\mathbf{x}', t'; \mathbf{x}, t) \\ & + \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 m_{jk}^\dagger(\mathbf{x}', t') \frac{\partial G_{ji}(\mathbf{x}', t'; \mathbf{x}, t)}{\partial x'_k}. \end{aligned} \quad (3.333)$$

Because the adjoint wavefield $u_i^\dagger(\mathbf{x}, t)$ satisfies the inhomogeneous adjoint initial/boundary conditions, the surface term in the generalized Green's identity (Eq. 3.216) is not zero. Because the Green's function $G_{ji}(\mathbf{x}', t'; \mathbf{x}, t)$ satisfies the homogeneous initial/boundary conditions, the first term is zero at $t = 0$ and the third term and the fourth term are zero. After removing the zero terms, the surface term in Eq. 3.216 becomes

$$\begin{aligned} & \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 a_j^\dagger(\mathbf{x}', T) \frac{\partial G_{ji}(\mathbf{x}', T; \mathbf{x}, t)}{\partial t} \\ & - \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 b_j^\dagger(\mathbf{x}', T) G_{ji}(\mathbf{x}', T; \mathbf{x}, t) \\ & + \int_0^T dt' \int_{\partial V} da(\mathbf{x}') \sum_{j=1}^3 G_{ji}(\mathbf{x}', t'; \mathbf{x}, t) T_j^\dagger(\mathbf{x}', t') \\ & - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k,l,m=1}^3 \left\{ \frac{\partial G_{ji}(\mathbf{x}', t'; \mathbf{x}, t)}{\partial x'_k} \right. \\ & \quad \left. \int_T^{+\infty} d\tau \frac{\partial a_l^\dagger(\mathbf{x}', \tau)}{\partial x'_m} \Psi_{jklm}(\mathbf{x}', t' - \tau) \right\}. \end{aligned} \quad (3.334)$$

Bring Eqs. 3.332, 3.333 and 3.334 into the generalized Green's identity (Eq. 3.159) and apply Eq. 3.297, we obtain the representation of $u_i^\dagger(\mathbf{x}, t)$ in terms of the adjoint Green's tensor,

$$\begin{aligned}
u_i^\dagger(\mathbf{x}, t) = & \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 f_j^\dagger(\mathbf{x}', t') G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', t') \\
& + \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 m_{jk}^\dagger(\mathbf{x}', t') \frac{\partial G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} \\
& + \int_0^T dt' \int_{\partial V} da(\mathbf{x}') \sum_{j=1}^3 G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', t') T_j^\dagger(\mathbf{x}', t') \\
& + \int_V dV(\mathbf{x}') \rho(\mathbf{x}') \sum_{j=1}^3 \left[a_j^\dagger(\mathbf{x}', T) \frac{\partial G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', T)}{\partial t} - b_j^\dagger(\mathbf{x}', T) G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', T) \right] \\
& - \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k,l,m=1}^3 \left\{ \frac{\partial G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} \right. \\
& \quad \left. \int_T^{+\infty} d\tau \frac{\partial a_l^\dagger(\mathbf{x}', \tau)}{\partial x'_m} \Psi'_{jklm}(\mathbf{x}', t' - \tau) \right\}. \tag{3.335}
\end{aligned}$$

Equation 3.335 is analogous to the representation theorem in Eq. 3.326, except that it is a representation of the adjoint wavefield $u_i^\dagger(\mathbf{x}, t)$ in terms of the adjoint Green's tensor $G_{ij}^\dagger(\mathbf{x}, t; \mathbf{x}', t')$. We may call Eq. 3.335 the *adjoint representation theorem* and its utility in solving the inverse problem will be discussed in Chap. 5.

The derivations above have shown that the representation theorem and the adjoint representation theorem are consequences of the generalized Green's identity applied to the case of inhomogeneous initial/boundary conditions and inhomogeneous adjoint initial/boundary conditions. Solutions of differential equations coupled with inhomogeneous initial/boundary conditions can be represented using Green's functions obtained from the differential equations accompanied by homogeneous initial/boundary conditions. Comparing Eqs. 3.316 and 3.326 with Eq. 3.262 and comparing Eq. 3.335 with Eq. 3.276, we can see that the effect of inhomogeneous initial/boundary conditions is to introduce a few extra terms into the solution of the corresponding homogeneous problem.

3.4 Receiver Green's Tensor (RGT)

In Sect. 2.2.7 and Fig. 2.12, I have shown a comparison between synthetic seismograms computed by solving the forward isotropic viscoelastic wave equation in a half-space structural model and those computed from the receiver Green's tensors (RGTs) by applying the reciprocity relation and Fig. 2.12 shows that the synthetics computed using both methods are identical. The foundation of computing synthetic seismograms using RGTs and reciprocity is the representation theorem in Eq. 3.326. If $u_i(\mathbf{x}, t)$ satisfies the quiescent past homogeneous initial conditions,

the free-surface boundary condition and the body-force sources are zero, Eq. 3.326 can be simplified to

$$u_i(\mathbf{x}, t) = \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ij}(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} m_{jk}(\mathbf{x}', t'). \quad (3.336)$$

Since $G_{ij}(\mathbf{x}, t; \mathbf{x}', t')$ is translation-invariant in time, we can re-write Eq. 3.336 as

$$\begin{aligned} u_i(\mathbf{x}, t) &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ij}(\mathbf{x}, t - t'; \mathbf{x}', 0)}{\partial x'_k} m_{jk}(\mathbf{x}', t') \\ &= \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ij}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x'_k} * m_{jk}(\mathbf{x}', t), \end{aligned} \quad (3.337)$$

where “*” denotes the temporal convolution operation. If we apply the reciprocity relation in Eqs. 3.306 to 3.337, we obtain

$$u_i(\mathbf{x}, t) = \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}', t; \mathbf{x}, 0)}{\partial x'_k} * m_{jk}(\mathbf{x}', t). \quad (3.338)$$

For a given receiver location \mathbf{x}_r , the synthetic displacement seismogram $u_i(\mathbf{x}_r, t)$ due to the moment source $m_{jk}(\mathbf{x}', t)$ can be represented as

$$\begin{aligned} u_i(\mathbf{x}_r, t) &= \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}', t; \mathbf{x}_r, 0)}{\partial x'_k} * m_{jk}(\mathbf{x}', t) \\ &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * m_{jk}(\mathbf{x}, t), \end{aligned} \quad (3.339)$$

where I have replaced the source coordinate \mathbf{x}' with \mathbf{x} in the second equality. For a spatial point source located at \mathbf{x}_s , i.e.,

$$m_{jk}(\mathbf{x}, t) = M_{jk}(t) \delta(\mathbf{x} - \mathbf{x}_s), \quad (3.340)$$

Eq. 3.339 can be written as

$$u_i(\mathbf{x}_r, t) = \sum_{j,k=1}^3 \frac{\partial G_{ji}}{\partial x_k}(\mathbf{x}_s, t; \mathbf{x}_r, 0) * M_{jk}(t). \quad (3.341)$$

The Fréchet derivative of the waveform $u_i(\mathbf{x}_r, t)$ with respect to the moment source $m_{jk}(\mathbf{x}, t)$ can be obtained by applying Eq. 3.145. We consider the action of the Fréchet derivative $\mathbf{D}u_i(\mathbf{x}_r, t)$ on $\delta \mathbf{m}(\mathbf{x}, t)$, or $\delta m_{jk}(\mathbf{x}, t)$ in index notation, whose

norm is sufficiently small so that it is in the neighborhood of the zero moment source. Using Eq. 3.339, we obtain

$$\begin{aligned}
[\mathbf{D}u_i(\mathbf{x}_r, t)] \delta \mathbf{m} &= \lim_{\alpha \rightarrow 0} \\
&\left\{ \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * [m_{jk}(\mathbf{x}, t) + \alpha \delta m_{jk}(\mathbf{x}, t)] \right. \\
&\quad \left. - \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * m_{jk}(\mathbf{x}, t) \right\} / \alpha \\
&= \lim_{\alpha \rightarrow 0} \left[\int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * \alpha \delta m_{jk}(\mathbf{x}, t) \right] / \alpha \\
&= \int d\tau \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r, 0)}{\partial x_k} \delta m_{jk}(\mathbf{x}, \tau). \tag{3.342}
\end{aligned}$$

Equation 3.342 shows that the Fréchet derivative $\mathbf{D}u_i(\mathbf{x}_r, t)$ is an integral operator and its integration kernel, which is called the *Fréchet kernel*, is given by $\partial G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r, 0) / \partial x_k$.

The Green's function $G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$, as a spatial-temporal displacement field of \mathbf{x} and t , can be computed by placing a unit body-force pointing in the direction $\hat{\mathbf{e}}_i$ at the receiver location \mathbf{x}_r and is named “receiver Green tensor (RGT)” in (Lee et al. 2011).

3.4.1 Time-Reversed Adjoint Green's Tensor

An alternative to using the reciprocity relation in Eq. 3.306 is to use the relation between the Green's function and the adjoint Green's function in Eq. 3.297. Bring Eq. 3.297 into Eq. 3.336, we obtain

$$u_i(\mathbf{x}, t) = \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t)}{\partial x'_k} m_{jk}(\mathbf{x}', t'). \tag{3.343}$$

It can be verified that the adjoint Green's tensor is also translation-invariant in time and we have

$$G_{ji}^\dagger(\mathbf{x}', t'; \mathbf{x}, t) = G_{ji}^\dagger(\mathbf{x}', t' - t; \mathbf{x}, 0). \tag{3.344}$$

If we define the time-reversed adjoint Green's function

$$G_{ji}^\ddagger(\mathbf{x}', t - t'; \mathbf{x}, 0) \equiv G_{ji}^\dagger(\mathbf{x}', t' - t; \mathbf{x}, 0), \tag{3.345}$$

for a given receiver location $\mathbf{x} = \mathbf{x}_r$, Eq. 3.343 can be re-written as

$$\begin{aligned}
 u_i(\mathbf{x}_r, t) &= \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ji}^{\ddagger}(\mathbf{x}', t - t'; \mathbf{x}_r, 0)}{\partial x'_k} m_{jk}(\mathbf{x}', t'). \\
 &= \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{ji}^{\ddagger}(\mathbf{x}', t; \mathbf{x}_r, 0)}{\partial x'_k} * m_{jk}(\mathbf{x}', t) \\
 &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}^{\ddagger}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * m_{jk}(\mathbf{x}, t), \tag{3.346}
 \end{aligned}$$

where I have replaced the source coordinate \mathbf{x}' with \mathbf{x} in the last equality. The time-reversed adjoint Green's tensor $G_{ji}^{\ddagger}(\mathbf{x}, t; \mathbf{x}_r, 0)$ can be computed by placing a unit body-force at the receiver location and solving the adjoint anisotropic viscoelastic wave equation (Eq. 3.215) coupled with the adjoint initial/boundary conditions (e.g., Eqs. 3.254, 3.255) backward in time. This calculation is stable because the anticausal rate-of-relaxation function in the adjoint operator and the adjoint free-surface boundary condition becomes causal after time-reversal.

As pointed out in the end of Sect. 3.3.2, the requirement for obtaining the reciprocity in Eq. 3.306 is more stringent than the requirement for obtaining Eq. 3.297. In obtaining the representation theorem in Eq. 3.326 and also in the alternative derivation given above I only used Eqs. 3.296 and 3.297 and did not use the reciprocity relation anywhere through the derivation. The time-reversed adjoint Green's tensor (Eq. 3.345) has similar benefits as the RGT (Sect. 3.4.2) and this alternative approach does not require the self-adjointness of the wave-equation operator, either before or after time-reversal, and may still be applicable when the reciprocity relation does not hold.

3.4.2 Receiver-Side Strain Green's Tensor (RSGT)

In practice, we compute and store the spatial gradient of the displacement $\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$ instead of the displacement $G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$. Since the spatial gradient of the displacement field is the strain field, in (Zhao et al. 2006) we called $\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$ the “receiver-side strain Green's tensor (RSGT)”. In many seismic wave-propagation simulation codes, including AWP-ODC, the strain is computed as an intermediate variable during the calculation of the stress field, or it can be obtained by converting the stress field using constitutive relations. For the isotropic elastic case, we have the constitutive relation

$$\boldsymbol{\epsilon} = \frac{1}{2\mu} \boldsymbol{\sigma} - \frac{\lambda}{2\mu(3\lambda + 2\mu)} \text{tr} \boldsymbol{\sigma} \mathbf{I}, \tag{3.347}$$

where $\boldsymbol{\epsilon}$ and $\boldsymbol{\sigma}$ are strain and stress tensors, $\text{tr} \boldsymbol{\sigma} = \sigma_{11} + \sigma_{22} + \sigma_{33}$ denotes the trace of the stress tensor, λ and μ are Lamé parameters and \mathbf{I} is the identity matrix.

The amount of addition CPU time needed for obtaining the RSGT during the wave-propagation simulation is negligible. If we are interested in the vertical-component synthetic seismograms at a receiver $u_3(\mathbf{x}_r, t)$, we only need one finite-difference simulation for a unit body-force source pointing in the vertical direction at the receiver location to obtain the strain field $\partial G_{j3}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$. If we are interested in horizontal-component synthetic seismograms also, we need two more simulations with two unit body-force sources at the receiver location pointing in the two orthogonal horizontal directions of the spatial Cartesian coordinate to obtain $\partial G_{j1}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$ and $\partial G_{j2}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$.

In most applications, we are interested in the synthetic seismograms within a certain frequency band. We can convolve a filter, whose time-domain expression is $h(t)$, on both sides of Eq. 3.339, the filtered synthetic seismogram, denoted as $\check{u}_i(\mathbf{x}_r, t)$, can be expressed as

$$\begin{aligned} \check{u}_i(\mathbf{x}_r, t) &= h(t) * u_i(\mathbf{x}_r, t) \\ &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial [h * G_{ji}]}{\partial x_k}(\mathbf{x}, t; \mathbf{x}_r, 0) * m_{jk}(\mathbf{x}, t) \\ &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial \check{G}_{ji}}{\partial x_k}(\mathbf{x}, t; \mathbf{x}_r, 0) * m_{jk}(\mathbf{x}, t), \end{aligned} \quad (3.348)$$

where the filtered RSGT

$$\frac{\partial \check{G}_{ji}}{\partial x_k}(\mathbf{x}, t; \mathbf{x}_r, 0) = h(t) * \frac{\partial G_{ji}}{\partial x_k}(\mathbf{x}, t; \mathbf{x}_r, 0) = \frac{\partial [h * G_{ji}]}{\partial x_k}(\mathbf{x}, t; \mathbf{x}_r, 0) \quad (3.349)$$

can be computed by using $h(t)$ as the source-time function of the body-forces acting at the receiver location. This operation is actually equivalent to using the first term on the RHS of the representation theorem (Eq. 3.326). For a point source located at \mathbf{x}_s (Eq. 3.340), we have

$$\check{u}_i(\mathbf{x}_r, t) = \sum_{j,k=1}^3 \frac{\partial \check{G}_{ji}}{\partial x_k}(\mathbf{x}_s, t; \mathbf{x}_r, 0) * M_{jk}(t). \quad (3.350)$$

The synthetic seismograms $\check{u}_i(\mathbf{x}_r, t)$ computed using Eqs. 3.348 or 3.350 are already filtered using $h(t)$, because $h(t)$ has already been convolved into the RSGTs (Eq. 3.349).

The advantage of computing synthetics using RSGTs by applying the reciprocity principle is that the number of forward wave-propagation simulations needed is independent of the number of seismic sources. This technique is particularly suitable when the number of seismic sources is significantly larger than the number of receivers. The RSGTs completely capture wave-propagation effects in complex three-dimensional earth structural models. And once the RSGTs for all receivers are computed and stored, no more expensive wave-propagation simulations are needed

in order to generate synthetics at those receivers for any sources lying inside the modeling volume.

In practice, it is usually unnecessary to store the RSGTs at all grid points used in the finite-difference simulations. It is often adequate to store the RSGTs at grid points on or close to pre-existing faults, since the majority of the earthquakes occur on and around pre-existing faults. This RSGT-based seismogram synthesis technique has been adopted by the Southern California Earthquake Center (SCEC) in its CyberShake platform (Graves et al. 2011), a physics-based seismic hazard analysis platform in which synthetic seismograms from hundreds and thousands of potential earthquake ruptures within a certain distance from a given site are computed and analyzed to give a ground-motion estimate at the site. Using the RSGTs and the reciprocity principle, only three wave-propagation simulations are needed for each site.

The RSGT-based technique also makes (near)-real-time seismic source parameter inversions in three-dimensional earth structural models feasible. Since synthetic seismograms for different source models can be simply calculated by extracting the RSGTs from a pre-computed and archived RSGT database for all receivers and applying Eq. 3.339, we can easily compare synthetic waveforms for different source models with observed waveforms and search for the optimal source model. Such a searching algorithm can be easily parallelized on distributed-memory computer clusters with minimal inter-processor communication overhead. In inversions for centroid moment tensors (CMT) (e.g., Zhao et al. 2006; Lee et al. 2011) or finite moment tensors (FMT) (Chen et al. 2005), the total number of unknown source parameters ranges from 8 to 20, which allows us to adopt grid-search algorithms or stochastic optimization algorithms such as the genetic algorithm (Goldberg 1989), simulated annealing (Kirkpatrick et al. 1983) and particle swarm algorithms (Kennedy 2010) to search for global optimal solutions. In (Lee et al. 2011), we have developed a hierarchical grid-search algorithm based on Bayesian inference that can provide not only the optimal CMT solution but also estimates of the uncertainties in the solution.

In inversions for fault-slip distributions (FSD), the total number of unknown source parameters can easily exceed 100 depending upon the parameterization and kinematic assumptions. In such situations, grid-search or stochastic optimization algorithms may not be suitable and we may resort to gradient-based iterative optimization algorithms. If we represent a finite source using the moment density function $m_{jk}(\mathbf{x}, t)$, Eq. 3.342 shows that the RSGT is actually the Fréchet kernel (Chap. 4) of the waveform with respect to the moment density function, which allows us to develop highly efficient gradient- or Hessian-based optimization algorithms for FSD inversions.

A potential disadvantage of RSGT-based source parameter inversions is that the disk storage cost can be high and the storage needs to be constantly on-line for rapid access. Traditional HPC platforms such as distributed-memory computer clusters are usually shared by many users and suffer from long delays in acquiring resources. The emerging Cloud computing platforms provide users with abundant storage capacity and allow users to quickly acquire and release large amounts of

computational resources on-demand with minimal scheduling overheads. In (Chen et al. 2013) we have successfully implemented a numerically efficient and also cost-effective synthetic seismogram generator based on Eq. 3.339 using the Windows Azure Cloud Application Framework and obtained highly encouraging overall performance.

3.4.3 RSGTs in the Half-Space Example

In Sect. 2.2 I gave examples on how to compute RSGTs and how to obtain synthetic seismograms from RSGTs using the software F3DWT. In this section, I give a brief summary on the calculations that are directly linked to RSGTs using the half-space example.

To compute the RSGTs, we need to specify the location \mathbf{x}_r (e.g., Eq. 3.339) at which the three orthogonal unit body-force sources are applied. This location does not have to be a receiver location, but in source-parameter inversions and in tomography we usually use RSGTs computed for specific receivers. In the CyberShake calculations (Graves et al. 2010), the body-force sources are applied at sites at which we would like to obtain seismic hazard estimates and may not correspond to receiver locations. In the source-receiver file (Listing 2.6), the first column should be set to 2 if we are applying the body-forces at a receiver location, and 1 if the location is not a receiver location. If we are interested in the RSGT of only one direction, say $\partial G_{j3}(\mathbf{x}, t; \mathbf{x}_r, 0) / \partial x_k$, which is needed if we are only interested in using vertical-component synthetic seismograms, we only need to specify one body-force source line in the source-receiver file. If we are interested in using horizontal-component synthetics, we must compute the full RSGT, which requires three orthogonal body-forces and therefore three body-force source lines in the source-receiver file. In Listing 2.6, line 3–5, I specified three body-force sources at the location of the receiver IN.RC01. The source IDs for these three body-force sources on the second column of Listing 2.6 are constructed by appending the receiver ID with “.” followed by a number indicating the direction of the body force. Other naming conventions can also be adopted here. Columns 3–6 on Line 3–5 of Listing 2.6 specify the location, either in terms of the global grid indices on column 4–6 if column 3 is 1, or in terms of geographic coordinates (i.e., longitude, latitude in degrees and depth in km) on column 4–6 if column 3 is -1. Since we are specifying body-force sources in the box coordinate, column 7 must be set to -2. The direction of the body force is specified on column 8–10. If the body force points to the x_1 direction of the box coordinate, column 8–10 must be set to 1 0 0; if the body force points to the x_2 direction of the box coordinate, column 8–10 must be set to 0 1 0; if the body force points to the x_3 direction, column 8–10 must be set to 0 0 1. Column 11 specifies the full-path file name for the source-time function of the body-force source.

The next step is to specify the locations \mathbf{x} (e.g., Eq. 3.339) at which we would like to record the strain fields. The script `set_ker_grd.py` can produce a uniform mesh decimated according to the numbers specified by the keywords NXSKP,

NYSKP and NZSKP in the parameter file (Listing 2.5, Table 2.2). This script generates the local grid indices of the recording grid points in each sub-mesh and store them in the files named as HSGRDRECprocID inside the directory specified by the MEDIA keyword (Listing 2.5, Table 2.2). Here procID is a 9-digit number used for identifying the sub-meshes (Sect. 2.2.2). The script set_sr_grd.py can generate nonuniformly-distributed recording grid points based on the source-receiver list (Listing 2.6) and also the numbers specified by the three SOURCE STENCIL SIZE keywords in the parameter file (Listing 2.5, Table 2.2). One way of adding additional recording grid points is to modify the source-receiver list by adding receivers at any locations of your choice. The format of a receiver line is specified in Sect. 2.2.4.1. A second way of adding nonuniformly-distributed recording points is to add source lines into the source-receiver list. If we would like to perform CMT inversions, we can add a source line (with the first column set to 1) specified using the reference CMT solution and set the source-stencil sizes in the parameter file to values larger than the estimated location errors in the reference CMT solution. The output files of this script are named as HSGRDRECRDMprocID and are also stored under the directory specified by the MEDIA keyword in the parameter file.

After setting up the correct directory structure for each source in the source-receiver list using the script setup_awp.py and executing the parallel wave-propagation simulation code awp we obtain two groups of binary output files inside the output_vlm sub-directory under each source directory. One group of the binary output files are named as SX96PVprocID and store the 6-component strain and the 3-component particle velocity at recording grid points specified by set_ker_grd.py. The other group of the binary output files are named as SX96PVRDMprocID and store the strain and particle velocity at recording grid points specified by the script set_sr_grd.py. The ordering of the binary data at each recording grid point and each recording time step is given in Eq. 2.125. These two sets of binary output files contain the RSGTs that can be used for computing synthetic seismograms, for source-parameter inversion and for tomography.

The script rgt2syn.py allows us to compute synthetic seismograms from the RSGTs for point sources using Eq. 3.350. Note that the RSGTs generated by awp have been filtered by the source-time function of the body-force sources applied at the receiver, i.e., the imp_hs.dat file (Listing 2.6, Fig. 2.11), and this source-time function was denoted as $h(t)$ in Eq. 3.349. Suppose the point moment source also has a source-time function, denoted as $s(t)$, i.e.,

$$M_{jk}(t) = \hat{M}_{jk}s(t), \quad (3.351)$$

then Eq. 3.350 can be re-written as

$$\check{u}_i(\mathbf{x}_r, t) = s(t) * \left[\sum_{j,k=1}^3 \frac{\partial \check{G}_{ji}}{\partial x_k}(\mathbf{x}_s, t; \mathbf{x}_r, 0) \hat{M}_{jk} \right]. \quad (3.352)$$

The script rgt2syn.py does not perform the convolution with the source-time function of the point moment source $s(t)$. It only produces the synthetics corresponding to the part inside the square bracket on the RHS of Eq. 3.352. When

generating Fig. 2.12, inside the MATLAB script `reciprocity_hs.m`, I explicitly convolved the synthetics generated by `rgt2syn.py` with the source-time function of the point moment source $s(t)$ stored in the file `src_hs.dat` (Listing 2.6, Fig. 2.7) and then plotted them in Fig. 2.12 as thin black solid lines. The synthetics generated through the forward simulation by running `awp` in the source directory 100001 already have the source-time function of the point moment source $s(t)$ built in, because the file `src_hs.dat` was used by the script `proc_srlst.py` for generating the source file `100001.src`, which was read by `awp`. However, the forward synthetics were not filtered by $h(t)$, which was built into the RSGT synthetics. Therefore inside the MATLAB script `reciprocity_hs.m`, I explicitly convolved the forward synthetics with $h(t)$ stored in `imp_hs.dat` and then plotted them in Fig. 2.12 as thick gray dash lines.

In Figure 3.2, I repeat the same experiment using the 29 Mar 2014 La Habra (M_w 5.1) earthquake and the three-dimensional structural model CVM-S4.26 (Sect. 2.3.3). The source-time function $h(t)$ used for computing the RSGTs is a band-pass Butterworth filter with corners at 0.02 Hz and 0.2 Hz and is identical to the source-time function used in the forward simulation $s(t)$. In this case the two sets of synthetics are identical to each other without additional convolution operations. More examples of numerical validations of the reciprocity principle using three-dimensional earth structure models can be found in e.g., (Zhao et al. 2006; Chen et al. 2007).

Back to the half-space example, suppose we would like to compute the RSGTs for two more receivers, both of which are located on the free surface and are named `SF.RC02` and `SF.RC03`, and we would like to extend the source stencil sizes to 5 in the vertical direction (x_3) and to 3 in the two horizontal directions (x_1 and x_2), and we would also like to change the decimation rate of the uniform recording grid in the x_1 and x_3 directions to 2. The locations of the two added receivers, as well as locations of the receiver `IN.RC01` and the source 100001, are shown in Figure 3.3.

The first step is to add the three orthogonal body-force sources at `SF.RC02` and `SF.RC03` into the source-receiver list (Listing 2.6). These six source lines (Listing 3.1) can be placed anywhere inside the source-receiver list. I placed them at the beginning of the file. We also need to add two more receiver lines for these two receivers. The updated source-receiver list is shown in Listing 3.1. The second step is to modify the parameter file shown in Listing 2.5. We need to change the numbers on line 18-20 from 1 1 2 to 3 3 5 and change the numbers on lines 21 and 23 from 4 to 2. The last 7 lines in the modified parameter file is shown in Listing 3.2. After making these changes, we need to run the following commands in the terminal

```
proc_srlst.py param_HS.dat 1 13
set_sr_grd.py param_HS.dat
set_ker_grd.py param_HS.dat
setup_awp.py param_HS.dat FWF 1 10
```

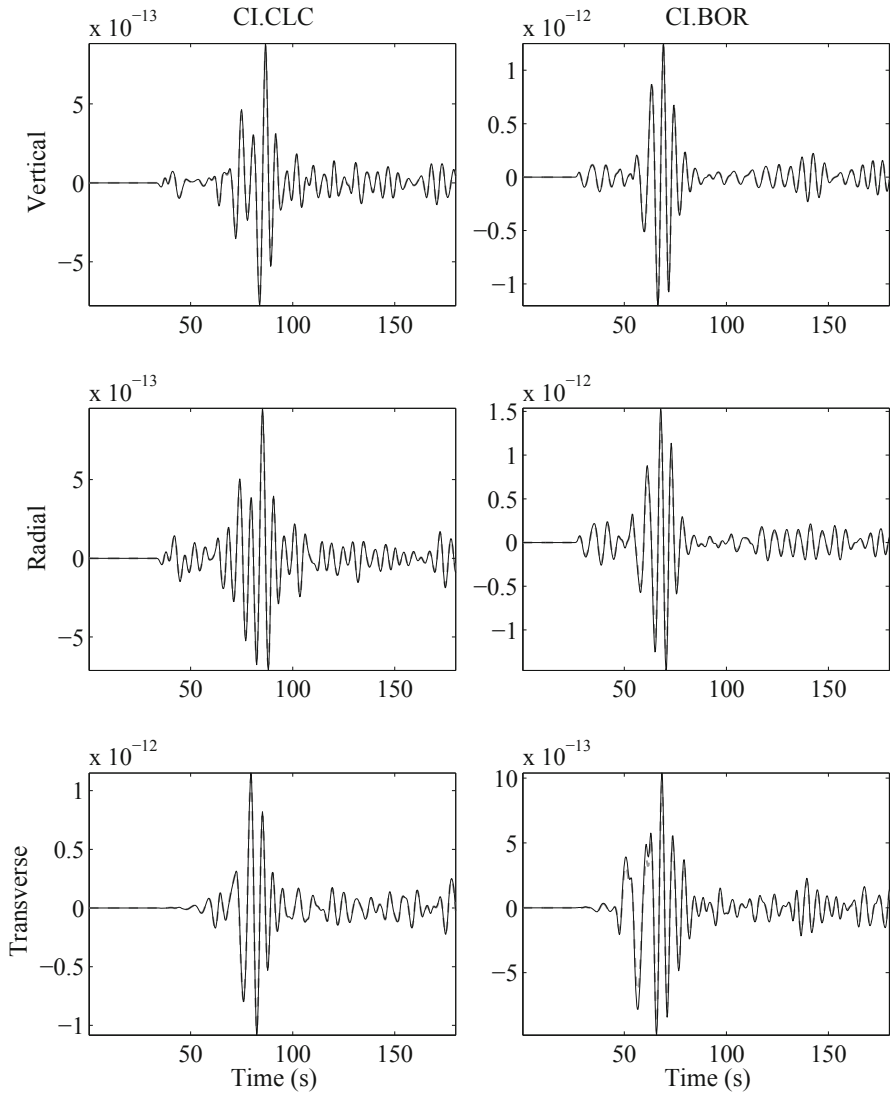


Fig. 3.2 Comparisons of vertical (*top row*), radial (*center row*) and transverse (*bottom row*) synthetic velocity seismograms for the La Habra earthquake (Sect. 2.3.3) to the station CLC (*left column*) and BOR (*right column*) calculated by forward finite-difference simulation (*thick gray dash lines*) and by reciprocity using the RSGTs for CLC and BOR (*thin black solid lines*). The seismic structural model used in the forward simulations and the RSGT calculations is the three-dimensional CVM-S4.26 (Sect. 2.3.3). Figure 2.15 shows the source-receiver paths.

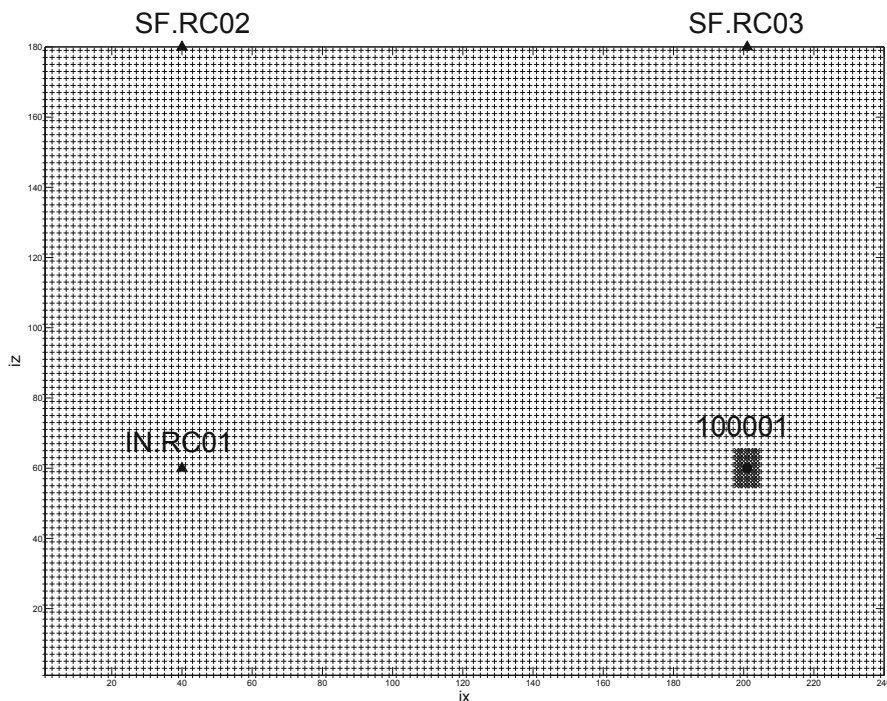


Fig. 3.3 A cross-section view of the recording grid points perpendicular to the source-receiver plane $i_y=40$. The explosive point source 100001 (*black circle*) is located at global grid index $(i_x, i_y, i_z) = (210, 40, 60)$, the receiver IN.RC01 (*black triangle*) is located at global grid index $(40, 40, 60)$, the receiver SF.RC02 (*black triangle*) is at $(40, 40, 180)$ and the receiver SF.RC03 (*black triangle*) is at $(210, 40, 180)$. Both SF.RC02 and SF.RC03 are located at the free surface. Locations marked with “+” are recording grid points generated by `set_ker_grd.py`. Locations marked with “x” are recording grid points generated by `set_sr_grd.py`. The “x” markers at the source and receiver locations are covered by the black circle and the black triangles, therefore not visible. The “x” markers for recording grid points of the source stencil can be seen around the hypocenter of 100001.

to re-build the correct running environment using the modified source-receiver list and the parameter file. The next step is to run `awp` inside each source sub-directory under the `FWF` directory in the same way that I have shown in Sect. 2.2.6 and in Listing 2.17. After the wavefield simulations are completed, we can generate the forward synthetics and the RSGT synthetics using the following commands in the terminal

```
fwf2syn.py param_HS.dat 10 10 11 13
rgt2syn.py param_HS.dat 1 9 10 10
```

```

1 2 SF.RC02.1 1 40 40 180 -2 1 0 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
2 2 SF.RC02.2 1 40 40 180 -2 0 1 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
3 2 SF.RC02.3 1 40 40 180 -2 0 0 1 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
4 2 SF.RC03.1 1 201 40 180 -2 1 0 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
5 2 SF.RC03.2 1 201 40 180 -2 0 1 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
6 2 SF.RC03.3 1 201 40 180 -2 0 0 1 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
7 2 IN.RC01.1 1 40 40 60 -2 1 0 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
8 2 IN.RC01.2 1 40 40 60 -2 0 1 0 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
9 2 IN.RC01.3 1 40 40 60 -2 0 0 1 /home/pochen/wrk/F3DWI/HS/JOBS
  /srinfo/imp_hs.dat
10 1 100001 -1 -117.59233339866256 33.815181833226426 24.0 1 1 1
    1 0 0 0 /home/pochen/wrk/F3DWI/HS/JOBS/srinfo/src_hs.dat
11 0 SF.RC02 1 40 40 180
12 0 SF.RC03 1 201 40 180
13 0 IN.RC01 1 40 40 60

```

Listing 3.1 The modified source-receiver list for the half-space example. The source lines for the 6 body-force sources at the 2 surface receivers are on line 1–6. The 2 receiver lines for the 2 surface receivers are on line 11–12.

```

1 SOURCE STENCIL SIZE IN X = 3
2 SOURCE STENCIL SIZE IN Y = 3
3 SOURCE STENCIL SIZE IN Z = 5
4 NXSKP = 2
5 NYSKP = 4
6 NZSKP = 2
7 NTSKP = 2

```

Listing 3.2 The last 7 lines of the modified parameter file shown in Listing 2.5. Other lines are identical to those shown in Listing 2.5.

and the synthetics generated by both scripts are stored in the 100001 sub-directory under the SYN directory. Synthetics generated by the forward simulations are named as `rID.rot`, where `rID` is the receiver ID, and the synthetics generated using RSGTs by applying the reciprocity relation is named as `rID.rgt.rot`. The synthetic seismograms for the receiver `SF.RC02` are shown in Fig. 3.4 and the synthetics for the receiver `SF.RC03` are shown in Fig. 3.5. Note that the receiver `SF.RC03` is located right on top of the source 100001, therefore we cannot rotate the two horizontal components into the radial and the transverse directions and Fig. 3.5 shows only 5 different components.

The RSGTs provide the Fréchet kernel of the waveform with respect to seismic source parameters. The Fréchet kernel of the waveform with respect to structural

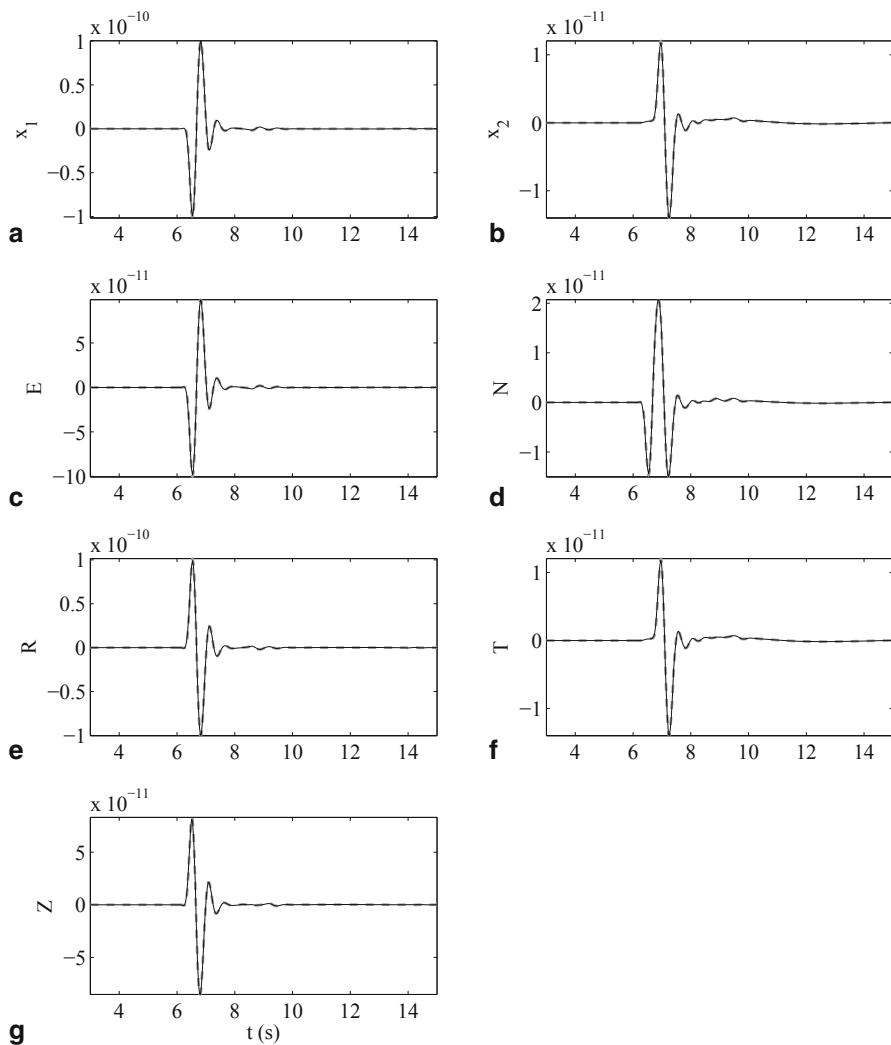


Fig. 3.4 Synthetic seismograms generated through the forward simulation and by using RSGTs. The source is 100001 and the receiver is SF.RC02. The format of this figure is identical to the format of Fig. 2.12.

parameters can also be obtained by using RSGTs, which is the topic of the next section.

3.5 Born Series

Green's functions allow us to convert differential equations into integral equations, which makes the problem amenable to solution techniques developed for integral equations.

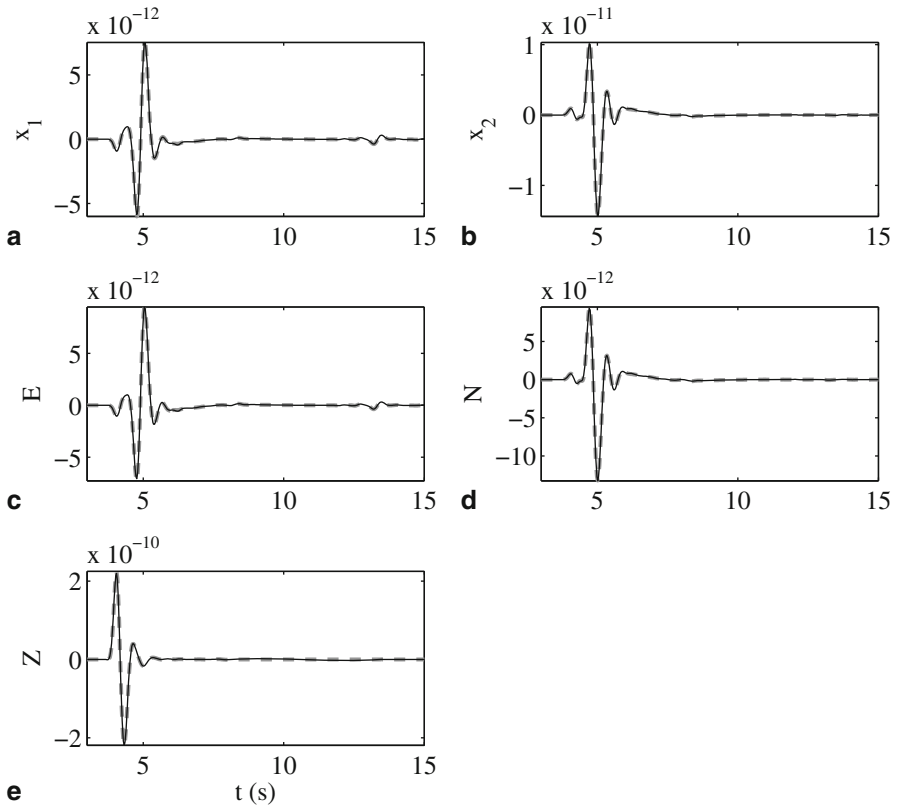


Fig. 3.5 Synthetics for receiver SF.RC03. The format of this figure is identical to the format of Fig. 2.12.

Integral equations are classified into two major categories. If the limits of the integration are constants, the integral equation is called a *Fredholm equation*; otherwise, the integral equation is called a *Volterra equation*. If the unknown function does not appear outside of the integral, the integral equation is of *the first kind*; otherwise, it is of *the second kind*. Suppose $g(x)$ is the unknown function, function $f(x)$ is given, the integration kernel is given by $K(x, y)$ and a and b are two constants, an example of the Fredholm equation of the first kind is

$$\int_a^b K(x, y)g(y)dy = f(x), \tag{3.353}$$

an example of the Fredholm equation of the second kind is

$$g(x) - \int_a^b K(x, y)g(y)dy = f(x), \tag{3.354}$$

an example of the Volterra equation of the first kind is

$$\int_a^x K(x, y)g(y)dy = f(x), \quad (3.355)$$

and an example of the Volterra equation of the second kind is

$$g(x) - \int_a^x K(x, y)g(y)dy = f(x). \quad (3.356)$$

For the purpose of this section, we are particularly interested in the Fredholm equation of the second kind.

In the vector form, Eq. 3.354 can be written as

$$(\mathbf{1} - \mathbf{K}) |g\rangle = |f\rangle, \quad (3.357)$$

and its solution can be expressed as

$$|g\rangle = (\mathbf{1} - \mathbf{K})^{-1} |f\rangle. \quad (3.358)$$

Using the representation of $(\mathbf{1} - \mathbf{K})^{-1}$ given in Eq. 3.124, the solution in Eq. 3.358 can be expressed as

$$|g\rangle = \sum_{n=0}^{\infty} \mathbf{K}^n |f\rangle, \text{ if } \|\mathbf{K}\| < 1, \quad (3.359)$$

which is also known as the Neumann series solution (Arfken and Weber 2005). Written in function form, Eq. 3.359 can be expressed as

$$\begin{aligned} g(x) = & f(x) + \int_a^b K(x, y)f(y)dy + \int_a^b \int_a^b K(x, y_1)K(y_1, y_2)f(y_2)dy_2dy_1 + \\ & \cdots + \int_a^b \cdots \int_a^b K(x, y_1) \cdots K(y_{n-1}, y_n)f(y_n)dy_n dy_{n-1} \cdots dy_1 + \cdots \end{aligned} \quad (3.360)$$

and the series converges if

$$\|\mathbf{K}\|^2 = \int_a^b \int_a^b |K(x, y)|^2 dx dy < 1. \quad (3.361)$$

The same procedure can also be generalized to multi-dimensional functions $f(\mathbf{x})$, $g(\mathbf{x})$ and the integration kernel $K(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, $m \geq 1$.

3.5.1 The Scalar Wave Equation

Consider the scalar wave equation operator given in Eq. 3.186,

$$\mathbf{L}_{\mathbf{x}, t} [u(\mathbf{x}, t)] = \left[\nabla^2 - \frac{1}{\alpha^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \right] u(\mathbf{x}, t) = f(\mathbf{x}, t), \quad (3.362)$$

where $f(\mathbf{x}, t)$ is a source term, and also the scalar wave equation operator defined by a reference velocity $\tilde{\alpha}(\mathbf{x})$,

$$\tilde{\mathbf{L}}_{\mathbf{x},t} [\tilde{u}(\mathbf{x}, t)] = \left[\nabla^2 - \frac{1}{\tilde{\alpha}^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} \right] \tilde{u}(\mathbf{x}, t) = f(\mathbf{x}, t). \quad (3.363)$$

We assume that both $u(\mathbf{x}, t)$ and $\tilde{u}(\mathbf{x}, t)$ satisfy homogeneous initial/boundary conditions. If the initial/boundary conditions are inhomogeneous, we only need to add some terms to the solution as discussed in Sect. 3.3.3. We define the *scattered wavefield* as

$$\Delta u(\mathbf{x}, t) \equiv u(\mathbf{x}, t) - \tilde{u}(\mathbf{x}, t), \quad (3.364)$$

and we have

$$u(\mathbf{x}, t) = \tilde{u}(\mathbf{x}, t) + \Delta u(\mathbf{x}, t). \quad (3.365)$$

We also define the refractive index at location \mathbf{x} as

$$r(\mathbf{x}) \equiv \frac{\tilde{\alpha}(\mathbf{x})}{\alpha(\mathbf{x})}. \quad (3.366)$$

Using Eq. 3.366, we can re-write $\mathbf{L}_{\mathbf{x},t}$ in Eq. 3.362 as

$$\mathbf{L}_{\mathbf{x},t} = \tilde{\mathbf{L}}_{\mathbf{x},t} - \frac{r^2(\mathbf{x}) - 1}{\tilde{\alpha}^2(\mathbf{x})} \frac{\partial^2}{\partial t^2}. \quad (3.367)$$

Bring Eqs. 3.367 and 3.365 into Eq. 3.362, considering Eq. 3.363, we obtain

$$\tilde{\mathbf{L}}_{\mathbf{x},t} [\Delta u(\mathbf{x}, t)] = \frac{r^2(\mathbf{x}) - 1}{\tilde{\alpha}^2(\mathbf{x})} \frac{\partial^2 u}{\partial t^2}(\mathbf{x}, t). \quad (3.368)$$

We therefore can represent the scattered wavefield using the Green's function for $\tilde{\mathbf{L}}_{\mathbf{x},t}$, denoted at $\tilde{G}(\mathbf{x}, t; \mathbf{x}', t')$ in the following,

$$\begin{aligned} \Delta u(\mathbf{x}, t) &= \int_V dV(\mathbf{x}') \int_0^T dt' \tilde{G}(\mathbf{x}, t; \mathbf{x}', t') \frac{r^2(\mathbf{x}') - 1}{\tilde{\alpha}^2(\mathbf{x}')} \frac{\partial^2 u}{\partial t'^2}(\mathbf{x}', t') \\ &= \int_V dV(\mathbf{x}') \int_0^T dt' \tilde{G}(\mathbf{x}, t - t'; \mathbf{x}', 0) \frac{r^2(\mathbf{x}') - 1}{\tilde{\alpha}^2(\mathbf{x}')} \frac{\partial^2 u}{\partial t'^2}(\mathbf{x}', t') \\ &= \int_V dV(\mathbf{x}') \frac{r^2(\mathbf{x}') - 1}{\tilde{\alpha}^2(\mathbf{x}')} \tilde{G}(\mathbf{x}, t; \mathbf{x}', 0) * \frac{\partial^2 u}{\partial t^2}(\mathbf{x}', t), \end{aligned} \quad (3.369)$$

where I have used the translation-invariance in time of the Green's function (because the material property does not vary with time) at the second equality. Using Eq. 3.364, we can re-write Eq. 3.369 as

$$u(\mathbf{x}, t) = \tilde{u}(\mathbf{x}, t) + \int_V dV(\mathbf{x}') \frac{r^2(\mathbf{x}') - 1}{\tilde{\alpha}^2(\mathbf{x}')} \tilde{G}(\mathbf{x}, t; \mathbf{x}', 0) * \frac{\partial^2 u}{\partial t^2}(\mathbf{x}', t), \quad (3.370)$$

which is in the form of the Fredholm integral equation of the second kind (Eq. 3.354) for the unknown function $u(\mathbf{x}, t)$. If we apply Fourier transform on both sides of Eq. 3.370 to convert the time coordinate t to the angular frequency ω , we obtain

$$\hat{u}(\mathbf{x}, \omega) = \hat{\tilde{u}}(\mathbf{x}, \omega) + \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \left[1 - r^2(\mathbf{x}') \right] \tilde{k}^2(\mathbf{x}') \hat{u}(\mathbf{x}', \omega). \quad (3.371)$$

Here we have introduced the wave-number in the reference media

$$\tilde{k}(\mathbf{x}) \equiv \frac{\omega}{\tilde{\alpha}(\mathbf{x})} = \frac{2\pi}{\tilde{\lambda}(\mathbf{x})}, \quad (3.372)$$

and $\tilde{\lambda}(\mathbf{x})$ is the wavelength in the reference media. If we introduce

$$o(\mathbf{x}, \omega) \equiv r^2(\mathbf{x}) \tilde{k}^2(\mathbf{x}), \quad (3.373)$$

$$\tilde{o}(\mathbf{x}, \omega) \equiv \tilde{k}^2(\mathbf{x}), \quad (3.374)$$

$$\Delta o(\mathbf{x}, \omega) \equiv \tilde{o} - o = \tilde{k}^2(\mathbf{x}) - r^2(\mathbf{x}) \tilde{k}^2(\mathbf{x}), \quad (3.375)$$

using Eq. 3.359 with the integration kernel

$$K(\mathbf{x}, \mathbf{x}', \omega) = \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \Delta o(\mathbf{x}', \omega), \quad (3.376)$$

we obtain

$$\begin{aligned} \hat{u}(\mathbf{x}, \omega) &= \hat{\tilde{u}}(\mathbf{x}, \omega) \\ &+ \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \Delta o(\mathbf{x}', \omega) \hat{u}(\mathbf{x}', \omega) \\ &+ \int_V dV(\mathbf{x}'_2) \left\{ \hat{G}(\mathbf{x}, \omega; \mathbf{x}'_2) \Delta o(\mathbf{x}'_2, \omega) \right. \\ &\quad \left. \left[\int_V dV(\mathbf{x}'_1) \hat{G}(\mathbf{x}'_2, \omega; \mathbf{x}'_1) \Delta o(\mathbf{x}'_1, \omega) \hat{u}(\mathbf{x}'_1, \omega) \right] \right\} + \dots \end{aligned} \quad (3.377)$$

Converting to the time domain, Eq. 3.377 can be expressed as

$$\begin{aligned} u(\mathbf{x}, t) &= \tilde{u}(\mathbf{x}, t) \\ &+ \int_V dV(\mathbf{x}') \frac{r^2(\mathbf{x}') - 1}{\tilde{\alpha}^2(\mathbf{x}')} \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}, t; \mathbf{x}', 0) * \tilde{u}(\mathbf{x}', t) \\ &+ \int_V dV(\mathbf{x}'_2) \left\{ \frac{r^2(\mathbf{x}'_2) - 1}{\tilde{\alpha}^2(\mathbf{x}'_2)} \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}, t; \mathbf{x}'_2, 0) * \right. \\ &\quad \left. \left[\int_V dV(\mathbf{x}'_1) \frac{r^2(\mathbf{x}'_1) - 1}{\tilde{\alpha}^2(\mathbf{x}'_1)} \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0) * \tilde{u}(\mathbf{x}'_1, t) \right] \right\} + \dots \end{aligned} \quad (3.378)$$

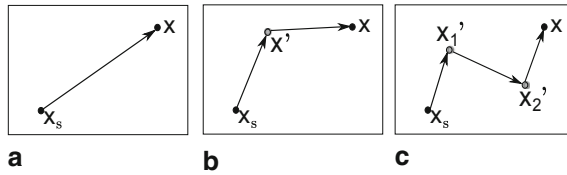


Fig. 3.6 The contributions to the full propagation from the source \mathbf{x}_s to the receiver \mathbf{x} in the media $\alpha(\mathbf{x})$ from the zeroth-order (*left*), first-order (*center*) and second-order (*right*) scattering terms. Left: free propagation from the source to the receiver in the reference media $\tilde{\alpha}$; center: a free propagation from the source to a scatterer located at \mathbf{x}' in the reference media and a free propagation from the scatterer to the receiver in the reference media; right: free propagations in the reference media that involve two scatterers located at \mathbf{x}'_1 and \mathbf{x}'_2 . Locations of the source and the receiver are shown as black dots and locations of the scatterers are shown as gray dots.

Equations 3.377 and 3.378 are the *Born series* in the frequency domain and in the time domain, respectively, for the scalar wave equation given in Eq. 3.362.

Equation 3.377 (or 3.378) has an intuitive interpretation due to Richard Feynman. Suppose the seismogram $\hat{u}(\mathbf{x}, \omega)$ is generated by a point source located at \mathbf{x}_s and recorded at receiver location \mathbf{x} . The Green's function $\hat{G}(\mathbf{x}, \omega; \mathbf{x}')$ can be considered as a *propagator* from \mathbf{x}' to \mathbf{x} in the reference media with velocity $\tilde{\alpha}(\mathbf{x})$. The first term on the RHS of Eq. 3.377 is simply a free propagation from the source \mathbf{x}_s to the receiver \mathbf{x} in the reference media (Fig. 3.6a). The second term on the RHS of Eq. 3.377 can be understood as a summation of all waves that have been scattered once. The diagram for one of such single-scattered waves is shown in Fig. 3.6b. It involves a free propagation from the source \mathbf{x}_s to the scatterer at \mathbf{x}' in the reference media, which is given by $\hat{u}(\mathbf{x}', \omega)$ in the second term, an interaction with the scatterer at \mathbf{x}' , which is given by $\Delta o(\mathbf{x}', \omega)$ in the second term, and a free propagation from \mathbf{x}' to the receiver at \mathbf{x} in the reference media, which is given by $\hat{G}(\mathbf{x}, \omega; \mathbf{x}')$ in the second term. The integral in the second term is over all scattering locations \mathbf{x}' in the volume V . The third term on the RHS of Eq. 3.377 is a summation of all the waves that have been scattered twice. One of such double-scattered waves is shown in Fig. 3.6c. The wave propagates in the reference media from the source to the first scatterer at \mathbf{x}'_1 , which is represented by $\hat{u}(\mathbf{x}'_1, \omega)$ in the third term, interacts with the scatterer at \mathbf{x}'_1 , which is $\Delta o(\mathbf{x}'_1, \omega)$, propagates from the first scatterer to the second scatterer at \mathbf{x}'_2 in the reference media, which is represented by $\hat{G}(\mathbf{x}'_2, \omega; \mathbf{x}'_1)$, interacts with the second scatterer, $\Delta o(\mathbf{x}'_2, \omega)$, then propagates from \mathbf{x}'_2 to the receiver \mathbf{x} in the reference media, which is represented by $\hat{G}(\mathbf{x}, \omega; \mathbf{x}'_2)$ in the third term. The integrals in the third term are over all first-scattering positions \mathbf{x}'_1 and all second-scattering positions \mathbf{x}'_2 . The fourth term on the RHS of Eq. 3.377, which is not shown, is a summation of all triple-scattered waves and the n th term is a summation of all the waves that have been scattered $n - 1$ times. Although Feynman's interpretation was originally applied to quantum electrodynamics (e.g., Arfken and Weber 2005), it is also applicable in many other areas including our discussion here on the Born series of mechanical waves.

3.5.1.1 Born Kernel

Comparing Eq. 3.377 with Eq. 3.154, we can see that the Born series has the form of the Taylor expansion of the waveform with respect to $o(\mathbf{x}, \omega)$ (Eq. 3.373) at the reference $\tilde{o}(\mathbf{x}, \omega)$ (Eq. 3.374). Since the Taylor expansion is unique, the Fréchet derivatives of the waveform with respect to $o(\mathbf{x}, \omega)$ can be directly read off from the Born series. For a fixed receiver location $\mathbf{x} = \mathbf{x}_r$ and a fixed frequency $\omega = \omega_i$, the action of the first-order Fréchet derivative $\mathbf{D}\hat{u}(\mathbf{x}_r, \omega_i)$ on $-\delta o(\mathbf{x}, \omega_i) = o(\mathbf{x}, \omega_i) - \tilde{o}(\mathbf{x}, \omega_i)$, whose norm is sufficiently small so that it is in the neighborhood of zero, can be directly read off from the second term on the RHS of Eq. 3.377,

$$\begin{aligned} & [\mathbf{D}\hat{u}(\mathbf{x}_r, \omega_i)] [-\delta o(\mathbf{x}, \omega_i)] \\ &= \int_V dV(\mathbf{x}') \left\{ \left[-\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x}') \hat{u}(\mathbf{x}', \omega_i) \right] [-\delta o(\mathbf{x}', \omega_i)] \right\}, \end{aligned} \quad (3.379)$$

where the integration kernel

$$K_B(\mathbf{x}|\mathbf{x}_r, \omega_i) = -\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x}) \hat{u}(\mathbf{x}', \omega_i) \quad (3.380)$$

is often called the *Born kernel*. In Eq. 3.380 I have replaced \mathbf{x}' with \mathbf{x} and separated the independent variable \mathbf{x} from the “parameters” \mathbf{x}_r and ω_i using “[|]”. Here I use the notation δo to represent an arbitrary small perturbation to the reference model \tilde{o} , while the notation Δo (Eq. 3.375) represents the (finite) difference between the actual and the reference structural models. Note that the Born kernel has a different expression from the Fredholm integration kernel given in Eq. 3.376.

The action of the second-order Fréchet derivative $\mathbf{D}^2\hat{u}(\mathbf{x}_r, \omega_i)$ on the 2-tuple $(-\delta o(\mathbf{x}, \omega_i), -\delta o(\mathbf{x}, \omega_i))$ can be read off from the third term on the RHS of Eq. 3.377,

$$\begin{aligned} & \left[\mathbf{D}^2\hat{u}(\mathbf{x}_r, \omega_i) \right] (-\delta o(\mathbf{x}, \omega_i), -\delta o(\mathbf{x}, \omega_i)) \\ &= 2 \int_V dV(\mathbf{x}'_2) \int_V dV(\mathbf{x}'_1) \left\{ \hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x}'_2) \hat{G}(\mathbf{x}'_2, \omega_i; \mathbf{x}'_1) \hat{u}(\mathbf{x}'_1, \omega_i) \right. \\ & \quad \left. [-\delta o(\mathbf{x}'_1, \omega_i)] [-\delta o(\mathbf{x}'_2, \omega_i)] \right\}. \end{aligned} \quad (3.381)$$

For the time-domain Born series in Eq. 3.378, we introduce

$$\delta R(\mathbf{x}) \equiv R(\mathbf{x}) - \tilde{R}(\mathbf{x}) = \frac{r^2(\mathbf{x})}{\tilde{\alpha}^2(\mathbf{x})} - \frac{1}{\tilde{\alpha}^2(\mathbf{x})}. \quad (3.382)$$

Then Eq. 3.378 gives the Taylor expansion of the time-domain waveform with respect to $R(\mathbf{x})$ at the reference $\tilde{R}(\mathbf{x})$. For a fixed receiver location $\mathbf{x} = \mathbf{x}_r$ and a fixed time $t = t_i$, the action of the first-order Fréchet derivative $\mathbf{D}u(\mathbf{x}_r, t_i)$ on $\delta R(\mathbf{x})$ can be read off from the second term on the RHS of Eq. 3.378,

$$\begin{aligned} & [\mathbf{D}u(\mathbf{x}_r, t_i)] \delta R(\mathbf{x}) \\ &= \int_V dV(\mathbf{x}') \left[\int d\tau \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}_r, t_i - \tau; \mathbf{x}', 0) \tilde{u}(\mathbf{x}', \tau) \right] \delta R(\mathbf{x}') \end{aligned} \quad (3.383)$$

$$= \int_V dV(\mathbf{x}') \left[\int d\tau \tilde{G}(\mathbf{x}_r, t_i - \tau; \mathbf{x}', 0) \frac{\partial^2 \tilde{u}}{\partial t^2}(\mathbf{x}', \tau) \right] \delta R(\mathbf{x}') \quad (3.384)$$

$$= \int_V dV(\mathbf{x}') \left[\int d\tau \frac{\partial \tilde{G}}{\partial t}(\mathbf{x}_r, t_i - \tau; \mathbf{x}', 0) \frac{\partial \tilde{u}}{\partial t}(\mathbf{x}', \tau) \right] \delta R(\mathbf{x}'). \quad (3.385)$$

Note that the expressions on 3.383, 3.384 and 3.385 are all equivalent and the time-domain Born kernel is given by the terms inside the square brackets. Here I use the expression on 3.383 and the time-domain Born kernel with respect to $R(\mathbf{x})$ at the reference $\tilde{R}(\mathbf{x})$ can be expressed as

$$K_B(\mathbf{x}|\mathbf{x}_r, t_i) = \int d\tau \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}_r, t_i - \tau; \mathbf{x}, 0) \tilde{u}(\mathbf{x}, \tau), \quad (3.386)$$

where I have replaced \mathbf{x}' with \mathbf{x} .

The action of the second-order Fréchet derivative $\mathbf{D}^2u(\mathbf{x}_r, t_i)$ on the 2-tuple $(\delta R(\mathbf{x}), \delta R(\mathbf{x}))$ can be read off from the third term on the RHS of Eq. 3.378,

$$\begin{aligned} & \left[\mathbf{D}^2u(\mathbf{x}_r, t_i) \right] (\delta R(\mathbf{x}), \delta R(\mathbf{x})) = 2 \int_V dV(\mathbf{x}'_2) \int_V dV(\mathbf{x}'_1) \\ & \left\{ \int dt \delta(t - t_i) \left[\frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}_r, t; \mathbf{x}'_2, 0) * \frac{\partial^2 \tilde{G}}{\partial t^2}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0) * \tilde{u}(\mathbf{x}'_1, t) \right] \right. \\ & \left. \delta R(\mathbf{x}'_1) \delta R(\mathbf{x}'_2) \right\}. \end{aligned} \quad (3.387)$$

To construct the Born kernel in Eqs. 3.380 or 3.386 we need access to both the forward wavefield from the source to all the scatterers $\hat{u}(\mathbf{x}, \omega_i)$ and also the Green's function from all the scatterers to the receiver $\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x})$. The forward wavefield $\hat{u}(\mathbf{x}, \omega_i)$ can be generated easily through one wave-propagation simulation from the source. To generate the Green's function $\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x})$, the number of wave-propagation simulations is the same as the number of scatterers, which is effectively the number of recording grid points in the finite-difference mesh (Sect. 2.2.5). The computational cost would be prohibitive if we adopt such a brute-force approach, especially when the reference structural model $\tilde{\alpha}(\mathbf{x})$ is three-dimensional in space and we have to resort to the finite-difference or other numerical techniques to solve the wave equation.

The technique implemented in (Zhao et al. 2005, 2006) and later named the *scattering-integral* (SI) method in (Chen et al. 2007a, b) is to use the reciprocity relation (Eq. 3.293) to replace $\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x})$ in the Born kernel with $\hat{G}(\mathbf{x}, \omega_i; \mathbf{x}_r)$.

The Green's function $\hat{G}(\mathbf{x}, \omega_i; \mathbf{x}_r)$ is the RGT (Sect. 3.4) and requires only one wave-propagation simulation from the receiver at \mathbf{x}_r to all the scatterers.

An alternative to using the reciprocity relation is to use Eq. 3.286 to replace $\hat{G}(\mathbf{x}_r, \omega_i; \mathbf{x})$ in the Born kernel with $\left[\hat{G}^\dagger(\mathbf{x}, \omega_i; \mathbf{x}_r)\right]^*$, which corresponds to the time-reversed⁶ adjoint Green's function (Sect. 3.4.1) and can be generated through one adjoint wave-propagation simulation from the receiver \mathbf{x}_r to all the scatterers by solving the adjoint scalar wave equation (Eq. 3.188) coupled with the adjoint initial/boundary conditions (Eqs. 3.247, 3.249 or 3.251).

For a given data set including N_s sources and N_r receivers, the total number of wave-propagation simulations (including the adjoint wave-propagation simulations if the alternative approach is adopted) needed to construct the Born kernels for all source-receiver pairs is $N_s + N_r$.

Consider the three-dimensional whole-space structural model with constant wave velocity $\tilde{\alpha}(\mathbf{x}) = c_0$, whose Green's function can be obtained by solving Eq. 3.363 analytically and is given by (e.g., Aki and Richards 2002)

$$\hat{G}(\mathbf{x}, \omega; \mathbf{x}_r) = -\frac{e^{-i\omega|\mathbf{x}-\mathbf{x}_r|/c_0}}{4\pi|\mathbf{x}-\mathbf{x}_r|}, \quad (3.388)$$

in the frequency domain and by

$$\tilde{G}(\mathbf{x}, t; \mathbf{x}_r) = -\frac{\delta(t - |\mathbf{x} - \mathbf{x}_r|/c_0)}{4\pi|\mathbf{x} - \mathbf{x}_r|} \quad (3.389)$$

in the time domain. Here $|\mathbf{x} - \mathbf{x}_r|$ represents the Euclidean distance between location \mathbf{x} and location \mathbf{x}_r and the origin time of the Dirac delta source is set at $t = 0$. Note that the differences between Eq. 3.389 above and Eq. 4.3 in (Aki and Richards 2002) are due to differences in the definition of the scalar wave equation, Eq. 3.186 above and Eq. 4.2 in (Aki and Richards 2002), and also the sign of the Dirac delta source. We assume that the forward wavefield in the reference media is generated by the source $\delta(\mathbf{x} - \mathbf{x}_s)S(t)$ and we have

$$\hat{u}(\mathbf{x}, \omega) = -\frac{e^{-i\omega|\mathbf{x}-\mathbf{x}_s|/c_0}}{4\pi|\mathbf{x}-\mathbf{x}_s|}\hat{S}(\omega), \quad (3.390)$$

$$\tilde{u}(\mathbf{x}, t) = -\frac{S(t - |\mathbf{x} - \mathbf{x}_s|/c_0)}{4\pi|\mathbf{x} - \mathbf{x}_s|}, \quad (3.391)$$

where $\hat{S}(\omega)$ is the Fourier transform of $S(t)$. Bring Eqs. 3.388 and 3.390 into Eq. 3.380, we obtain the frequency-domain Born kernel for the whole-space structural model

$$K_B(\mathbf{x}|\mathbf{x}_r, \mathbf{x}_s, \omega_i) = -\frac{e^{-i\omega_i(|\mathbf{x}-\mathbf{x}_r|+|\mathbf{x}-\mathbf{x}_s|)/c_0}}{(4\pi)^2|\mathbf{x}-\mathbf{x}_r||\mathbf{x}-\mathbf{x}_s|}\hat{S}(\omega_i). \quad (3.392)$$

⁶ Conjugation in the frequency domain is equivalent to reversing the direction of the time axis in the time domain.

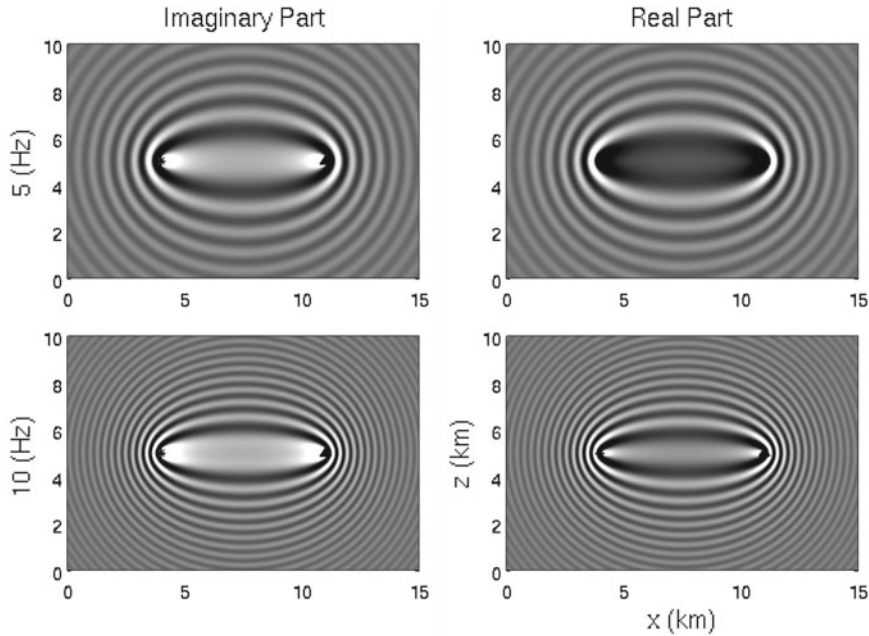


Fig. 3.7 Cross-section views of the frequency-domain Born kernels for the whole-space structural model. *Left column:* imaginary part of the kernel; *right column:* real part of the kernel. *Top row:* the kernel at 5 Hz; *bottom row:* the kernel at 10 Hz. Black star: source location; black triangle: receiver location. The wave velocity $\tilde{\alpha} = 6.0$ km/s. The source is located at coordinate (4 km, 0 km, 5 km) and the receiver is located at coordinate (11 km, 0 km, 5 km). The cross-section views are perpendicular to the source-receiver $x - z$ plane. *Gray scale* shows the value (darker: positive; lighter: negative) of the real or imaginary part of the kernel at every spatial location \mathbf{x} on the cross-section.

In Fig. 3.7, I have plotted examples of the real and imaginary parts of $K_B(\mathbf{x}|\mathbf{x}_r, \mathbf{x}_s, \omega_i)$ at two different frequencies. Bring Eqs. 3.389 and 3.391 into Eq. 3.386, we obtain the time-domain Born kernel for the whole-space model,

$$K_B(\mathbf{x}|\mathbf{x}_r, \mathbf{x}_s, t_i) = \frac{\ddot{S}(t_i - (|\mathbf{x} - \mathbf{x}_r| + |\mathbf{x} - \mathbf{x}_s|)/c_0)}{(4\pi)^2 |\mathbf{x} - \mathbf{x}_r| |\mathbf{x} - \mathbf{x}_s|}, \quad (3.393)$$

where $\ddot{S}(t)$ is the second-order time derivative of $S(t)$. Examples of the time-domain Born kernel $K_B(\mathbf{x}|\mathbf{x}_r, \mathbf{x}_s, t_i)$ for four different t_i values are shown in Fig. 3.8. The whole-space Born kernels shown here were named the *Born wavepaths* in (Woodward 1992) and they have utilities in *diffraction tomography*, which have been discussed extensively in the literature. For general three-dimensional reference structural models, we cannot use Eqs. 3.392 or 3.393 for obtaining the Born kernel and must resort to purely numerical solutions of the wave equation.

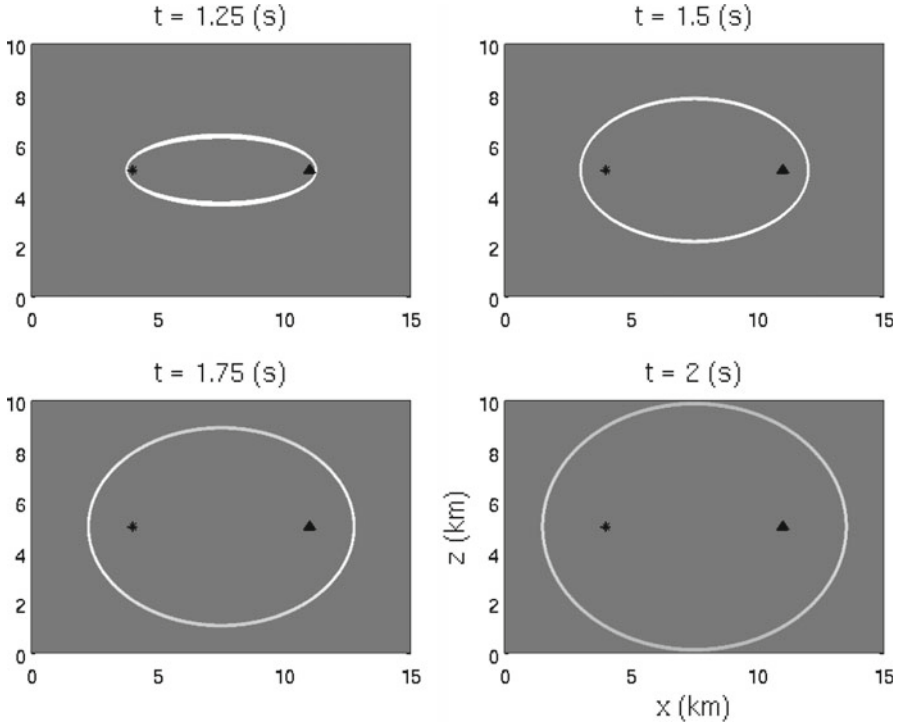


Fig. 3.8 Cross-section views of the time-domain Born kernels for the whole-space structural model. From *top to bottom, left to right*, are time-domain kernel amplitudes at $t_i = 1.25$ s, 1.5 s, 1.75 s and 2.0 s. Black star: source location; black triangle: receiver location. Gray scale shows kernel amplitude (lighter: positive; darker: negative) at every spatial location \mathbf{x} on the cross-section. Note that in 3D the kernel is shaped as an ellipsoid with the source and the receiver located at its two focal points.

3.5.1.2 Born Approximation

Whether the Born series in Eq. 3.377 converges or not depends upon the norm of the Fredholm integration kernel $K(\mathbf{x}, \mathbf{x}', \omega)$ given in Eq. 3.376. The Born series converges if the condition in Eq. 3.361 is satisfied. Unlike the Born kernel (Eq. 3.380), the Fredholm kernel involves the difference between the actual and the reference structural models $\Delta o(\mathbf{x}, \omega)$, which is usually unknown in tomography studies. Instead of analyzing the norm of the Fredholm kernel, we take a more intuitive approach and try to give qualitative conditions under which the first two terms on the RHS of Eqs. 3.377 or 3.378 can provide a good approximation to the LHS of Eqs. 3.377 or 3.378. This first-order approximation

$$\hat{u}(\mathbf{x}, \omega) \approx \hat{u}(\mathbf{x}, \omega) + \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \Delta o(\mathbf{x}', \omega) \hat{u}(\mathbf{x}', \omega) \quad (3.394)$$

$$\Rightarrow \Delta \hat{u}(\mathbf{x}, \omega) \approx \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \Delta o(\mathbf{x}', \omega) \hat{u}(\mathbf{x}', \omega) \quad (3.395)$$

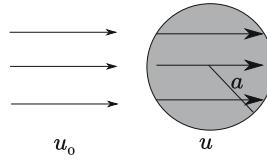


Fig. 3.9 An incident plane wave in the whole-space structure model $\hat{u}(\mathbf{x}, \omega)$, a cylindrical object with radius a and the wavefield inside the cylinder $\hat{u}(\mathbf{x}, \omega)$.

is often called the *Born approximation*. A time-domain formula can be obtained by ignoring the second- and higher-order terms on the RHS of Eq. 3.378. Note that Eqs. 3.394 or 3.395 is called an “approximation” because we have neglected higher-order terms on the RHS of the Born series (Eq. 3.377). The Born kernel (Eq. 3.380), which appears on the RHS of Eq. 3.395 still provides the exact first-order Fréchet derivative of the waveform with respect to $o(\mathbf{x}, \omega)$ (Sect. 3.5.1.1) and there is no approximation to this first-order Fréchet derivative itself. In practice the Green’s function and the forward wavefield in the Born kernel can be computed using approximate solutions to the wave equation, e.g., ray-theoretic solutions as in the generalized Radon transform (e.g., Beylkin 1985), but this type of approximations are different from the Born approximation discussed here.

Consider a cylindrical object with radius a embedded inside a whole-space structural model with constant wave velocity c_0 (Fig. 3.9). The difference between the refractive index inside the cylinder and that outside the cylinder is Δr . The incident wave is a plane wave propagating in the direction of the unit vector $\hat{\mathbf{n}}$ and we have

$$\hat{u}(\mathbf{x}, \omega) = A e^{-i\tilde{k}\hat{\mathbf{n}}\cdot\mathbf{x}}, \tag{3.396}$$

where “ \cdot ” represents the Euclidean dot product. The wavefield inside the cylinder propagates either slower or faster than the incident wave depending upon Δr . We can approximate the wavefield inside the cylinder as

$$\hat{u}(\mathbf{x}, \omega) \approx A e^{-i\tilde{k}(1+\Delta r)\hat{\mathbf{n}}\cdot\mathbf{x}}. \tag{3.397}$$

The accumulative phase change through the cylinder is therefore

$$\Delta\phi = \tilde{k}\Delta r \times 2a = \frac{2\pi}{\tilde{\lambda}}\Delta r \times 2a = \frac{4\pi a \Delta r}{\tilde{\lambda}}. \tag{3.398}$$

A necessary condition for the Born approximation to be valid is that the accumulative phase change is small, i.e., we require

$$\Delta\phi = \frac{4\pi a \Delta r}{\tilde{\lambda}} \ll \pi \Rightarrow a \Delta r \ll \frac{\tilde{\lambda}}{4}. \tag{3.399}$$

For a fixed wavelength $\tilde{\lambda}$ there are three scenarios under which the condition in Eq. 3.399 can be satisfied. One possibility is to have small scatterers (i.e., small a)

with possibly large material contrast (i.e., large Δr). A second possibility is to have very small material contrast (i.e., small Δr) but the size of the scatterer might be large (i.e., large a). And a third possibility is that both the size of the scatterer and the material contrast are small. For fixed a and Δr , Eq. 3.399 can also be satisfied by increasing the wavelength $\tilde{\lambda}$, which can often be achieved through low-pass filtering (e.g., Bunks et al. 1995; Sirgue and Pratt 2004).

If the approximation in Eqs. 3.394 or 3.395 is valid, we have established an approximate linear relation between the waveform difference $\Delta \hat{u}(\mathbf{x}, \omega)$ and the difference in material property $\Delta o(\mathbf{x}', \omega)$. The waveform difference can be computed by simply subtracting the model-predicted (i.e., synthetic) waveform computed in the reference structural model $\hat{u}(\mathbf{x}, \omega)$ from the observed waveform $\hat{u}(\mathbf{x}, \omega)$ at locations of the receivers. The Born kernel (Eqs. 3.380 or 3.386) can be computed by convolving the forward wavefield from the source with either the RGTs or the time-reversed adjoint Green's functions (Sect. 3.5.1.1), which are also computed in the reference structural model. For a given data set and a given reference structural model, it is therefore possible to set up a linear inverse problem based on the Born approximation (Eq. 3.395) to solve for the difference in material property between the actual and the reference structural models.

3.5.2 The Anisotropic Viscoelastic Wave Equation

Consider the anisotropic viscoelastic wave-equation operator defined in Eq. 3.204,

$$\begin{aligned} \mathbf{L}_{\mathbf{x},t} [\mathbf{u}(\mathbf{x}, t)]_i &= \rho(\mathbf{x}) \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right] = f_i(\mathbf{x}, t), \end{aligned} \quad (3.400)$$

and the equation for the reference media defined by $\tilde{\rho}(\mathbf{x})$ and $\tilde{\Psi}_{ijkl}(\mathbf{x}, t)$,

$$\begin{aligned} \tilde{\mathbf{L}}_{\mathbf{x},t} [\tilde{\mathbf{u}}(\mathbf{x}, t)]_i &= \tilde{\rho}(\mathbf{x}) \frac{\partial^2 \tilde{u}_i(\mathbf{x}, t)}{\partial t^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \tilde{\Psi}_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial \tilde{u}_k(\mathbf{x}, \tau)}{\partial x_l} \right] = f_i(\mathbf{x}, t), \end{aligned} \quad (3.401)$$

where $\tilde{u}_i(\mathbf{x}, t)$ is the wavefield in the reference media. The material differences are given by

$$\begin{bmatrix} \Delta \rho(\mathbf{x}) \\ \Delta \Psi_{ijkl}(\mathbf{x}, t) \end{bmatrix} \equiv \begin{bmatrix} \rho(\mathbf{x}) - \tilde{\rho}(\mathbf{x}) \\ \Psi_{ijkl}(\mathbf{x}, t) - \tilde{\Psi}_{ijkl}(\mathbf{x}, t) \end{bmatrix}, \quad (3.402)$$

and the wavefield difference is

$$\Delta u_i(\mathbf{x}, t) \equiv u_i(\mathbf{x}, t) - \tilde{u}_i(\mathbf{x}, t). \quad (3.403)$$

Subtract Eq. 3.401 from Eq. 3.400, considering the definitions in Eqs. 3.402 and 3.403, we obtain

$$\begin{aligned} \tilde{\mathbf{L}}_{\mathbf{x},t} [\Delta \mathbf{u}(\mathbf{x}, t)]_i &= \tilde{\rho}(\mathbf{x}) \frac{\partial^2 \Delta u_i(\mathbf{x}, t)}{\partial t^2} \\ &- \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \tilde{\Psi}_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial \Delta u_k(\mathbf{x}, \tau)}{\partial x_l} \right] = -\Delta \rho(\mathbf{x}) \frac{\partial^2 u_i(\mathbf{x}, t)}{\partial t^2} \\ &+ \sum_{j,k,l=1}^3 \frac{\partial}{\partial x_j} \left[\int_{-\infty}^{+\infty} d\tau \Delta \Psi_{ijkl}(\mathbf{x}, t - \tau) \frac{\partial u_k(\mathbf{x}, \tau)}{\partial x_l} \right]. \end{aligned} \quad (3.404)$$

Using the representation theorem in Eq. 3.326, we obtain

$$\begin{aligned} u_i(\mathbf{x}, t) &= \tilde{u}_i(\mathbf{x}, t) \\ &+ \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j=1}^3 \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}', t') \left[-\Delta \rho(\mathbf{x}') \frac{\partial^2 u_j(\mathbf{x}', t')}{\partial t'^2} \right] \\ &+ \int_0^T dt' \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}', t')}{\partial x'_k} \\ &\sum_{l,m=1}^3 \left[-\int_{-\infty}^{+\infty} d\tau \Delta \Psi_{jklm}(\mathbf{x}', t' - \tau) \frac{\partial u_l(\mathbf{x}', \tau)}{\partial x'_m} \right], \end{aligned} \quad (3.405)$$

where \tilde{G}_{ij} is the Green's tensor computed in the reference media. We can apply Eq. 3.359 to the integral Eq. 3.405 and obtain the Born series for the anisotropic viscoelastic wave equation

$$\begin{aligned} u_i(\mathbf{x}, t) &= \tilde{u}_i(\mathbf{x}, t) \quad (3.406) \\ &- \int_V dV(\mathbf{x}') \sum_{j=1}^3 \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}', 0) * \frac{\partial^2 \tilde{u}_j(\mathbf{x}', t)}{\partial t^2} \Delta \rho(\mathbf{x}') \\ &- \int_V dV(\mathbf{x}') \sum_{j,k,l,m=1}^3 \frac{\partial \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x'_k} * \frac{\partial \tilde{u}_l(\mathbf{x}', t)}{\partial x'_m} * \Delta \Psi_{jklm}(\mathbf{x}', t) \\ &+ \int_V dV(\mathbf{x}'_2) \sum_{j=1}^3 \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}'_2, 0) * \left[\Delta \rho(\mathbf{x}'_2) \frac{\partial^2}{\partial t^2} \right] \\ &\left\{ \int_V dV(\mathbf{x}'_1) \sum_{k=1}^3 \tilde{G}_{jk}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0) * \left[\Delta \rho(\mathbf{x}'_1) \frac{\partial^2 \tilde{u}_k(\mathbf{x}'_1, t)}{\partial t^2} \right] \right. \\ &\left. + \int_V dV(\mathbf{x}'_1) \sum_{n,o,p,q=1}^3 \frac{\partial \tilde{G}_{jn}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0)}{\partial x'_{1o}} * \left[\Delta \Psi_{nopq}(\mathbf{x}'_1, t) * \frac{\partial \tilde{u}_p(\mathbf{x}'_1, t)}{\partial x'_{1q}} \right] \right\} \end{aligned}$$

$$\begin{aligned}
& + \int_V dV(\mathbf{x}'_2) \sum_{j,k,l,m=1}^3 \frac{\partial \tilde{G}_{ij}(\mathbf{x}, t; \mathbf{x}'_2, 0)}{\partial x'_{2k}} * \left[\Delta \Psi_{jklm}(\mathbf{x}'_2, t) * \frac{\partial}{\partial x'_{2m}} \right] \\
& \quad \left\{ \int_V dV(\mathbf{x}'_1) \sum_{k=1}^3 \tilde{G}_{lk}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0) * \left[\Delta \rho(\mathbf{x}'_1) \frac{\partial^2 \tilde{u}_k(\mathbf{x}'_1, t)}{\partial t^2} \right] \right. \\
& + \int_V dV(\mathbf{x}'_1) \sum_{n,o,p,q=1}^3 \frac{\partial \tilde{G}_{ln}(\mathbf{x}'_2, t; \mathbf{x}'_1, 0)}{\partial x'_{1o}} * \left[\Delta \Psi_{nopq}(\mathbf{x}'_1, t) * \frac{\partial \tilde{u}_p(\mathbf{x}'_1, t)}{\partial x'_{1q}} \right] \left. \right\} \\
& \quad + \dots
\end{aligned}$$

Here I have used the translation-invariance in time of the Green's tensor to convert the integrations over time to temporal convolutions, denoted using “*” in Eq. 3.406.

The first- and higher-order Fréchet derivatives of the waveform $u_i(\mathbf{x}, t)$ with respect to material properties $\rho(\mathbf{x})$ and $\Psi_{jklm}(\mathbf{x}, t)$ at the reference structural model $\tilde{\rho}(\mathbf{x})$ and $\tilde{\Psi}_{jklm}(\mathbf{x}, t)$ can be obtained from Eq. 3.406. For a fixed receiver location $\mathbf{x} = \mathbf{x}_r$ and time t , the Born kernel for the density $\rho(\mathbf{x})$ can be expressed as

$$K_B^\rho(\mathbf{x}|\mathbf{x}_r, t) = - \sum_{j=1}^3 \tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0) * \frac{\partial^2 \tilde{u}_j(\mathbf{x}, t)}{\partial t^2} \quad (3.407)$$

$$= - \sum_{j=1}^3 \frac{\partial^2 \tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0)}{\partial t^2} * \tilde{u}_j(\mathbf{x}, t) \quad (3.408)$$

$$= - \sum_{j=1}^3 \frac{\partial \tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0)}{\partial t} * \frac{\partial \tilde{u}_j(\mathbf{x}, t)}{\partial t}. \quad (3.409)$$

Here I have replaced \mathbf{x}' with \mathbf{x} and the expressions in 3.407, 3.408 and 3.409 are equivalent. The Born kernel for the rate of relaxation at a given time τ , $\Psi_{jklm}(\mathbf{x}, \tau)$, can be expressed as

$$K_B^{\Psi_{jklm}}(\mathbf{x}|\mathbf{x}_r, t, \tau) = - \int dt' \frac{\partial \tilde{G}_{ij}(\mathbf{x}_r, t - t'; \mathbf{x}, 0)}{\partial x_k} \frac{\partial \tilde{u}_l(\mathbf{x}, t' - \tau)}{\partial x_m}. \quad (3.410)$$

Here I have replaced \mathbf{x}' with \mathbf{x} . If the material is perfectly elastic, we have

$$\Psi_{jklm}(\mathbf{x}, \tau) = c_{jklm}(\mathbf{x})\delta(\tau), \quad (3.411)$$

the Born kernel for $c_{jklm}(\mathbf{x})$ can then be expressed as

$$K_B^{c_{jklm}}(\mathbf{x}|\mathbf{x}_r, t) = - \frac{\partial \tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0)}{\partial x_k} * \frac{\partial \tilde{u}_l(\mathbf{x}, t)}{\partial x_m}. \quad (3.412)$$

The Green's tensor $\tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0)$ inside the Born kernels can be replaced with the RGT $\tilde{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$ if we use the reciprocity relation in Eq. 3.306 and we obtain

$$K_B^\rho(\mathbf{x}|\mathbf{x}_r, t) = - \sum_{j=1}^3 \frac{\partial \tilde{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial \tilde{u}_j(\mathbf{x}, t)}{\partial t}, \quad (3.413)$$

for the Born density kernel,

$$K_B^{\Psi jklm}(\mathbf{x}|\mathbf{x}_r, t, \tau) = - \int dt' \frac{\partial \tilde{G}_{ji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial \tilde{u}_l(\mathbf{x}, t' - \tau)}{\partial x_m}, \quad (3.414)$$

for the viscoelastic rate-of-relaxation and

$$K_B^{c jklm}(\mathbf{x}|\mathbf{x}_r, t) = - \frac{\partial \tilde{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * \frac{\partial \tilde{u}_l(\mathbf{x}, t)}{\partial x_m}, \quad (3.415)$$

for the elastic moduli. If we use Eq. 3.297 instead, we can replace $\tilde{G}_{ij}(\mathbf{x}_r, t; \mathbf{x}, 0)$ with the time-reversed adjoint Green's tensor $\tilde{G}_{ji}^{\ddagger}(\mathbf{x}, t; \mathbf{x}_r, 0)$ (Eq. 3.345, Sect. 3.4.1) and we obtain

$$K_B^{\rho}(\mathbf{x}|\mathbf{x}_r, t) = - \sum_{j=1}^3 \frac{\partial \tilde{G}_{ji}^{\ddagger}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial \tilde{u}_j(\mathbf{x}, t)}{\partial t}, \quad (3.416)$$

$$K_B^{\Psi jklm}(\mathbf{x}|\mathbf{x}_r, t, \tau) = - \int dt' \frac{\partial \tilde{G}_{ji}^{\ddagger}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial \tilde{u}_l(\mathbf{x}, t' - \tau)}{\partial x_m}, \quad (3.417)$$

$$K_B^{c jklm}(\mathbf{x}|\mathbf{x}_r, t) = - \frac{\partial \tilde{G}_{ji}^{\ddagger}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * \frac{\partial \tilde{u}_l(\mathbf{x}, t)}{\partial x_m}. \quad (3.418)$$

If the second- and higher-order terms on the RHS of the Born series (Eq. 3.406) are ignored, we obtain the Born approximation for the anisotropic viscoelastic wave equation.

References

- Aki, K., & Richards, P. G. (2002). *Quantitative seismology* (2nd ed.). University Science Books.
- Arfken, G., & Weber, H. (2005). *Mathematical methods for physicists international student edition*. Elsevier Science.
- Axler, S. (1997). *Linear algebra done right*. Lecture Notes in Statistics. Springer.
- Backus, G., & Mulcahy, M. (1976a). Moment tensors and other phenomenological descriptions of seismic sources—I. continuous displacements. *Geophysical Journal International*, 46(2), 341–361.
- Backus, G., & Mulcahy, M. (1976b). Moment tensors and other phenomenological descriptions of seismic sources—II. discontinuous displacements. *Geophysical Journal International*, 47(2), 301–329.
- Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26(1), 99–108.
- Bunks, C., Saleck, F. M., Zaleski, S., & Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5), 1457–1473.
- Chen, P., Jordan, T., & Zhao, L. (2005). Finite-moment tensor of the 3 september 2002 Yorba Linda earthquake. *Bulletin of the Seismological Society of America*, 95(3), 1170–1180.

- Chen, P., Jordan, T. H., & Zhao, L. (2007a). Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1), 175–181.
- Chen, P., Zhao, L., & Jordan, T. H. (2007b). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, 97(4), 1094–1120.
- Chen, P., Lee, E.-J., & Wang, L. (2013). A cloud-based synthetic seismogram generator implemented using Windows Azure. *Earthquake Science*, 26(5), 321–329.
- Dahlen, F. A., & Tromp, J. (1998). *Theoretical global seismology*. Princeton University Press.
- DeVito, C. (1990). *Functional analysis and linear operator theory*. Advanced book program: Addison-Wesley. Addison-Wesley Publishing Company, Advanced Book Program.
- Friedman, A. (1970). *Foundations of modern analysis*. Dover Books on Mathematics Series. Dover.
- Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Artificial Intelligence. Addison-Wesley.
- Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., et al. (2011). Cybershake: A physics-based seismic hazard model for southern California. *Pure and Applied Geophysics*, 168(3–4), 367–381.
- Hassani, S. (1999). *Mathematical physics: A modern introduction to its foundations*. Springer.
- Hassani, S. (2008). *Mathematical methods: For students of physics and related fields*. Lecture Notes in Mathematics; 720. Springer Science+Business Media.
- Kennedy, J. (2010). Particle swarm optimization. *Encyclopedia of Machine Learning*, 760–766. Springer.
- Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., et al. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.
- Kubrusly, C. (2012). *Spectral theory of operators on Hilbert spaces*. Springer.
- Lee, E.-J., Chen, P., Jordan, T. H., & Wang, L. (2011). Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. *Geophysical Journal International*, 186(1), 311–330.
- MacCluer, B. (2008). *Elementary functional analysis*. Graduate Texts in Mathematics. Springer.
- Prudnikov, A. (1986). *Integrals and series: Volume 1: Elementary functions; Volume 2: Special functions*. Taylor & Francis.
- Schröder, B. (2008). *Mathematical analysis: A concise introduction*. Wiley.
- Sirgue, L., & Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1), 231–248.
- Sunder, V. (1997). *Functional analysis: Spectral theory* (Birkhäuser advanced texts: Basler Lehrbücher). Springer.
- Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128, 365–399.
- Trèves, F. (1967). *Topological vector spaces, distributions and kernels*. Pure and Applied Mathematics. Elsevier Science.
- Woodward, M. J. (1992). Wave-equation tomography. *Geophysics*, 57(1), 15–26.
- Zeidler, E. (1995). *Applied functional analysis: Applications to mathematical physics*. Number v. 108 in Applied Mathematical Sciences. Springer New York.
- Zhao, L., Chen, P., & Jordan, T. H. (2006). Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. *Bulletin of the Seismological Society of America*, 96(5), 1753–1763.
- Zhao, L., Jordan, T. H., Olsen, K. B., & Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, 95(6), 2066–2080.

Chapter 4

Data Sensitivity Kernels

The previous chapter is centered around Green’s functions and adjoint Green’s functions. I have shown that the solutions of the forward problems can be represented using Green’s functions and explicit formulas are given for computing Fréchet derivatives of the waveform with respect to the moment tensor density of the seismic source and the rate-of-relaxation and density of the structural model using either receiver Green’s tensors (RGTs) or time-reversed adjoint Green’s tensors. For the anisotropic viscoelastic wave equation, we can combine Eqs. 3.342, 3.413 and 3.414 and write

$$\begin{aligned}
 \delta u_i^s(\mathbf{x}_r, t) = & \int d\tau \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r, 0)}{\partial x_k} \delta m_{jk}^s(\mathbf{x}, \tau) \\
 & - \int_V dV(\mathbf{x}) \sum_{j=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t} \delta \rho(\mathbf{x}) \\
 & - \int d\tau \int_V dV(\mathbf{x}) \sum_{j,k,l,m=1}^3 \left\{ \int dt' \frac{\partial G_{ji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - \tau)}{\partial x_m} \right. \\
 & \quad \left. \delta \Psi_{jklm}(\mathbf{x}, \tau) \right\}, \tag{4.1}
 \end{aligned}$$

where I have applied the reciprocity principle (Eq. 3.306), the Green’s tensor and the forward wavefield are computed using the reference source and structural models¹ and the superscript ^s is an index for the seismic source that generates the forward wavefield. In Eq. 4.1, we can also replace the Green’s tensor with the time-reversed adjoint Green’s tensor by applying Eqs. 3.297 and 3.345. This formulation based on Green’s functions will allow us to analyze the computational costs of different implementations more easily and will make explicit both the connections and the differences between the adjoint-wavefield method (e.g., Tarantola 1988b; Pratt

¹ Here I have dropped the \sim on top of the Green’s tensor and the forward wavefield used in Eqs. 3.413 and 3.414 to reduce clutter.

1999; Pratt and Shipp 1999; Tromp et al. 2005; Tape et al. 2009; Fichtner 2010; Liu and Gu 2012) and the scattering-integral method (e.g., Zhao et al. 2005, 2006; Chen et al. 2007a, b; Chen 2011) for full-3D waveform tomography (F3DT).

Unlike travel-time observations, seismic waveforms are highly nonlinear with respect to seismic velocity models and this strong nonlinearity often exhibits itself as numerous spurious local minima on objective functions defined in terms of waveform differences. An early numerical example of such nonlinearity was presented in (Gauthier et al. 1986), in which the objective function was defined as the energy of the time-domain differences between the observed and the corresponding synthetic waveforms, i.e., $\|\bar{u}(t) - u(t)\|_{L_2}$, where $\bar{u}(t)$ is the observed waveform and $u(t)$ is the corresponding synthetic waveform. For such strongly nonlinear optimization problems, gradient-based descent algorithms can stagnate at a local minimum without converging to a useful model, unless the starting model is already close enough to the global minimum. In such a situation, the Fréchet derivatives, even though they can be computed accurately using purely numerical solutions of the 3D wave equation, are not useful for achieving global convergence. It is therefore important to consider other types of misfit measures that are more linear with respect to seismic velocity models.

In this chapter, I will focus on an issue that is highly important for solving seismic inverse problems: the construction of Fréchet kernels for different types of waveform misfit measures and for different types of parameterizations of source and structural models. I will show that the Fréchet kernels for any types of waveform misfit measures with respect to any model parameters can be constructed using the *primary kernels* derived in the previous chapter and summarized in Eq. 4.1. The calculations involved in this chapter will be demonstrated using the software F3DWT.

4.1 Definition

A certain parameterization of the unknown (target) earth structure and seismic source models can be combined into a model vector \mathbf{m} . The vector-valued displacement field computed using \mathbf{m} for the s -th seismic source can be denoted as $\mathbf{u}^s(\mathbf{x}, t)$. The synthetic displacement field computed using a reference source and structural model $\tilde{\mathbf{m}}$ can be denoted as $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$. For each pair of $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and $\mathbf{u}^s(\mathbf{x}, t)$ we can introduce a finite set of scalars, indexed using n , that quantify the misfits between $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and $\mathbf{u}^s(\mathbf{x}, t)$,

$$d^{sn} \equiv D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t))[\mathbf{u}^s(\mathbf{x}, t)]. \quad (4.2)$$

Here D_n is a map defined using the reference displacement field $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and applied on the target displacement field $\mathbf{u}^s(\mathbf{x}, t)$. Since it maps $\mathbf{u}^s(\mathbf{x}, t)$, which is a vector-valued function of space and time, to a scalar $d^{sn} \in \mathbb{C}$, D_n can be called a *data functional*. A data functional must satisfy the condition

$$D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t))[\tilde{\mathbf{u}}^s(\mathbf{x}, t)] = 0, \quad (4.3)$$

i.e., it must be zero when $\mathbf{u}^s(\mathbf{x}, t) = \tilde{\mathbf{u}}^s(\mathbf{x}, t)$. The observed displacement wavefield can be denoted as $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and when $\mathbf{u}^s(\mathbf{x}, t) = \tilde{\mathbf{u}}^s(\mathbf{x}, t)$, we have

$$\bar{d}^{sn} \equiv D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)]. \quad (4.4)$$

Using Eqs. 4.2 and 4.4, we can define the difference

$$\Delta d^{sn} \equiv \bar{d}^{sn} - d^{sn} = D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] - D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\mathbf{u}^s(\mathbf{x}, t)]. \quad (4.5)$$

This difference takes on the value of \bar{d}^{sn} when the target wavefield is identical to the reference wavefield, i.e.,

$$\begin{aligned} \Delta d^{sn} &= D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] - D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] \\ &= D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] = \bar{d}^{sn}, \end{aligned} \quad (4.6)$$

considering Eq. 4.3. And when the target wavefield is identical to the observed wavefield, we have

$$\Delta d^{sn} = D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] - D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\tilde{\mathbf{u}}^s(\mathbf{x}, t)] = 0. \quad (4.7)$$

We can evaluate the Fréchet derivative of the data functional d^{sn} with respect to \mathbf{m} at the reference model $\tilde{\mathbf{m}}$ and we have

$$\delta d^{sn} = \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{K}(\mathbf{x}, t | \tilde{\mathbf{m}}) \cdot \delta \mathbf{m}(\mathbf{x}, t), \quad (4.8)$$

where $\tilde{\mathbf{m}}(\mathbf{x}, t)$ denotes the reference source and structural models, “ \cdot ” denotes the Euclidean dot product and the integration kernel of the Fréchet derivative $\mathbf{K}(\mathbf{x}, t | \tilde{\mathbf{m}})$ evaluated at the reference source and structural models $\tilde{\mathbf{m}}(\mathbf{x}, t)$ is called the *data sensitivity kernel* (Backus and Gilbert 1968). Since \bar{d}^{sn} has no dependence on $\mathbf{m}(\mathbf{x}, t)$, the Fréchet kernel of Δd^{sn} with respect to $\mathbf{m}(\mathbf{x}, t)$ at the reference model $\tilde{\mathbf{m}}(\mathbf{x}, t)$ can be expressed as

$$\delta \Delta d^{sn} = -\delta d^{sn} = - \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{K}(\mathbf{x}, t | \tilde{\mathbf{m}}) \cdot \delta \mathbf{m}(\mathbf{x}, t). \quad (4.9)$$

In Sect. 4.2, I will discuss the construction of the data sensitivity kernels for any data functionals by applying the chain rule of the Fréchet derivative (Eq. 3.152). In Sect. 4.3, I will discuss how to construct the data sensitivity kernels for model parameterizations other than the rate-of-relaxation tensor or the stiffness tensor through linear combinations of the primary kernels.

4.2 Data Functionals

The discrepancies between the reference wavefield $\tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and the corresponding target wavefield $\mathbf{u}^s(\mathbf{x}, t)$ can be quantified in infinitely many different ways. For example, we can use the difference between the time-domain waveforms on the i -th component of the r -th receiver at time t_n as a data functional, i.e.,

$$\begin{aligned} d^{sn} &= D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t))[\mathbf{u}^s(\mathbf{x}, t)] \\ &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 \delta(\mathbf{x} - \mathbf{x}_r) \delta(t - t_n) \delta_{ip} \left[u_p^s(\mathbf{x}, t) - \tilde{u}_p^s(\mathbf{x}, t) \right] \\ &= u_i^s(\mathbf{x}_r, t_n) - \tilde{u}_i^s(\mathbf{x}_r, t_n), \end{aligned} \quad (4.10)$$

where \mathbf{x}_r is the location of the r th receiver, δ_{ip} is the Kronecker delta and $\delta(\cdot)$ denotes the Dirac Delta function.

Multiple components on the same receiver can be involved in a single misfit measurement. In the example above, suppose the difference is taken on the radial component, then we have

$$\begin{aligned} d^{sn} &= D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t))[\mathbf{u}^s(\mathbf{x}, t)] \\ &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 \delta(\mathbf{x} - \mathbf{x}_r) \delta(t - t_n) \sum_{i=1}^3 \gamma_i \delta_{ip} \left[u_p^s(\mathbf{x}, t) - \tilde{u}_p^s(\mathbf{x}, t) \right] \\ &= \sum_{i=1}^3 \gamma_i u_i^s(\mathbf{x}_r, t_n) - \sum_{i=1}^3 \gamma_i \tilde{u}_i^s(\mathbf{x}_r, t_n), \end{aligned} \quad (4.11)$$

where $\gamma_1 = \sin(\theta)$, $\gamma_2 = \cos(\theta)$, $\gamma_3 = 0$ and θ is the angle between the radial and the x_2 (e.g., north) directions.

A single data functional can also involve the displacement at multiple spatial locations, multiple time samples and/or multiple frequency samples. In practice the data functional can also involve operations that are nonlinear with respect to $\mathbf{u}^s(\mathbf{x}, t)$ and much more complex than the simple subtraction used in Eqs. 4.10 and 4.11. But no matter how complex the data functional is, it is always possible to construct the Fréchet derivative of the data functional with respect to the target wavefield as long as all the operations involved in the data functional are given. The integration kernel of this Fréchet derivative was named *the seismogram perturbation kernel* (SPK) in (Chen et al. 2007a). In (Chen et al. 2007a), I defined the SPK as the time-dependent, space-independent Fréchet kernel of the data functional with respect to the target waveform on a fixed component at a fixed receiver location. The same concept can be generalized to account for data functionals that involve the full target wavefield. Using Eq 3.145, the action of the Fréchet derivative $\mathbf{D}d^{sn}$ on a small wavefield

perturbation $\delta \mathbf{u}^s(\mathbf{x}, t)$ can be computed as

$$[\mathbf{D}d^{sn}] \delta \mathbf{u}^s(\mathbf{x}, t) = \lim_{\alpha \rightarrow 0} \frac{D_n(\tilde{\mathbf{u}}^s) [\mathbf{u}^s + \alpha \delta \mathbf{u}^s] - D_n(\tilde{\mathbf{u}}^s) [\mathbf{u}^s]}{\alpha}. \quad (4.12)$$

The integration kernel of the operator $\mathbf{D}d^{sn}$ is defined as the vector-valued, space- and time-dependent *wavefield perturbation kernel* (WPK) $\mathbf{J}^{sn}(\mathbf{x}, t)$ and we have

$$\begin{aligned} \delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}, t) \cdot \delta \mathbf{u}^s(\mathbf{x}, t) \\ &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \delta u_p^s(\mathbf{x}, t). \end{aligned} \quad (4.13)$$

For the example shown in Eq. 4.11, the WPK for this data functional can be expressed as

$$J_p^{sn}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_r) \delta(t - t_n) \gamma_p. \quad (4.14)$$

Because this data functional is linear with respect to the target wavefield $\mathbf{u}^s(\mathbf{x}, t)$, the WPK does not depend on $\mathbf{u}^s(\mathbf{x}, t)$. For data functionals that are nonlinear with respect to $\mathbf{u}^s(\mathbf{x}, t)$, the corresponding WPKs should have dependence upon $\mathbf{u}^s(\mathbf{x}, t)$.

4.2.1 Data Functionals for $\nabla \times \mathbf{u}$ and $\nabla \cdot \mathbf{u}$

For sensors that measure a certain transform of the displacement field, for instance, the rotational sensors that measure the curl of the displacement $\nabla \times \mathbf{u}$ at fixed locations (e.g., Cochard et al. 2006, Lee et al. 2009a, b), the data functional can be defined using the reference rotational field $\tilde{\mathbf{w}}^s(\mathbf{x}, t) \equiv \nabla \times \tilde{\mathbf{u}}^s(\mathbf{x}, t)$ and applied on the target rotational field $\mathbf{w}^s(\mathbf{x}, t) \equiv \nabla \times \mathbf{u}^s(\mathbf{x}, t)$, i.e.,

$$d^{sn} \equiv D_n(\tilde{\mathbf{w}}^s(\mathbf{x}, t)) [\mathbf{w}^s(\mathbf{x}, t)]. \quad (4.15)$$

The action of the Fréchet derivative $\mathbf{D}d^{sn}$ on $\delta \mathbf{u}^s(\mathbf{x}, t)$ can therefore be computed as

$$\begin{aligned} [\mathbf{D}d^{sn}] \delta \mathbf{u}^s(\mathbf{x}, t) &= \lim_{\alpha \rightarrow 0} \{ D_n(\nabla \times \tilde{\mathbf{u}}^s) [\nabla \times (\mathbf{u}^s + \alpha \delta \mathbf{u}^s)] \\ &\quad - D_n(\nabla \times \tilde{\mathbf{u}}^s) [\nabla \times \mathbf{u}^s] \} / \alpha. \end{aligned} \quad (4.16)$$

The WPK can be constructed in two steps. In the first step, we introduce the integration kernel $\mathbf{J}'(\mathbf{x}, t)$ such that

$$\delta d^{sn} = \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{J}'(\mathbf{x}, t) \cdot [\nabla \times \delta \mathbf{u}^s(\mathbf{x}, t)], \quad (4.17)$$

where I have used the linearity of the curl operator $\nabla \times (\mathbf{u}^s + \delta \mathbf{u}^s) = \nabla \times \mathbf{u}^s + \nabla \times \delta \mathbf{u}^s = \mathbf{w}^s + \delta \mathbf{w}^s$ and the integration kernel $\mathbf{J}'(\mathbf{x}, t)$ is the Fréchet kernel of the data

functional with respect to the target rotational field $\mathbf{w}^s(\mathbf{x}, t)$. In the second step, we use Eq. 3.199 to transfer the curl operator from $\delta\mathbf{u}^s(\mathbf{x}, t)$ to $\mathbf{J}'(\mathbf{x}, t)$,

$$\begin{aligned}\delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt [\nabla \times \mathbf{J}'(\mathbf{x}, t)] \cdot \delta\mathbf{u}^s(\mathbf{x}, t) \\ &= \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}, t) \cdot \delta\mathbf{u}^s(\mathbf{x}, t).\end{aligned}\quad (4.18)$$

Here I have defined the WPK for the data functional d^{sn} with respect to $\mathbf{u}^s(\mathbf{x}, t)$ as

$$\mathbf{J}^{sn}(\mathbf{x}, t) \equiv \nabla \times \mathbf{J}'(\mathbf{x}, t), \quad (4.19)$$

and I have assumed that the surface term $\int_{\partial V} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da(\mathbf{x}) = 0$, which can be satisfied by choosing the boundary of the volume ∂V away from all receiver locations.

For sensors that measure the divergence of the displacement field $\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)$, the data functional can be defined using the reference divergence field and applied on the target divergence field, i.e.,

$$d^{sn} \equiv D_n (\nabla \cdot \tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)]. \quad (4.20)$$

We denote the Fréchet kernel of d^{sn} with respect to the divergence field $\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)$ as $J'(\mathbf{x}, t)$, then we have

$$\delta d^{sn} = \int_V dV(\mathbf{x}) \int_0^T dt J'(\mathbf{x}, t) [\nabla \cdot \delta\mathbf{u}^s(\mathbf{x}, t)]. \quad (4.21)$$

Using Eq. 3.196, we obtain

$$\begin{aligned}\delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt [-\nabla J'(\mathbf{x}, t)] \cdot \delta\mathbf{u}^s(\mathbf{x}, t) \\ &= \int_V dV(\mathbf{x}) \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}, t) \cdot \delta\mathbf{u}^s(\mathbf{x}, t),\end{aligned}\quad (4.22)$$

where we have defined the WPK of d^{sn} with respect to $\mathbf{u}^s(\mathbf{x}, t)$ as

$$\mathbf{J}^{sn}(\mathbf{x}, t) \equiv -\nabla J'(\mathbf{x}, t), \quad (4.23)$$

and the surface term $\int_{\partial V} \mathbf{Q} \cdot \hat{\mathbf{e}}_n da(\mathbf{x})$ can always be made zero.

4.2.2 Filtered RGT

In practice, the RGTs used in constructing the primary kernels in Eq. 4.1 are usually filtered using a finite-bandwidth source-time function. From a theoretical point-of-view, using a finite-bandwidth source-time function instead of a Dirac delta function may improve the numerical behavior of the wave-equation solver in certain

numerical algorithms for solving differential equations. From a practical point-of-view, before measuring the discrepancies between the observed seismogram and the corresponding synthetic seismogram, both seismograms are usually low-pass or band-pass filtered using the same filter. To account for this filtering step in the corresponding data sensitivity kernels, we need to convolve the same filter to the primary kernels and derive the WPK from the filtered synthetic seismogram. Convolution of the filter to the primary kernels is equivalent to using this filter as the source-time function of the RGT.

In Eq. 3.349, the filtered RGT is denoted as $\check{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$ and the source-time function used for computing the RGT (or RSGT) is denoted as $h(t)$. If we replace the RGT on the RHS of Eq. 4.1 $G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$ with the filtered RGT $\check{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$, the LHS of Eq. 4.1 $\delta u_i^s(\mathbf{x}_r, t)$ must also be replaced with $\delta \check{u}_i^s(\mathbf{x}_r, t)$, where

$$\check{u}_i^s(\mathbf{x}_r, t) = u_i^s(\mathbf{x}_r, t) * h(t) \quad (4.24)$$

is the displacement seismogram filtered using $h(t)$ (Eq. 3.348). In this situation, the data functional can be constructed to quantify the discrepancies between the filtered reference wavefield and the filtered target wavefield,

$$\check{d}^{sn} = D_n \left(\check{\mathbf{u}}^s(\mathbf{x}, t) \right) \left[\check{\mathbf{u}}^s(\mathbf{x}, t) \right], \quad (4.25)$$

where the filtered reference wavefield is

$$\check{u}_i^s(\mathbf{x}, t) = \tilde{u}_i^s(\mathbf{x}, t) * h(t). \quad (4.26)$$

The WPK corresponding to the data functional in Eq. 4.25 must also be constructed using the filtered reference wavefield $\check{\mathbf{u}}^s(\mathbf{x}, t)$. If this WPK is denoted as $\check{\mathbf{J}}^{sn}(\mathbf{x}, t)$, then we have

$$\delta \check{d}^{sn} = \int_V dV(\mathbf{x}) \int_0^T dt \check{\mathbf{J}}^{sn}(\mathbf{x}, t) \cdot \delta \check{\mathbf{u}}^s(\mathbf{x}, t), \quad (4.27)$$

where the filtered wavefield perturbation can be expressed as

$$\begin{aligned} \delta \check{u}_i^s(\mathbf{x}_r, t) &= h(t) * \delta u_i^s(\mathbf{x}_r, t) \quad (4.28) \\ &= \int d\tau \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial \check{G}_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r, 0)}{\partial x_k} \delta m_{jk}^s(\mathbf{x}, \tau) \\ &\quad - \int_V dV(\mathbf{x}) \sum_{j=1}^3 \frac{\partial \check{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t} \delta \rho(\mathbf{x}) \\ &\quad - \int d\tau \int_V dV(\mathbf{x}) \sum_{j,k,l,m=1}^3 \left\{ \int dt' \frac{\partial \check{G}_{ji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - \tau)}{\partial x_m} \right. \\ &\quad \left. \delta \Psi_{jklm}(\mathbf{x}, \tau) \right\}. \end{aligned}$$

Equation 4.28 gives the filtered primary kernels. Note that the filtered RGT $\check{G}_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$ is also computed using the reference structural model and I have dropped the “ $\check{\cdot}$ ” used in Chap. 3 to reduce clutter. The RGTs used in constructing the data sensitivity kernels in this chapter are all filtered RGTs and I will drop the “ $\check{\cdot}$ ” in the following to reduce clutter.

4.2.3 The Chain Rule

The data functional maps the target wavefield $\mathbf{u}^s(\mathbf{x}, t)$ to a scalar in \mathbb{C} and the anisotropic viscoelastic wave-equation solver allows us to map the source and structural models $m_{jk}^s(\mathbf{x}, \tau)$, $\rho(\mathbf{x})$ and $\Psi_{jklm}(\mathbf{x}, \tau)$ to the target wavefield $\mathbf{u}^s(\mathbf{x}, t)$. The Fréchet kernel of the target wavefield with respect to source and structural models are given by the primary kernels derived in Chap. 3 and summarized in Eq. 4.1. The Fréchet kernel of the data functional with respect to the target wavefield is given by the WPK (Eq. 4.13). The Fréchet kernel of the data functional with respect to source and structural models can be obtained by composing the primary kernels with the WPK (Eq. 3.152).

For the moment tensor density of the source we have

$$\begin{aligned} \delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \\ &= \int d\tau \int_V dV(\mathbf{x}') \sum_{j,k=1}^3 \frac{\partial G_{jp}(\mathbf{x}', t - \tau; \mathbf{x}, 0)}{\partial x'_k} \delta m_{jk}^s(\mathbf{x}', \tau) \\ &= \sum_{j,k=1}^3 \int d\tau \int_V dV(\mathbf{x}') \left\{ \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \right. \\ &\quad \left. \frac{\partial G_{jp}(\mathbf{x}', t - \tau; \mathbf{x}, 0)}{\partial x'_k} \right\} \delta m_{jk}^s(\mathbf{x}', \tau), \end{aligned} \quad (4.29)$$

and the Fréchet kernel of the data functional with respect to $m_{jk}^s(\mathbf{x}, t)$ can be expressed as

$$K_{d^{sn}}^{m_{jk}^s}(\mathbf{x}, t) = \int_V dV(\mathbf{x}') \int_0^T d\tau \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', \tau) \frac{\partial G_{jp}(\mathbf{x}, \tau - t; \mathbf{x}', 0)}{\partial x_k}, \quad (4.30)$$

where I have switched the roles of \mathbf{x} and \mathbf{x}' (i.e., $\mathbf{x} \leftrightarrow \mathbf{x}'$) and the roles of t and τ (i.e., $t \leftrightarrow \tau$).

For the density $\rho(\mathbf{x})$, we have

$$\begin{aligned}
 \delta d^{sn} &= - \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \\
 &\left[\int_V dV(\mathbf{x}') \sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}', t; \mathbf{x}, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}', t)}{\partial t} \delta \rho(\mathbf{x}') \right] \\
 &= - \int_V dV(\mathbf{x}') \left\{ \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \right. \\
 &\quad \left. \sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}', t; \mathbf{x}, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}', t)}{\partial t} \right\} \delta \rho(\mathbf{x}'), \tag{4.31}
 \end{aligned}$$

and the corresponding Fréchet kernel can be expressed as

$$\begin{aligned}
 K_{d^{sn}}^\rho(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{j,p=1}^3 J_p^{sn}(\mathbf{x}', t) \\
 &\quad \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t}, \tag{4.32}
 \end{aligned}$$

where I have switched the roles of \mathbf{x} and \mathbf{x}' (i.e., $\mathbf{x} \leftrightarrow \mathbf{x}'$).

For the rate-of-relaxation tensor $\Psi_{jklm}(\mathbf{x}, t)$ we have

$$\begin{aligned}
 \delta d^{sn} &= - \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \int d\tau \int_V dV(\mathbf{x}') \\
 &\sum_{j,k,l,m=1}^3 \left[\int dt' \frac{\partial G_{jp}(\mathbf{x}', t-t'; \mathbf{x}, 0)}{\partial x'_k} \frac{\partial u_l^s(\mathbf{x}', t'-\tau)}{\partial x'_m} \delta \Psi_{jklm}(\mathbf{x}', \tau) \right] \\
 &= - \sum_{j,k,l,m=1}^3 \int_V dV(\mathbf{x}') \int d\tau \left\{ \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \right. \\
 &\quad \left. \int dt' \frac{\partial G_{jp}(\mathbf{x}', t-t'; \mathbf{x}, 0)}{\partial x'_k} \frac{\partial u_l^s(\mathbf{x}', t'-\tau)}{\partial x'_m} \right\} \delta \Psi_{jklm}(\mathbf{x}', \tau), \tag{4.33}
 \end{aligned}$$

and the corresponding Fréchet kernel can be expressed as

$$\begin{aligned}
 K_{d^{sn}}^{\Psi_{jklm}}(\mathbf{x}, t) &= - \int_V dV(\mathbf{x}') \int_0^T d\tau \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', \tau) \\
 &\quad \int dt' \frac{\partial G_{jp}(\mathbf{x}, \tau-t'; \mathbf{x}', 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t'-t)}{\partial x_m}, \tag{4.34}
 \end{aligned}$$

where I have made the changes $\mathbf{x} \leftrightarrow \mathbf{x}'$ and $t \leftrightarrow \tau$. If the material is elastic, the Fréchet kernel with respect to $c_{jklm}(\mathbf{x})$ can be expressed as

$$K_{d^{sn}}^{c_{jklm}}(\mathbf{x}) = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m}. \quad (4.35)$$

Combining Eqs. 4.29–4.34, the perturbation of the data functional δd^{sn} is related to the perturbations in the source and structural models, parameterized in terms of the moment tensor density, the mass density and the rate-of-relaxation tensor, as

$$\begin{aligned} \delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{j,k=1}^3 K_{d^{sn}}^{m_{jk}^s}(\mathbf{x}, t) \delta m_{jk}^s(\mathbf{x}, t) \\ &+ \int_V dV(\mathbf{x}) K_{d^{sn}}^\rho(\mathbf{x}) \delta \rho(\mathbf{x}) \\ &+ \int_V dV(\mathbf{x}) \int_0^T dt \sum_{j,k,l,m=1}^3 K_{d^{sn}}^{\Psi_{jklm}}(\mathbf{x}, t) \delta \Psi_{jklm}(\mathbf{x}, t). \end{aligned} \quad (4.36)$$

For elastic models, Eq. 4.36 can be modified as

$$\begin{aligned} \delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{j,k=1}^3 K_{d^{sn}}^{m_{jk}^s}(\mathbf{x}, t) \delta m_{jk}^s(\mathbf{x}, t) \\ &+ \int_V dV(\mathbf{x}) K_{d^{sn}}^\rho(\mathbf{x}) \delta \rho(\mathbf{x}) \\ &+ \int_V dV(\mathbf{x}) \sum_{j,k,l,m=1}^3 K_{d^{sn}}^{c_{jklm}}(\mathbf{x}) \delta c_{jklm}(\mathbf{x}). \end{aligned} \quad (4.37)$$

Note that in Eqs. 4.29–4.35, the Green's tensor G_{jp} and the forward wavefield from the s th source u_l^s are all computed using the reference source and structural model $\tilde{\mathbf{m}}$ and I have dropped the \sim on top of the Green's tensor and the forward wavefield to reduce clutter.

4.2.4 Time-Domain Waveform Differences

Consider the data functional defined in Eq. 4.11, whose WPK is given in Eq. 4.14. This data functional was used in e.g., (Gauthier et al. 1986; Tarantola 1988b). Bring

Eq. 4.14 into Eq 4.32, we obtain the Fréchet kernel of d^{sn} with respect to the density $\rho(\mathbf{x})$, i.e.,

$$\begin{aligned} K_{d^{sn}}^\rho(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{j,p=1}^3 \delta(\mathbf{x}' - \mathbf{x}_r) \delta(t - t_n) \gamma_p \\ &\quad \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t} \\ &= - \sum_{j,p=1}^3 \gamma_p \int_0^T dt \delta(t - t_n) \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t}. \end{aligned} \quad (4.38)$$

Bring Eq. 4.14 into Eqs. 4.34 or 4.35, we obtain the Fréchet kernel of d^{sn} with respect to the rate-of-relaxation or the elastic moduli, i.e.,

$$\begin{aligned} K_{d^{sn}}^{\Psi_{jklm}}(\mathbf{x}, t) &= - \int_V dV(\mathbf{x}') \int_0^T d\tau \sum_{p=1}^3 \delta(\mathbf{x}' - \mathbf{x}_r) \delta(\tau - t_n) \gamma_p \\ &\quad \int dt' \frac{\partial G_{jp}(\mathbf{x}, \tau - t'; \mathbf{x}', 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - t)}{\partial x_m} \\ &= - \sum_{p=1}^3 \gamma_p \int dt' \frac{\partial G_{jp}(\mathbf{x}, t_n - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - t)}{\partial x_m}, \end{aligned} \quad (4.39)$$

$$\begin{aligned} K_{d^{sn}}^{C_{jklm}}(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \delta(\mathbf{x}' - \mathbf{x}_r) \delta(t - t_n) \gamma_p \\ &\quad \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m} \\ &= - \sum_{p=1}^3 \gamma_p \int_0^T dt \delta(t - t_n) \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m}. \end{aligned} \quad (4.40)$$

In the case that the waveform difference is taken on the velocity seismogram instead of the displacement seismogram, we have

$$\begin{aligned} \delta d^{sn} &= \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}, t) \delta \dot{u}_p^s(\mathbf{x}, t) \\ &= - \int_V dV(\mathbf{x}) \int_0^T dt \sum_{p=1}^3 j_p^{sn}(\mathbf{x}, t) \delta u_p^s(\mathbf{x}, t), \end{aligned} \quad (4.41)$$

where the “ $\dot{\cdot}$ ” indicates time derivative and I have applied integration-by-parts at the second equality. Equation 4.41 indicates that we only need to replace the WPK

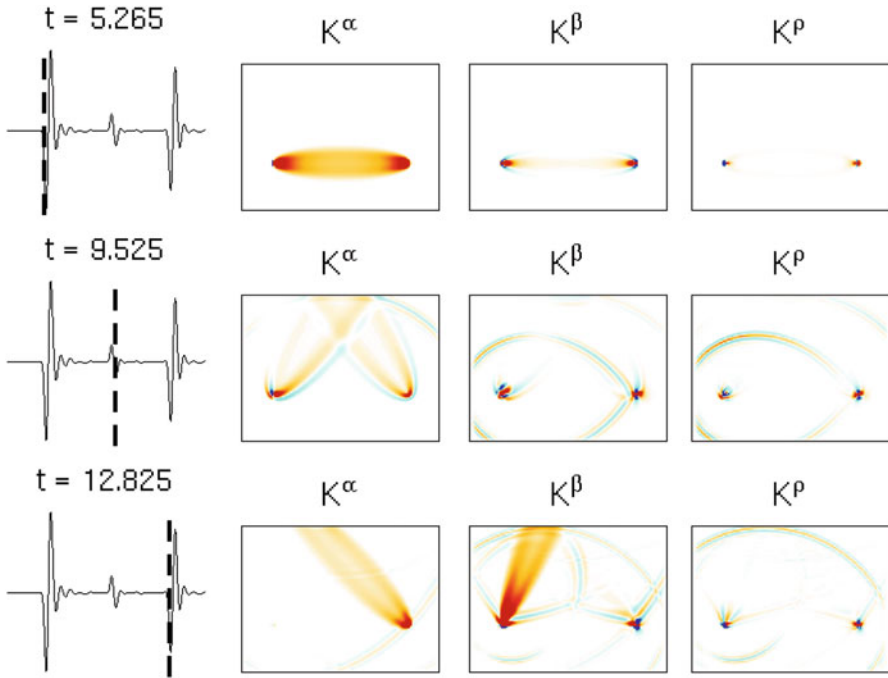


Fig. 4.1 Cross-section views of the data sensitivity kernels for the time-domain waveform differences at three different time samples on the x_1 component of the half-space example. The location of the cross-section in this and the following half-space kernel figures is identical to that in Fig. 3.3. *First column:* synthetic seismogram (black solid lines) and the location of the time sample (vertical dash lines and title of the plot); *second column:* data sensitivity kernel with respect to the relative perturbation in P-velocity $\delta\alpha/\alpha$; *third column:* data sensitivity kernel with respect to the relative perturbation in S-velocity $\delta\beta/\beta$; *fourth column:* data sensitivity kernel with respect to the relative perturbation in density $\delta\rho/\rho$. The locations of the source 100001 and the receiver IN.RC01 are the same as those shown in Fig. 3.3 and the cross-section views are perpendicular to the source-receiver plane at $y=40$. The synthetic seismogram shown on the first column is identical to that in Fig. 2.12a.

for the displacement seismogram $J_p^n(\mathbf{x}, t)$ with $-J_p^n(\mathbf{x}, t)$ to obtain the WPK for the same data functional applied on the corresponding velocity seismogram.

Figure 4.1 shows cross-section views of the data sensitivity kernels at three different time samples on the x_1 -component synthetic seismogram of the half-space example. The computation of the synthetic seismograms, the forward wavefields for the explosive source 100001 and the RSGTs for the receiver IN.RC01 were discussed in Sects. 2.2.8 and 3.4.3. The x_1 -component synthetic velocity seismogram shown on the first column of Fig. 4.1 is identical to that shown in Fig. 2.12a. The data sensitivity kernels shown in Fig. 4.1 are not for the rate-of-relaxation tensor or the elastic moduli as shown in Eqs. 4.39 and 4.40, but for relative perturbations in the P- and S-velocities, $\delta\alpha/\alpha$ and $\delta\beta/\beta$, and the relative perturbation of the density $\delta\rho/\rho$. The conversion from the kernels in Eqs. 4.38–4.40 to those shown in Fig. 4.1

will be discussed in Sect. 4.3. The data sensitivity kernel is a measure of the *sensitivity* of the data functional to the model parameters. If the kernel with respect to some model parameter is zero at a certain spatial location, any perturbations of the model parameter at that location will not change the misfit measurement produced by the data functional. On the other hand, if the kernel with respect to some model parameter has a nonzero value at a spatial location, perturbations to the model parameter at that location will change the misfit measurement produced by the data functional. Therefore the spatial pattern of the kernel depends upon both the data functional and the particular model parameter. The first time sample (top row, Fig. 4.1) is located on the direct-arriving P-wave, therefore the sensitivity with respect to P-velocity perturbation $\delta\alpha/\alpha$ is much stronger than the sensitivity with respect to S-velocity $\delta\beta/\beta$. The second time sample (middle row, Fig. 4.1) lies on the surface-reflected P-wave (PP) and the sensitivity kernel with respect to $\delta\alpha/\alpha$ clearly indicates the propagation path of this arrival. The third time sample (bottom row, Fig. 4.1) lies on the P-to-S converted wave. The leg from the source to the free surface is a P-wave, therefore the sensitivity to $\delta\alpha/\alpha$ is much stronger (second column, bottom row, Fig. 4.1). The leg from the free surface to the receiver is an S-wave, therefore the sensitivity to $\delta\beta/\beta$ is much stronger (third column, bottom row, Fig. 4.1).

4.2.5 Frequency-Domain Waveform Differences

It is also possible to measure the waveform differences in the frequency-domain (e.g., Pratt et al. 1998; Pratt 1999; Sirgue and Pratt 2004; Brenders and Pratt 2007). The Fourier transform is usually applied on a selected segment of the time-domain seismogram. The time window used for isolating the segment can be denoted as $W_n(t)$, the data functional at frequency ω_n can be expressed as

$$\begin{aligned}
 d^{sn} &= D_n(\tilde{\mathbf{u}}^s(\mathbf{x}, t)) [\mathbf{u}^s(\mathbf{x}, t)] \\
 &= \int_V dV(\mathbf{x}) \int dt \sum_{p=1}^3 \delta(\mathbf{x} - \mathbf{x}_r) e^{-i\omega_n t} \sum_{i=1}^3 \gamma_i \delta_{ip} W_n(t) [u_p^s(\mathbf{x}, t) - \tilde{u}_p^s(\mathbf{x}, t)] \\
 &= \sum_{i=1}^3 \gamma_i \hat{u}_i^s(\mathbf{x}_r, \omega_n) - \sum_{i=1}^3 \gamma_i \hat{\tilde{u}}_i^s(\mathbf{x}_r, \omega_n), \tag{4.42}
 \end{aligned}$$

where $\hat{u}_i^s(\mathbf{x}_r, \omega_n)$ and $\hat{\tilde{u}}_i^s(\mathbf{x}_r, \omega_n)$ are Fourier transforms of the windowed target and reference seismograms, i.e.,

$$\hat{u}_i^s(\mathbf{x}_r, \omega_n) = \int dt e^{-i\omega_n t} [W_n(t) u_i^s(\mathbf{x}_r, t)], \tag{4.43}$$

$$\hat{\tilde{u}}_i^s(\mathbf{x}_r, \omega_n) = \int dt e^{-i\omega_n t} [W_n(t) \tilde{u}_i^s(\mathbf{x}_r, t)]. \tag{4.44}$$

This data functional is also linear with respect to the time-domain target wavefield and the corresponding WPK is given by

$$J_p^{sn}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_r) e^{-i\omega_n t} W_n(t) \gamma_p. \quad (4.45)$$

Bring Eq. 4.45 into Eqs. 4.32, 4.34 and 4.35, we obtain

$$\begin{aligned} K_{d^{sn}}^\rho(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{j,p=1}^3 \delta(\mathbf{x} - \mathbf{x}_r) e^{-i\omega_n t} W_n(t) \gamma_p \\ &\quad \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t} \\ &= - \sum_{j,p=1}^3 \gamma_p \int_0^T dt e^{-i\omega_n t} W_n(t) \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial t} * \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t}, \end{aligned} \quad (4.46)$$

$$\begin{aligned} K_{d^{sn}}^{\Psi_{jklm}}(\mathbf{x}, t) &= - \int_V dV(\mathbf{x}') \int_0^T d\tau \sum_{p=1}^3 \delta(\mathbf{x}' - \mathbf{x}_r) e^{-i\omega_n \tau} W_n(\tau) \gamma_p \\ &\quad \int dt' \frac{\partial G_{jp}(\mathbf{x}, \tau - t'; \mathbf{x}', 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - t)}{\partial x_m} \\ &= - \sum_{p=1}^3 \gamma_p \int_0^T d\tau e^{-i\omega_n \tau} W_n(\tau) \\ &\quad \int dt' \frac{\partial G_{jp}(\mathbf{x}, t_n - t'; \mathbf{x}_r, 0)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, t' - t)}{\partial x_m}, \end{aligned} \quad (4.47)$$

$$\begin{aligned} K_{d^{sn}}^{c_{jklm}}(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \delta(\mathbf{x}' - \mathbf{x}_r) e^{-i\omega_n t} W_n(t) \gamma_p \\ &\quad \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m} \\ &= - \sum_{p=1}^3 \gamma_p \int_0^T dt e^{-i\omega_n t} W_n(t) \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m}. \end{aligned} \quad (4.48)$$

An example of the WPK (Eq. 4.45) for the half-space example is shown in Fig. 4.2. It is for the direct-arriving P-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01. The time window is a Tukey (tapered cosine) window (Harris 1978), which is 3-s (100 samples) long with the first and last 0.375 s (12.5%) equal to parts of a cosine function. The frequency ω_n is fixed at 2 Hz (4π rad/s). If the frequency-domain waveform difference is measured on velocity seismograms instead of displacement seismograms, we can apply Eq. 4.41 and replace the WPK given in Eq. 4.45 $J_p^{sn}(\mathbf{x}, t)$ with $-J_p^{sn}(\mathbf{x}, t)$ and obtain the corresponding data sensitivity kernels.

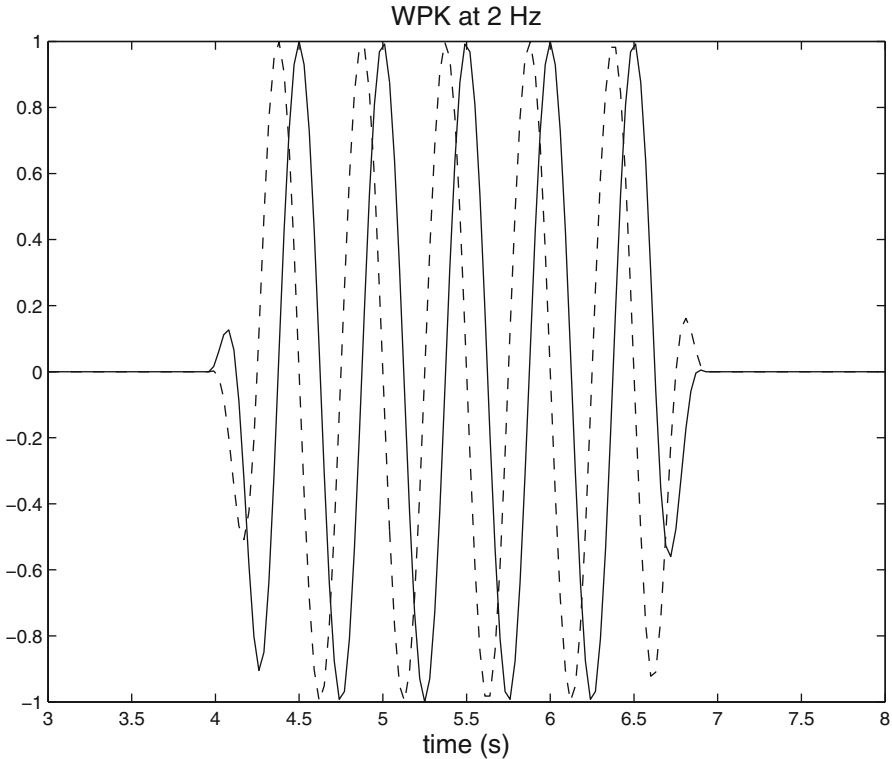


Fig. 4.2 The WPK, as given in Eq. 4.45, for the direct-arriving P-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01. *Solid line*: the real part; *dash line*: the imaginary part. The time window $W_n(t)$ is a Tukey window and the frequency $\omega_n = 4\pi$.

Examples of the data sensitivity kernels for the direct-arriving P, the surface-reflected PP and the surface-converted PS waves on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01 in the half-space example are shown in Figs. 4.3, 4.4 and 4.5. Spatial variations of the sensitivity depend upon both the propagation paths of the different seismic arrivals and also the frequency ω_n . Different from the frequency-domain Born kernels shown in Fig. 3.7, the data sensitivity kernels shown in Figs. 4.3, 4.4 and 4.5 have compact spatial extent, which is a result of the finite-width time window $W_n(t)$ applied on the synthetic seismogram. Consider a scatterer located very far off the propagation path of, say, the direct-arriving P-wave, since the time window has a finite duration (3 s in our example), the wave scattered by this scatterer will not be able to reach the receiver within the finite-width time window because the scattered wave has to travel a much longer distance than the direct-arriving P-wave, therefore this scatterer has zero contribution to the sensitivity of the direct-arriving P-wave. The frequency-domain Born kernels shown in Fig. 3.7 do not have the restrictions imposed by the

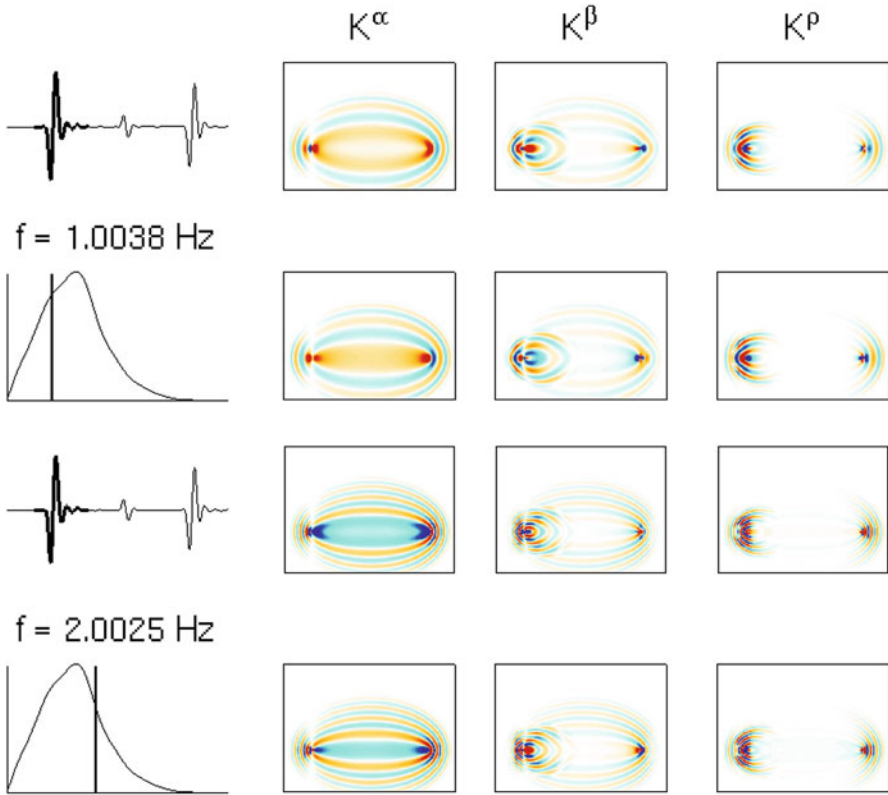


Fig. 4.3 Cross-section views of the real and imaginary parts of the data sensitivity kernels for the frequency-domain waveform differences measured at two different frequencies on the P-wave of the x_1 -component synthetic velocity seismogram from the source 100001 to the receiver IN.RC01. *First row*: the real-part kernels at 1 Hz; *second row*: the imaginary-part kernels at 1 Hz; *third row*: the real-part kernels at 2 Hz; *bottom row*: the imaginary-part kernels at 2 Hz. The *first column* shows the time-domain synthetic seismogram (row 1 and 3) and the amplitude spectrum of the selected waveform (row 2 and 4). The frequencies at which the kernels were computed are indicated with the vertical lines on the amplitude spectrum. The amplitudes of the kernels shown in Figs. 4.3, 4.4 and 4.5 have been normalized by the frequency and the maximum amplitudes of the selected waveforms.

finite-width time window, therefore every scatterer inside the volume can contribute to the sensitivity at a single frequency.

4.2.6 Broadband Cross-Correlation Measurements

The time shift between an isolated waveform on the observed displacement seismogram $\bar{u}_i^s(\mathbf{x}_r, t)$ and the corresponding waveform on the reference synthetic

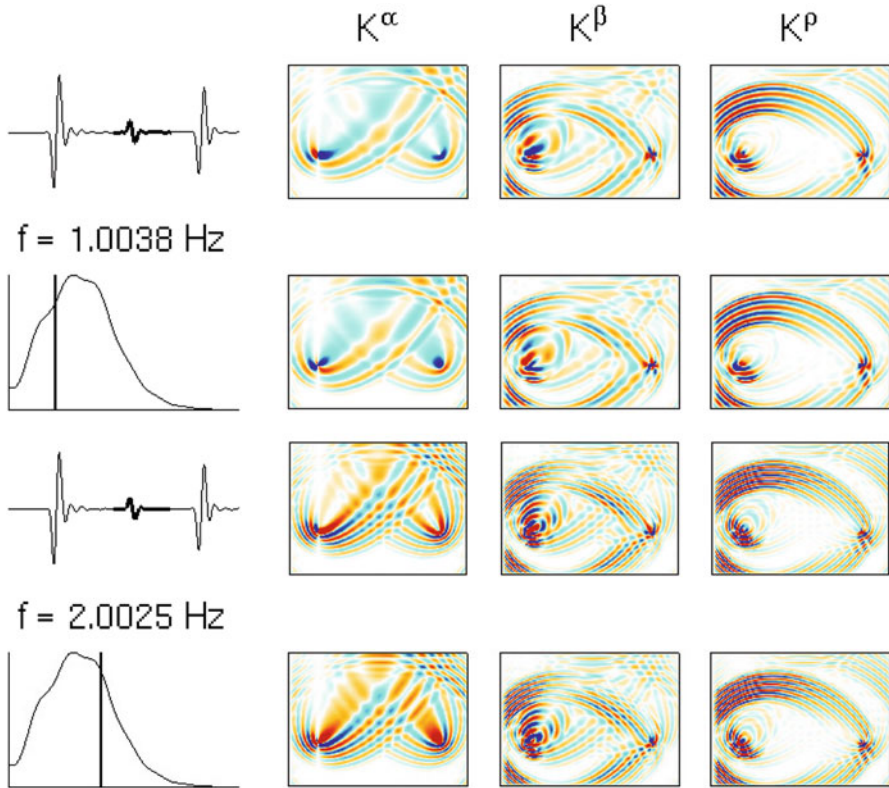


Fig. 4.4 The format is identical to that in Fig. 4.3. The data sensitivity kernels shown here are for the PP-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01.

displacement seismogram $\tilde{u}_i^s(\mathbf{x}_r, t)$ can be estimated by maximizing the cross-correlation between the two waveforms (e.g., Woodward and Masters 1991). The same procedure can also be applied on velocity seismograms. An example of the measurement process is illustrated in Fig. 4.6 using two velocity seismograms. To make the misfit measurement, one first isolates the waveforms on the reference synthetic seismogram and the observed seismogram through time-domain windowing and obtains the windowed waveforms $W_n(t)\dot{\tilde{u}}_i^s(\mathbf{x}_r, t)$ and $W_n(t)\dot{\tilde{u}}_i^o(\mathbf{x}_r, t)$, where the dot denotes the time derivative. The auto-correlation of $W_n(t)\dot{\tilde{u}}_i^s(\mathbf{x}_r, t)$ has its maximum at delay time zero, i.e., $\Delta t = 0$. The maximum of the cross-correlation between the windowed observed velocity $W_n(t)\dot{\tilde{u}}_i^o(\mathbf{x}_r, t)$ and the windowed reference synthetic velocity $W_n(t)\dot{\tilde{u}}_i^s(\mathbf{x}_r, t)$ gives the broadband cross-correlation delay time ΔT^{sn} . For the example shown in Fig. 4.6, the observed P-wave arrives earlier than the reference synthetic P-wave, therefore the broadband cross-correlation delay time ΔT^{sn} is negative. For band-limited signals, the broadband cross-correlation

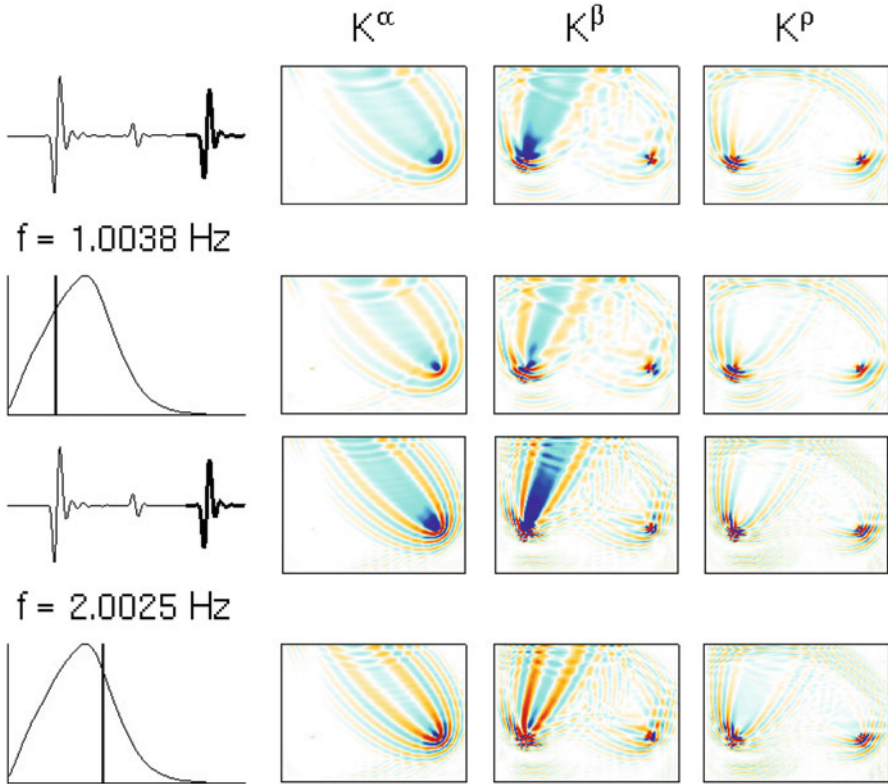


Fig. 4.5 The format is identical to that in Fig. 4.3. The data sensitivity kernels shown here are for the PS-wave on the x_1 -component synthetic seismogram from the source 100001 to the receiver IN.RC01.

delay time approximates the phase delay at the dominant frequency of the signal (e.g., Sipkin and Jordan 1980; Gee and Jordan 1992).

The derivation of the WPK for ΔT^{sn} measured on displacement seismograms has been given in e.g., (Luo and Shuster 1991; Dahlen et al. 2000). Denoting the cross-correlation between $W_n(t)\tilde{u}_i^s(\mathbf{x}_r, t)$ and $W_n(t)u_i^s(\mathbf{x}_r, t)$ as $C_{u\tilde{u}}(t)$, at the maximum of the cross-correlation $t = \delta T$ we have

$$\left. \frac{dC_{u\tilde{u}}(t)}{dt} \right|_{t=\delta T} = 0. \quad (4.49)$$

In the frequency-domain, Eq. 4.49 can be written as

$$\int d\omega(i\omega)\hat{u}^*(\omega)\hat{u}(\omega)e^{i\omega\delta T} = 0, \quad (4.50)$$

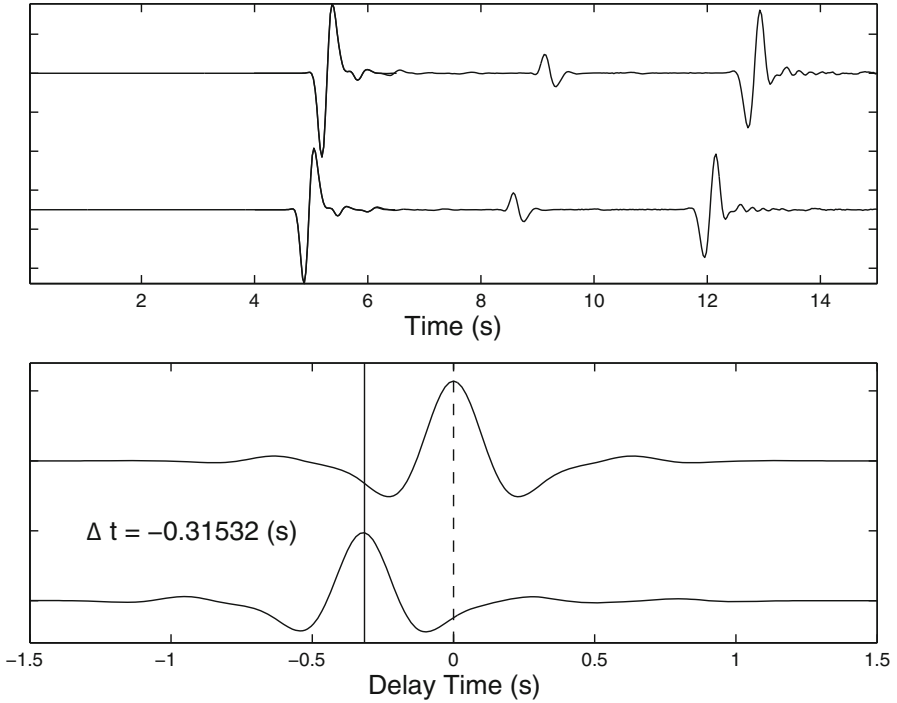


Fig. 4.6 Upper plot: the reference synthetic velocity seismogram (top) and the observed velocity seismogram (bottom) and the waveforms selected for making the cross-correlation measurement (black thick lines). Lower plot: the auto-correlation of the selected waveform on the reference synthetic (top) and the cross-correlation between the selected waveforms on the reference synthetic and the observed waveforms (bottom). The maximum on the auto-correlation, which is located at $\Delta t = 0$ s, is marked using a black vertical dash line. The maximum on the cross-correlation, which is at around $\Delta t = -0.31532$ s, is marked using a black vertical solid line.

where

$$\hat{\tilde{u}}(\omega) = \int dt e^{-i\omega t} W_n(t) \tilde{u}_i^s(\mathbf{x}_r, t), \quad \hat{u}(\omega) = \int dt e^{-i\omega t} W_n(t) u_i^s(\mathbf{x}_r, t) \quad (4.51)$$

are the Fourier transforms of the windowed displacement seismograms. Consider

$$\hat{u}(\omega) = \hat{\tilde{u}}(\omega) + \delta\hat{u}(\omega) \quad (4.52)$$

and bring Eq. 4.52 into to Eq. 4.50, we obtain

$$\int d\omega (i\omega) \hat{\tilde{u}}^*(\omega) \left[\hat{\tilde{u}}(\omega) + \delta\hat{u}(\omega) \right] e^{i\omega\delta T} = 0. \quad (4.53)$$

In Eq. 4.53, we can replace $e^{i\omega\delta T}$ with its first-order Taylor expansion $1 + i\omega\delta T$ and obtain

$$\begin{aligned} & \int d\omega(i\omega)\hat{u}^*(\omega) \left[\hat{u}(\omega) + \delta\hat{u}(\omega) \right] (1 + i\omega\delta T) = 0 \\ & \Rightarrow \int d\omega(i\omega)\hat{u}^*(\omega)\hat{u}(\omega) + \int d\omega(i\omega)\hat{u}^*(\omega)\delta\hat{u}(\omega) \\ & + \delta T \int d\omega(i\omega)^2\hat{u}^*(\omega)\hat{u}(\omega) + \int d\omega(i\omega)^2\hat{u}^*(\omega)\delta\hat{u}(\omega)\delta T = 0. \end{aligned} \quad (4.54)$$

The first term on the LHS of Eq. 4.54 can be re-written as

$$\int d\omega(i\omega)\hat{u}^*(\omega)\hat{u}(\omega) = \int d\omega(i\omega)\hat{u}^*(\omega)\hat{u}(\omega)e^{i\omega 0} = \left. \frac{dC_{\hat{u}\hat{u}}(t)}{dt} \right|_{t=0} = 0, \quad (4.55)$$

which is the derivative of the auto-correlation of the windowed reference synthetic waveform at delay time zero. Bring Eq. 4.55 into Eq. 4.54 and drop the second-order term on the LHS of Eq. 4.54, we obtain

$$\begin{aligned} \delta T &= -\frac{\int d\omega(i\omega)\hat{u}^*(\omega)\delta\hat{u}(\omega)}{\int d\omega(i\omega)^2\hat{u}^*(\omega)\hat{u}(\omega)} = -\frac{\int d\omega \left[i\omega\hat{u}(\omega) \right]^* \delta\hat{u}(\omega)}{\int d\omega \left[i\omega\hat{u}(\omega) \right]^* \left[i\omega\hat{u}(\omega) \right]} \\ &= -\frac{\int dt \dot{\hat{u}}(t)\delta u(t)}{\int dt \dot{\hat{u}}^2(t)}, \end{aligned} \quad (4.56)$$

where the dot “ $\dot{}$ ” denotes the time derivative and I have used the Parseval’s theorem (Eq. 3.101) in the last equality. Equation 4.56 shows that the WPK for the broadband cross-correlation delay time is proportional to the reference synthetic velocity seismogram and is given by

$$J_p^{sn}(\mathbf{x}, t) = -\delta(\mathbf{x} - \mathbf{x}_r) \frac{\dot{\hat{u}}_i^s(\mathbf{x}, t)}{\int dt \left[\dot{\hat{u}}_i^s(\mathbf{x}, t) \right]^2} \gamma_p, \quad (4.57)$$

where $\dot{\hat{u}}_i^s(\mathbf{x}, t)$ is the time derivative of the windowed reference synthetic displacement (i.e., synthetic velocity) on the i th component. Note that here the direction of the i th component can have oblique angles to the coordinate axes (e.g., the radial or transverse direction) and γ_p is simply the cosine of the angle between the direction of the i th component and the direction of the p th coordinate axis. As discussed in Sect. 4.2.2, when the RGTs used for constructing the kernels are filtered with a source-time function $h(t)$, we need to filter the reference synthetic seismogram with $h(t)$ before using it in constructing the WPK in Eq. 4.57.

A similar analysis can be applied on the cross-correlation delay-time measurement made on the velocity seismograms (e.g., Fig. 4.6). Denote the windowed

velocity seismograms as $v(t)$ and $\tilde{v}(t)$ and their Fourier transforms as $\hat{v}(\omega)$ and $\hat{\tilde{v}}(\omega)$, Eq. 4.56 can be written in terms of the velocity seismograms as

$$\delta T = -\frac{\int d\omega \left[i\omega \hat{v}(\omega) \right]^* \delta \hat{v}(\omega)}{\int d\omega \left[i\omega \hat{\tilde{v}}(\omega) \right]^* \left[i\omega \hat{v}(\omega) \right]} = -\frac{\int dt \dot{\tilde{v}}(t) \delta v(t)}{\int dt \dot{\tilde{v}}^2(t)}. \quad (4.58)$$

Since $\delta v(t) = \delta \dot{u}(t)$ and $\dot{\tilde{v}}(t)$ is compactly supported, apply integration-by-parts on the numerator of Eq. 4.58, we obtain

$$\delta T = -\frac{\int dt \dot{\tilde{v}}(t) \delta v(t)}{\int dt \dot{\tilde{v}}^2(t)} = \frac{\int dt \ddot{\tilde{v}}(t) \delta u(t)}{\int dt \dot{\tilde{v}}^2(t)}, \quad (4.59)$$

and the WPK in Eq. 4.57 can be changed accordingly.

The amplitude anomaly of a selected waveform can also be estimated using the cross-correlation technique (e.g., Ritsema et al. 2002; Tromp et al. 2005) and the WPK for the relative amplitude anomaly was derived in (Dahlen and Baig 2002). The amplitude misfit between the observed waveform and the reference synthetic waveform is defined as

$$\Delta U^{sn} = \frac{\bar{A} - \tilde{A}}{\tilde{A}}, \quad (4.60)$$

where the amplitudes are defined as

$$\bar{A} = \sqrt{\frac{1}{t'_n - t_n} \int \bar{u}^2(t) dt}, \quad (4.61)$$

$$\tilde{A} = \sqrt{\frac{1}{t'_n - t_n} \int \tilde{u}^2(t) dt}. \quad (4.62)$$

Here $\bar{u}(t)$ and $\tilde{u}(t)$ are observed and reference synthetic displacement seismograms multiplied by a time window $W_n(t)$ with its support on $[t_n, t'_n]$. Consider the target displacement seismogram $u(t) = \tilde{u}(t) + \delta u(t)$, the perturbation to the relative amplitude anomaly can be expressed as

$$\delta U^{sn} = \sqrt{\frac{\int [\tilde{u}(t) + \delta u(t)]^2 dt}{\int \tilde{u}^2(t) dt}} - 1. \quad (4.63)$$

Ignore the second-order term in Eq. 4.63, we have

$$\delta U^{sn} = \sqrt{1 + 2 \frac{\int \tilde{u}(t) \delta u(t) dt}{\int \tilde{u}^2(t) dt}} - 1. \quad (4.64)$$

Expand the first term on the RHS of Eq. 4.64 in Taylor series and truncate to first order, we obtain

$$\delta U^{sn} = \frac{\int \tilde{u}(t) \delta u(t) dt}{\int \tilde{u}^2(t) dt}. \quad (4.65)$$

The WPK for the broadband cross-correlation amplitude anomaly is then proportional to the reference synthetic displacement, i.e.,

$$J_p^{sn}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_r) \frac{\tilde{u}_i^s(\mathbf{x}, t)}{\int dt [\tilde{u}_i^s(\mathbf{x}, t)]^2} \gamma_p. \quad (4.66)$$

The same analysis can also be applied to velocity seismograms, and Eq. 4.65 can be modified to

$$\delta U^{sn} = \frac{\int \tilde{v}(t) \delta v(t) dt}{\int \tilde{v}^2(t) dt}. \quad (4.67)$$

Apply integration-by-parts on the numerator of Eq. 4.67, we obtain

$$\delta U^{sn} = - \frac{\int \dot{\tilde{v}}(t) \delta u(t) dt}{\int \tilde{v}^2(t) dt}. \quad (4.68)$$

The WPK in Eq. 4.66 can be modified accordingly. Note that the definition of the amplitude misfit data functional (Eq. 4.60) is different from that used in (Zhao et al. 2005, 2006). In particular, Eq. 4.65 above differs from Eq. (3) in (Zhao et al. 2005) by a factor of $-1/\omega_a$, where ω_a is the dominant angular frequency of the waveform. One consequence of this difference is that the sign of the amplitude data sensitivity kernels shown in Fig. 4.8 here is different from that shown in (Zhao et al. 2005, 2006).

The WPKs for the broadband cross-correlation delay-time and amplitude anomaly computed using Eqs. 4.59 and 4.68 for the three arrivals on the x_1 -component of the reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown in Fig. 4.7. The corresponding data sensitivity kernels are shown in Fig. 4.8. The kernels for cross-correlation delay-time measurements show the counter-intuitive “banana-doughnut” behavior (i.e., vanishing sensitivity on the ray path and nonzero sensitivity off the ray path) observed and explained in e.g., (Marquering et al. 1999; Dahlen et al. 2000; Hung et al. 2000; Zhao et al. 2000).

The shape of the data sensitivity kernels depend upon several factors (Fig. 4.9). For a fixed forward wavefield from the source, changing the source-time function (filter) used for computing the RGTs can change the shape of the kernel. Kernels shown in the top row in Fig. 4.9 were computed using the RGT whose source-time function was given by the sixth-order low-pass Butterworth filter with the corner at 1.5 Hz (Fig. 2.11) and have a wider first Fresnel zone due to a longer wavelength. Kernels shown in the center and bottom rows in Fig. 4.9 were computed using the RGT whose source-time function was given by a fourth-order low-pass Butterworth filter with the corner at 3.5 Hz and have a narrower first Fresnel zone due to a shorter wavelength. Whether the measurement is made on the velocity seismogram (top and center rows in Fig. 4.9) or the displacement seismogram (bottom row in Fig. 4.9) can have effects on the shape and strength of the side-bands of the kernels. For the measurement made on the velocity seismogram, the WPK is given by Eq. 4.59, while for the measurement made on the displacement seismogram, the WPK is given by Eq. 4.57.

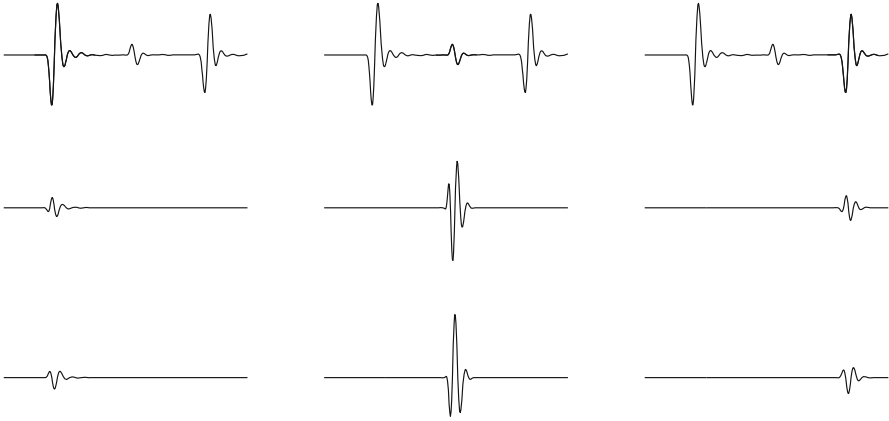


Fig. 4.7 The WPKs for the broadband cross-correlation delay-time (*middle row*) and the broadband cross-correlation amplitude anomaly (*bottom row*) for the direct-arriving P-wave (*left column*), the PP-wave (*middle column*) and the PS-wave (*right column*) on the x_1 -component velocity synthetic seismogram from 100001 to IN.RC01 in the half-space example. The waveforms used for computing the WPKs of each column are plotted in *thick black lines* on the *top row*. The synthetic seismograms shown on the *top row* are identical to that shown in Fig. 2.12a.

4.2.7 Frequency-Dependent Phase and Amplitude Misfits

For band-limited signals, the broadband cross-correlation delay time and amplitude anomaly provide good estimates around the dominant frequency of the waveforms, but they do not capture the differences in the details of the waveform shapes very well. One possibility to better account for discrepancies in waveforms for broadband signals is to measure the phase and amplitude differences at multiple frequencies across the frequency band. Consider the Fourier transforms of the windowed seismograms in Eq. 4.51, the reference synthetic waveform can be mapped into the corresponding observed waveform by an exponential operator

$$\hat{\tilde{u}}(\omega) = \hat{\tilde{u}}(\omega)e^{-i\omega[\Delta\tau_p(\omega) - i\Delta\tau_q(\omega)]}, \quad (4.69)$$

where the frequency-dependent *phase-delay time* $\Delta\tau_p(\omega)$ and *amplitude-reduction time* $\Delta\tau_q(\omega)$ ² can completely capture the waveform discrepancies between the observed waveform $\tilde{u}(t)$ and the corresponding reference synthetic waveform $u(t)$ within the time window $W_n(t)$.

There are a number of methods for estimating $\Delta\tau_{p,q}(\omega)$ at different frequencies. The Generalized Seismological Data Functionals (GSDF) is an approach for obtaining approximations of $\Delta\tau_{p,q}(\omega)$ through time-windowing and narrow-band

² An increase in $\Delta\tau_q$ leads to a decrease in waveform amplitude and $\Delta\tau_q$ has the unit of time, therefore it is named the amplitude-reduction time.

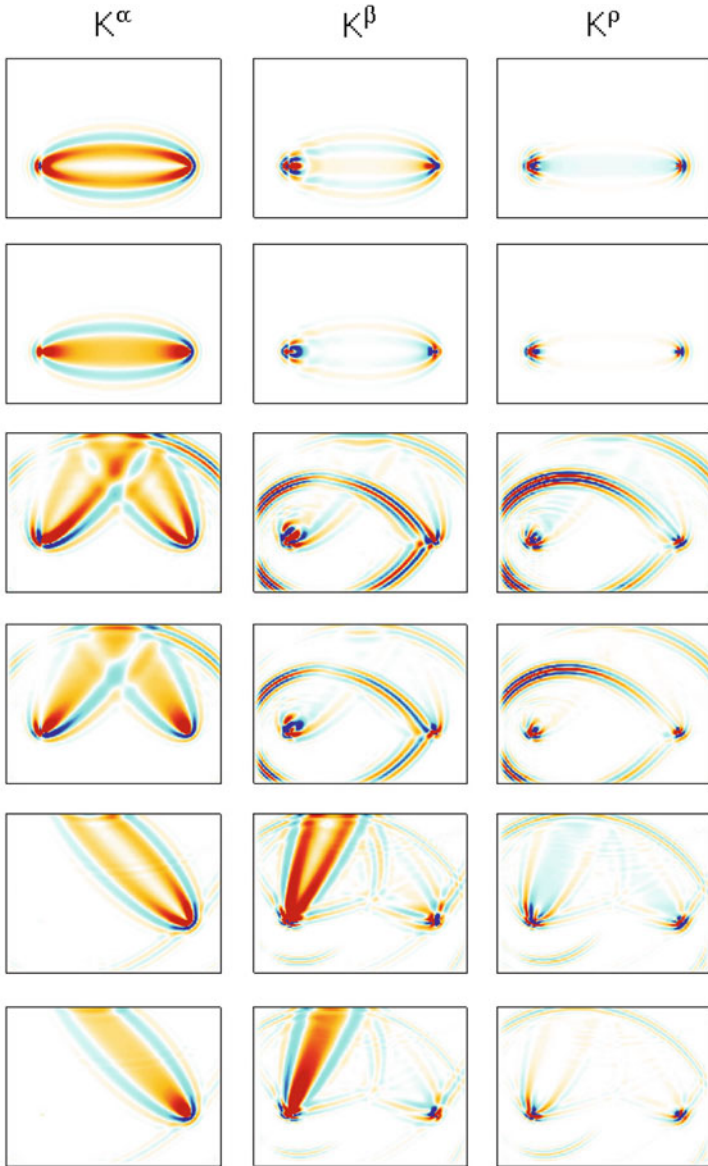


Fig. 4.8 The data sensitivity kernels of the broadband cross-correlation delay time (row 1, 3, 5) and the broadband cross-correlation amplitude anomaly (row 2, 4, 6) for the direct-arriving P-wave (row 1-2), the PP-wave (row 3-4) and the PS-wave (row 5-6) on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example. The synthetic seismogram and the WPKs used for computing these kernels are shown in Fig. 4.7. The kernels shown here do not have the numerical artifacts discussed in (Zhao et al. 2005).

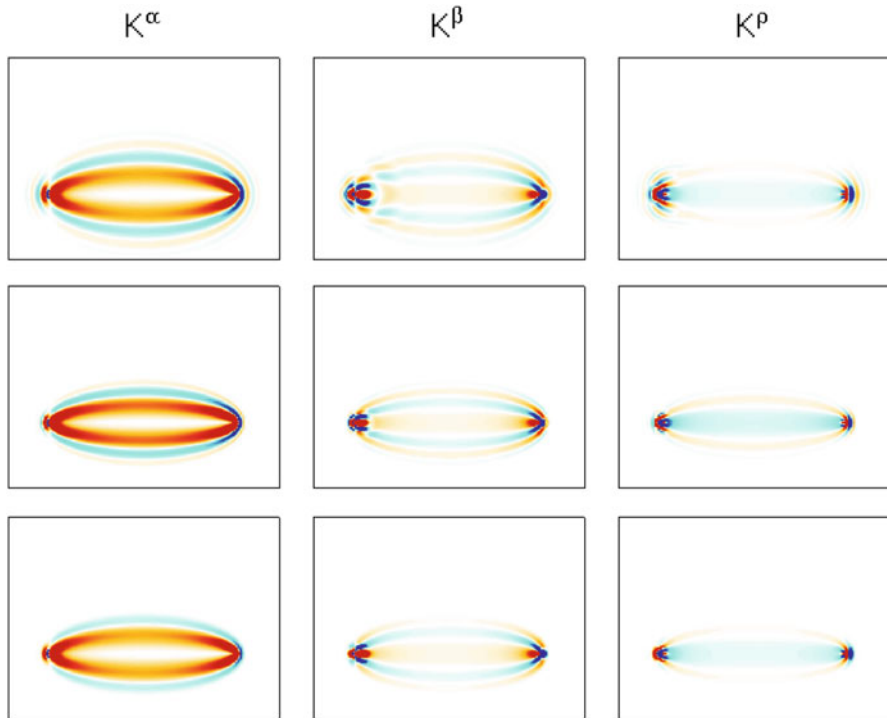


Fig. 4.9 Comparisons of data sensitivity kernels for the broadband cross-correlation delay time measured for the P-wave on the velocity seismograms (*top* and *center* rows) and the displacement seismogram (*bottom* row), filtered with a low-pass Butterworth filter with corner frequency at 1.5 Hz (*top* row) and 3.5 Hz (*center* and *bottom* rows). From *left* to *right*: kernels with respect to relative perturbations in P- and S-velocity and density. The source-receiver setup is identical to that in Fig. 4.8.

filtering of the waveform cross-correlations. It was formalized in (Gee and Jordan 1992) and later extended to broadband waveforms in (Chen et al. 2010a). The procedure for making GSDF misfit measurements was documented in (Gee and Jordan 1992) and (Chen et al. 2007b) and the WPKs for GSDF measurements were derived in (Chen et al. 2010a). The GSDF-type analysis was extended to the continuous time-frequency case using the Gabor transform in (Fichtner et al. 2008), in which the frequency-dependent phase and amplitude differences are quantified using phase and envelope misfit functions that are continuous in the time-frequency domain. A technique based on the continuous wavelet transform was presented in (Holschneider et al. 2005; Kulesh et al. 2005). This technique extracts frequency-dependent phase and amplitude misfits by modeling the deformation of the wavelet transforms of the seismograms. The continuous wavelet transform allows localization in the time-frequency domain for selective fitting of portions of the seismogram (Lee and Chen 2013). In (Bozdağ et al. 2011) the Hilbert transform was applied to derive the Fréchet kernels of the instantaneous phase and envelope misfit measurements based

on the adjoint-wavefield method. The literature listed above is far from exhaustive and new techniques are being actively developed and automated (e.g., Maggi et al. 2009; Lee and Chen 2013).

In the software F3DWT, I adopted an approach similar to that in (Holschneider et al. 2005). For a pair of frequency-domain observed and synthetic waveforms $\hat{u}(\omega)$ and $\hat{\tilde{u}}(\omega)$, we can parameterize their frequency-dependent phase difference $\Delta\phi(\omega)$ and amplitude discrepancy $\Delta\psi(\omega)$ using B-spline functions $B(\omega)$ of order four (e.g., Gallier 2000),

$$\Delta\phi(\omega, \mathbf{p}) = \sum_{i=0}^{P-1} p_i B\left(\frac{\omega - \omega_0 - i\Delta\phi}{\Delta\phi} + 3\right), \quad (4.70)$$

$$\Delta\psi(\omega, \mathbf{q}) = \sum_{j=0}^{Q-1} q_j B\left(\frac{\omega - \omega_0 - j\Delta\psi}{\Delta\psi} + 3\right). \quad (4.71)$$

Here the phase difference is related to $\Delta\tau_p(\omega)$ through

$$\Delta\tau_p(\omega) = \frac{\Delta\phi(\omega)}{\omega}, \quad (4.72)$$

and the amplitude discrepancy is related to $\Delta\tau_q(\omega)$ through

$$\Delta\tau_q(\omega) = \frac{\Delta\psi(\omega)}{\omega}. \quad (4.73)$$

The frequency band, in which Eqs. 4.70 and 4.71 are applied, is $[\omega_0, \omega_1]$ and we have

$$\Delta\phi = \frac{\omega_1 - \omega_0}{P - 3}, \quad \Delta\psi = \frac{\omega_1 - \omega_0}{Q - 3}. \quad (4.74)$$

Finding a set of acceptable model parameters $\mathbf{p} = (p_0, p_1, \dots, p_{P-1})$ and $\mathbf{q} = (q_0, q_1, \dots, q_{Q-1})$ is equivalent to solving an optimization problem, in which we minimize the objective function

$$\chi^2(\mathbf{p}, \mathbf{q}) = \int_0^T \left\{ \tilde{u}(t) - \mathcal{F}^{-1} \left[e^{-i\Delta\phi(\omega, \mathbf{p}) - \Delta\psi(\omega, \mathbf{q})} \hat{\tilde{u}}(\omega) \right] (t) \right\}^2 dt \quad (4.75)$$

by searching for the optimal \mathbf{p} and \mathbf{q} . Here $\mathcal{F}^{-1}[\]$ denotes the inverse Fourier transform. The Levenberg-Marquardt algorithm (i.e., damped least-squares) (Marquardt 1963; Nocedal and Wright 2006; Press 2007) is adopted here to solve this optimization problem. Once the optimal \mathbf{p} and \mathbf{q} are found, the frequency-dependent phase difference $\Delta\phi(\omega)$ and amplitude discrepancy $\Delta\psi(\omega)$ can be obtained from Eqs. 4.70 and 4.71 and the phase-delay time and amplitude reduction time $\Delta\tau_p(\omega)$ and $\Delta\tau_q(\omega)$ can be obtained using Eqs. 4.72 and 4.73. The frequency-dependent group-delay time $\Delta\tau_g(\omega)$ can also be recovered from the phase difference $\Delta\phi(\omega)$.

The Taylor expansion of the phase difference around a given frequency ω_n can be expressed as

$$\Delta\phi(\omega) = \omega_n \Delta\tau_p(\omega_n) + (\omega - \omega_n) \Delta\tau_g(\omega_n) + \dots \quad (4.76)$$

and the group-delay time at ω_n can be expressed as

$$\Delta\tau_g(\omega_n) = \frac{d\Delta\phi}{d\omega}(\omega_n). \quad (4.77)$$

Therefore the frequency-dependent group-delay $\Delta\tau_g(\omega)$ can be obtained through numerical differentiation of the frequency-dependent phase difference.

In Fig. 4.10 I illustrate the measurement process using a seismogram from the half-space example. The reference synthetic waveform is the PS-wave on the x_1 -component from source 100001 to receiver IN.RC01 and the “observed” seismogram was computed using a velocity model different from the uniform half-space model constructed in Listing 2.4 (Fig. 4.10a). Once we have found the optimal \mathbf{p} and \mathbf{q} parameters, we can apply a phase and amplitude correction, $\Delta\phi(\omega, \mathbf{p})$ and $\Delta\psi(\omega, \mathbf{q})$, to the reference synthetic waveform through

$$u(t) = \mathcal{F}^{-1} \left[e^{-i\Delta\phi(\omega, \mathbf{p}) - \Delta\psi(\omega, \mathbf{q})} \hat{u}(\omega) \right] (t). \quad (4.78)$$

The corrected reference synthetic waveform $u(t)$ and the observed waveform $\bar{u}(t)$ are shown in Fig. 4.10b. The amplitude spectra of the observed and reference synthetic waveforms are shown in Fig. 4.10c. I have marked the frequencies at which either the observed or the reference synthetic amplitude falls below 20% of its peak amplitude. The reason for doing this is that uncertainties in phase measurements can become large at frequencies where the signal-to-noise ratios in amplitudes are low (e.g., Fichtner et al. 2008; Maggi et al. 2009; Lee and Chen 2013). The frequency-dependent phase-delay time $\Delta\tau_p(\omega)$ and group-delay time $\Delta\tau_g(\omega)$ measurements are shown in Fig. 4.10d and the frequency-dependent amplitude-reduction time $\Delta\tau_q(\omega)$ measurements are shown in Fig. 4.10e. It is possible to have cycle-skipping errors in the phase-delay time $\Delta\tau_p(\omega)$ measurements. In (Chen et al. 2007b) I corrected those cycle-skipping errors by bootstrapping the phase from low frequencies to high frequencies (e.g., Ekström et al. 1997). In (Lee and Chen 2013; Lee et al. 2014b) we corrected cycle-skipping errors by bootstrapping the phase from the dominant frequency bilaterally towards lower and higher frequencies using the broadband cross-correlation delay-time measurement ΔT^{sn} as the anchor. The frequency-dependent group-delay time $\Delta\tau_g(\omega)$ measurements usually do not have cycle-skipping errors and can also be used as a reference when choosing the correct cycles for the phase-delay time measurements.

The WPK for $\Delta\tau_{p,q}(\omega)$ measurements can be derived by considering the target displacement

$$u(t) = \tilde{u}(t) + \delta u(t), \quad (4.79)$$

in the time-domain or

$$\hat{u}(\omega) = \hat{\tilde{u}}(\omega) + \delta\hat{u}(\omega) \quad (4.80)$$

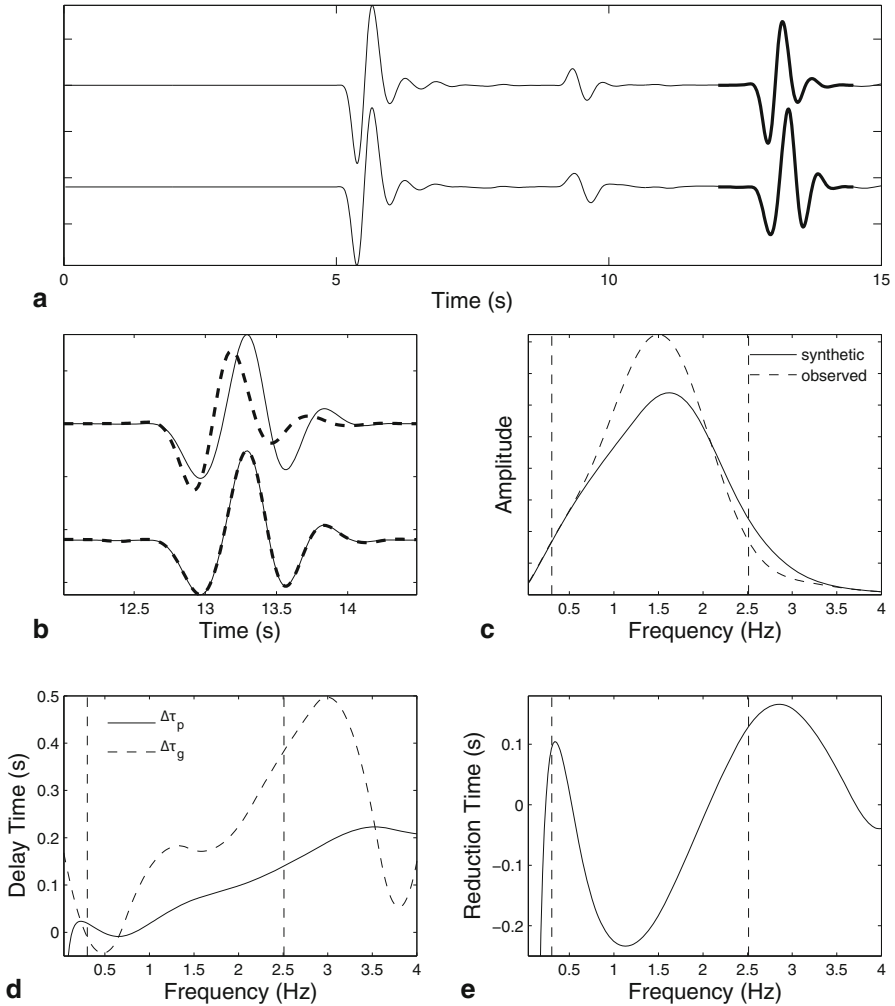


Fig. 4.10 An example of measuring frequency-dependent phase-delay time $\Delta\tau_p(\omega)$, group-delay time $\Delta\tau_g(\omega)$ and amplitude-reduction time $\Delta\tau_a(\omega)$ using the x_1 -component seismogram from 100001 to IN.RC01 in the half-space example. **a** Reference synthetic seismogram (*upper*) and observed seismogram (*lower*), the waveforms selected for making the measurements are plotted in *thick black lines*. **b** The upper two traces show the fit between the selected reference synthetic waveform (*dash line*) and the selected observed waveform (*solid line*), the lower two traces show the fit between the selected observed waveform (*solid line*) and the corrected synthetic waveform (*dash line*) obtained by applying the optimized phase and amplitude corrections $\Delta\phi(\omega)$ and $\Delta\psi(\omega)$ to the reference synthetic waveform. **c** The amplitude spectrum of the observed (*dash*) and reference synthetic (*solid*) waveforms. The two *vertical dash lines* in (**c**), (**d**) and (**e**) indicate the frequencies at which either the observed or the reference synthetic amplitude falls below 20% of its peak amplitude. **d** The phase-delay (*solid*) and group-delay (*dash*) time measurements. **e** The amplitude-reduction time measurements.

in the frequency-domain. The reference synthetic displacement waveform can be mapped into the target displacement waveform through

$$\hat{u}(\omega) = \hat{u}(\omega) e^{-i\omega[\delta\tau_p(\omega) - i\delta\tau_q(\omega)]}. \quad (4.81)$$

Bring Eq. 4.80 into the LHS of Eq. 4.81 and expand the exponential on the RHS of Eq. 4.81 in Taylor series to first-order, we obtain

$$\begin{aligned} \hat{u}(\omega) + \delta\hat{u}(\omega) &= \hat{u}(\omega) [1 - i\omega\delta\tau_p(\omega) - \omega\delta\tau_q(\omega)] \\ \Rightarrow [i\delta\tau_p(\omega) + \delta\tau_q(\omega)] &= -\frac{\delta\hat{u}(\omega)}{\omega\hat{u}(\omega)} \\ \Rightarrow \delta\tau_p(\omega) &= -\Im \left[\frac{\delta\hat{u}(\omega)}{\omega\hat{u}(\omega)} \right], \quad \delta\tau_q(\omega) = -\Re \left[\frac{\delta\hat{u}(\omega)}{\omega\hat{u}(\omega)} \right], \end{aligned} \quad (4.82)$$

where $\Im[\cdot]$ and $\Re[\cdot]$ represent the imaginary part and the real part, respectively. Considering Eq. 4.51, we have

$$\delta\hat{u}(\omega) = \int dt e^{-i\omega t} W_n(t) \delta u_i^s(\mathbf{x}_r, t). \quad (4.83)$$

Bring Eq. 4.83 into Eq. 4.82, we obtain

$$\delta\tau_p(\omega) = -\int \Im \left[\frac{e^{-i\omega t} W_n(t)}{\omega\hat{u}(\omega)} \right] \delta u_i^s(\mathbf{x}_r, t) dt, \quad (4.84)$$

$$\delta\tau_q(\omega) = -\int \Re \left[\frac{e^{-i\omega t} W_n(t)}{\omega\hat{u}(\omega)} \right] \delta u_i^s(\mathbf{x}_r, t) dt. \quad (4.85)$$

For a given frequency $\omega = \omega_n$, the WPKs for $\Delta\tau_{p,q}$ measured on displacement seismograms are functions of time t . Denoting the WPK for $\Delta\tau_p$ as $J_p^{sn,p}$ and the WPK for $\Delta\tau_q$ as $J_p^{sn,q}$, we have

$$J_p^{sn,p}(\mathbf{x}, t|\omega_n) = -\delta(\mathbf{x} - \mathbf{x}_r) \Im \left[\frac{e^{-i\omega_n t} W_n(t)}{\omega_n \hat{u}(\omega_n)} \right] \gamma_p, \quad (4.86)$$

$$J_p^{sn,q}(\mathbf{x}, t|\omega_n) = -\delta(\mathbf{x} - \mathbf{x}_r) \Re \left[\frac{e^{-i\omega_n t} W_n(t)}{\omega_n \hat{u}(\omega_n)} \right] \gamma_p, \quad (4.87)$$

where $\hat{u}(\omega_n)$ is given by Eq. 4.51, i.e.,

$$\hat{u}(\omega_n) = \int dt e^{-i\omega_n t} W_n(t) \tilde{u}_i^s(\mathbf{x}_r, t), \quad (4.88)$$

and can be computed by the Fourier transform of the windowed i th-component reference synthetic displacement seismogram, and γ_p is the cosine of the angle

between the direction of the i th component and the direction of the p th coordinate axis.

The misfit measurements $\Delta\tau_{p,q}$ can also be made on the velocity seismograms $\tilde{v}_i^s(\mathbf{x}_r, t)$ and $\tilde{v}_i^s(\mathbf{x}_r, t)$, in which case Eqs. 4.84 and 4.85 can be modified to

$$\delta\tau_p(\omega) = - \int \Im \left[\frac{e^{-i\omega t} W_n(t)}{\omega \hat{v}(\omega)} \right] \delta v_i^s(\mathbf{x}_r, t) dt, \quad (4.89)$$

$$\delta\tau_q(\omega) = - \int \Re \left[\frac{e^{-i\omega t} W_n(t)}{\omega \hat{v}(\omega)} \right] \delta v_i^s(\mathbf{x}_r, t) dt, \quad (4.90)$$

where $\hat{v}(\omega)$ is given by

$$\hat{v}(\omega) = \int dt e^{-i\omega t} W_n(t) \tilde{v}_i^s(\mathbf{x}_r, t). \quad (4.91)$$

Apply integration-by-parts on the RHS of Eqs. 4.89 and 4.90, we obtain

$$\delta\tau_p(\omega) = \int \frac{d}{dt} \Im \left[\frac{e^{-i\omega t} W_n(t)}{\omega \hat{v}(\omega)} \right] \delta u_i^s(\mathbf{x}_r, t) dt, \quad (4.92)$$

$$\delta\tau_q(\omega) = \int \frac{d}{dt} \Re \left[\frac{e^{-i\omega t} W_n(t)}{\omega \hat{v}(\omega)} \right] \delta u_i^s(\mathbf{x}_r, t) dt. \quad (4.93)$$

And the expressions for the WPKs in Eqs. 4.86 and 4.87 can be modified according to Eqs. 4.92 and 4.93.

The WPK for group-delay measurements $\Delta\tau_g(\omega)$, denoted as $J^{sn,g}(t|\omega)$, can be obtained by numerically differentiating $\omega J^{sn,p}(t|\omega)$ with respect to ω (Xu et al. 2013). Here I have drop the dependence of the WPKs on the spatial coordinate \mathbf{x} and the component index (i.e., the subscript p) to reduce clutter. If we adopt a central differencing scheme (i.e., the three-point formula) (Press 2007), the WPK for group-delay at frequency $\omega = \omega_n$ can be approximated as

$$J^{sn,g}(t|\omega_n) \approx \left[(\omega_n + \Delta\omega) J^{sn,p}(t|\omega_n + \Delta\omega), \right. \\ \left. - (\omega_n - \Delta\omega) J^{sn,p}(t|\omega_n - \Delta\omega) \right] / (2\Delta\omega), \quad (4.94)$$

where $\Delta\omega$ is a sufficiently small frequency interval. The WPK for phase-delay time $J^{sn,p}(t|\omega)$ is given in Eq. 4.86 for measurements made on displacement seismograms and can be obtained from Eq. 4.92 for measurements made on velocity seismograms.

Examples of the WPKs for $\Delta\tau_{p,q,g}$ at two different frequencies for the three arrivals on the x_1 -component reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown in Fig. 4.11. Compared with the WPKs for the broadband cross-correlation delay-time and amplitude anomaly shown in Fig. 4.7, the WPKs for $\Delta\tau_{p,q,g}$ have frequency-dependence and are also compactly supported in time, which is a result of the finite-width time window used to isolate the different arrivals. The corresponding data sensitivity kernels are shown on Figs. 4.12, 4.13 and 4.14.

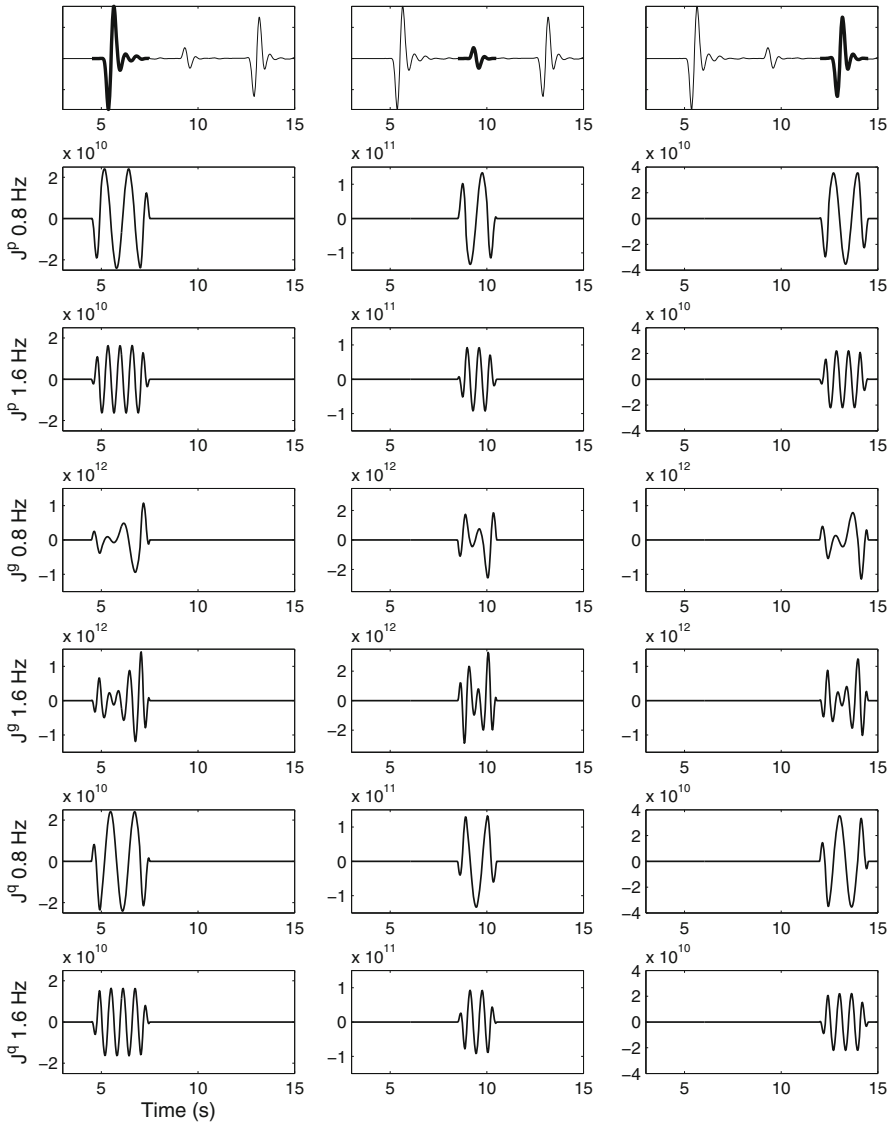


Fig. 4.11 Examples of the WPKs of the $\Delta\tau_{p,q,g}$ measurements at 0.8 and 1.6 Hz for the three arrivals on the x_1 -component reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The second and third rows are WPKs for $\Delta\tau_p$, the fourth and fifth rows are WPKs for $\Delta\tau_g$ and the sixth and seventh rows are WPKs for $\Delta\tau_q$. The second, fourth and sixth rows are WPKs at 0.8 Hz and the third, fifth and seventh rows are WPKs at 1.6 Hz.

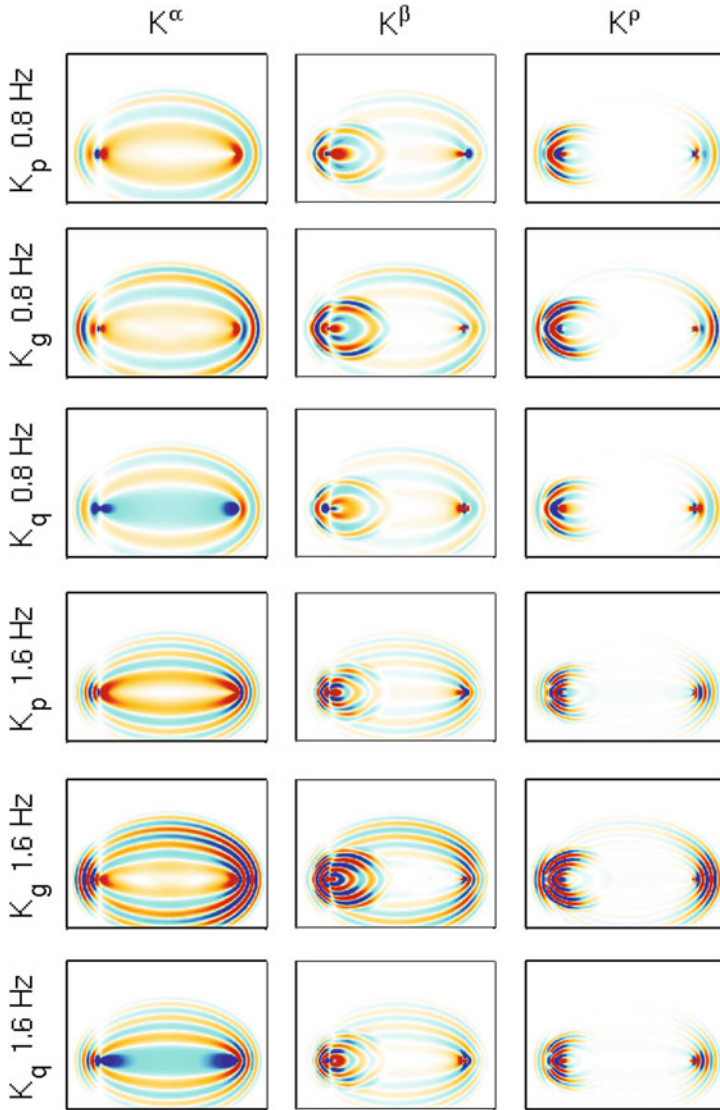


Fig. 4.12 Data sensitivity kernels of $\Delta\tau_p$ (first and fourth rows), $\Delta\tau_g$ (second and fifth rows) and $\Delta\tau_q$ (third and sixth rows) at 0.8 Hz (first to third row) and 1.6 Hz (fourth to sixth row) for the direct-arriving P-wave on the x_1 -component reference synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The WPKs for these kernels are shown on the first column of Fig. 4.11.

Compared with the data sensitivity kernels for the broadband cross-correlation phase-delay time shown in Fig. 4.8, the kernels of the frequency-dependent phase-delay time $\Delta\tau_p$ and group-delay time $\Delta\tau_g$ also exhibit the counter-intuitive “banana-doughnut” phenomena of vanishing sensitivities on the ray path. This

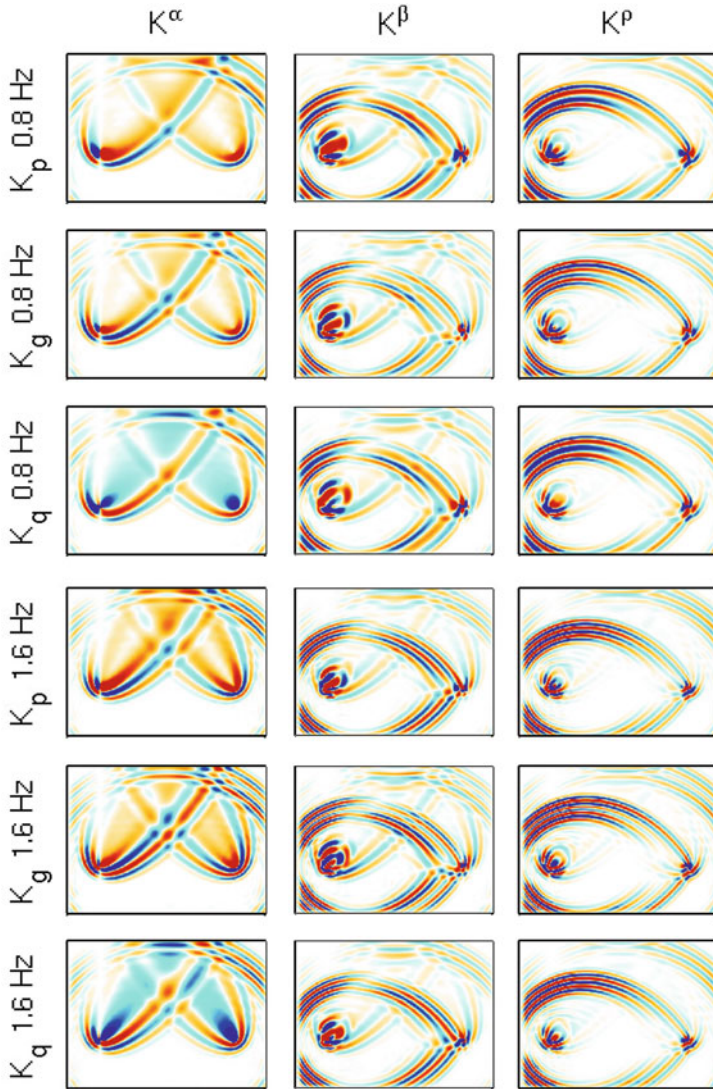


Fig. 4.13 Data sensitivity kernels for the PP-wave on the same seismogram as that in Fig. 4.12. The format is identical to that in Fig. 4.12. The WPKs for these kernels are shown on the second column of Fig. 4.11.

phenomena for the single-frequency phase-delay time kernels will be analyzed in Sect. 4.2.8. All data sensitivity kernels are spatially compact, which is also a result of the finite-width time windows used for isolating the different arrivals. The widths of the first Fresnel zones and the spatial oscillatory patterns of the kernels depend upon the frequency ω_n at which the misfit measurements are made. The widths

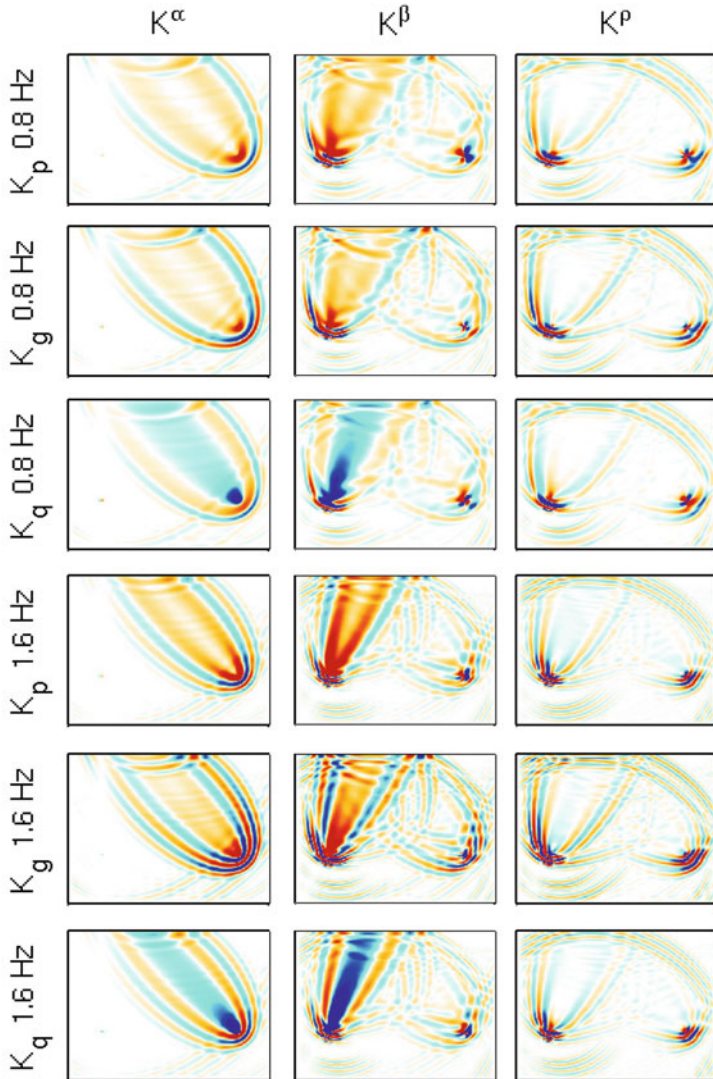


Fig. 4.14 Data sensitivity kernels for the PS-wave on the same seismogram as that in Fig. 4.12. The format is identical to that in Fig. 4.12. The WPKs for these kernels are shown on the third column of Fig. 4.11.

of the first Fresnel zones for the PP-wave and the PS-wave are larger than that for the direct-arriving P-wave because the propagation distances for the PP-wave and the PS-wave are longer than that for the direct P-wave. Compared with the phase-delay kernels, the group-delay kernels show stronger side-bands, which was also observed in full-wave group-delay kernels for surface waves (Dahlen and Zhou

2006). The kernels for amplitude-reduction time $\Delta\tau_q$ show positive sensitivity on the ray path, suggesting an increase in the wave speeds leads to a positive perturbation in amplitude-reduction time, therefore a decrease in amplitude. This result can be understood in terms of the de-focusing effect, i.e., the background wave is diffracted away from the ray path when the wave speed there is increased, which results in a reduction in the amplitude at the receiver.

4.2.8 Rytov Approximation

In Sects. 4.2.4 and 4.2.5, waveform discrepancies are quantified using additive data functionals, while in Sect. 4.2.7, waveform discrepancies are quantified using multiplicative data functionals. The linearization of the data functionals defined in Sects. 4.2.4 and 4.2.5 is essentially the Born approximation (Sect. 3.5.1.2). In (Chen et al. 2007a, b), we pointed out the connection between the linearization of the frequency-dependent phase and amplitude misfits defined in Eq. 4.69 and the Rytov approximation (e.g., Chernov 1960; Woodward 1992a; Snieder and Lomax 1996). The accuracy of the linearization of the data functionals is highly important in solving seismic inverse problems, which are often formulated as optimization problems.

Because of the large dimension of the model space and the significant computational cost involved in solving three-dimensional wave equations, it is often mandatory to use local optimization algorithms, in which case spurious local minima in the objective function can prevent global convergence of a local optimization algorithm if the starting model is not already very close to the global minimum. The accuracy of the linearization determines the shape of the objective function. If the data functionals are exactly linear (or affine) with respect to the model parameters, the linearization becomes exact and the objective function defined in terms of the least-squares of the misfits is quadratic and a Newton-type local optimization algorithm might be able to reach the global minimum in one (or a few) iteration(s) from any starting model. In realistic seismic inverse problems, the data functionals are usually nonlinear with respect to model parameters, but the shape of the objective function can still be improved by selecting the data functionals whose linearizations are more accurate than others.

The accuracy of the Born approximation was analyzed in Sect. 3.5.1.2. It is a good approximation when the accumulative phase shift between the observed and the synthetic waveforms is small, i.e., $\omega_0\Delta T^{sn} \ll 1$, where ω_0 is the dominant frequency of the waveforms. In transmission tomography that involves long propagation distances (long as compared with the dominant wavelength), the accumulative phase shift can be large even when the average perturbation of the wave speeds along the source-receiver path is relatively small. In such situations, the Born approximation might not be valid and the linearization of the data functionals defined in Sects. 4.2.4 and 4.2.5 can be poor. The objective function defined in terms

of those data functionals can have numerous local minima and a local optimization algorithm can stagnate at a local minimum without converging to a useful model, even though the exact gradient of the objective function can be computed using the Born kernels.

The *Rytov approximation* has a validity regime different from that of the Born approximation (e.g., Beydoun and Tarantola 1988; Wu and Aki 1988; Wu 1989; Woodward 1992a; Wu 2003). It does not require the accumulative phase shift to be small, it only requires the phase shift *per wavelength* to be small (e.g., Chernov 1960; Snieder and Lomax 1996). I will use the scalar wave equation to demonstrate the main ideas and adopt the notations in Sect. 3.5.1 in the following. The derivation follows that in (Wu 2003) with some additional explanations about the intermediate steps. In the frequency-domain, the ratio between the perturbed wavefield $\hat{u}(\mathbf{x}, \omega)$ and the reference wavefield $\hat{u}(\mathbf{x}, \omega)$ can be defined in terms of a complex phase function $\varphi(\mathbf{x}, \omega)$,

$$\hat{u}(\mathbf{x}, \omega) = e^{\varphi(\mathbf{x}, \omega)} \hat{u}(\mathbf{x}, \omega). \quad (4.95)$$

Equation 4.95 is called the Rytov transform (e.g., Tatarskiĭ et al. 1971; Ishimaru 1978). This complex phase function $\varphi(\mathbf{x}, \omega)$ can be related to the frequency-dependent phase-delay time and amplitude-reduction time in Eq. 4.69 via

$$\Delta\tau_q(\mathbf{x}, \omega) = -\frac{\Re[\varphi(\mathbf{x}, \omega)]}{\omega}, \quad \Delta\tau_p(\mathbf{x}, \omega) = -\frac{\Im[\varphi(\mathbf{x}, \omega)]}{\omega}. \quad (4.96)$$

I will drop the dependence on ω in the following to reduce clutter. Considering Eq. 3.372, in the frequency-domain Eq. 3.363 can be written as

$$\left(\nabla^2 + \tilde{k}^2\right) \hat{u}(\mathbf{x}) = \hat{f}(\mathbf{x}), \quad (4.97)$$

where $\hat{f}(\mathbf{x})$ is the Fourier transform of the source term $f(\mathbf{x}, t)$. Considering Eq. 3.367, in the frequency-domain Eq. 3.362 can be written as

$$\left(\nabla^2 + \tilde{k}^2\right) \hat{u}(\mathbf{x}) = \hat{f}(\mathbf{x}) - \tilde{k}^2(r^2 - 1)\hat{u}(\mathbf{x}). \quad (4.98)$$

Bring Eq. 4.95 into Eq. 4.98, we obtain

$$\nabla^2 \left(\hat{u}e^\varphi\right) + \tilde{k}^2\hat{u}e^\varphi = \hat{f} - \tilde{k}^2(r^2 - 1)\hat{u}e^\varphi. \quad (4.99)$$

Using the identity³

$$\nabla^2 [a(\mathbf{x})b(\mathbf{x})] = b\nabla^2 a + a\nabla^2 b + 2\nabla a \cdot \nabla b, \quad (4.100)$$

³ This identity can be verified by expressing $\nabla^2 [a(\mathbf{x})b(\mathbf{x})] = \nabla \cdot [\nabla (ab)] = \nabla \cdot (a\nabla b + b\nabla a) = \nabla \cdot (a\nabla b) + \nabla \cdot (b\nabla a) = a\nabla^2 b + \nabla a \cdot \nabla b + \nabla b \cdot \nabla a + b\nabla^2 a = a\nabla^2 b + b\nabla^2 a + 2\nabla a \cdot \nabla b$.

the expression $\nabla^2 \left(\hat{u} e^\varphi \right)$ in Eq. 4.99 can be expressed as

$$\nabla^2 \left(\hat{u} e^\varphi \right) = e^\varphi \nabla^2 \hat{u} + \hat{u} \nabla^2 e^\varphi + 2 \nabla \hat{u} \cdot \nabla e^\varphi. \quad (4.101)$$

The expression ∇e^φ in Eq. 4.101 can be expressed as

$$\nabla e^\varphi = e^\varphi \nabla \varphi, \quad (4.102)$$

and the expression $\nabla^2 e^\varphi$ in Eq. 4.101 can be expressed as

$$\begin{aligned} \nabla^2 e^\varphi &= \nabla \cdot \nabla e^\varphi = \nabla \cdot (e^\varphi \nabla \varphi) \\ &= e^\varphi \nabla^2 \varphi + \nabla e^\varphi \cdot \nabla \varphi \\ &= e^\varphi \nabla^2 \varphi + e^\varphi \nabla \varphi \cdot \nabla \varphi. \end{aligned} \quad (4.103)$$

Bring Eqs. 4.101, 4.102 and 4.103 into Eq. 4.99, re-arrange the terms, we obtain

$$\left(\nabla^2 + \tilde{k}^2 \right) \hat{u} + \hat{u} (\nabla \varphi \cdot \nabla \varphi) + \hat{u} \nabla^2 \varphi + 2 \nabla \hat{u} \cdot \nabla \varphi = \hat{f} - \tilde{k}^2 (r^2 - 1) \hat{u}. \quad (4.104)$$

Considering Eq. 4.97, Eq. 4.104 can be simplified to

$$\hat{u} \nabla^2 \varphi + 2 \nabla \hat{u} \cdot \nabla \varphi = -\hat{u} \left[\tilde{k}^2 (r^2 - 1) + (\nabla \varphi \cdot \nabla \varphi) \right]. \quad (4.105)$$

Considering the identity

$$\nabla^2 \left(\hat{u} \varphi \right) = \hat{u} \nabla^2 \varphi + 2 \nabla \hat{u} \cdot \nabla \varphi + \varphi \nabla^2 \hat{u}, \quad (4.106)$$

Eq. 4.105 can be re-written as

$$\nabla^2 \left(\hat{u} \varphi \right) - \varphi \nabla^2 \hat{u} = -\hat{u} \left[\tilde{k}^2 (r^2 - 1) + (\nabla \varphi \cdot \nabla \varphi) \right]. \quad (4.107)$$

In the absence of the external force, i.e., $\hat{f}(\mathbf{x}) = 0$ in Eq. 4.97, the term $\nabla^2 \hat{u}$ in Eq. 4.107 can be replaced with $-\tilde{k}^2 \hat{u}$ and Eq. 4.107 can be re-written as

$$\left(\nabla^2 + \tilde{k}^2 \right) \left(\hat{u} \varphi \right) = -\hat{u} \left[\tilde{k}^2 (r^2 - 1) + (\nabla \varphi \cdot \nabla \varphi) \right]. \quad (4.108)$$

Under the *Rytov validity condition*

$$|\nabla \varphi \cdot \nabla \varphi| \ll \left| \tilde{k}^2 (r^2 - 1) \right|, \quad (4.109)$$

we can ignore the term $(\nabla\varphi \cdot \nabla\varphi)$ on the RHS of Eq. 4.108 and obtain an approximate solution of $\varphi(\mathbf{x})$ using the Green's function $\hat{G}(\mathbf{x}, \omega; \mathbf{x}')$ in the reference media,

$$\begin{aligned}\varphi(\mathbf{x}) &\approx \frac{1}{\hat{u}(\mathbf{x})} \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}; \mathbf{x}') \hat{u}(\mathbf{x}') \tilde{k}^2(\mathbf{x}') [1 - r^2(\mathbf{x}')] \\ &= \frac{1}{\hat{u}(\mathbf{x})} \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}; \mathbf{x}') \hat{u}(\mathbf{x}') \Delta o(\mathbf{x}').\end{aligned}\quad (4.110)$$

Equation 4.110 is called the *Rytov approximation*. The relation between the complex phase function $\varphi(\mathbf{x})$ and the model parameter $\Delta o(\mathbf{x})$ in Eq. 4.110 is linear and this linear relation holds when the Rytov validity condition in Eq. 4.109 is satisfied.

The term $\nabla\varphi$ in Eq. 4.109 is the change in the complex phase function per unit distance. The wave-number in the reference media \tilde{k} can be expressed in terms of the wavelength using Eq. 3.372 and the Rytov validity condition can be written as

$$\begin{aligned}\left| \frac{\nabla\varphi \cdot \nabla\varphi}{\tilde{k}^2} \right| &= \left| \frac{(\tilde{\lambda}\nabla\varphi) \cdot (\tilde{\lambda}\nabla\varphi)}{4\pi^2} \right| \ll |r^2 - 1| \\ \Rightarrow \|\tilde{\lambda}\nabla\varphi\| &\ll 2\pi\sqrt{\|r^2 - 1\|}.\end{aligned}\quad (4.111)$$

Compare with the Born validity condition in Eq. 3.399, the Rytov validity condition in Eq. 4.111 does not depend on the size of the scatterer. The term $\tilde{\lambda}\nabla\varphi$ in Eq. 4.111 is roughly the phase difference over the distance of one wavelength, therefore the Rytov approximation is valid when the phase difference per wavelength is small. Compared with the Born validity condition that requires the accumulative phase difference to be small, the Rytov validity condition is less restrictive.

If we consider an incident plane wave

$$\hat{u}(\mathbf{x}) = \tilde{A}e^{i\tilde{\mathbf{k}}\cdot\mathbf{x}} \quad (4.112)$$

and assume the total wavefield after scattering is still a plane wave

$$\hat{u}(\mathbf{x}) = Ae^{i\mathbf{k}\cdot\mathbf{x}}, \quad (4.113)$$

the complex phase function can then be expressed as

$$\varphi(\mathbf{x}) = \ln \frac{A}{\tilde{A}} + i(\mathbf{k} - \tilde{\mathbf{k}}) \cdot \mathbf{x}, \quad (4.114)$$

and its spatial gradient is given by

$$\nabla\varphi(\mathbf{x}) = \nabla \ln \frac{A}{\tilde{A}} + i(\mathbf{k} - \tilde{\mathbf{k}}). \quad (4.115)$$

The size of the first term on the RHS of Eq. 4.115 is usually much smaller than the size of the second term. Therefore we have

$$\nabla\varphi \cdot \nabla\varphi \approx \left| \mathbf{k} - \tilde{\mathbf{k}} \right|^2 = 4\tilde{k}^2 \sin^2 \frac{\theta}{2}, \quad (4.116)$$

where θ is the scattering angle between the directions of \mathbf{k} and $\tilde{\mathbf{k}}$. Bring Eq. 4.116 into the Rytov validity condition in Eq. 4.109, we obtain

$$\sin \frac{\theta}{2} \ll \frac{1}{2} \sqrt{|(r^2 - 1)|} = \frac{\sqrt{\tilde{\alpha}^2(\mathbf{x}) - \alpha^2(\mathbf{x})}}{2\alpha(\mathbf{x})}. \quad (4.117)$$

Equation 4.117 indicates that the Rytov approximation is valid when the scattering angle is small and is called the *small-angle scattering* condition. In tomography that uses transmitted or forward-scattered waves, the scattering angle is usually small and Eq. 4.117 is often satisfied, therefore the Rytov approximation provides a good linearization for the complex phase and also the frequency-dependent phase and amplitude misfits (Eq. 4.96).

The connection between the Rytov and the Born representations can be analyzed by expanding the exponential term in Eq. 4.95 in Taylor series and comparing with the Born series in Eq. 3.377 (Heidbreder 1967). Equation 4.95 can be re-written as

$$\begin{aligned} \hat{u}(\mathbf{x}) &= \hat{u}(\mathbf{x}) \left[1 + \varphi(\mathbf{x}) + \frac{\varphi^2(\mathbf{x})}{2} + \cdots + \frac{\varphi^n(\mathbf{x})}{n!} + \cdots \right] \\ &= \hat{u}(\mathbf{x}) \left[1 + \sum_{n=1}^{\infty} r'_n(\mathbf{x}) \right], \end{aligned} \quad (4.118)$$

where $r'_n(\mathbf{x})$ can be represented using the Rytov approximation in Eq. 4.110 as,

$$r'_n(\mathbf{x}) = \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}; \mathbf{x}') \Delta o(\mathbf{x}') \frac{\hat{u}(\mathbf{x}')}{\hat{u}(\mathbf{x})} \frac{\varphi^{n-1}(\mathbf{x}')}{n!}. \quad (4.119)$$

The Born series in Eq. 3.377 can be re-written as

$$\hat{u}(\mathbf{x}) = \hat{u}(\mathbf{x}) \left[1 + \sum_{n=1}^{\infty} r_n(\mathbf{x}) \right], \quad (4.120)$$

where $r_n(\mathbf{x})$ can be expressed recursively as

$$r_n(\mathbf{x}) = \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}; \mathbf{x}') \Delta o(\mathbf{x}') \frac{\hat{u}(\mathbf{x}')}{\hat{u}(\mathbf{x})} r_{n-1}(\mathbf{x}'), \text{ and } r_0 = 1. \quad (4.121)$$

The zeroth- and first-order terms in Eqs. 4.118 and 4.120 are identical. The second-order term in the Rytov representation $r'_2(\mathbf{x})$ is generated by using $\varphi(\mathbf{x})/2$ in Eq. 4.119, while the second-order term in the Born series $r_2(\mathbf{x})$ is generated by

using $r_1(\mathbf{x})$ in Eq. 4.121. Since $\varphi(\mathbf{x}) = r_1(\mathbf{x})$, the second-order term in the Rytov representation is obtained by approximating the first-order scattered field throughout the whole scattering volume $V(\mathbf{x}')$ using half of the first-order scattered field at the observation point \mathbf{x} . Similarly, the n th-order term in the Rytov representation is obtained by approximating the $(n - 1)$ th-order scattered field through the whole scattering volume using $\varphi^{n-1}(\mathbf{x})/n!$ at the observation point. Therefore the linear relation given by the Rytov approximation (Eq. 4.110) provides an approximate treatment of the multiple scattering effects that exist in the second- and higher-order (i.e., the nonlinear) terms in the Born series. The accuracy of this approximate treatment of multiple scattering is certainly not uniform through the whole scattering volume. As analyzed above and also in e.g., (Wu 2003), the Rytov approximation is inappropriate in the backward-scattering regime, where the scattering angle θ is very large. An intuitive explanation about the differences between the Born and the Rytov approximations is given in (Wu 2003). The Born approximation is a superposition of the scattered waves from all parts of the scattering volume. In the forward-scattering regime, the incident wave propagates together with the scattered waves and they coherently interfere, leading to a linear increase of the total wavefield. The Born approximation does not account for this coherent interference effect, therefore has a very limited validity range in the forward direction. However, in the backward direction, there is no incident wave and the range of coherent interference among back-scattered (i.e., reflected) waves is limited to $\tilde{\lambda}/4$, where $\tilde{\lambda}$ is the dominant wavelength, due to the two-way travel-time difference. Therefore the Born approximation has a larger validity range in the back-scattering regime.

A disadvantage of the Rytov approximation is that it needs to be applied to one wave at a time (e.g., Keller 1969; Woodward 1992a). When it is applied to a superposition of multiple seismic arrivals, the nonlinear interference effect among the multiple seismic arrivals can destroy the linear relation between the complex phase and the structural parameters provided by the Rytov approximation. A numerical example of this effect is presented in (Holschneider et al. 2005) using two interfering Ricker pulses with different dispersion and attenuation characteristics. In (Lee and Chen 2013) we have developed an automated seismogram segmentation and waveform selection technique based on the continuous wavelet transform and the topological watershed algorithm (Couprie and Bertrand 1997; Couprie et al. 2005). In our technique, the time-domain seismogram is transformed to the time-frequency domain through the continuous wavelet transform. The time-frequency domain scalogram (i.e., the squared modulus of the continuous wavelet transform) is then automatically segmented using the topological watershed algorithm, an image segmentation algorithm used in e.g., automatic extraction of geographic features from satellite images (e.g., Quackenbush 2004; Chen et al. 2005b). Each wave packet obtained from segmenting the observed seismogram is paired with a wave packet obtained from segmenting the corresponding synthetic seismogram based on a collection of criteria that measure the similarities among the two wave packets in the time-frequency domain. On a selected pair of observed and synthetic wave packets, we make the frequency-dependent phase-delay time and amplitude-reduction

time misfit measurements (Fig. 4.10). Compared with a purely time-domain waveform selection technique (e.g., Maggi et al. 2009) an important advantage of our technique is that it allows us to separate wave packets arriving at overlapping time windows but have disjoint frequency-domain or time-frequency-domain supports. Like many other seismic data processing techniques, our technique also has a handful of parameters that control the behavior of our workflow and need to be adjusted manually for different data sets or different types of applications. However, our recent experiments have shown that even this step can be fully automated with a nearly zero failure rate by using an importance-aided artificial neural network (Diersen et al. 2011).

4.2.9 “Banana-Doughnut” Structure of Rytov Phase Kernels

The data sensitivity kernels for the frequency-dependent phase-delay time (Figs. 4.12, 4.13 and 4.14) also exhibit the counter-intuitive “banana-doughnut” phenomena (i.e., vanishing sensitivity along the ray path and nonzero sensitivity off the ray path) as the data sensitivity kernels for the broadband cross-correlation delay-time (Fig. 4.8). To give a qualitative explanation about this phenomena, I will use the whole-space scalar wave equation (Eqs. 3.388 and 3.390) as an example. The linear Rytov approximation in Eq. 4.110 also gives the Fréchet kernel of the complex phase function $\varphi(\mathbf{x}, \omega)$ with respect to the model parameter $-\delta o(\mathbf{x}, \omega) = o(\mathbf{x}, \omega) - \tilde{o}(\mathbf{x}, \omega)$,

$$\delta\varphi(\mathbf{x}, \omega) = -\frac{1}{\hat{u}(\mathbf{x}, \omega)} \int_V dV(\mathbf{x}') \hat{G}(\mathbf{x}, \omega; \mathbf{x}') \hat{u}(\mathbf{x}', \omega) [-\delta o(\mathbf{x}', \omega)]. \quad (4.122)$$

Bring the whole-space Green’s function (Eq. 3.388) and the forward wavefield (Eq. 3.390) into Eq. 4.122, we obtain

$$\delta\varphi(\mathbf{x}_r, \omega) = \frac{1}{4\pi} \int_V dV(\mathbf{x}) \frac{|\mathbf{x}_r - \mathbf{x}_s| e^{-i\frac{\omega}{c_0}(|\mathbf{x}-\mathbf{x}_r|+|\mathbf{x}-\mathbf{x}_s|-|\mathbf{x}_r-\mathbf{x}_s|)}}{|\mathbf{x} - \mathbf{x}_r| |\mathbf{x} - \mathbf{x}_s|} [-\delta o(\mathbf{x}, \omega)], \quad (4.123)$$

where \mathbf{x}_r and \mathbf{x}_s are locations of the receiver and the point source, respectively and I have replaced \mathbf{x}' with \mathbf{x} and replaced \mathbf{x} with \mathbf{x}_r . If we denote

$$L = \frac{|\mathbf{x}_r - \mathbf{x}_s|}{|\mathbf{x} - \mathbf{x}_r| |\mathbf{x} - \mathbf{x}_s|}, \quad \Delta l = |\mathbf{x} - \mathbf{x}_r| + |\mathbf{x} - \mathbf{x}_s| - |\mathbf{x}_r - \mathbf{x}_s|, \quad (4.124)$$

considering Eq. 4.96, we have

$$\delta\tau_p(\mathbf{x}_r, \omega) = \frac{1}{4\pi\omega} \int_V dV(\mathbf{x}) L \sin(\tilde{k}\Delta l) [-\delta o(\mathbf{x}, \omega)], \quad (4.125)$$

$$\delta\tau_q(\mathbf{x}_r, \omega) = -\frac{1}{4\pi\omega} \int_V dV(\mathbf{x}) L \cos(\tilde{k}\Delta l) [-\delta o(\mathbf{x}, \omega)], \quad (4.126)$$

where $\tilde{k} = \omega/c_0$. The Fréchet kernel of the frequency-dependent phase-delay time is then given by

$$K_p(\mathbf{x}|\mathbf{x}_s, \mathbf{x}_r, \omega) = \frac{L \sin(\tilde{k} \Delta l)}{4\pi\omega}, \quad (4.127)$$

and the kernel of the amplitude-reduction time is given by

$$K_q(\mathbf{x}|\mathbf{x}_s, \mathbf{x}_r, \omega) = -\frac{L \cos(\tilde{k} \Delta l)}{4\pi\omega}. \quad (4.128)$$

Consider the source-receiver configuration shown in Fig. 4.15. We assume that

$$\eta^2 = y^2 + z^2 \ll x^2. \quad (4.129)$$

The distance between the source and the scatterer can then be expanded in Taylor series and approximated as

$$|\mathbf{x} - \mathbf{x}_s| = x\sqrt{1 + \frac{\eta^2}{x^2}} = x \left[1 + \frac{\eta^2}{2x^2} - \frac{1}{8} \left(\frac{\eta^2}{x^2} \right)^2 + \dots \right] \approx x + \frac{\eta^2}{2x}. \quad (4.130)$$

The distance between the receiver and the scatterer can be approximated as

$$|\mathbf{x} - \mathbf{x}_r| \approx (d - x) + \frac{\eta^2}{2(d - x)}. \quad (4.131)$$

Bring Eqs. 4.130 and 4.131 into the expression for Δl in Eq. 4.124, we obtain

$$\Delta l \approx \frac{\eta^2}{2} \left[\frac{d}{x(d - x)} \right] = \frac{\eta^2}{2\xi}, \quad (4.132)$$

where

$$\xi = \frac{x(d - x)}{d}. \quad (4.133)$$

If the scatterer is located on the unperturbed ray path, we have $|\mathbf{x} - \mathbf{x}_r| + |\mathbf{x} - \mathbf{x}_s| = |\mathbf{x}_r - \mathbf{x}_s|$ and $\Delta l = 0$, therefore the Rytov phase kernel $K_p = L \sin(\tilde{k}0)/(4\pi\omega) = 0$. The phase kernel reaches a maximum at $\tilde{k}\Delta l = \pi/2$. Considering Eqs. 3.372 and 4.132, the condition $\tilde{k}\Delta l = \pi/2$ is equivalent to

$$\frac{2\pi}{\tilde{\lambda}} \frac{\eta^2}{2\xi} \approx \frac{\pi}{2} \Rightarrow \eta \approx \sqrt{\frac{\xi \tilde{\lambda}}{2}}. \quad (4.134)$$

When $x = d/2$, we have $\xi = d/4$ and

$$\eta \approx \sqrt{\frac{\tilde{\lambda}d}{8}}, \quad (4.135)$$

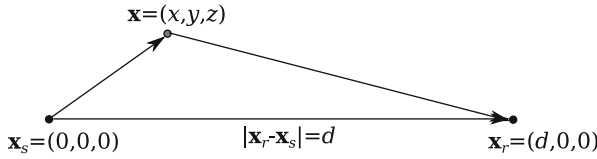


Fig. 4.15 Locations of the source \mathbf{x}_s , the receiver \mathbf{x}_r and a scatterer \mathbf{x} . The source is located at the origin. The receiver is located on the x -axis of the coordinate system at a distance d from the source. The scatterer is located at (x, y, z) . The source-receiver distance is therefore d . The distance between the source and the scatterer is $|\mathbf{x} - \mathbf{x}_s| = \sqrt{x^2 + y^2 + z^2}$ and the distance between the receiver and the scatterer is $|\mathbf{x} - \mathbf{x}_r| = \sqrt{(d - x)^2 + y^2 + z^2}$.

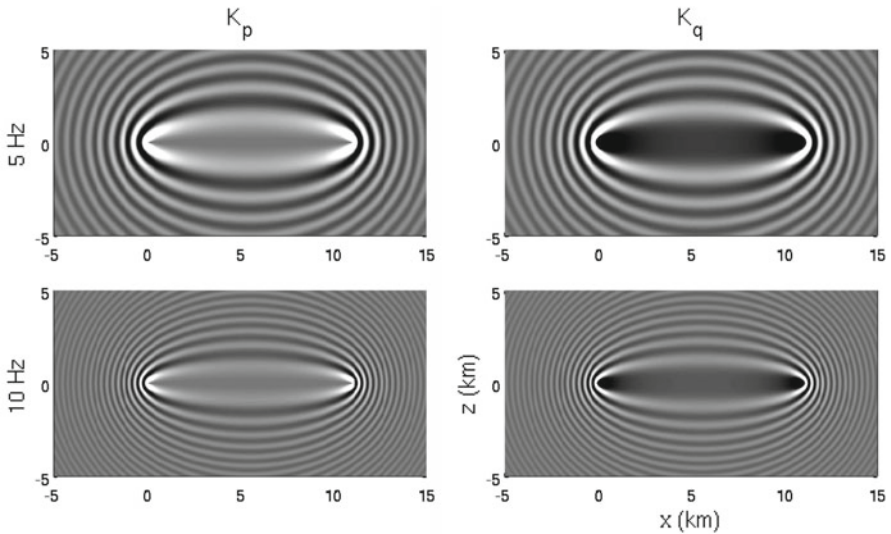


Fig. 4.16 Cross-section views of the frequency-domain Rytov kernels for the scalar wave equation in the whole-space structure model. *Left column*: Rytov phase kernels (Eq. 4.127) at 5 Hz (*top row*) and 10 Hz (*bottom row*); *right column*: Rytov amplitude kernels (Eq. 4.128) at 5 Hz (*top row*) and 10 Hz (*bottom row*). The source is located at coordinate $(0, 0, 0)$ and the receiver is located at $(11, 0, 0)$. The cross-section view is perpendicular to the source-receiver $x - z$ plane.

which is the width of the first *Fresnel zone*.

Examples of the Rytov kernels for the scalar wave equation in the whole-space structural model (Eqs. 4.127 and 4.128) are shown in Fig. 4.16. This type of kernels were named the *Rytov wavepaths* in (Woodward 1992b). Unlike the data sensitivity kernels for the frequency-dependent phase-delay time and amplitude-reduction time shown in Figs. 4.12, 4.13 and 4.14, the single-frequency Rytov kernels shown in Fig. 4.16 have infinite spatial extent. In Figs. 4.12, 4.13 and 4.14, the effect of the finite-width time window used for isolating the particular arrival is built into the kernel through the corresponding WPK. The spatial extent of the single-frequency Rytov kernels shown in Fig. 4.16 is not restricted by any windowing operations

in the time domain. For the Rytov kernels shown in Fig. 4.16, the source-receiver distance is 11 km, the wave speed is 6 km/s, therefore the width of the first Fresnel zone (Eq. 4.135) is about 1.3 km at 5 Hz and about 0.9 km at 10 Hz.

4.3 Model Parameterization

In Sect. 4.2.3 I have shown that by using the chain rule for Fréchet derivatives we can obtain the Fréchet kernels of the data functional with respect to the source and structural models parameterized using the moment tensor density function and the rate-of-relaxation tensor. The data sensitivity (Fréchet) kernels are the composition of the corresponding WPK of the data functional and the primary kernels (Eq. 4.1). Examples of the WPKs for some widely used data functionals and the corresponding data sensitivity kernels are shown in Sects. 4.2.4–4.2.7.

In practice, the seismic inverse problem is often under-determined and the amount and the nature of seismic observations usually do not allow us to resolve all 21 components of the elastic tensor simultaneously. It is therefore important to choose parameterizations judiciously such that most of the data can be explained using a few parameters. In this section I will show that by using the chain rule for partial derivatives, we can obtain the Fréchet kernels of the data functional with respect to other types of parameterizations of the source and structural models. Some of these parameterizations for the structural model might be more meaningful or more amenable to interpretation than the rate-of-relaxation tensor or elastic tensor. The appropriate choice of the parameterization might also help reduce the nonlinearity of the inverse problem.

The full fourth-order elastic stiffness tensor c_{ijkl} has 81 components. But because of symmetry conditions similar to those in Eq. 3.201, i.e.,

$$c_{ijkl} = c_{kjlm} = c_{jkml} = c_{lmjk}, \quad (4.136)$$

the total number of independent components is reduced to 21. The number of independent components can be reduced even further in materials with higher symmetry.

4.3.1 Bulk Modulus

In fluids, the fourth-order stiffness tensor c_{ijkl} is reduced to one elastic parameter, the bulk modulus κ , and we have

$$c_{ijkl} = \kappa \delta_{jk} \delta_{lm}, \quad (4.137)$$

where δ_{jk} and δ_{lm} are the Kronecker delta. The perturbation δc_{ijklm} can therefore be expressed in terms of the perturbation $\delta\kappa$ as

$$\delta c_{ijklm} = \frac{\partial c_{ijklm}}{\partial \kappa} \delta\kappa = \delta_{jk} \delta_{lm} \delta\kappa. \quad (4.138)$$

Bring the representation of δc_{ijklm} in Eq. 4.138 into the last term on the RHS of Eq. 4.37, we obtain

$$\begin{aligned} & \int_V dV(\mathbf{x}) \sum_{j,k,l,m=1}^3 K_{d^{sn}}^{c_{ijklm}}(\mathbf{x}) \delta c_{ijklm}(\mathbf{x}) \\ &= \int_V dV(\mathbf{x}) \sum_{j,k,l,m=1}^3 K_{d^{sn}}^{c_{ijklm}}(\mathbf{x}) \delta_{jk} \delta_{lm} \delta\kappa(\mathbf{x}) \\ &= \int_V dV(\mathbf{x}) \sum_{j,l=1}^3 K_{d^{sn}}^{c_{jjll}}(\mathbf{x}) \delta\kappa(\mathbf{x}). \end{aligned} \quad (4.139)$$

The data sensitivity kernel of the data functional d^{sn} with respect to the bulk modulus $\kappa(\mathbf{x})$ is therefore given by

$$K_{d^{sn}}^{\kappa}(\mathbf{x}) = \sum_{j,l=1}^3 K_{d^{sn}}^{c_{jjll}}(\mathbf{x}). \quad (4.140)$$

Considering the representation of $K_{d^{sn}}^{c_{ijklm}}(\mathbf{x})$ given in Eq. 4.35, we can express $K_{d^{sn}}^{\kappa}(\mathbf{x})$ as

$$\begin{aligned} K_{d^{sn}}^{\kappa}(\mathbf{x}) &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \sum_{j,l=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \\ &= - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \\ &= - \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \}. \end{aligned} \quad (4.141)$$

In practice, we often invert for the relative perturbation, e.g., $\delta\kappa/\kappa$, instead of the absolute perturbation $\delta\kappa$. In such a situation, Eq. 4.138 can be modified to

$$\delta c_{ijklm} = \delta_{jk} \delta_{lm} \kappa \frac{\delta\kappa}{\kappa}, \quad (4.142)$$

which can also be written as (e.g., Tromp et al. 2005)

$$\delta c_{ijklm} = \delta_{jk} \delta_{lm} \kappa \frac{\delta\kappa}{\kappa} = \delta_{jk} \delta_{lm} \kappa \delta \ln \kappa. \quad (4.143)$$

Equation 4.140 can then be modified to

$$K_{d^{sn}}^{\ln \kappa}(\mathbf{x}) = \kappa(\mathbf{x}) \sum_{j,l=1}^3 K_{d^{sn}}^{cjjll}(\mathbf{x}), \quad (4.144)$$

and Eq. 4.141 can be modified to

$$\begin{aligned} K_{d^{sn}}^{\ln \kappa}(\mathbf{x}) &= -\kappa(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\ &\quad \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \\ &= -\kappa(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\ &\quad \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \}. \end{aligned} \quad (4.145)$$

Note that the Green's tensor \mathbf{G} and the forward wavefield \mathbf{u}^s in Eqs. 4.141 and 4.145 are computed using the reference structural model and $\kappa(\mathbf{x})$ is the reference bulk modulus. I have dropped the $\tilde{\cdot}$ on top of the Green's tensor, the forward wavefield and the reference bulk modulus to reduce clutter.

4.3.2 Bulk and Shear Moduli

In isotropic solids, we have two independent elastic parameters. One possibility is to use the bulk modulus κ and the shear modulus μ and the stiffness tensor can be expressed as

$$c_{jklm} = \left(\kappa - \frac{2}{3}\mu \right) \delta_{jk} \delta_{lm} + \mu (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}). \quad (4.146)$$

The perturbation δc_{jklm} can be expressed in terms of the perturbations in the two moduli as

$$\begin{aligned} \delta c_{jklm} &= \frac{\partial c_{jklm}}{\partial \kappa} \delta \kappa + \frac{\partial c_{jklm}}{\partial \mu} \delta \mu \\ &= \delta_{jk} \delta_{lm} \delta \kappa + \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) \delta \mu. \end{aligned} \quad (4.147)$$

Equation 4.147 can also be written in terms of the relative perturbations

$$\delta c_{jklm} = \delta_{jk} \delta_{lm} \kappa \delta \ln \kappa$$

$$+ \left(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl} - \frac{2}{3}\delta_{jk}\delta_{lm} \right) \mu \delta \ln \mu. \quad (4.148)$$

Bring Eq. 4.148 into the last term on the RHS of Eq. 4.37, we obtain the data sensitivity kernels with respect to the relative perturbations in the two moduli, i.e.,

$$K_{d^{sn}}^{\ln \kappa}(\mathbf{x}) = \kappa(\mathbf{x}) \sum_{j,l=1}^3 K_{d^{sn}}^{c_{jll}}(\mathbf{x}), \quad (4.149)$$

$$K_{d^{sn}}^{\ln \mu}(\mathbf{x}) = \mu(\mathbf{x}) \left[\sum_{j,k=1}^3 K_{d^{sn}}^{c_{jkjk}}(\mathbf{x}) + \sum_{j,k=1}^3 K_{d^{sn}}^{c_{jkkj}}(\mathbf{x}) - \frac{2}{3} \sum_{j,l=1}^3 K_{d^{sn}}^{c_{jll}}(\mathbf{x}) \right]. \quad (4.150)$$

Bring Eq. 4.35 into the RHS of Eq. 4.149, we recover Eq. 4.145. Bring Eq. 4.35 into the RHS of Eq. 4.150, we obtain

$$\begin{aligned} K_{d^{sn}}^{\ln \mu}(\mathbf{x}) &= -\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\ &\left\{ \sum_{j,k=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \left[\frac{\partial u_j^s(\mathbf{x}, t)}{\partial x_k} + \frac{\partial u_k^s(\mathbf{x}, t)}{\partial x_j} \right] \right. \\ &\left. - \frac{2}{3} \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \right\} \\ &= -\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\ &\cdot \left\{ \nabla \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0) : \left[\nabla \mathbf{u}^s(\mathbf{x}, t) + [\nabla \mathbf{u}^s(\mathbf{x}, t)]^T \right] \right. \\ &\left. - \frac{2}{3} [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \right\}, \quad (4.151) \end{aligned}$$

where “:” denotes the temporal convolution “*” and also the summation over both j and k .

Equation 4.151 can be further simplified if we introduce the deviatoric strain (Eq. 2.49). We denote the forward strain field from the s th source as

$$\epsilon^s \equiv \frac{1}{2} \left[\nabla \mathbf{u}^s + (\nabla \mathbf{u}^s)^T \right], \text{ or } \epsilon_{jk}^s \equiv \frac{1}{2} \left[\frac{\partial u_j^s}{\partial x_k} + \frac{\partial u_k^s}{\partial x_j} \right], \quad (4.152)$$

which is a second-order tensor field. The gradient of the Green's tensor $\nabla \mathbf{G}$ or $\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)/\partial x_k$ is a third-order tensor field. If we introduce

$$\epsilon_{jkp}^G \equiv \frac{1}{2} \left[\frac{\partial G_{jp}}{\partial x_k} + \frac{\partial G_{kp}}{\partial x_j} \right], \quad (4.153)$$

then for every fixed p , Eq. 4.153 defines a strain field induced by the displacement field G_{jp} . The second term inside the square bracket in Eq. 4.153, $\partial G_{kp}/\partial x_j$, is a transposition of the first two indices (i.e., j and k) in the first term, $\partial G_{jp}/\partial x_k$. If we denote this transposition using the superscript ²¹³ (e.g., Zhao et al. 2005, 2006), we can write Eq. 4.153 as

$$\epsilon^G \equiv \frac{1}{2} \left[\nabla \mathbf{G} + (\nabla \mathbf{G})^{213} \right]. \quad (4.154)$$

The corresponding deviatoric strain fields can then be defined as

$$\check{\epsilon}^s \equiv \epsilon^s - \frac{1}{3} \text{tr}(\epsilon^s) \mathbf{I} \text{ or } \check{\epsilon}_{jk}^s \equiv \epsilon_{jk}^s - \frac{\delta_{jk}}{3} \sum_{l=1}^3 \epsilon_{ll}^s, \quad (4.155)$$

$$\check{\epsilon}^G \equiv \epsilon^G - \frac{1}{3} \text{tr}(\epsilon^G) \mathbf{I} \text{ or } \check{\epsilon}_{jkp}^G \equiv \epsilon_{jkp}^G - \frac{\delta_{jk}}{3} \sum_{l=1}^3 \epsilon_{llp}^G, \quad (4.156)$$

where \mathbf{I} is a 3-by-3 identity tensor, $\text{tr}(\epsilon^s) = \sum_{l=1}^3 \epsilon_{ll}^s$ is the trace of the second-order tensor ϵ^s and $\text{tr}(\epsilon^G) = \sum_{l=1}^3 \epsilon_{llp}^G$ is the trace for the first two indices of the third-order tensor ϵ^G . It can be verified that

$$\begin{aligned} \nabla \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0) &: \left[\nabla \mathbf{u}^s(\mathbf{x}, t) + [\nabla \mathbf{u}^s(\mathbf{x}, t)]^T \right] \\ &= 2\check{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) : \epsilon^s(\mathbf{x}, t), \end{aligned} \quad (4.157)$$

and from the definition of the divergence we have

$$\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0) = \text{tr} \left[\epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) \right], \quad \nabla \cdot \mathbf{u}^s(\mathbf{x}, t) = \text{tr} \left[\epsilon^s(\mathbf{x}, t) \right]. \quad (4.158)$$

Bring Eqs. 4.157 and 4.158 into Eq. 4.151, we obtain

$$\begin{aligned} K_{d^{sn}}^{\text{ln}\mu}(\mathbf{x}) &= -2\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\ &\cdot \left\{ \epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) : \epsilon^s(\mathbf{x}, t) - \frac{1}{3} \text{tr} \left[\epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) \right] * \text{tr} \left[\epsilon^s(\mathbf{x}, t) \right] \right\}. \end{aligned} \quad (4.159)$$

It can be verified that the term inside $\{ \}$ in Eq. 4.159 is identical to $\check{\epsilon}^G : \check{\epsilon}^s$, i.e.,

$$\begin{aligned} \epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) : \epsilon^s(\mathbf{x}, t) - \frac{1}{3} \text{tr} \left[\epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) \right] * \text{tr} \left[\epsilon^s(\mathbf{x}, t) \right] \\ = \check{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) : \check{\epsilon}^s(\mathbf{x}, t). \end{aligned} \quad (4.160)$$

Combining Eqs. 4.157, 4.158 and 4.160, we have

$$\begin{aligned}
 & \nabla \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0): \left[\nabla \mathbf{u}^s(\mathbf{x}, t) + [\nabla \mathbf{u}^s(\mathbf{x}, t)]^T \right] \\
 & \quad - \frac{2}{3} [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \\
 = & 2 \left\{ \boldsymbol{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0): \boldsymbol{\epsilon}^s(\mathbf{x}, t) - \frac{1}{3} \text{tr} \left[\boldsymbol{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) \right] * \text{tr} \left[\boldsymbol{\epsilon}^s(\mathbf{x}, t) \right] \right\} \\
 = & 2 \check{\boldsymbol{\epsilon}}^G(\mathbf{x}, t; \mathbf{x}', 0): \check{\boldsymbol{\epsilon}}^s(\mathbf{x}, t). \tag{4.161}
 \end{aligned}$$

Bring Eq. 4.161 into Eq. 4.151, we obtain

$$\begin{aligned}
 K_{d^{sn}}^{\ln \mu}(\mathbf{x}) = & -2\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\
 & \cdot \left[\check{\boldsymbol{\epsilon}}^G(\mathbf{x}, t; \mathbf{x}', 0): \check{\boldsymbol{\epsilon}}^s(\mathbf{x}, t) \right], \tag{4.162}
 \end{aligned}$$

where “:” has the same meaning as in Eq. 4.151. Equation 4.162 is consistent with (Tromp et al. 2005). Note that in Eq. 4.162 both $\check{\boldsymbol{\epsilon}}^G$ and $\check{\boldsymbol{\epsilon}}^s$ are computed in the reference structural model and $\mu(\mathbf{x})$ is the reference shear modulus.

The data sensitivity kernels of the broadband cross-correlation delay time (Sect. 4.2.6) with respect to the relative perturbations in the two moduli and the density for the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown on the top rows in Figs. 4.17, 4.18 and 4.19. The expressions for the kernels with respect to the relative perturbations in the bulk modulus and the shear modulus are given in Eqs. 4.145 and 4.151, respectively. The expression for the kernel with respect to the relative perturbation in the density is given in Eq. 4.178 in Sect. 4.3.4.

4.3.3 Lamé Parameters

In isotropic solids, the fourth-order stiffness tensor can also be represented using the two Lamé parameters through

$$c_{jklm} = \lambda \delta_{jk} \delta_{lm} + \mu (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}), \tag{4.163}$$

where the first Lamé parameter λ is related to the bulk and shear moduli as

$$\lambda = \kappa - \frac{2}{3}\mu. \tag{4.164}$$

The perturbation δc_{jklm} can be expressed in terms of the perturbations in the Lamé parameters as

$$\delta c_{jklm} = \frac{\partial c_{jklm}}{\partial \lambda} \delta \lambda + \frac{\partial c_{jklm}}{\partial \mu} \delta \mu$$

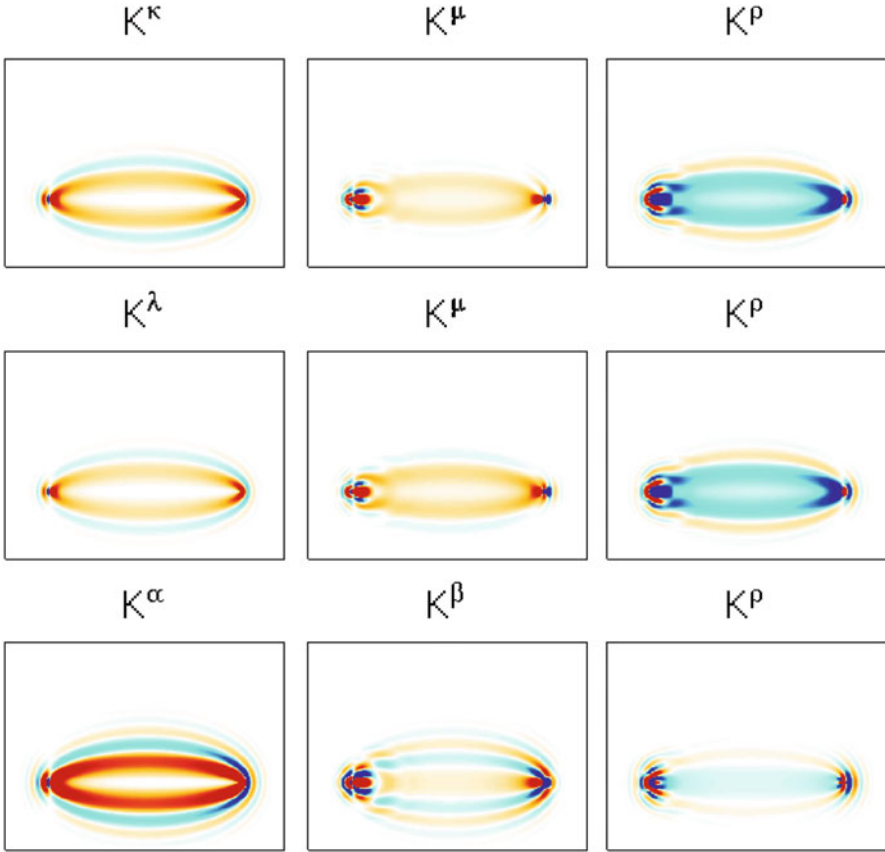


Fig. 4.17 Data sensitivity kernels of the broadband cross-correlation delay time with respect to relative perturbations in the bulk modulus κ (left column, top row) (Eq. 4.145), shear modulus μ (center column, top row) (Eq. 4.151) and density ρ (right column, top row) (Eq. 4.178), the first Lamé parameter λ (left column, center row) (Eq. 4.167), the second Lamé parameter μ (center column, center row) (Eq. 4.168), the density ρ (right column, center row) (Eq. 4.178), the P-wave velocity α (left column, bottom row) (Eq. 4.185), the S-wave velocity β (center column, bottom row) (Eq. 4.186) and the density ρ (right column, bottom row) (Eq. 4.175). The data functional is applied on the P-wave on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example.

$$= \delta_{jk} \delta_{lm} \delta \lambda + (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \delta \mu, \quad (4.165)$$

and in terms of the relative perturbations in the Lamé parameters as

$$\begin{aligned} \delta c_{ijklm} &= \delta_{jk} \delta_{lm} \lambda \delta \ln \lambda \\ &+ (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \mu \delta \ln \mu. \end{aligned} \quad (4.166)$$

Bring Eq. 4.166 into the last term on the RHS of Eq. 4.37, we obtain the data sensitivity kernels with respect to the relative perturbations in the two Lamé parameters,

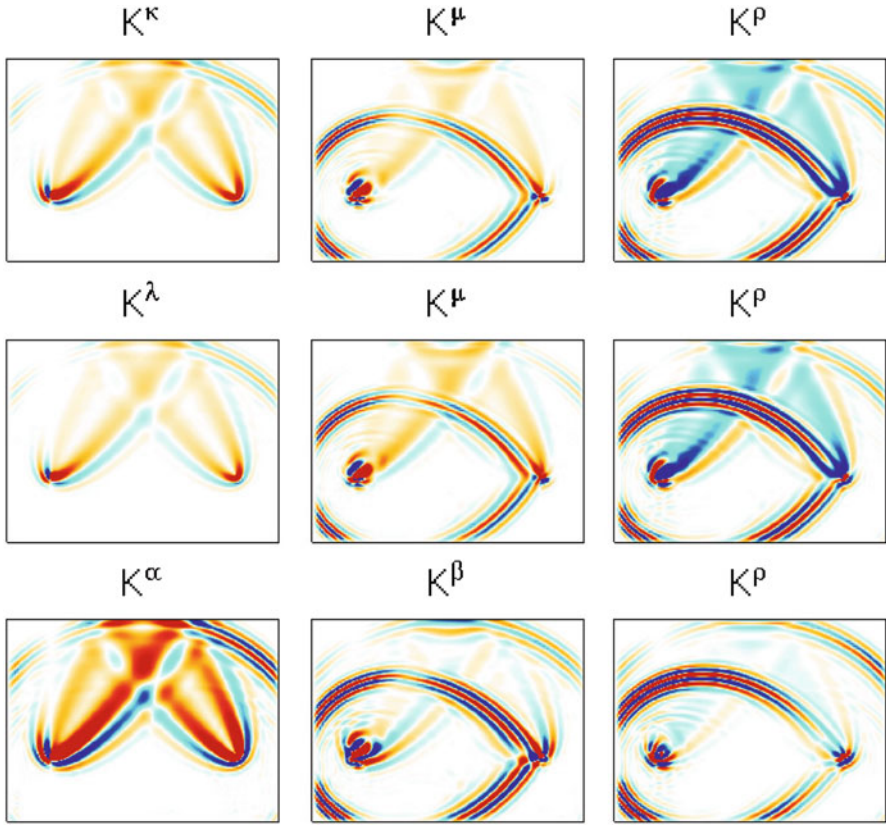


Fig. 4.18 The format is identical to that in Fig. 4.17, except that the data functional is applied on the PP-wave.

i.e.,

$$\begin{aligned}
 K_{d^{sn}}^{\ln \lambda}(\mathbf{x}) &= \lambda(\mathbf{x}) \sum_{j,l=1}^3 K_{d^{sn}}^{c_{jll}}(\mathbf{x}), \\
 &= -\lambda(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\
 &\quad \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \}. \tag{4.167}
 \end{aligned}$$

$$\begin{aligned}
 \check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x}) &= \mu(\mathbf{x}) \left[\sum_{j,k=1}^3 K_{d^{sn}}^{c_{jkk}}(\mathbf{x}) + \sum_{j,k=1}^3 K_{d^{sn}}^{c_{jkk}}(\mathbf{x}) \right] \\
 &= -\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \cdot \left\{ \nabla \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0) \right\}
 \end{aligned}$$

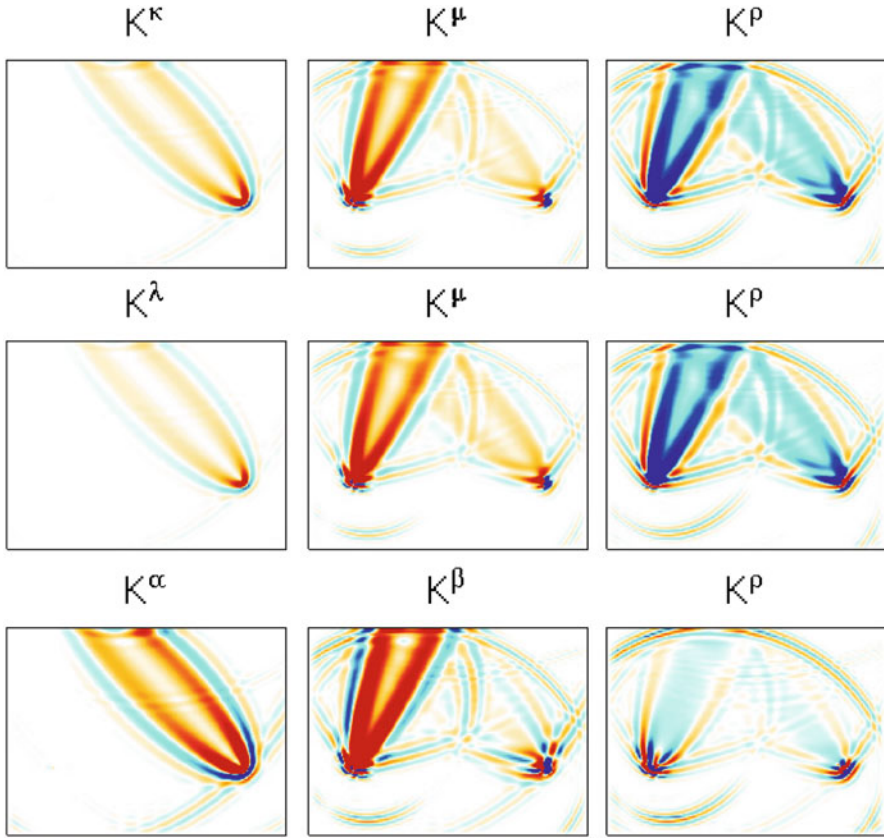


Fig. 4.19 The format is identical to that in Fig. 4.17, except that the data functional is applied on the PS wave.

$$: \left[\nabla \mathbf{u}^s(\mathbf{x}, t) + [\nabla \mathbf{u}^s(\mathbf{x}, t)]^T \right], \quad (4.168)$$

where the “:” has the same meaning as in Eq. 4.151. Considering Eqs. 4.149 and 4.164, we can express $K_{d^{sn}}^{\ln \lambda}(\mathbf{x})$ in Eq. 4.167 as

$$K_{d^{sn}}^{\ln \lambda}(\mathbf{x}) = \frac{\kappa(\mathbf{x}) - \frac{2}{3}\mu(\mathbf{x})}{\kappa(\mathbf{x})} K_{d^{sn}}^{\ln \kappa}(\mathbf{x}). \quad (4.169)$$

Considering Eqs. 4.149 and 4.150, we can also express $\check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x})$ in Eq. 4.168 in terms of $K_{d^{sn}}^{\ln \kappa}(\mathbf{x})$ and $K_{d^{sn}}^{\ln \mu}(\mathbf{x})$ as

$$\check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x}) = K_{d^{sn}}^{\ln \mu}(\mathbf{x}) + \frac{2\mu(\mathbf{x})}{3\kappa(\mathbf{x})} K_{d^{sn}}^{\ln \kappa}(\mathbf{x}). \quad (4.170)$$

Equations 4.169 and 4.170 indicate that the data sensitivity kernels with respect to the relative perturbations in the Lamé parameters can be expressed as linear combinations of the data sensitivity kernels with respect to the relative perturbations in the two moduli. By re-arranging the terms in Eqs. 4.169 and 4.170 we can also express the kernels of the two moduli in terms of the kernels of the two Lamé parameters. Note that both $K_{dsn}^{\ln\mu}(\mathbf{x})$ (Eq. 4.162) and $\check{K}_{dsn}^{\ln\mu}(\mathbf{x})$ (Eq. 4.170) are the data sensitivity kernels with respect to the relative perturbation in the shear modulus. But depending upon the choice of the other parameter (i.e., either the bulk modulus κ or the first Lamé parameter λ), the expressions for $K_{dsn}^{\ln\mu}(\mathbf{x})$ and $\check{K}_{dsn}^{\ln\mu}(\mathbf{x})$ are different.

The data sensitivity kernels of the broadband cross-correlation delay time with respect to the relative perturbations in the two Lamé parameters and the density for the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown on the center rows in Figs. 4.17, 4.18 and 4.19. The expressions for the kernels with respect to the relative perturbations in the two Lamé parameters are given by Eqs. 4.167 and 4.168, respectively. The expression for the kernel with respect to the relative perturbation in the density is given in Eq. 4.178 in Sect. 4.3.4.

For the half-space example, the factor $(\kappa - \frac{2}{3}\mu)/\kappa$ in Eq. 4.169 has a constant value of about 0.6849. Compared with the kernel $K_{dsn}^{\ln\kappa}(\mathbf{x})$ shown on the left column, top row in Figs. 4.17, 4.18 and 4.19, the kernel $K_{dsn}^{\ln\lambda}(\mathbf{x})$ (left column, center row) has slightly reduced amplitudes, as predicted by Eq. 4.169. For our half-space example, the factor $(2\mu)/(3\kappa)$ in Eq. 4.170 has a constant value of about 0.3151. Compared with the kernel $K_{dsn}^{\ln\mu}(\mathbf{x})$ shown on the center column, top row in Figs. 4.17, 4.18 and 4.19, the kernel $\check{K}_{dsn}^{\ln\mu}(\mathbf{x})$ (center column, center row) has some slight imprint having the shape of $K_{dsn}^{\ln\kappa}(\mathbf{x})$, which is most obvious for the PS wave kernels (Fig. 4.19), as predicted by Eq. 4.170.

4.3.4 P- and S-wave Velocities

In isotropic solids, the fourth-order stiffness tensor can also be expressed in terms of the mass density ρ , the P-wave velocity α and S-wave velocity β , i.e.,

$$\alpha = \sqrt{\frac{\kappa + \frac{4}{3}\mu}{\rho}}, \quad \beta = \sqrt{\frac{\mu}{\rho}}, \quad (4.171)$$

and we have

$$c_{jklm} = \left(\rho\alpha^2 - 2\rho\beta^2\right) \delta_{jk}\delta_{lm} + \rho\beta^2(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl}). \quad (4.172)$$

The perturbation δc_{jklm} can then be expressed as

$$\delta c_{jklm} = \frac{\partial c_{jklm}}{\partial \rho} \delta \rho + \frac{\partial c_{jklm}}{\partial \alpha} \delta \alpha + \frac{\partial c_{jklm}}{\partial \beta} \delta \beta$$

$$\begin{aligned}
&= \left[(\alpha^2 - 2\beta^2) \delta_{jk} \delta_{lm} + \beta^2 (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \right] \delta \rho \\
&+ 2\rho \alpha (\delta_{jk} \delta_{lm}) \delta \alpha \\
&+ 2\rho \beta (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - 2\delta_{jk} \delta_{lm}) \delta \beta.
\end{aligned} \tag{4.173}$$

or in terms of the relative perturbations as

$$\begin{aligned}
\delta c_{jklm} &= \left[\kappa \delta_{jk} \delta_{lm} + \mu \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) \right] \delta \ln \rho \\
&+ 2\rho \alpha^2 (\delta_{jk} \delta_{lm}) \delta \ln \alpha \\
&+ 2\rho \beta^2 (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - 2\delta_{jk} \delta_{lm}) \delta \ln \beta.
\end{aligned} \tag{4.174}$$

Bring Eq. 4.174 into the RHS of Eq. 4.37, we obtain the data sensitivity kernels with respect to the relative perturbations in the mass density, P- and S-wave velocities,

$$\begin{aligned}
\check{K}_{dsn}^{\ln \rho}(\mathbf{x}) &= \rho(\mathbf{x}) K_{dsn}^{\rho}(\mathbf{x}) + \kappa(\mathbf{x}) \sum_{j,l=1}^3 K_{dsn}^{cjjll}(\mathbf{x}) \\
&+ \mu(\mathbf{x}) \left[\sum_{j,k=1}^3 K_{dsn}^{cjkjk}(\mathbf{x}) + \sum_{j,k=1}^3 K_{dsn}^{cjkkj}(\mathbf{x}) - \frac{2}{3} \sum_{j,l=1}^3 K_{dsn}^{cjjll}(\mathbf{x}) \right],
\end{aligned} \tag{4.175}$$

$$K_{dsn}^{\ln \alpha}(\mathbf{x}) = 2\rho \alpha^2 \sum_{j,l=1}^3 K_{dsn}^{cjjll}(\mathbf{x}), \tag{4.176}$$

$$K_{dsn}^{\ln \beta}(\mathbf{x}) = 2\rho \beta^2 \left[\sum_{j,k=1}^3 K_{dsn}^{cjkjk}(\mathbf{x}) + \sum_{j,k=1}^3 K_{dsn}^{cjkkj}(\mathbf{x}) - 2 \sum_{j,l=1}^3 K_{dsn}^{cjjll}(\mathbf{x}) \right]. \tag{4.177}$$

If we define

$$K_{dsn}^{\ln \rho}(\mathbf{x}) = \rho(\mathbf{x}) K_{dsn}^{\rho}(\mathbf{x}), \tag{4.178}$$

where the expression for $K_{dsn}^{\rho}(\mathbf{x})$ is given in Eq. 4.32, considering Eqs. 4.149 and 4.150, Eq. 4.175 can be expressed as

$$\check{K}_{dsn}^{\ln \rho}(\mathbf{x}) = K_{dsn}^{\ln \rho}(\mathbf{x}) + K_{dsn}^{\ln \kappa}(\mathbf{x}) + K_{dsn}^{\ln \mu}(\mathbf{x}). \tag{4.179}$$

Considering Eq. 4.149, Eq. 4.176 can be expressed as

$$K_{dsn}^{\ln \alpha}(\mathbf{x}) = \frac{2 \left(\kappa + \frac{4}{3} \mu \right)}{\kappa} K_{dsn}^{\ln \kappa}(\mathbf{x}). \tag{4.180}$$

Considering Eqs. 4.149 and 4.150, we can express $K_{dsn}^{\ln \beta}(\mathbf{x})$ in Eq. 4.177 as

$$K_{dsn}^{\ln \beta}(\mathbf{x}) = 2 \left[K_{dsn}^{\ln \mu}(\mathbf{x}) - \frac{4\mu}{3\kappa} K_{dsn}^{\ln \kappa}(\mathbf{x}) \right]. \tag{4.181}$$

Therefore the data sensitivity kernels with respect to the relative perturbations in P- and S-wave velocities can be expressed as linear combinations of the data sensitivity kernels with respect to the relative perturbations in the bulk and shear moduli. Note that the expression for $\check{K}_{d^{sn}}^{\ln \rho}(\mathbf{x})$ (Eq. 4.179) is different from the expression for $K_{d^{sn}}^{\ln \rho}(\mathbf{x})$ (Eqs. 4.32 and 4.178) even though both of them are the data sensitivity kernels with respect to the relative perturbation in the mass density.

Considering Eqs. 4.167 and 4.168, the data sensitivity kernels $\check{K}_{d^{sn}}^{\ln \rho}(\mathbf{x})$, $K_{d^{sn}}^{\ln \alpha}(\mathbf{x})$ and $K_{d^{sn}}^{\ln \beta}(\mathbf{x})$ can also be expressed in terms of $K_{d^{sn}}^{\ln \lambda}(\mathbf{x})$ and $\check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x})$, i.e.,

$$\check{K}_{d^{sn}}^{\ln \rho}(\mathbf{x}) = K_{d^{sn}}^{\ln \rho}(\mathbf{x}) + K_{d^{sn}}^{\ln \lambda}(\mathbf{x}) + \check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x}), \quad (4.182)$$

$$K_{d^{sn}}^{\ln \alpha}(\mathbf{x}) = \frac{2(\lambda + 2\mu)}{\lambda} K_{d^{sn}}^{\ln \lambda}(\mathbf{x}), \quad (4.183)$$

$$K_{d^{sn}}^{\ln \beta}(\mathbf{x}) = 2 \left[\check{K}_{d^{sn}}^{\ln \mu}(\mathbf{x}) - \frac{2\mu}{\lambda} K_{d^{sn}}^{\ln \lambda}(\mathbf{x}) \right]. \quad (4.184)$$

On the other hand, the data sensitivity kernels with respect to the relative perturbations in the two Lamé parameters or the two moduli can also be expressed in terms of the data sensitivity kernels with respect to the relative perturbations in P- and S-wave velocities.

Consider Eq. 4.176, we can obtain the expression for the data sensitivity kernel with respect to the relative perturbation in P-wave velocity by bringing Eq. 4.35 into Eq. 4.176, i.e.,

$$\begin{aligned} K_{d^{sn}}^{\ln \alpha}(\mathbf{x}) &= -2\rho\alpha^2 \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\ &\quad \sum_{j,l=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} * \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \\ &= -2\rho\alpha^2 \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\ &\quad \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \\ &= -2\rho(\mathbf{x})\alpha^2(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\ &\quad \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \}, \end{aligned} \quad (4.185)$$

which is consistent with Eqs. (16) and (17) in (Zhao et al. 2005)⁴.

Bring Eq. 4.35 into Eq. 4.177, we can obtain an explicit expression for the data sensitivity kernel with respect to the relative perturbation in the S-wave velocity, i.e.,

$$\begin{aligned}
K_{dsn}^{\ln\beta}(\mathbf{x}) &= -2\rho\beta^2 \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\
&\left\{ \sum_{j,k=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \left[\frac{\partial u_j^s(\mathbf{x}, t)}{\partial x_k} + \frac{\partial u_k^s(\mathbf{x}, t)}{\partial x_j} \right] \right. \\
&\quad \left. - 2 \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \right\} \\
&= -2\rho\beta^2 \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\
&\cdot \left\{ \nabla \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0) : \left[\nabla \mathbf{u}^s(\mathbf{x}, t) + [\nabla \mathbf{u}^s(\mathbf{x}, t)]^T \right] \right. \\
&\quad \left. - 2 [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \right\}, \tag{4.186}
\end{aligned}$$

where “:” denotes the temporal convolution “*” and the summation over both j and k . Equation 4.186 can be compared with Eqs. (18) and (19) in (Zhao et al. 2005)⁵.

Using the definition of the deviatoric strain tensors (Eqs. 4.155 and 4.156) and the identities in Eqs. 4.157, 4.158, 4.160 and 4.161, we can re-write Eq. 4.186 as

$$\begin{aligned}
K_{dsn}^{\ln\beta}(\mathbf{x}) &= -2\mu(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\
&\cdot \left\{ 2 \left[\check{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) : \check{\epsilon}^s(\mathbf{x}, t) \right] - \frac{4}{3} \text{tr} \left[\epsilon^G(\mathbf{x}, t; \mathbf{x}', 0) \right] * \text{tr} \left[\epsilon^s(\mathbf{x}, t) \right] \right\}. \tag{4.187}
\end{aligned}$$

⁴ Note that the Green’s tensor \mathbf{G} used in Eq. 4.185 here is the RGT, i.e., the Green’s tensor from the receiver to the scatterer, while the one used in Eqs. (16)–(17) in (Zhao et al. 2005) is the Green’s tensor from the scatterer to the receiver. Therefore no transpose operation is applied on the RGT in Eq. 4.185. The kernel given in Eq. 4.185 is for the relative perturbation $\delta\alpha/\alpha$, while the kernels in Eqs. (16)–(17) in (Zhao et al. 2005) are for the absolute perturbation $\delta\alpha$. Therefore Eq. 4.185 has the factor $\alpha^2(\mathbf{x})$, while Eqs. (16)–(17) in (Zhao et al. 2005) have the factor $\alpha(\mathbf{x})$.

⁵ Note that no transpose operation is needed on the RGT \mathbf{G} in Eq. 4.186 here, while the Green’s tensor used in Eqs. (18)–(19) in (Zhao et al. 2005), which is from the scatterer to the receiver, needs to be transposed. The kernel given in Eq. 4.186 here is for the relative perturbation $\delta\beta/\beta$, while those in Eqs. (18)–(19) in (Zhao et al. 2005) are for the absolute perturbation $\delta\beta$. Therefore Eq. 4.186 here has the factor $\beta^2(\mathbf{x})$, while Eqs. (18)–(19) in (Zhao et al. 2005) have the factor $\beta(\mathbf{x})$. Equation 4.186 here and Eqs. (18)–(19) in (Zhao et al. 2005) differ by a minus sign.

Considering Eqs. 4.162 and 4.145, Eq. 4.187 can be expressed in terms of $K_{d^{sn}}^{\ln\kappa}(\mathbf{x})$ and $K_{d^{sn}}^{\ln\mu}(\mathbf{x})$ as

$$K_{d^{sn}}^{\ln\beta}(\mathbf{x}) = 2 \left[K_{d^{sn}}^{\ln\mu}(\mathbf{x}) - \frac{4\mu(\mathbf{x})}{3\kappa(\mathbf{x})} K_{d^{sn}}^{\ln\kappa}(\mathbf{x}) \right], \quad (4.188)$$

which verifies Eq. 4.181.

The data sensitivity kernels of the broadband cross-correlation delay time with respect to the relative perturbations in the P- and S-wave velocities and the density for the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown on the bottom rows in Figs. 4.17, 4.18 and 4.19. The expressions for the kernels with respect to the relative perturbations in the P- and S-wave velocities are given by Eqs. 4.185 and 4.186, respectively. The expression for the kernel with respect to the relative perturbation in the density is given in Eq. 4.175.

For the half-space example, the factor $2 \left(\kappa + \frac{4}{3}\mu \right) / \kappa$ in Eq. 4.180 is a constant with the value at about 3.2605. As shown in Figs. 4.17, 4.18 and 4.19, the amplitudes of the kernels with respect to the relative perturbation in α are about three times larger than the amplitudes of the kernels with respect to κ , as predicted by Eq. 4.180. The factor $4\mu/(3\kappa)$ in Eq. 4.181 is a constant of about 0.6302. We can examine the kernels for the PS wave on Figs. 4.19. For the S-wave leg from the free surface to the receiver, the amplitude of $K_{d^{sn}}^{\ln\beta}$ (center column, bottom row) is about twice as large as $K_{d^{sn}}^{\ln\mu}$ (center column, top row) and for the P-wave leg from the source to the free surface, $K_{d^{sn}}^{\ln\beta}$ has a similar shape with $K_{d^{sn}}^{\ln\kappa}$ (left column, top row) but with the opposite sign, which is predicted by Eq. 4.181. Unlike the density kernels shown on the top and center rows in Figs. 4.17, 4.18 and 4.19, the density kernel $\check{K}_{d^{sn}}^{\ln\rho}$ shown on the right column, bottom row in Figs. 4.17, 4.18 and 4.19 has much smaller amplitudes, as predicted by Eq. 4.179.

4.3.5 Love's Elastic Parameters (Hexagonal Symmetry)

For anisotropic material with hexagonal symmetry, there are five independent elastic coefficients, which are often denoted as A , C , L , N , F in Love's notation (Dziewonowski and Anderson 1981; Love 2013). The five Love's parameters can be related to the components of the fourth-order stiffness tensor once the orientation of the symmetry axis with respect to the geographical coordinate system is determined. In the seismological literature, a widely used Cartesian coordinate system is the one used in e.g., (Aki and Richards 2002), with the x_1 axis pointing to the North, the x_2 axis pointing to the East and the x_3 axis oriented downward. We will call this coordinate system the A&R coordinate. The coordinate system used in our software package F3DWI has the x_3 axis pointing up, away from the center of the Earth. The two horizontal axes in F3DWI do not have to point to the East and North directions, but for

Table 4.1 The Voigt convention for assigning components in the fourth-order tensor c_{jklm} to components in the 6-by-6 Voigt matrix C_{ab} . As a and b vary from 1 to 6, the indices (j, k) and (l, m) take on values (1,1), (2,2), (3,3), (2,3) or (3,2), (1,3) or (3,1), (1,2) or (2,1)

jk	lm	11	22	33	23 = 32	13 = 31	12 = 21
a	b	1	2	3	4	5	6

the sake of simplicity we assume that the x_1 axis points to the East and the x_2 axis points to the North in Sects. 4.3.5 and 4.3.6. The coordinate transform between the A&R coordinate and the F3DWI coordinate is therefore given by the transformation matrix

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (4.189)$$

For the fourth-order elastic stiffness tensor c'_{nopq} given in the A&R coordinate system, its representation in the F3DWI coordinate system is given by

$$c_{jklm} = \sum_{n,o,p,q=1}^3 R_{jn} R_{ko} R_{lp} R_{mq} c'_{nopq}, \quad (4.190)$$

where R_{jn} is the index notation of \mathbf{R} in Eq. 4.189. In tensor notation, we have

$$\mathbf{c} = \mathbf{R} \cdot \mathbf{R} \cdot \mathbf{c}' \cdot \mathbf{R}^T \cdot \mathbf{R}^T, \quad (4.191)$$

where “ \cdot ” represents tensor contraction, \mathbf{R}^T is the transpose of \mathbf{R} in Eq. 4.189, \mathbf{c} and \mathbf{c}' are the fourth-order stiffness tensor in the F3DWI coordinate system and the A&R coordinate system respectively. If the coordinate transformation matrix is different from that shown in Eq. 4.189, Eqs. 4.190 and 4.191 are still applicable. For an arbitrary coordinate rotation, the matrix \mathbf{R} is an orthogonal matrix, i.e., $\mathbf{R}^T \mathbf{R} = \mathbf{I}$, and we can obtain \mathbf{c}' from a given \mathbf{c} by reversing the relation given in Eqs. 4.190 and 4.191, i.e.,

$$\mathbf{c}' = \mathbf{R}^T \cdot \mathbf{R}^T \cdot \mathbf{c} \cdot \mathbf{R} \cdot \mathbf{R}. \quad (4.192)$$

Because of the symmetry conditions (Eq. 4.136), the stiffness of an anisotropic medium can be described using a 6-by-6 matrix, often denoted as C_{ab} , instead of the full fourth-order tensor c_{jklm} . The relation between the components of the C_{ab} matrix and the components in c_{jklm} is often established following the Voigt convention (Voigt 1928), as shown in Table 4.1. For a stiffness tensor \mathbf{c}' given in the A&R

coordinate, the corresponding Voigt matrix is then given by

$$\mathbf{C} = \begin{bmatrix} c'_{1111} & c'_{1122} & c'_{1133} & c'_{1123} & c'_{1113} & c'_{1112} \\ c'_{2211} & c'_{2222} & c'_{2233} & c'_{2223} & c'_{2213} & c'_{2212} \\ c'_{3311} & c'_{3322} & c'_{3333} & c'_{3323} & c'_{3313} & c'_{3312} \\ c'_{2311} & c'_{2322} & c'_{2333} & c'_{2323} & c'_{2313} & c'_{2312} \\ c'_{1311} & c'_{1322} & c'_{1333} & c'_{1323} & c'_{1313} & c'_{1312} \\ c'_{1211} & c'_{1222} & c'_{1233} & c'_{1223} & c'_{1213} & c'_{1212} \end{bmatrix}. \quad (4.193)$$

Considering the symmetry condition $c'_{jklm} = c'_{lmjk}$ (Eq. 4.136), the Voigt matrix in Eq. 4.193 is symmetric, which allows us to reduce the number of independent components in the Voigt matrix from 36 to 21. Considering the relation between \mathbf{c}' and \mathbf{c} given in Eq. 4.192, the Voigt matrix in Eq. 4.193 can also be expressed in terms of the components of \mathbf{c} in the F3DWI coordinate as

$$\mathbf{C} = \begin{bmatrix} c_{2222} & c_{2211} & c_{2233} & -c_{2213} & -c_{2223} & c_{2221} \\ c_{1122} & c_{1111} & c_{1133} & -c_{1113} & -c_{1123} & c_{1121} \\ c_{3322} & c_{3311} & c_{3333} & -c_{3313} & -c_{3323} & c_{3321} \\ -c_{1322} & -c_{1311} & -c_{1333} & c_{1313} & c_{1323} & -c_{1321} \\ -c_{2322} & -c_{2311} & -c_{2333} & c_{2313} & c_{2323} & -c_{2321} \\ c_{2122} & c_{2111} & c_{2133} & -c_{2113} & -c_{2123} & c_{2121} \end{bmatrix}. \quad (4.194)$$

The data sensitivity kernels with respect to the 21 independent components in the Voigt matrix in Eq. 4.194 can then be obtained from the data sensitivity kernels with respect to the corresponding components in the fourth-order stiffness tensor given by Eq. 4.35.

In the hexagonal case, if the vertical axis x_3 in the A&R coordinate is the symmetry axis, the Voigt matrix in Eq. 4.194 can be expressed in terms of the five Love's elastic parameters as (e.g., Babuska and Cara 1991)

$$\mathbf{C} = \begin{bmatrix} A & A - 2N & F & 0 & 0 & 0 \\ A - 2N & A & F & 0 & 0 & 0 \\ F & F & C & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & L & 0 \\ 0 & 0 & 0 & 0 & 0 & N \end{bmatrix}. \quad (4.195)$$

Comparing Eq. 4.195 with Eq. 4.194, we obtain

$$c_{2222} = c_{1111} = A, \quad (4.196)$$

$$c_{3333} = C, \quad (4.197)$$

$$c_{1313} = c_{3113} = c_{1331} = c_{3131} = c_{2323} = c_{3223} = c_{2332} = c_{3232} = L, \quad (4.198)$$

$$c_{2121} = c_{1221} = c_{2112} = c_{1212} = N, \quad (4.199)$$

$$c_{2211} = c_{1122} = A - 2N, \quad (4.200)$$

$$c_{2233} = c_{3322} = c_{1133} = c_{3311} = F. \quad (4.201)$$

The perturbations in the relevant c_{ijklm} can then be expressed in terms of the perturbations in the five Love's elastic parameters as

$$\delta c_{2222} = \delta c_{1111} = \delta A, \quad (4.202)$$

$$\delta c_{3333} = \delta C, \quad (4.203)$$

$$\begin{aligned} \delta c_{1313} &= \delta c_{3113} = \delta c_{1331} = \delta c_{3131} \\ &= \delta c_{2323} = \delta c_{3223} = \delta c_{2332} = \delta c_{3232} = \delta L, \end{aligned} \quad (4.204)$$

$$\delta c_{2121} = \delta c_{1221} = \delta c_{2112} = \delta N, \quad (4.205)$$

$$\delta c_{2211} = \delta c_{1122} = \delta A - 2\delta N, \quad (4.206)$$

$$\delta c_{2233} = \delta c_{3322} = \delta c_{1133} = \delta c_{3311} = \delta F. \quad (4.207)$$

Bring Eqs. 4.202–4.207 into the last term on the RHS of Eq. 4.37 and collect all the terms containing δA , we obtain the data sensitivity kernel with respect to A , i.e.,

$$K_{dsn}^A(\mathbf{x}) = K_{dsn}^{c_{1111}}(\mathbf{x}) + K_{dsn}^{c_{2222}}(\mathbf{x}) + K_{dsn}^{c_{2211}}(\mathbf{x}) + K_{dsn}^{c_{1122}}(\mathbf{x}). \quad (4.208)$$

Collect all the terms containing δC , we obtain the data sensitivity kernel with respect to C , i.e.,

$$K_{dsn}^C(\mathbf{x}) = K_{dsn}^{c_{3333}}(\mathbf{x}). \quad (4.209)$$

Collect all the terms containing δL , we obtain the data sensitivity kernel with respect to L , i.e.,

$$\begin{aligned} K_{dsn}^L(\mathbf{x}) &= K_{dsn}^{c_{1313}}(\mathbf{x}) + K_{dsn}^{c_{3113}}(\mathbf{x}) + K_{dsn}^{c_{1331}}(\mathbf{x}) + K_{dsn}^{c_{3131}}(\mathbf{x}) \\ &\quad + K_{dsn}^{c_{2323}}(\mathbf{x}) + K_{dsn}^{c_{3223}}(\mathbf{x}) + K_{dsn}^{c_{2332}}(\mathbf{x}) + K_{dsn}^{c_{3232}}(\mathbf{x}). \end{aligned} \quad (4.210)$$

Collect all the terms containing δN , we obtain the data sensitivity kernel with respect to N , i.e.,

$$\begin{aligned} K_{dsn}^N(\mathbf{x}) &= K_{dsn}^{c_{2121}}(\mathbf{x}) + K_{dsn}^{c_{1221}}(\mathbf{x}) + K_{dsn}^{c_{2112}}(\mathbf{x}) + K_{dsn}^{c_{1212}}(\mathbf{x}) \\ &\quad - 2K_{dsn}^{c_{2211}}(\mathbf{x}) - 2K_{dsn}^{c_{1122}}(\mathbf{x}). \end{aligned} \quad (4.211)$$

Collect all the terms containing δF , we obtain the data sensitivity kernel with respect to F , i.e.,

$$K_{dsn}^F(\mathbf{x}) = K_{dsn}^{c2233}(\mathbf{x}) + K_{dsn}^{c3322}(\mathbf{x}) + K_{dsn}^{c1133}(\mathbf{x}) + K_{dsn}^{c3311}(\mathbf{x}). \quad (4.212)$$

If we bring Eq. 4.35 into the RHS of Eqs. 4.208–4.212, we can verify the following relations between the data sensitivity kernels with respect to the P- and S-wave velocities in the isotropic medium (Eqs. 4.185 and 4.186 in Sect. 4.3.4) and the data sensitivity kernels with respect to the five Love's parameters (e.g., Sieminski et al. 2007),

$$K_{dsn}^{\ln \alpha}(\mathbf{x}) = 2\rho\alpha^2 \left[K_{dsn}^A(\mathbf{x}) + K_{dsn}^C(\mathbf{x}) + K_{dsn}^F(\mathbf{x}) \right], \quad (4.213)$$

$$K_{dsn}^{\ln \beta}(\mathbf{x}) = 2\rho\beta^2 \left[K_{dsn}^L(\mathbf{x}) + K_{dsn}^N(\mathbf{x}) - 2K_{dsn}^F(\mathbf{x}) \right]. \quad (4.214)$$

The hexagonal symmetry case is of considerable practical interest because many realistic anisotropy problems can be approximated using this relatively simple case. Because the elastic properties do not vary in the plane perpendicular to the symmetry axis, hexagonal symmetry is also called *transverse isotropy*. If the symmetry axis is vertical, this type of anisotropy is also called *azimuthal isotropy*. An example of transverse isotropy with a vertical symmetry axis is the apparent anisotropy produced by fine layering of isotropic sediments inside a sedimentary basin. In such a situation, the amount of anisotropy is proportional to the vertical variance in the elastic moduli (Backus 1962). In such a structure, the P-wave velocity depends upon the incident angle of the wave. The velocity of a refracted P-wave propagating horizontally, denoted as α_H , is often larger than the velocity of a reflected P-wave propagating vertically, denoted as α_V . The two P-wave velocities are related to the Love's parameters via

$$\alpha_H = \sqrt{\frac{A}{\rho}}, \quad \alpha_V = \sqrt{\frac{C}{\rho}}, \quad (4.215)$$

and the two types of P-waves are often called P_H and P_V waves, respectively. Horizontally propagating S-waves can have different velocities depending upon their polarization. The velocity of a horizontally propagating S-wave with horizontal polarization, denoted as β_H is larger than the velocity of a horizontally propagating S-wave with vertical polarization, denoted as β_V . The two S-wave velocities are related to the Love's parameters via

$$\beta_H = \sqrt{\frac{N}{\rho}}, \quad \beta_V = \sqrt{\frac{L}{\rho}}, \quad (4.216)$$

and the two S-waves are often called S_H and S_V waves, respectively. For vertically propagating S-waves, the S-wave velocity does not depend upon the polarization and equals to β_V . Data sensitivity kernels with respect to the two P-wave velocities and the two S-wave velocities can also be obtained by using the data sensitivity kernels with respect to the Love's parameters and the chain rule.

Table 4.2 The 21 asymptotic parameters as denoted in (Chen and Tromp 2007). The angle ψ is the local azimuth along the geometric ray path. The 21 asymptotic parameters are grouped according to the azimuthal dependence and wave types they control. Subscripts c,s correspond to azimuthal dependence on $\cos(l\psi)$ and $\sin(l\psi)$, respectively, where $l = 1, 2, 3, 4$

	0ψ	1ψ	2ψ	3ψ	4ψ
Quasi-Rayleigh	A, C, L, N, F		$B_{c,s}, G_{c,s}, H_{c,s}$		$E_{c,s}$
Quasi-Love	L, N		$G_{c,s}$		$E_{c,s}$
Quasi-P	A, C, L, F	$J_{c,s}, K_{c,s}$	$B_{c,s}, G_{c,s}, H_{c,s}$	$D_{c,s}$	$E_{c,s}$
Quasi-S	A, C, L, N, F	$K_{c,s}, M_{c,s}$	$B_{c,s}, G_{c,s}, H_{c,s}$	$D_{c,s}$	$E_{c,s}$

It is also possible to use other types of parameterizations, such as Thomsen’s parameters (Thomsen 1986; Chevrot 2006; Panning and Nolet 2008), which can also be expressed in terms of components of the Voigt matrix C_{ab} . The data sensitivity kernels with respect to Thomsen’s parameters can also be obtained by using the partial derivatives of C_{ab} with respect to the five Thomsen’s parameters and the chain rule.

4.3.6 Asymptotic Parameters (Triclinic Symmetry)

In general anisotropic media with triclinic symmetry, we have 21 independent elastic parameters. One possible parameterization is to use the 21 independent components in the Voigt matrix C_{ab} and it is relatively straightforward to obtain the data sensitivity kernels with respect to C_{ab} by using the relations between C_{ab} and components of the fourth-order stiffness tensor c_{ijklm} as shown in Eq. 4.194. A different parameterization that may have more physical meaning is to use the 21 “asymptotic parameters” (Chen and Tromp 2007), which are related to the azimuthal dependence of the phase velocity for asymptotic wave propagation in weakly anisotropic media (Smith and Dahlen 1973; Maupin 1985; Montagner and Nataf 1986; Romanowicz and Snieder 1988; Larson et al. 1998; Chen and Tromp 2007). The 21 asymptotic parameters used in (Chen and Tromp 2007) are listed in Table 4.2 according to the azimuthal dependences and the wave types these parameters control. The five Love’s parameters A, C, L, N, F only control the transverse isotropy with a vertical symmetry axis (Sect. 4.3.5), therefore they do not introduce azimuthal anisotropy. For surface-wave anisotropy studies, the 13 parameters $A, C, L, N, F, B_{c,s}, G_{c,s}, H_{c,s}, E_{c,s}$ determine the transverse isotropy and the azimuthal anisotropy with 2ψ - and 4ψ -dependence. For body-wave studies, 8 more parameters $J_{c,s}, K_{c,s}, M_{c,s}, D_{c,s}$ need to be introduced to determine the azimuthal anisotropy with 1ψ - and 3ψ -dependence.

The 21 asymptotic parameters used in (Chen and Tromp 2007) can be related to the 21 independent components of the Voigt matrix in Eq. 4.193 through the following equations for the 0ψ terms,

$$A = 3(C_{11} + C_{22})/8 + C_{12}/4 + C_{66}/2, \quad (4.217)$$

$$C = C_{33}, \quad (4.218)$$

$$N = (C_{11} + C_{22})/8 - C_{12}/4 + C_{66}/2, \quad (4.219)$$

$$L = (C_{44} + C_{55})/2, \quad (4.220)$$

$$F = (C_{13} + C_{23})/2, \quad (4.221)$$

the 1ψ terms,

$$J_c = 3C_{15}/8 + C_{25}/8 + C_{46}/4, \quad (4.222)$$

$$J_s = -C_{14}/8 - 3C_{24}/8 - C_{56}/4, \quad (4.223)$$

$$K_c = 3C_{15}/8 + C_{25}/8 + C_{46}/4 - C_{35}/2, \quad (4.224)$$

$$K_s = -C_{14}/8 - 3C_{24}/8 - C_{56}/4 + C_{34}/2, \quad (4.225)$$

$$M_c = (C_{15} - C_{25})/4 + C_{46}/2, \quad (4.226)$$

$$M_s = (-C_{14} + C_{24})/4 + C_{56}/2, \quad (4.227)$$

the 2ψ terms,

$$G_c = (C_{55} - C_{44})/2, \quad (4.228)$$

$$G_s = C_{45}, \quad (4.229)$$

$$B_c = (C_{11} - C_{22})/2, \quad (4.230)$$

$$B_s = C_{16} + C_{26}, \quad (4.231)$$

$$H_c = (C_{13} - C_{23})/2, \quad (4.232)$$

$$H_s = C_{36}, \quad (4.233)$$

the 3ψ terms,

$$D_c = (C_{15} - C_{25})/4 - C_{46}/2, \quad (4.234)$$

$$D_s = (-C_{14} + C_{24})/4 - C_{56}/2, \quad (4.235)$$

and the 4ψ terms.

$$E_c = (C_{11} + C_{22})/8 - C_{12}/4 - C_{66}/2, \quad (4.236)$$

$$E_s = (C_{16} - C_{26})/2. \quad (4.237)$$

The equations given above are consistent with those in (Babuska and Cara 1991; Chen and Tromp 2007)⁶. This set of equations can be solved for the 21 independent components of the Voigt matrix C_{ab} and we can obtain the expressions for C_{ab} in terms of the 21 asymptotic parameters. Considering the relation between C_{ab} and

⁶ Note that the expressions for $J_s, K_s, M_s, G_s, B_s, H_s, D_s, E_s$ given here differ from the corresponding equations in (Chen and Tromp 2007) by a minus sign, which is due to differences in the coordinate system. The coordinate system used here is the A&R coordinate, which is consistent with the one used in Sect. 2.5.3 of (Babuska and Cara 1991), therefore expressions for G_s, B_s, H_s, E_s given here are identical to the corresponding equations given in Sect. 2.5.3 of (Babuska and Cara 1991).

the components of the fourth-order stiffness tensor c_{ijklm} in the $F3DWI$ coordinate system given in Eq. 4.194, we can obtain the expressions for c_{ijklm} in terms of the 21 asymptotic parameters, i.e.,

$$c_{2222} = C_{11} = A + B_c + E_c, \quad (4.238)$$

$$c_{1111} = C_{22} = A - B_c + E_c, \quad (4.239)$$

$$c_{3333} = C_{33} = C, \quad (4.240)$$

$$c_{1313} = c_{3113} = c_{1331} = c_{3131} = C_{44} = L - G_c, \quad (4.241)$$

$$c_{2323} = c_{3223} = c_{2332} = c_{3232} = C_{55} = L + G_c, \quad (4.242)$$

$$c_{2121} = c_{1221} = c_{2112} = c_{1212} = C_{66} = N - E_c, \quad (4.243)$$

$$c_{2211} = c_{1122} = C_{12} = A - 2N - E_c, \quad (4.244)$$

$$c_{2233} = c_{3322} = C_{13} = F + H_c, \quad (4.245)$$

$$c_{2213} = c_{2231} = c_{1322} = c_{3122} = -C_{14} = 2J_s + 2M_s + D_s, \quad (4.246)$$

$$c_{2223} = c_{2232} = c_{2322} = c_{3222} = -C_{15} = -2J_c - D_c, \quad (4.247)$$

$$c_{2221} = c_{2212} = c_{2122} = c_{1222} = C_{16} = B_s/2 + E_s, \quad (4.248)$$

$$c_{1133} = c_{3311} = C_{23} = F - H_c, \quad (4.249)$$

$$c_{1113} = c_{1131} = c_{1311} = c_{3111} = -C_{24} = -D_s + 2J_s, \quad (4.250)$$

$$c_{1123} = c_{1132} = c_{2311} = c_{3211} = -C_{25} = D_c - 2J_c + 2M_c, \quad (4.251)$$

$$c_{1121} = c_{1112} = c_{2111} = c_{1211} = C_{26} = B_s/2 - E_s, \quad (4.252)$$

$$c_{3313} = c_{3331} = c_{1333} = c_{3133} = -C_{34} = 2J_s - 2K_s, \quad (4.253)$$

$$c_{3323} = c_{3332} = c_{2333} = c_{3233} = -C_{35} = -2J_c + 2K_c, \quad (4.254)$$

$$c_{3321} = c_{3312} = c_{2133} = c_{1233} = C_{36} = H_s, \quad (4.255)$$

$$c_{1323} = c_{3123} = c_{1332} = c_{3132}$$

$$= c_{2313} = c_{2331} = c_{3213} = c_{3231} = C_{45} = G_s, \quad (4.256)$$

$$c_{1321} = c_{3121} = c_{1312} = c_{3112}$$

$$= c_{2113} = c_{2131} = c_{1213} = c_{1231} = -C_{46} = D_c - M_c \quad (4.257)$$

$$c_{2321} = c_{3221} = c_{2312} = c_{3212}$$

$$= c_{2123} = c_{2132} = c_{1223} = c_{1232} = -C_{56} = D_s - M_s. \quad (4.258)$$

Perturbations of the fourth-order stiffness tensor components δc_{ijklm} can now be expressed in terms of the perturbations of the 21 asymptotic parameters using this set of equations. Bringing the perturbations δc_{ijklm} into the last term on the RHS of Eq. 4.37 and collect the terms containing the perturbation of each asymptotic parameter, we can obtain the data sensitivity kernel with respect to each asymptotic

parameter. The data sensitivity kernels with respect to the five 0ψ terms are identical to the data sensitivity kernels with respect to the Love's parameters (Eqs. 4.208–4.212),

$$K_{d^{sn}}^A = K_{d^{sn}}^{C1111} + K_{d^{sn}}^{C2222} + K_{d^{sn}}^{C1122} + K_{d^{sn}}^{C2211}, \quad (4.259)$$

$$K_{d^{sn}}^C = K_{d^{sn}}^{C3333}, \quad (4.260)$$

$$\begin{aligned} K_{d^{sn}}^N &= K_{d^{sn}}^{C2121} + K_{d^{sn}}^{C1221} + K_{d^{sn}}^{C2112} + K_{d^{sn}}^{C1212} \\ &\quad - 2K_{d^{sn}}^{C2211} - 2K_{d^{sn}}^{C1122}, \end{aligned} \quad (4.261)$$

$$\begin{aligned} K_{d^{sn}}^L &= K_{d^{sn}}^{C1313} + K_{d^{sn}}^{C3113} + K_{d^{sn}}^{C1331} + K_{d^{sn}}^{C3131} \\ &\quad + K_{d^{sn}}^{C2323} + K_{d^{sn}}^{C3223} + K_{d^{sn}}^{C2332} + K_{d^{sn}}^{C3232}, \end{aligned} \quad (4.262)$$

$$K_{d^{sn}}^F = K_{d^{sn}}^{C2233} + K_{d^{sn}}^{C3322} + K_{d^{sn}}^{C1133} + K_{d^{sn}}^{C3311}. \quad (4.263)$$

Here I have drop the kernel's dependence on \mathbf{x} to reduce clutter. The data sensitivity kernels with respect to the 1ψ terms, $J_{c,s}$, $K_{c,s}$, $M_{c,s}$, can be expressed as

$$\begin{aligned} K_{d^{sn}}^{Jc} &= -2[K_{d^{sn}}^{C1123} + K_{d^{sn}}^{C1132} + K_{d^{sn}}^{C2311} + K_{d^{sn}}^{C3211} \\ &\quad + K_{d^{sn}}^{C2223} + K_{d^{sn}}^{C2232} + K_{d^{sn}}^{C2322} + K_{d^{sn}}^{C3222} \\ &\quad + K_{d^{sn}}^{C3323} + K_{d^{sn}}^{C3332} + K_{d^{sn}}^{C2333} + K_{d^{sn}}^{C3233}], \end{aligned} \quad (4.264)$$

$$\begin{aligned} K_{d^{sn}}^{Js} &= 2[K_{d^{sn}}^{C1113} + K_{d^{sn}}^{C1131} + K_{d^{sn}}^{C1311} + K_{d^{sn}}^{C3111} \\ &\quad + K_{d^{sn}}^{C2213} + K_{d^{sn}}^{C2231} + K_{d^{sn}}^{C1322} + K_{d^{sn}}^{C3122} \\ &\quad + K_{d^{sn}}^{C3313} + K_{d^{sn}}^{C3331} + K_{d^{sn}}^{C1333} + K_{d^{sn}}^{C3133}], \end{aligned} \quad (4.265)$$

$$K_{d^{sn}}^{Kc} = 2[K_{d^{sn}}^{C3323} + K_{d^{sn}}^{C3332} + K_{d^{sn}}^{C2333} + K_{d^{sn}}^{C3233}], \quad (4.266)$$

$$K_{d^{sn}}^{Ks} = -2[K_{d^{sn}}^{C3313} + K_{d^{sn}}^{C3331} + K_{d^{sn}}^{C1333} + K_{d^{sn}}^{C3133}], \quad (4.267)$$

$$\begin{aligned} K_{d^{sn}}^{Mc} &= 2[K_{d^{sn}}^{C1123} + K_{d^{sn}}^{C1132} + K_{d^{sn}}^{C2311} + K_{d^{sn}}^{C3211}] \\ &\quad - [K_{d^{sn}}^{C1321} + K_{d^{sn}}^{C3121} + K_{d^{sn}}^{C1312} + K_{d^{sn}}^{C3112} \\ &\quad + K_{d^{sn}}^{C2113} + K_{d^{sn}}^{C2131} + K_{d^{sn}}^{C1213} + K_{d^{sn}}^{C1231}], \end{aligned} \quad (4.268)$$

$$\begin{aligned} K_{d^{sn}}^{Ms} &= 2[K_{d^{sn}}^{C2213} + K_{d^{sn}}^{C2231} + K_{d^{sn}}^{C1322} + K_{d^{sn}}^{C3122}] \\ &\quad - [K_{d^{sn}}^{C2321} + K_{d^{sn}}^{C3221} + K_{d^{sn}}^{C2312} + K_{d^{sn}}^{C3212} \\ &\quad + K_{d^{sn}}^{C2123} + K_{d^{sn}}^{C2132} + K_{d^{sn}}^{C1223} + K_{d^{sn}}^{C1232}]. \end{aligned} \quad (4.269)$$

The data sensitivity kernels with respect to the 2ψ terms, $G_{c,s}$, $B_{c,s}$, $H_{c,s}$, can be expressed as

$$K_{dsn}^{G_c} = - \left[K_{dsn}^{c1313} + K_{dsn}^{c3113} + K_{dsn}^{c1331} + K_{dsn}^{c3131} \right] + \left[K_{dsn}^{c2323} + K_{dsn}^{c3223} + K_{dsn}^{c2332} + K_{dsn}^{c3232} \right], \quad (4.270)$$

$$K_{dsn}^{G_s} = K_{dsn}^{c1323} + K_{dsn}^{c3123} + K_{dsn}^{c1332} + K_{dsn}^{c3132} + K_{dsn}^{c2313} + K_{dsn}^{c3213} + K_{dsn}^{c2331} + K_{dsn}^{c3231}, \quad (4.271)$$

$$K_{dsn}^{B_c} = K_{dsn}^{c2222} - K_{dsn}^{c1111}, \quad (4.272)$$

$$K_{dsn}^{B_s} = \left[K_{dsn}^{c2221} + K_{dsn}^{c2212} + K_{dsn}^{c2122} + K_{dsn}^{c1222} + K_{dsn}^{c1121} + K_{dsn}^{c1112} + K_{dsn}^{c2111} + K_{dsn}^{c1211} \right] / 2, \quad (4.273)$$

$$K_{dsn}^{H_c} = K_{dsn}^{c2233} + K_{dsn}^{c3322} - K_{dsn}^{c1133} - K_{dsn}^{c3311}, \quad (4.274)$$

$$K_{dsn}^{H_s} = K_{dsn}^{c3321} + K_{dsn}^{c3312} + K_{dsn}^{c2133} + K_{dsn}^{c1233}. \quad (4.275)$$

The data sensitivity kernels with respect to the 3ψ terms, $D_{c,s}$, can be expressed as

$$K_{dsn}^{D_c} = - \left[K_{dsn}^{c2223} + K_{dsn}^{c2232} + K_{dsn}^{c2322} + K_{dsn}^{c3222} \right] + K_{dsn}^{c1123} + K_{dsn}^{c1132} + K_{dsn}^{c2311} + K_{dsn}^{c3211} + K_{dsn}^{c1321} + K_{dsn}^{c3121} + K_{dsn}^{c1312} + K_{dsn}^{c3112} + K_{dsn}^{c2113} + K_{dsn}^{c2131} + K_{dsn}^{c1213} + K_{dsn}^{c1231}, \quad (4.276)$$

$$K_{dsn}^{D_s} = K_{dsn}^{c2213} + K_{dsn}^{c2231} + K_{dsn}^{c1322} + K_{dsn}^{c3122} - \left[K_{dsn}^{c1113} + K_{dsn}^{c1131} + K_{dsn}^{c1311} + K_{dsn}^{c3111} \right] + K_{dsn}^{c2321} + K_{dsn}^{c3221} + K_{dsn}^{c2312} + K_{dsn}^{c3212} + K_{dsn}^{c2123} + K_{dsn}^{c2132} + K_{dsn}^{c1223} + K_{dsn}^{c1232}. \quad (4.277)$$

The data sensitivity kernels with respect to the 4ψ terms, $E_{c,s}$, can be expressed as

$$K_{dsn}^{E_c} = K_{dsn}^{c1111} + K_{dsn}^{c2222} - \left[K_{dsn}^{c2121} + K_{dsn}^{c1221} + K_{dsn}^{c2112} + K_{dsn}^{c1212} + K_{dsn}^{c2211} + K_{dsn}^{c1122} \right], \quad (4.278)$$

$$K_{dsn}^{E_s} = K_{dsn}^{c2221} + K_{dsn}^{c2212} + K_{dsn}^{c2122} + K_{dsn}^{c1222} - \left[K_{dsn}^{c1121} + K_{dsn}^{c1112} + K_{dsn}^{c2111} + K_{dsn}^{c1211} \right]. \quad (4.279)$$

Expressions above have shown that the data sensitivity kernels with respect to the 21 asymptotic parameters are linear combinations of the data sensitivity kernels with respect to the components of the fourth-order stiffness tensor. By bringing Eq. 4.35 into the RHS of Eqs. 4.259–4.279, we can obtain explicit expressions for the data sensitivity kernels with respect to the asymptotic parameters in terms of the RGT and the forward wavefield.

The data sensitivity kernels of the broadband cross-correlation delay time with respect to the perturbations in the 21 asymptotic parameters for the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example are shown in Figs. 4.20, 4.21 and 4.22. The data sensitivity kernels of other types of data functionals, such as shear-wave splitting measurements, can be obtained by replacing the WPK for the broadband cross-correlation delay time with the WPK for the splitting measurements.

In realistic tomographic inversions, it is usually not possible to constrain all 21 asymptotic parameters simultaneously. For a given data set (i.e., a collection of waveforms that will be used in the inversion), the principle component analysis (Sieminski et al. 2009) can be used to identify a few asymptotic parameters or a few linear combinations of the asymptotic parameters that can explain most of the data.

4.3.7 Anelastic Attenuation Quality Factors

The rate-of-relaxation tensor $\Psi_{jklm}(\mathbf{x}, t)$ (Eq. 3.200) allows us to account for fully anisotropic and also anelastic material properties. However, in practice anisotropic anelasticity is usually negligible and it is often sufficient to consider isotropic anelasticity (e.g., Dahlen and Tromp 1998), which is often characterized using the bulk and shear quality factors $Q_\kappa(\mathbf{x}, \omega)$, $Q_\mu(\mathbf{x}, \omega)$, or the P- and S-wave quality factors $Q_\alpha(\mathbf{x}, \omega)$, $Q_\beta(\mathbf{x}, \omega)$. These two sets of quality factors are related through (Dahlen and Tromp 1998; Dahlen and Zhou 2006)

$$Q_\alpha^{-1} = \left(1 - \frac{4\beta^2}{3\alpha^2}\right) Q_\kappa^{-1} + \left(\frac{4\beta^2}{3\alpha^2}\right) Q_\mu^{-1}, \quad (4.280)$$

$$Q_\beta^{-1} = Q_\mu^{-1}, \quad (4.281)$$

where α and β are P- and S-wave velocities (Eq. 4.171). For the constant-Q absorption-band model (e.g., Liu et al. 1976; Kanamori and Anderson 1977; Dahlen and Tromp 1998), the quality factors can be considered as frequency-independent over a wide frequency band, i.e.,

$$Q_\kappa^{-1}(\mathbf{x}, \omega) \approx Q_\kappa^{-1}(\mathbf{x}), \quad Q_\mu^{-1}(\mathbf{x}, \omega) \approx Q_\mu^{-1}(\mathbf{x}). \quad (4.282)$$

In such a case, the frequency-domain bulk and shear moduli are complex, frequency-dependent functions and can be expressed as (e.g., Dahlen and Tromp 1998; Tromp

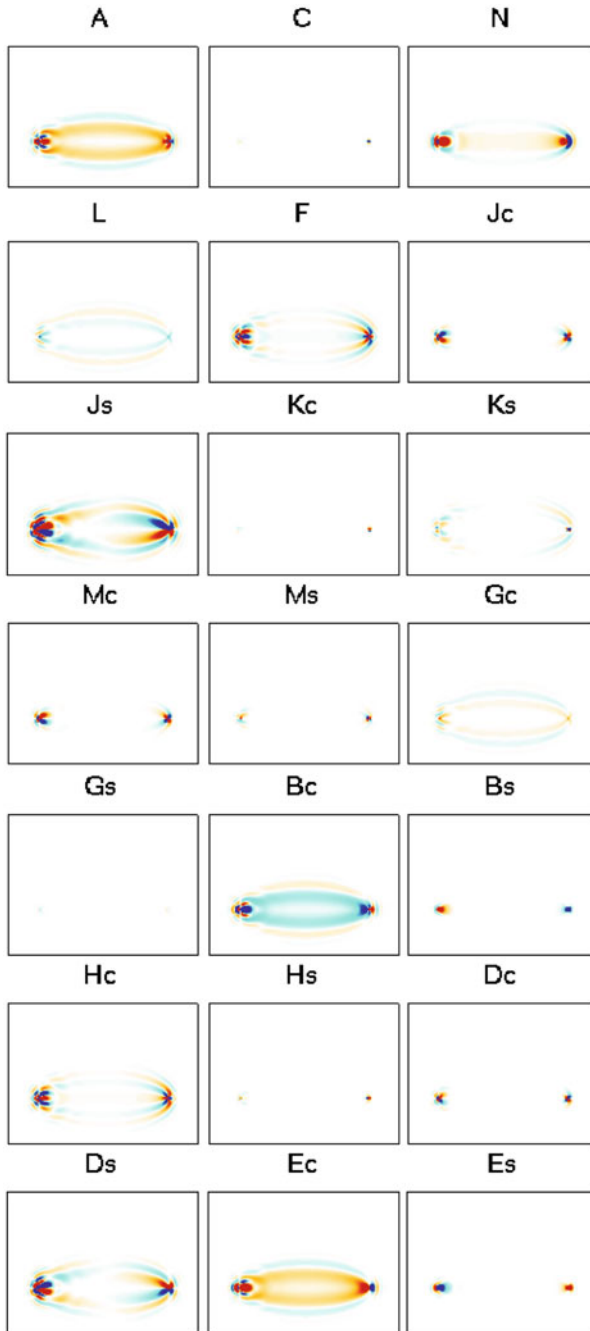


Fig. 4.20 Data sensitivity kernels of the broadband cross-correlation delay time with respect to the perturbations in the 21 asymptotic parameters. The data functional is applied on the P-wave on the x_1 -component synthetic seismogram from 100001 to IN.RC01 in the half-space example.

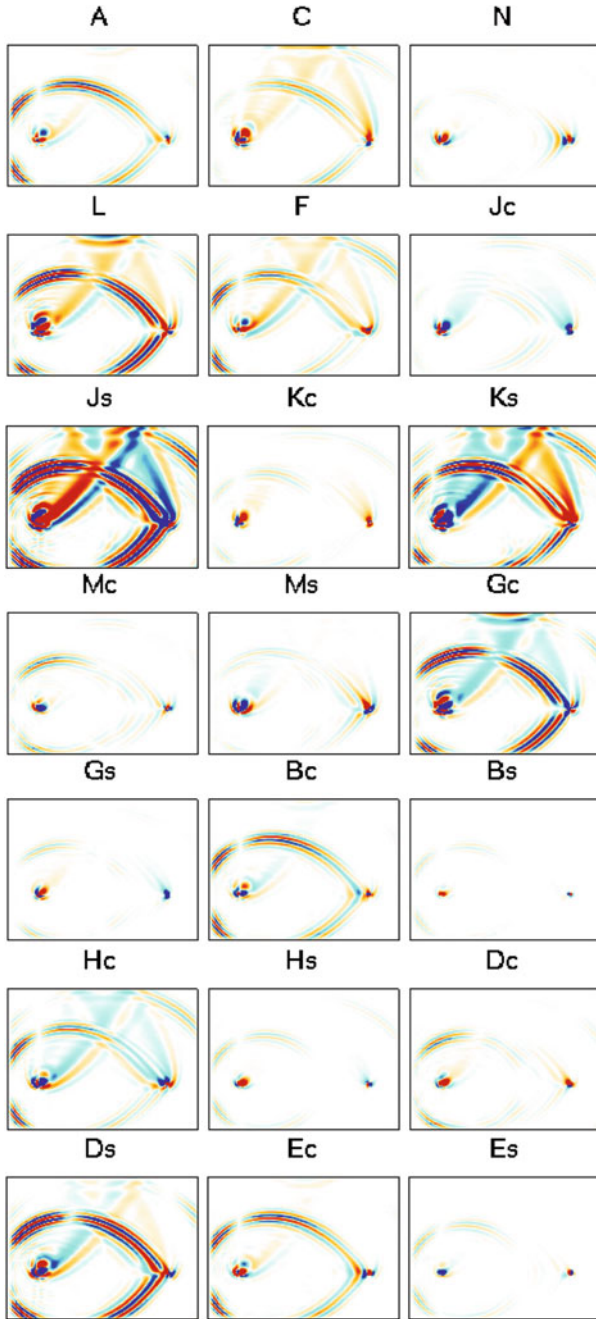


Fig. 4.21 The format is identical to that in Fig. 4.20, except that the data functional is applied on the PP wave.

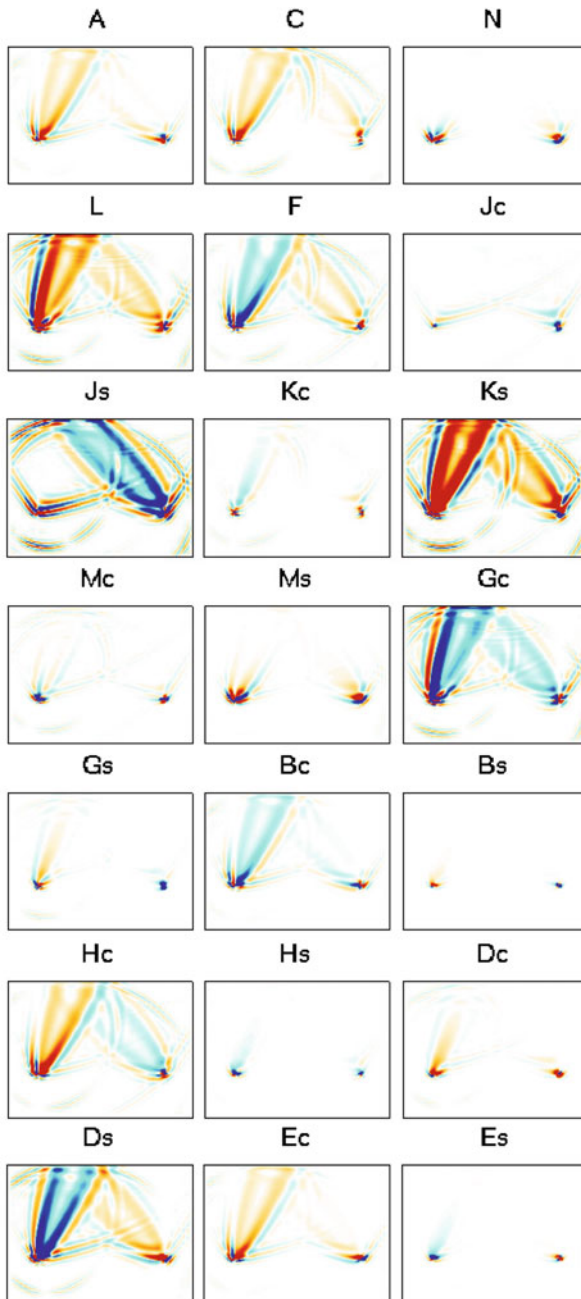


Fig. 4.22 The format is identical to that in Fig. 4.20, except that the data functional is applied on the PS wave.

et al. 2005)

$$\hat{\kappa}(\mathbf{x}, \omega) = \hat{\kappa}(\mathbf{x}, \omega_0) \left[1 + \frac{2}{\pi} Q_{\kappa}^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_{\kappa}^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right], \quad (4.283)$$

$$\hat{\mu}(\mathbf{x}, \omega) = \hat{\mu}(\mathbf{x}, \omega_0) \left[1 + \frac{2}{\pi} Q_{\mu}^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_{\mu}^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right], \quad (4.284)$$

where ω_0 is the reference frequency and $\operatorname{sgn}(\cdot)$ is the sign function. The frequency-dependent complex bulk and shear moduli in Eqs. 4.283–4.284 satisfy $\hat{\kappa}(\mathbf{x}, -\omega) = \hat{\kappa}^*(\mathbf{x}, \omega)$ and $\hat{\mu}(\mathbf{x}, -\omega) = \hat{\mu}^*(\mathbf{x}, \omega)$. The perturbations in the bulk and shear moduli $\delta\hat{\kappa}$ and $\delta\hat{\mu}$ and the perturbations in the inverses of the quality factors δQ_{κ}^{-1} and δQ_{μ}^{-1} can be related through

$$\begin{aligned} \delta\hat{\kappa}(\mathbf{x}, \omega) &= \hat{\kappa}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \delta Q_{\kappa}^{-1}(\mathbf{x}) \\ &\quad + \left[1 + \frac{2}{\pi} Q_{\kappa}^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_{\kappa}^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right] \delta\hat{\kappa}_0, \end{aligned} \quad (4.285)$$

$$\begin{aligned} \delta\hat{\mu}(\mathbf{x}, \omega) &= \hat{\mu}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \delta Q_{\mu}^{-1}(\mathbf{x}) \\ &\quad + \left[1 + \frac{2}{\pi} Q_{\mu}^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_{\mu}^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right] \delta\hat{\mu}_0, \end{aligned} \quad (4.286)$$

where $\hat{\kappa}_0(\mathbf{x}) = \hat{\kappa}(\mathbf{x}, \omega_0)$, $\hat{\mu}_0(\mathbf{x}) = \hat{\mu}(\mathbf{x}, \omega_0)$, $\delta\hat{\kappa}_0$ and $\delta\hat{\mu}_0$ are the perturbations to the bulk and shear moduli at the reference frequency.

The frequency-domain rate-of-relaxation tensor $\hat{\Psi}_{ijklm}(\mathbf{x}, \omega)$ can be expressed in the form

$$\begin{aligned} \hat{\Psi}_{ijklm}(\mathbf{x}, \omega) &= \left[\hat{\kappa}(\mathbf{x}, \omega) - \frac{2}{3} \hat{\mu}(\mathbf{x}, \omega) \right] \delta_{jk} \delta_{lm} \\ &\quad + \hat{\mu}(\mathbf{x}, \omega) (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) + \Psi'_{ijklm}(\mathbf{x}), \end{aligned} \quad (4.287)$$

where we assume that the anisotropic term $\Psi'_{ijklm}(\mathbf{x})$ is purely elastic. The perturbation in the rate-of-relaxation tensor $\delta\hat{\Psi}_{ijklm}$ and the perturbations in the quality factors $\delta Q_{\kappa, \mu}^{-1}$ can be related through

$$\begin{aligned} \delta\hat{\Psi}_{ijklm}(\mathbf{x}, \omega) &= \frac{\partial \hat{\Psi}_{ijklm}}{\partial \hat{\kappa}} \delta\hat{\kappa} + \frac{\partial \hat{\Psi}_{ijklm}}{\partial \hat{\mu}} \delta\hat{\mu} + \delta\Psi'_{ijklm}(\mathbf{x}) \\ &= \delta\hat{\kappa} \delta_{jk} \delta_{lm} + \delta\hat{\mu} \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) + \delta\Psi'_{ijklm}(\mathbf{x}). \end{aligned} \quad (4.288)$$

Bring Eqs. 4.285 and 4.286 into Eq. 4.288, we obtain

$$\delta\hat{\Psi}_{ijklm}(\mathbf{x}, \omega) = \hat{\kappa}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \delta_{jk} \delta_{lm} \delta Q_{\kappa}^{-1}(\mathbf{x})$$

$$\begin{aligned}
& +\hat{\mu}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) \delta Q_\mu^{-1}(\mathbf{x}) \\
& + \left[1 + \frac{2}{\pi} Q_\kappa^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_\kappa^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right] \delta_{jk} \delta_{lm} \delta \hat{\kappa}_0 \\
& + \left[1 + \frac{2}{\pi} Q_\mu^{-1}(\mathbf{x}) \ln \frac{|\omega|}{\omega_0} + i Q_\mu^{-1}(\mathbf{x}) \operatorname{sgn}(\omega) \right] (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \delta \hat{\mu}_0 \\
& + \delta \Psi'_{jklm}(\mathbf{x}). \tag{4.289}
\end{aligned}$$

In the frequency-domain, Eq. 4.33 can be expressed as

$$\begin{aligned}
\delta d^{sn} = - \sum_{j,k,l,m=1}^3 \int_V dV(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \\
\mathcal{F}^{-1} \left[\frac{\partial \hat{G}_{jp}(\mathbf{x}, \omega; \mathbf{x}', 0)}{\partial x_k} \frac{\partial \hat{u}_l^s(\mathbf{x}, \omega)}{\partial x_m} \delta \hat{\Psi}_{jklm}(\mathbf{x}, \omega) \right], \tag{4.290}
\end{aligned}$$

where $\mathcal{F}^{-1}[\]$ represents the inverse Fourier transform of the angular frequency ω to time t and I have switched the roles of \mathbf{x} and \mathbf{x}' (i.e., $\mathbf{x} \leftrightarrow \mathbf{x}'$). Bring Eq. 4.289 into Eq. 4.290 and collect the terms containing $\delta Q_\kappa^{-1}(\mathbf{x})$, we obtain the data sensitivity kernel with respect to Q_κ^{-1} ,

$$\begin{aligned}
K_{d^{sn}}^{Q_\kappa^{-1}}(\mathbf{x}) &= -\hat{\kappa}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \sum_{p=1}^3 \check{J}_p^{sn}(\mathbf{x}', t) \\
& \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \\
& = -\hat{\kappa}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \check{\mathbf{J}}^{sn}(\mathbf{x}, t) \\
& \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \} \tag{4.291}
\end{aligned}$$

where the WPK $\check{J}_p^{sn}(\mathbf{x}', t)$ has the form

$$\check{J}_p^{sn}(\mathbf{x}', t) = \int_0^T J_p^{sn}(\mathbf{x}', t') \psi(t' - t) dt', \tag{4.292}$$

and $\psi(t)$ is the inverse Fourier transform of $\hat{\psi}(\omega) = \frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega)$, i.e.,

$$\psi(t) = \mathcal{F}^{-1} \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right]. \tag{4.293}$$

If we introduce the function

$$\psi'(t) = \mathcal{F}^{-1} \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} - i \operatorname{sgn}(\omega) \right], \tag{4.294}$$

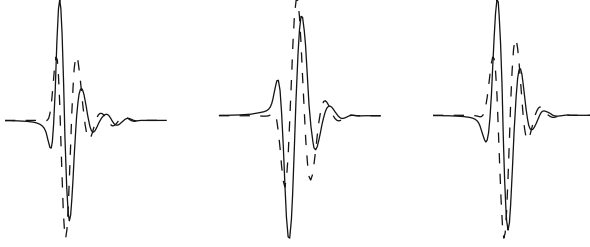


Fig. 4.23 Dash lines: the WPKs of the broadband cross-correlation amplitude anomalies measured on the P (left), PP (center) and PS (right) waves on the x_1 -component of the synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example; solid lines: the WPKs for the anelastic attenuation kernels obtained by convolving the WPKs shown in dash lines with the function $\psi'(t)$ using Eq. 4.296.

whose Fourier transform $\hat{\psi}'(\omega)$ satisfies $\hat{\psi}'(\omega) = \hat{\psi}^*(\omega)$, then we have

$$\psi'(t - t') = \psi(t' - t), \quad (4.295)$$

and Eq. 4.292 can be re-written as

$$\check{J}_p^{sn}(\mathbf{x}', t) = \int_0^T J_p^{sn}(\mathbf{x}', t') \psi'(t - t') dt' = J_p^{sn}(\mathbf{x}', t) * \psi'(t). \quad (4.296)$$

As pointed out in (Tromp et al. 2005), when the physical dispersion, represented by the term $\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0}$ in Eq. 4.294, is ignored, $\check{J}_p^{sn}(\mathbf{x}', t)$ is the Hilbert transform of $J_p^{sn}(\mathbf{x}', t)$. We use the WPK for the broadband cross-correlation amplitude anomaly as an example. Figure 4.23 shows the $\check{J}_p^{sn}(\mathbf{x}', t)$ functions and also the $J_p^{sn}(\mathbf{x}', t)$ functions in Eq. 4.296 for the broadband cross-correlation amplitude anomaly measured on the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The reference frequency $\omega_0 = 2\pi$ was used in the calculation. The $J_p^{sn}(\mathbf{x}', t)$ functions (dash lines) shown in Fig. 4.23 are identical to those shown on the bottom row in Fig. 4.7.

Bring Eq. 4.289 into Eq. 4.290 and collect the terms containing $\delta Q_\mu^{-1}(\mathbf{x})$, we obtain the data sensitivity kernel with respect to Q_μ^{-1} ,

$$\begin{aligned} K_{d^{sn}}^{Q_\mu^{-1}}(\mathbf{x}) &= -\hat{\mu}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \sum_{p=1}^3 \check{J}_p^{sn}(\mathbf{x}', t) \\ &\left\{ \sum_{j,k=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_k} * \left[\frac{\partial u_j^s(\mathbf{x}, t)}{\partial x_k} + \frac{\partial u_k^s(\mathbf{x}, t)}{\partial x_j} \right] \right. \\ &\quad \left. - \frac{2}{3} \left[\sum_{j=1}^3 \frac{\partial G_{jp}(\mathbf{x}, t; \mathbf{x}', 0)}{\partial x_j} \right] * \left[\sum_{l=1}^3 \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_l} \right] \right\} \\ &= -2\hat{\mu}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \check{\mathbf{J}}^{sn}(\mathbf{x}, t) \cdot \left[\check{\boldsymbol{\epsilon}}^G(\mathbf{x}, t; \mathbf{x}', 0) : \check{\boldsymbol{\epsilon}}^s(\mathbf{x}, t) \right], \quad (4.297) \end{aligned}$$

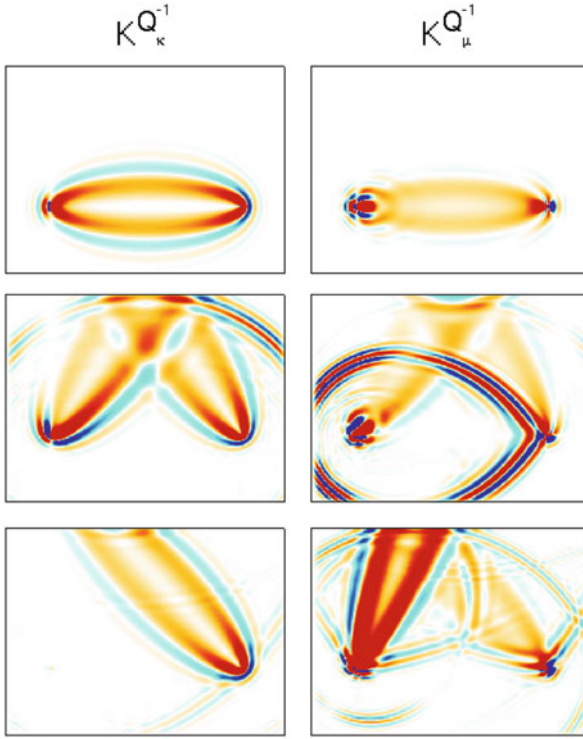


Fig. 4.24 Data sensitivity kernels of the broadband cross-correlation amplitude anomaly with respect to Q_κ^{-1} (left column) and Q_μ^{-1} (right column) for the P (top row), PP (center row) and PS (bottom row) waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example.

where $\check{J}_p^{sn}(\mathbf{x}', t)$ is given in Eq. 4.292, $\check{\epsilon}^s$ and $\check{\epsilon}^G$ are deviatoric strain fields given in Eqs. 4.155 and 4.156 and “ $\check{\cdot}$ ” has the same meaning as in Eq. 4.162. Comparing Eq. 4.291 with 4.145 and Eq. 4.297 with 4.151, we can see that except for the differences in the WPK $\check{J}_p^{sn}(\mathbf{x}', t)$ and the reference bulk and shear moduli $\hat{\kappa}_0$ and $\hat{\mu}_0$, the data sensitivity kernels with respect to the bulk and shear quality factors $K_{d^{sn}}^{Q_\kappa^{-1}}$ and $K_{d^{sn}}^{Q_\mu^{-1}}$ are identical to the data sensitivity kernels with respect to the relative perturbations in the two moduli $K_{d^{sn}}^{\ln \kappa}$ and $K_{d^{sn}}^{\ln \mu}$, respectively. Figure 4.24 shows examples of $K_{d^{sn}}^{Q_\kappa^{-1}}$ and $K_{d^{sn}}^{Q_\mu^{-1}}$ for the broadband cross-correlation amplitude anomaly (Sect. 4.2.6) measured on the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. In this calculation, the reference structural model was elastic.

When the reference model is purely elastic, the reference quality factors satisfy $Q_{\kappa,\mu}^{-1}(\mathbf{x}) = 0$ and the reference bulk and shear moduli $\kappa(\mathbf{x})$ and $\mu(\mathbf{x})$ are real, frequency-independent functions. Equation 4.289 can be simplified to

$$\begin{aligned} \delta \hat{\Psi}_{jklm}(\mathbf{x}, \omega) = & \kappa(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \delta_{jk} \delta_{lm} \delta Q_{\kappa}^{-1}(\mathbf{x}) \\ & + \mu(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) \delta Q_{\mu}^{-1}(\mathbf{x}) \\ & + \delta_{jk} \delta_{lm} \delta \kappa + (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \delta \mu + \delta \Psi'_{jklm}(\mathbf{x}). \end{aligned} \quad (4.298)$$

Bring Eq. 4.298 into Eq. 4.290 and collect the terms containing $\delta \kappa$ and $\delta \mu$, respectively, we obtain the data sensitivity kernels with respect to the two moduli $K_{d^{sn}}^{\ln \kappa}$ and $K_{d^{sn}}^{\ln \mu}$ evaluated at the reference moduli. The expressions for $K_{d^{sn}}^{\ln \kappa}$ and $K_{d^{sn}}^{\ln \mu}$ are identical to those given in Eqs. 4.145 and 4.151. When inverting amplitude misfit measurements for quality factors using the kernels in Eqs. 4.291 and 4.297, it is often necessary to account for focusing/de-focusing effects caused by elastic heterogeneities not included in the reference model. A joint inversion that combines $K_{d^{sn}}^{Q_{\kappa}^{-1}}$ and $K_{d^{sn}}^{Q_{\mu}^{-1}}$ with $K_{d^{sn}}^{\ln \kappa}$ and $K_{d^{sn}}^{\ln \mu}$ may allow us to recover the quality factors unbiased by elastic focusing/de-focusing effects.

When the reference model is anelastic, $Q_{\kappa,\mu}^{-1}(\mathbf{x})$ in Eq. 4.289 are the inverses of the reference quality factors and $\hat{\kappa}_0(\mathbf{x})$ and $\hat{\mu}_0(\mathbf{x})$ are the reference moduli at the reference frequency ω_0 . Since we often have $Q_{\kappa,\mu}^{-1}(\mathbf{x}) \ll 1$, Eq. 4.289 might be approximated as

$$\begin{aligned} \delta \hat{\Psi}_{jklm}(\mathbf{x}, \omega) \approx & \hat{\kappa}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \delta_{jk} \delta_{lm} \delta Q_{\kappa}^{-1}(\mathbf{x}) \\ & + \hat{\mu}_0(\mathbf{x}) \left[\frac{2}{\pi} \ln \frac{|\omega|}{\omega_0} + i \operatorname{sgn}(\omega) \right] \left(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl} - \frac{2}{3} \delta_{jk} \delta_{lm} \right) \delta Q_{\mu}^{-1}(\mathbf{x}) \\ & + \delta_{jk} \delta_{lm} \delta \hat{\kappa}_0 + (\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}) \delta \hat{\mu}_0 + \delta \Psi'_{jklm}(\mathbf{x}). \end{aligned} \quad (4.299)$$

In such a case, data sensitivity kernels with respect to $\hat{\kappa}_0$ and $\hat{\mu}_0$ are identical to Eqs. 4.145 and 4.151, respectively. When the condition $Q_{\kappa,\mu}^{-1}(\mathbf{x}) \ll 1$ does not hold, we need to compute the data sensitivity kernels with respect to $\hat{\kappa}_0$ and $\hat{\mu}_0$ using Eq. 4.289. For the relative perturbation in the bulk modulus at the reference frequency $\hat{\kappa}_0$ we have

$$\begin{aligned} K_{d^{sn}}^{\ln \hat{\kappa}_0}(\mathbf{x}) = & -\hat{\kappa}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \mathbf{J}^{sn}(\mathbf{x}', t) \\ & \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \} \\ & - \hat{\kappa}_0(\mathbf{x}) Q_{\kappa}^{-1}(\mathbf{x}) \int_V dV(\mathbf{x}') \int_0^T dt \check{\mathbf{J}}^{sn}(\mathbf{x}', t) \\ & \cdot \{ [\nabla \cdot \mathbf{G}(\mathbf{x}, t; \mathbf{x}', 0)] * [\nabla \cdot \mathbf{u}^s(\mathbf{x}, t)] \}, \end{aligned} \quad (4.300)$$

where $\check{\mathbf{J}}^{sn}(\mathbf{x}', t)$ is given in Eq. 4.292. And for the relative perturbation in the shear modulus at the reference frequency $\hat{\mu}_0$ we have

$$\begin{aligned} K_{d^{sn}}^{\ln \hat{\mu}_0}(\mathbf{x}) &= -2\hat{\mu}_0(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \mathbf{J}^{sn}(\mathbf{x}', t) \cdot \left[\check{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) : \check{\epsilon}^s(\mathbf{x}, t) \right] \\ &- 2\hat{\mu}_0(\mathbf{x}) \mathcal{Q}_{\mu}^{-1}(\mathbf{x}) \int_V dV(\mathbf{x}') \int dt \check{\mathbf{J}}^{sn}(\mathbf{x}', t) \cdot \left[\check{\epsilon}^G(\mathbf{x}, t; \mathbf{x}', 0) : \check{\epsilon}^s(\mathbf{x}, t) \right], \end{aligned} \quad (4.301)$$

where $\check{\epsilon}^s$ and $\check{\epsilon}^G$ are deviatoric strain fields given in Eqs. 4.155 and 4.156 and “:” has the same meaning as in Eq. 4.162.

4.3.8 Finite Moment Tensor (FMT)

For a general moment source $m_{jk}(\mathbf{x}, t)$, Eqs. 3.339 or 3.346 gives the synthetic displacement seismogram at receiver location \mathbf{x}_r and time t . As discussed in Sect. 3.4.2, it is possible to invert for $m_{jk}(\mathbf{x}, t)$, as a tensor-valued function of both space and time, using either equation. Since both equations are linear with respect to $m_{jk}(\mathbf{x}, t)$, the Fréchet kernel of the data functional with respect to $m_{jk}(\mathbf{x}, t)$ can be obtained by directly composing the integration kernel in Eqs. 3.339 or 3.346 with the corresponding WPK (Eq. 4.30). In practice, we may choose to invert for certain integral properties of the seismic source. One possibility is to use the spatial- and temporal polynomial moments of source-space-time distribution to characterize the source (e.g., Backus 1977a, b). If we assume that the moment density $m_{jk}(\mathbf{x}, t)$ is everywhere proportional to a constant seismic moment tensor M_{jk} , i.e.,

$$m_{jk}(\mathbf{x}, t) = M_{jk} \int_{-\infty}^t f(\mathbf{x}, t') dt', \quad (4.302)$$

where $f(\mathbf{x}, t)$ is the source-space-time function and M_{jk} is independent of space and time, we can then define the spatial and temporal polynomial moments of $f(\mathbf{x}, t)$ as

$$\mu^{(0,0)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) \equiv M_0, \quad (4.303)$$

$$\mu^{(0,1)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (t - t_0) \equiv t_1, \quad (4.304)$$

$$\boldsymbol{\mu}^{(1,0)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (\mathbf{x} - \mathbf{x}_0) \equiv \mathbf{x}_1, \quad (4.305)$$

$$\mu^{(0,2)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (t - t_0)^2, \quad (4.306)$$

$$\boldsymbol{\mu}^{(1,1)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (t - t_0) (\mathbf{x} - \mathbf{x}_0), \quad (4.307)$$

$$\boldsymbol{\mu}^{(2,0)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (\mathbf{x} - \mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0)^T, \quad (4.308)$$

where t_0 is the reference time and \mathbf{x}_0 is the reference location. Note that the source-space-time function $f(\mathbf{x}, t)$ defined in Eq. 4.302 is the moment-rate function in time, therefore it is compactly supported in time. Equation 4.304 defines the centroid time t_1 relative to the reference time t_0 and Eq. 4.305 defines the centroid location \mathbf{x}_1 relative to the reference location \mathbf{x}_0 . If we replace t_0 and \mathbf{x}_0 in Eqs. 4.306–4.308 with the centroid time and centroid location t_1 and \mathbf{x}_1 , respectively, we obtain

$$\hat{\boldsymbol{\mu}}^{(0,2)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (t - t_1)^2 \equiv \left(\frac{T_c}{2}\right)^2, \quad (4.309)$$

$$\hat{\boldsymbol{\mu}}^{(1,1)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (t - t_1) (\mathbf{x} - \mathbf{x}_1) \equiv \hat{\boldsymbol{\mu}}^{(0,2)} \mathbf{v}_d, \quad (4.310)$$

$$\hat{\boldsymbol{\mu}}^{(2,0)} = \int dV(\mathbf{x}) \int dt f(\mathbf{x}, t) (\mathbf{x} - \mathbf{x}_1) (\mathbf{x} - \mathbf{x}_1)^T \equiv \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T, \quad (4.311)$$

where T_c is the *characteristic duration*, \mathbf{v}_d is the *directivity velocity vector* (Ben-Menahem 1961; McGuire et al. 2002), the 3-by-3 diagonal matrix $\boldsymbol{\Lambda}$ contains the eigenvalues of the second-order spatial moment $\hat{\boldsymbol{\mu}}^{(2,0)}$ and \mathbf{U} is an orthogonal matrix of eigenvectors. For a planar rectangular rupture, the eigenvalue matrix $\boldsymbol{\Lambda}$ can be expressed as

$$\boldsymbol{\Lambda} = \begin{bmatrix} L_c/2 & 0 & 0 \\ 0 & W_c/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (4.312)$$

where L_c is the characteristic length of the rupture and W_c is the characteristic width of the rupture. The directivity velocity vector \mathbf{v}_d should lie on the plane of a simple planar rupture. We can define a dimensionless directivity parameter

$$D \equiv |\mathbf{v}_d| \frac{T_c}{L_c} \quad (4.313)$$

to measure the magnitude of the directivity effect. For a perfectly symmetric rupture (e.g., a bilateral rupture) we have $D = 0$ and for a perfectly unilateral rupture we have $D = 1$. In (Chen et al. 2005) we named the spatial and temporal polynomial moments up to degree-2 the *finite moment tensor* (FMT).

The FMTs for a set of global large earthquakes ($M_w \geq 7$) were inverted in (McGuire et al. 2001, 2002) using low-frequency teleseismic data and it was shown that the global large earthquakes are predominantly unilateral. The FMTs for local small- to medium-sized earthquakes ($2.5 \leq M_w \leq 5$) were estimated in (Chen et al. 2005, 2010b) using local broadband waveform data. In particular, in (Chen et al. 2010b) we have shown the possibility of resolving the fault plane ambiguity of local small earthquakes by detecting the constructive/destructive interference effects due

to $\hat{\boldsymbol{\mu}}^{(2,0)}$. The fault plane ambiguity is intrinsic to the point-source centroid moment tensor (CMT) representation. For large earthquakes with many aftershocks, it is possible to resolve this ambiguity by using aftershock hypocenters. This ambiguity can also be resolved using the directivity vector if the source has detectable directivity effects (e.g., Mori and Hartzell 1990; Mori 1996; Hellweg and Boatwright 1999; Okada et al. 2001; McGuire 2004). For small earthquakes without significant directivity effects, the technique based on $\hat{\boldsymbol{\mu}}^{(2,0)}$ as documented in (Chen et al. 2010b) provides a viable approach for resolving this ambiguity.

To obtain the partial derivatives of the data functional with respect to the FMT parameters, we bring Eq. 4.302 into Eq. 3.339 and obtain

$$\begin{aligned}
 u_i(\mathbf{x}_r, t) &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 \frac{\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)}{\partial x_k} * m_{jk}(\mathbf{x}, t) \\
 &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 M_{jk} \mathcal{F}^{-1} \left\{ \frac{\partial \hat{G}_{ji}(\mathbf{x}, \omega; \mathbf{x}_r, 0)}{\partial x_k} \left[\frac{\hat{f}(\mathbf{x}, \omega)}{i\omega} + c_1(\mathbf{x})\delta(\omega) \right] \right\} \\
 &= \int_V dV(\mathbf{x}) \sum_{j,k=1}^3 M_{jk} \mathcal{F}^{-1} \left\{ \left[\frac{1}{i\omega} \frac{\partial \hat{G}_{ji}(\mathbf{x}, \omega; \mathbf{x}_r, 0)}{\partial x_k} + c_2(\mathbf{x})\delta(\omega) \right] \hat{f}(\mathbf{x}, \omega) \right. \\
 &\quad \left. + c_1(\mathbf{x}) \frac{\partial \hat{G}_{ji}(\mathbf{x}, \omega; \mathbf{x}_r, 0)}{\partial x_k} \delta(\omega) - c_2(\mathbf{x}) \hat{f}(\mathbf{x}, \omega) \delta(\omega) \right\} \\
 &= \int_V dV(\mathbf{x}) \int dt' \sum_{j,k=1}^3 M_{jk} H_{kji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0) f(\mathbf{x}, t'), \tag{4.314}
 \end{aligned}$$

where $\hat{f}(\mathbf{x}, \omega)$ is the Fourier transform of the source-space-time function $f(\mathbf{x}, t)$ (Eq. 4.302), $\hat{G}_{ji}(\mathbf{x}, \omega; \mathbf{x}_r, 0)$ is the Fourier transform of the RGT $G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)$, $\mathcal{F}^{-1}\{\}$ represents the inverse Fourier transform and the third-order tensor $H_{kji}(\mathbf{x}, t; \mathbf{x}_r, 0)$ is defined as (e.g., Backus 1977a)

$$\begin{aligned}
 H_{kji}(\mathbf{x}, t; \mathbf{x}_r, 0) &\equiv \mathcal{F}^{-1} \left[\frac{1}{i\omega} \frac{\partial \hat{G}_{ji}(\mathbf{x}, \omega; \mathbf{x}_r, 0)}{\partial x_k} + c_2(\mathbf{x})\delta(\omega) \right] \\
 &= \int_{-\infty}^t \frac{\partial G_{ji}(\mathbf{x}, \tau; \mathbf{x}_r, 0)}{\partial x_k} d\tau. \tag{4.315}
 \end{aligned}$$

In Eqs. 4.314 and 4.315, $c_1(\mathbf{x})$ is the DC component of $f(\mathbf{x}, t)$ and $c_2(\mathbf{x})$ is the DC component of $\partial G_{ji}(\mathbf{x}, t; \mathbf{x}_r, 0)/\partial x_k$ and I have used the integration property of the Fourier transform. We can expand $H_{kji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)$ in Taylor series at the reference location $\mathbf{x} = \mathbf{x}_0$ and the reference time $t' = t_0$ and truncate the Taylor series to second order,

$$H_{kji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0) \approx H_{kji}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0) - (t' - t_0) \frac{\partial H_{kji}}{\partial t}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0)$$

$$\begin{aligned}
& +(\mathbf{x} - \mathbf{x}_0) \cdot \nabla H_{kji}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0) + \frac{(t' - t_0)^2}{2} \frac{\partial^2 H_{kji}}{\partial t^2}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0) \\
& \quad - (t' - t_0)(\mathbf{x} - \mathbf{x}_0) \cdot \nabla \frac{\partial H_{kji}}{\partial t}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0) \\
& \quad + \frac{1}{2} [(\mathbf{x} - \mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)^T] : \nabla \nabla H_{kji}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0). \tag{4.316}
\end{aligned}$$

Bring Eq. 4.316 into Eq. 4.314, we obtain

$$\begin{aligned}
u_i(\mathbf{x}_r, t) \approx & \left[M_0 - \mu^{(0,1)} \frac{\partial}{\partial t} + \boldsymbol{\mu}^{(1,0)} \cdot \nabla \right. \\
& \left. + \frac{\mu^{(0,2)}}{2} \frac{\partial^2}{\partial t^2} - \boldsymbol{\mu}^{(1,1)} \cdot \nabla \frac{\partial}{\partial t} + \frac{\boldsymbol{\mu}^{(2,0)}}{2} : \nabla \nabla \right] \tilde{u}_i(\mathbf{x}_0, t - t_0; \mathbf{x}_r), \tag{4.317}
\end{aligned}$$

where

$$\tilde{u}_i(\mathbf{x}_0, t - t_0; \mathbf{x}_r) = \sum_{j,k=1}^3 M_{jk} H_{kji}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0) \tag{4.318}$$

is the synthetic displacement seismogram due to a point source located at $\mathbf{x} = \mathbf{x}_0$, $t' = t_0$ with the seismic moment tensor M_{jk} . Note that the gradient operator ∇ in Eqs. 4.316 and 4.317 is with respect to the source coordinate \mathbf{x} in $H_{kji}(\mathbf{x}, t - t'; \mathbf{x}_r, 0)$ and the gradient is evaluated at the reference location $\mathbf{x} = \mathbf{x}_0$. Equation 4.317 is consistent with Eq. 2.11 in (Backus 1977a) and also the frequency-domain form, Eq. (11) in (McGuire et al. 2001), which was derived in (Dahlen and Tromp 1998). The partial derivatives of the displacement waveform due to a finite source with respect to the FMT parameters can then be read directly from Eq. 4.317. The change in the waveform $\delta u_i(\mathbf{x}_r, t)$ due to the FMT parameters can be expressed as

$$\begin{aligned}
\delta u_i(\mathbf{x}_r, t) = u_i(\mathbf{x}_r, t) - \tilde{u}_i(\mathbf{x}_0, t - t_0; \mathbf{x}_r) \approx & \left[M_0 - 1 - \mu^{(0,1)} \frac{\partial}{\partial t} + \boldsymbol{\mu}^{(1,0)} \cdot \nabla \right. \\
& \left. + \frac{\mu^{(0,2)}}{2} \frac{\partial^2}{\partial t^2} - \boldsymbol{\mu}^{(1,1)} \cdot \nabla \frac{\partial}{\partial t} + \frac{\boldsymbol{\mu}^{(2,0)}}{2} : \nabla \nabla \right] \tilde{u}_i(\mathbf{x}_0, t - t_0; \mathbf{x}_r). \tag{4.319}
\end{aligned}$$

Bring Eq. 4.319 into Eq. 4.13, we obtain the change of the data functional d^{sn} due to perturbations in the FMT parameters,

$$\begin{aligned}
\delta d^{sn} & = \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \delta u_p^s(\mathbf{x}', t) \\
& = \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \left[M_0 - 1 - \mu^{(0,1)} \frac{\partial}{\partial t} \right. \\
& \quad \left. + \boldsymbol{\mu}^{(1,0)} \cdot \nabla + \frac{\mu^{(0,2)}}{2} \frac{\partial^2}{\partial t^2} - \boldsymbol{\mu}^{(1,1)} \cdot \nabla \frac{\partial}{\partial t} \right]
\end{aligned}$$

$$+ \frac{\boldsymbol{\mu}^{(2,0)}}{2} : \nabla \nabla \left] \tilde{u}_p(\mathbf{x}_0, t - t_0; \mathbf{x}'). \quad (4.320)$$

The partial derivatives of d^{sn} with respect to the FMT parameters are then given by

$$\frac{\partial d^{sn}}{\partial M_0} = \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \tilde{u}_p(\mathbf{x}_0, t - t_0; \mathbf{x}'), \quad (4.321)$$

$$\frac{\partial d^{sn}}{\partial \mu^{(0,1)}} = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial \tilde{u}_p}{\partial t}(\mathbf{x}_0, t - t_0; \mathbf{x}'), \quad (4.322)$$

$$\frac{\partial d^{sn}}{\partial \mu_i^{(1,0)}} = \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial \tilde{u}_p}{\partial x_i}(\mathbf{x}_0, t - t_0; \mathbf{x}'), \quad (4.323)$$

$$\frac{\partial d^{sn}}{\partial \mu^{(0,2)}} = \frac{1}{2} \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial^2 \tilde{u}_p}{\partial t^2}(\mathbf{x}_0, t - t_0; \mathbf{x}'), \quad (4.324)$$

$$\frac{\partial d^{sn}}{\partial \mu_i^{(1,1)}} = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial^2 \tilde{u}_p}{\partial x_i \partial t}(\mathbf{x}_0, t - t_0; \mathbf{x}'), \quad (4.325)$$

$$\frac{\partial d^{sn}}{\partial \mu_{ij}^{(2,0)}} = \frac{1}{2} \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \frac{\partial^2 \tilde{u}_p}{\partial x_i \partial x_j}(\mathbf{x}_0, t - t_0; \mathbf{x}'). \quad (4.326)$$

The partial derivatives applied on $\tilde{u}_p(\mathbf{x}, t - t_0; \mathbf{x}')$ on the RHS of Eqs. 4.322–4.326 can be approximated using a differencing scheme. If we express $\tilde{u}_p(\mathbf{x}, t - t_0; \mathbf{x}')$ using the RSGTs (Sect. 3.4.2), the spatial partial derivatives on the RHS of Eqs. 4.322–4.326 involve the first- and second-order spatial derivatives with respect to the source coordinate \mathbf{x} , evaluated at the reference location \mathbf{x}_0 . The RSGTs around the source location can be stored during the wave-propagation simulations in F3DWI (Sect. 3.4.3). The first- and second-order spatial derivatives of the RSGTs can therefore be approximated through a differencing scheme.

Here I use the whole-space synthetic wavefield due to a moment source, e.g., Eq. 4.29 in (Aki and Richards 2002), to calculate some examples of the partial derivatives of the frequency-dependent phase-delay time $\Delta\tau_p(\omega)$ and amplitude-reduction time $\Delta\tau_q(\omega)$ (Sect. 4.2.7) with respect to the FMT parameters. In Fig. 4.25, I show the spatial variations of the partial derivatives of $\Delta\tau_{p,q}$ measured on the direct-arriving P wave at 2 Hz with respect to $\mu_1^{(1,0)}$, $\mu_1^{(1,1)}$, $\mu_{11}^{(2,0)}$, $\mu_{12}^{(2,0)}$ and $\mu_{13}^{(2,0)}$ on a spherical receiver surface centered on the reference point source, which is located at the origin of the coordinate system (i.e., x -east, y -north, z -up). The reference point source has an explosive source mechanism and a source-time function given by Eq. 2.122 (Fig. 2.7). The spherical receiver surface has a radius of

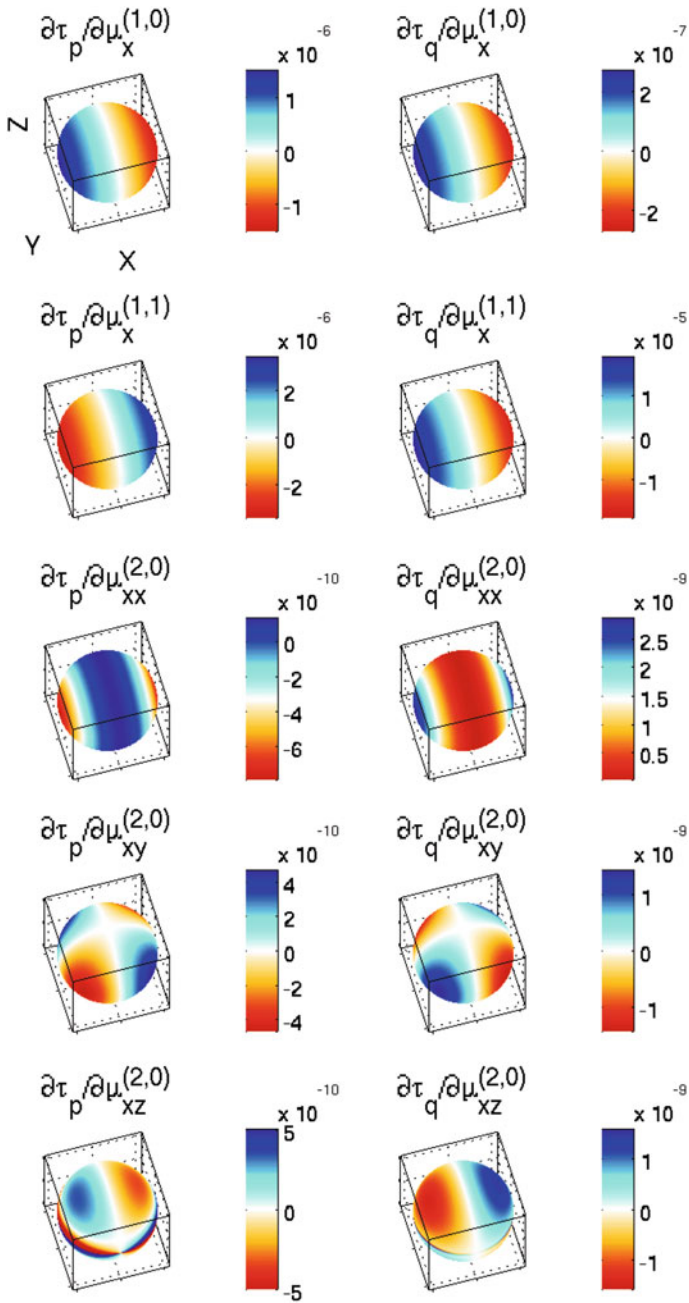


Fig. 4.25 Spatial variations of the partial derivatives of the frequency-dependent phase-delay time $\Delta\tau_p(\omega)$ (left column) and amplitude-reduction time $\Delta\tau_q(\omega)$ (right column) with respect to $\mu_1^{(1,0)}$ (top row), $\mu_1^{(1,1)}$ (second row), $\mu_{11}^{(2,0)}$ (third row), $\mu_{12}^{(2,0)}$ (fourth row), $\mu_{13}^{(2,0)}$ (bottom row).

11.1803 km. The P- and S-velocity and density of the whole-space structural model is 6.5 km/s, 3.5 km/s and 3.0 g/cm³.

A perturbation in $\mu_1^{(1,0)}$ produces an infinitesimal shift of the centroid location along the x -axis towards the east, which reduces the travel-time (hence a decrease in phase-delay time) and increases the amplitude (hence a decrease in amplitude-reduction time) for receivers on the east side of the reference point source. The decrease in travel-time on receivers to the east is caused by reduced source-receiver distances and the increase in amplitude on those receivers is due to the geometric spreading effect. A perturbation in $\mu_1^{(1,1)}$ is linked to the effect of source directivity along the x -axis. Suppose we have an infinitesimal-length unilateral line source lying along the x -axis, centered at the origin and propagating from west to east. The travel-time is reduced (hence a decrease in phase-delay time) on receivers to the west and increased (hence an increase in phase-delay time) because the rupture initiates from the west and propagates to the east. The amplitude is increased (hence a decrease in amplitude-reduction time) on receivers to the east because of the constructive interference effect in the direction of rupture propagation. The spatial variations in the partial derivatives with respect to components of $\boldsymbol{\mu}^{(2,0)}$ can be explained by the constructive and destructive interference effect (Chen et al. 2005, 2010b). For finite-size vertical planar ruptures, the amplitude-reduction time relative to the point-source synthetic waveform has a $\cos(2\theta)$ dependence on azimuth θ with two minima located at 90° and 270° from the strike. A more careful analysis of the partial derivative's azimuthal and also frequency dependence can be found in (Chen et al. 2010b).

Note that the expressions of the partial derivatives given in Eqs. 4.321–4.326 are different from those given in (McGuire et al., 2001; Chen et al., 2005, 2010b), but they are in fact mutually consistent. In (McGuire et al. 2001; Chen et al. 2005, 2010b) the partial derivatives are expressed in terms of the $\tau_{p,q}$ of the reference point-source synthetic seismogram, while the same partial derivatives are expressed in terms of the time-domain point-source synthetic waveform in Eqs. 4.321–4.326 above. The advantage of the expressions in Eqs. 4.321–4.326 is that the effects of localization in the time and frequency domain (e.g., windowing in time domain) are explicitly accounted for through the use of the WPK of the corresponding data functional. And expressions in Eqs. 4.321–4.326 can be easily generalized to other types of data functionals.

In practice, it is often necessary to invert for the CMT, which is a subset of the FMT parameters, before inverting for the second-order polynomial moments. In Sect. 3.4 I have shown that the RSGT gives the Fréchet kernel of the displacement with respect to the spatial-temporal moment tensor density function $m_{jk}(\mathbf{x}, t)$. And in Sect. 3.4.3 I have shown that the RSGT around the source region can be stored during the wave-propagation simulation in F3DWI. For the point-source CMT representation, if we truncate the RHS of Eq. 4.317 to first-order and bring in Eq. 4.318, we obtain

$$u_i(\mathbf{x}_r, t) \approx \left[M_0 - \mu^{(0,1)} \frac{\partial}{\partial t} + \boldsymbol{\mu}^{(1,0)} \cdot \nabla \right]$$

$$\sum_{j,k=1}^3 M_{jk} H_{kji}(\mathbf{x}_0, t - t_0; \mathbf{x}_r, 0), \quad (4.327)$$

which gives the partial derivative of the displacement $u_i(\mathbf{x}_r, t)$ with respect to the scalar moment M_0 , the six independent elements of the symmetric seismic moment tensor M_{jk} and also the centroid time $\mu^{(0,1)}$ and the centroid location $\boldsymbol{\mu}^{(1,0)}$ relative to the reference time t_0 and reference location \mathbf{x}_0 . The partial derivatives of any data functional with respect to the CMT parameters can therefore be obtained by composing the corresponding WPK with the partial derivatives of the displacement with the CMT parameters. These partial derivatives can be used in a gradient- or Hessian-based iterative optimization algorithm to search for the optimal CMT (e.g., Zhao et al. 2006).

Depending upon the restrictions applied on the CMT, the total number of independent CMT parameters is limited to 8–10, which allows applications of global optimization algorithms such as the Bayesian-based grid-search algorithm given in (Lee et al. 2011). In (Lee et al. 2011) the rapid calculation of synthetic seismograms due to a large number of possible CMT parameters in the model space was enabled by the use of RSGTs and reciprocity, which is encapsulated in the script `rgt2syn.py` under the `py` directory in `F3DWT`. The usage of `rgt2syn.py` has been discussed in Sects. 2.2.7 and 3.4.3. The calculation can be easily parallelized on a variety of parallel computing platforms with minimal inter-processor communication overhead. The technique demonstrated in (Lee et al. 2011) provides not only estimates of the CMT parameters in a 3D structural model, but also the uncertainties in the estimated CMT parameters, in (near-)real-time when combined with real-time access to telemetered digital data streams from the seismic network.

4.4 Software

The calculation of data sensitivity kernels involves the script `setup_ker.py` in the `py` directory and the parallel binary executable `ker` in the `bin` directory. The output files of `ker` are in binary format and the sequential binary executable `rdker` in the `bin` directory can be used to convert them into ASCII text files, which can be plotted using other data visualization software such as MATLAB. A critical input to the kernel calculation is the WPK for the specific data functional used in the inversion. Several MATLAB scripts are provided to generate the WPKs for the data functionals discussed in Sect. 4.2. Other inputs to the kernel calculations include the forward wavefield, i.e., spatial-temporal strain and particle velocity fields (Eq. 2.125), from the source and also the RSGTs from the receiver, whose calculations have been discussed in Sects. 2.2 and 3.4.3. In the following I will use examples to demonstrate the basic usage of the codes related to kernel calculations.

I will start with re-producing the kernels shown in Fig. 4.8, which are the data sensitivity kernels of the broadband cross-correlation delay time and amplitude anomaly (Sect. 4.2.6) with respect to the relative perturbations in P- and S-velocity

and density (Sect. 4.3.4) for the P, PP and PS waves on the synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. For this example, the parameter file `param_HS.dat` (Sect. 2.2.3) is identical to that shown in Listing 2.5. The source-receiver list `srlst` (Sect. 2.2.4.1) consists of the first three lines shown in Listing 2.6, i.e., the receiver line for IN.RC01, the source line for 100001 and the source line for the body-force source IN.RC01.1 in the x_1 direction and at the location of IN.RC01. The procedure for generating the forward wavefield from 100001 and the RGT from IN.RC01.1 is identical to that shown in Listing 2.17. We need to execute the following commands in the terminal under the project JOBS directory,

```
proc_srlst.py param_HS.dat 1 3
set_sr_grd.py param_HS.dat
set_ker_grd.py param_HS.dat
setup_awp.py param_HS.dat FWF 1 3
```

and then run the parallel binary executable `awp` inside the sub-directory `FWF/100001` to generate the forward wavefield for 100001

```
cd ../FWF/100001
mpiexec -n 8 -f ~/tools/mpi/share/myhosts ./awp
```

and then run `awp` in the sub-directory `FWF/IN.RC01.1` to generate the RSGT for IN.RC01.1.

```
cd ../IN.RC01.1
mpiexec -n 8 -f ~/tools/mpi/share/myhosts ./awp
```

After the forward wave-propagation simulation for 100001 is completed, the synthetic velocity seismograms from 100001 to IN.RC01 can be obtained by executing the following command under the project JOBS directory,

```
fwf2syn.py param_HS.dat 2 2 1 1
```

and the synthetic seismograms are stored in the file `SYN/100001/IN.RC01.rot`.

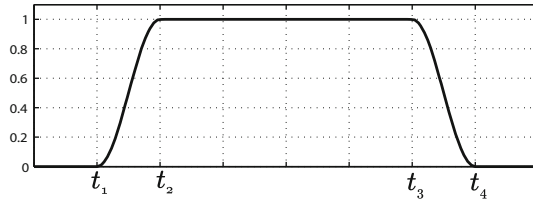


Fig. 4.26 The time window used for selecting waveforms from the complete synthetic seismogram. At time t_1 and t_4 , the window tapers to zero and at time t_2 and t_3 , the window increases to 1. The tapering parts of the window between t_1 and t_2 and between t_3 and t_4 follow a \cos^2 function.

4.4.1 WPK File

We can calculate the WPK for the broadband cross-correlation delay time and amplitude anomaly discussed in Sect. 4.2.6 using the MATLAB function `wpk_bbxcorr.m`, which is included in F3DWTI, together with a few other MATLAB functions for computing the WPKs of other data functionals discussed in Sect. 4.2. The function `wpk_bbxcorr.m` is written according to Eqs. 4.59 and 4.68 in Sect. 4.2.6. The definition of this function is in the following,

```
function [sconv,wpkt,wpka]=wpk_bbxcorr(syn,stf_grn,dt,ntskp,w)
```

It takes four inputs. The first input `syn` is a simple one-column vector containing the synthetic velocity seismogram at every time sample. The second input `stf_grn` is the source-time function used for computing the RGT and has the same format as shown in Listing 2.7 (without the comments). For the half-space example, this source-time function is plotted in Fig. 2.11. The third input `dt` is simply the time-sampling interval used in the wave-propagation simulations and can be found in the parameter file (Line 12, Listing 2.5). The fourth input `ntskp` is the decimation rate in time and can also be found in the parameter file (Line 24, Listing 2.5). The fifth input `w` is a four-element vector for the time window used to select the target seismic waveform on the complete synthetic seismogram. The four numbers inside `w` are t_1 , t_2 , t_3 and t_4 , as shown in Fig. 4.26.

The first output of this function `sconv` is a two-column vector. Its second column is the input synthetic seismogram `syn` convolved with the source-time function used in computing the RGT `stf_grn`. The reason for computing this convolution was explained in Sect. 4.2.2. The first column of `sconv` is the time axis of the convolved synthetic. The second output `wpkt` is a one-column vector containing the WPK for the broadband cross-correlation delay time. The first two numbers in `wpkt` are the starting and ending indices of the non-zero part of the WPK. They should coincide with the indices corresponding to t_1 and t_4 (Fig. 4.26). The time-series for the WPK starts at the third number in `wpkt`. The third output `wpka` is a one-column vector containing the WPK for the broadband cross-correlation amplitude anomaly and has the same format as `wpkt`. Both `wpkt` and `wpka` are derived from the convolved synthetic `sconv`, as discussed in Sect. 4.2.2.

```

1 6 998
2 1.510000e+02 1.510000e+02 2.840000e+02 2.840000e+02 4.010000e+02
  4.010000e+02
3 2.510000e+02 2.510000e+02 3.570000e+02 3.570000e+02 4.840000e+02
  4.840000e+02
4 0.000000e+00 -0.000000e+00 0.000000e+00 -0.000000e+00 0.000000e
  +00 -0.000000e+00
5 ...

```

Listing 4.1 The first four lines inside an example WPK file. This WPK file is for the broadband cross-correlation delay time and amplitude anomaly measured on the P, PP and PS waves of the half-space example.

Table 4.3 The ordering of the WPKs inside the WPK file for the half-space example shown in Listing 4.1

Column number	WPK
1	Broadband cross-correlation delay time for P
2	Broadband cross-correlation amplitude anomaly for P
3	Broadband cross-correlation delay time for PP
4	Broadband cross-correlation amplitude anomaly for PP
5	Broadband cross-correlation delay time for PS
6	Broadband cross-correlation amplitude anomaly for PS

The WPKs need to be written into an ASCII text file, which will be used as an input to `ker`. An example of the WPK file is shown in Listing 4.1. The two numbers shown on line 1 are the number of WPKs inside the file and the number of time samples of each WPK. For the example shown in Listing 4.1, there are six WPKs inside the file and each WPK has 998 time samples. The six WPKs correspond to the 6 columns from line 2 to the end of the file. The six WPKs shown in Listing 4.1 are for the two types of data functionals (i.e., the broadband cross-correlation delay time and amplitude anomaly) applied on the three different arrivals (P, PP, PS) on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. In practice it is often a good idea to put all the WPKs derived from the same synthetic seismogram into a single file. The kernel calculation in `ker` involves reading the forward wavefield and the RGT from disk into processor's memory. For the different WPKs derived from the same synthetic seismogram, `ker` reads the same forward wavefield and the same RGT. By combing those WPKs into a single file, `ker` will read the forward wavefield and the RGT only once for all those WPKs, which reduces the average I/O overhead.

An example MATLAB driver script `wpk_bbxcorr_hs.m` for computing the WPKs using the function `wpk_bbxcorr.m` and writing out the ASCII WPK file for the half-space example is included in F3DWI. A snippet of the output file generated by this driver script is shown in Listing 4.1. The ordering of the 6 WPK columns is listed in Table 4.3.

Note that the MATLAB function and script discussed in this section can be easily re-written using a compiled language (e.g., C/C++ and FORTRAN) and parallelized for efficiency. In practice, the total number of WPKs used in a realistic tomographic inversion can easily exceed millions and the calculation of WPKs needs to be repeated for every iteration of the inversion algorithm. The computational efficiency of this step is therefore highly important.

4.4.2 Kernel Calculation

The data sensitivity kernels are usually not as oscillatory as the wavefields used for computing the kernels and in practice we often regularize our inversion through damping, which allows us to sample the kernels on a mesh that is sparser than the mesh used for the wave-propagation simulations. The grid decimation rate in the x_1 , x_2 and x_3 direction can be adjusted by changing the values assigned to `NXSKP`, `NYSKP` and `NZSKP`, respectively, in the parameter file (Table 2.2, Listing 2.5, Sect. 2.2.3). For the half-space example, the mesh used for the wave-propagation simulations has a grid spacing of 200 meters (Listing 2.1, Listing 2.3, Sect. 2.2.1). For the examples in this section, I have set `NXSKP`, `NYSKP` and `NZSKP` to 4, which means that the kernels are sampled on a uniform mesh with 800-meter grid spacing.

The kernel calculations involve temporal convolutions between the forward wavefield and the RGT at every spatial grid point (e.g., Eqs. 4.32 and 4.35). The accuracy of this temporal convolution operation depends upon the time sampling interval for storing the forward wavefield and the RGT, which is controlled by the value assigned to `NTSKP` in the parameter file (Table 2.2, Listing 2.5, Sect. 2.2.3). It is usually sufficient to have ten time samples per minimum period. For the half-space example, the time sampling interval for the wave-propagation simulations is 0.015 s (Listing 2.5) and the maximum frequency is around 1.5 Hz. I have set `NTSKP` to 2 such that the time sampling interval for storing the forward wavefield and the RGT is 0.03 s, which gives more than 20 time samples per minimum period. The parameter file used in this section is identical to that shown in Listing 2.5. Every time we make a change to `NXSKP`, `NYSKP` and `NZSKP` in the parameter file, we need to re-run the script

```
set_ker_grd.py param_HS.dat
```

in the terminal under the project `JOBS` directory and re-compute all the forward wavefields and the RGTs by running `awp`.

Note that the kernels can also be sampled on a non-uniform mesh by modifying the script `set_ker_grd.py` (Sect. 2.2.5). For example, in certain applications we may wish to sample the kernels on a denser mesh at shallower depths and gradually

Table 4.4 The meanings of the first nine command-line arguments to the script `setup_ker.py`

Argument	Meaning
1	The full-path of the parameter file
2	The line number in the source-receiver list for the source
3	The line number in the source-receiver list for the receiver
4	Component number, 1: x_1 , 2: x_2 , 3: x_3 (Up), 4: Radial, 5: Transverse
5	The full-path of the WPK file
6	The kernel thresholding factor <code>kthresh</code> (see text)
7	Spatial grid buffer size <code>gbuff</code> (see text)
8	Temporal buffer size <code>tbuff</code> (see text)
9	Model parameterization flag (Table 4.5)

increase the spatial sampling interval at larger depths. In such situations, the script `set_ker_grd.py` can be modified to output non-uniformly distributed recording grid points.

The python script `setup_ker.py` in the `py` directory can be used to set up the correct environment for running the parallel binary kernel calculation code `ker` in the `bin` directory. For the half-space example, the following command can be executed in the terminal under the project `JOBS` directory.

```
setup_ker.py param_HS.dat 2 1 1~/wrk/F3DWI/HS/KER0/
100001_IN.RC01.1/bbxcor.wpk 1e4 6900 484 2 1 2 3
```

The complete command must be typed on a single line in the terminal. This script takes a variable number of command-line arguments. Depending upon the model parameterization used, the number of command-line arguments can range from 10 to 31. In practice it is often a good idea to write a driver script that calls `setup_ker.py` with automatically generated command-line arguments. The meanings of the first nine arguments are listed in Table 4.4.

The component number specified on the fourth argument (Table 4.4) can take on five different values in the current implementation. The number 1, 2 or 3 specifies the x_1 -, x_2 - or the x_3 -component of the box coordinate. The number 4 specifies the radial component, which is defined as from the source to the receiver. The number 5 specifies the transverse component, which is defined as clockwise from the radial direction by 90° and forms a left-hand system with the x_3 and the radial directions. In the half-space example, we are using the WPKs derived from the x_1 -component synthetic seismogram for the kernel calculation, therefore the fourth argument is set to 1.

The data sensitivity kernel, as a function of space, has highly variable amplitudes and usually has the largest amplitudes at the vicinity of the source and the receiver.

The kernel threshold factor `kthresh` on the sixth argument (Table 4.4) allows us to apply a threshold based on kernel amplitudes. Suppose the maximum amplitude of a kernel $K(\mathbf{x})$ is K_{\max} . If at a certain spatial location \mathbf{x} we have $|K(\mathbf{x})| < K_{\max}/kthresh$, then the kernel value at \mathbf{x} is treated as zero when `ker` writes out the kernel output file in a sparse matrix format.

Inside the kernel calculation code `ker`, the reading of the forward wavefield and the RGT is buffered to reduce the I/O overhead. The spatial and temporal buffer sizes `gbuff` and `tbuff` specified on the seventh and eighth argument (Table 4.4) determine the total amount of data each reading operation will load from disk into memory. To reduce the total number of reads, it is desirable to set `gbuff` and `tbuff` to large numbers. However, on modern distributed-memory parallel computing systems (Sect. 1.4), the total amount of memory accessible by each processor is usually limited and it takes some experimentation to find the optimal buffer sizes for different parallel computing systems.

For `gbuff`, a good number to start with is the number of kernel recording grid points in each sub-mesh (Sect. 2.2.5). For the half-space example, this number can be found on the first line of the `HSGRDRECprocID` files inside the project `VM` directory (i.e., the value assigned to the key `MEDIA` in the parameter file, Line 7, Listing 2.5), where `procID` is the 9-digit ID composed of (`ipx`, `ipy`, `ipz`) for identifying the different sub-meshes (Sect. 2.2.2). This number can be different for different sub-meshes and we can choose the largest one as our first try for `gbuff`. The Linux command for finding this number is

```
head -q -n 1 ../VM/HSGRDREC0* | sort -n | tail -n 1
```

which should be executed in the terminal under the `JOBS` directory. In this example, I have set `gbuff` to 6900, which works on my Linux box. If `ker` gives `Segmentation Fault` during the run on your system, you may want to reduce the `gbuff` value until `ker` does not complain.

For `tbuff`, a good first try is the largest number on line 3 of the `WPK` file (Listing 4.1). For the half-space example, the largest number shown on line 3 in Listing 4.1 is 484, therefore I have set the eighth argument to 484 as well. If `ker` gives `Segmentation Fault` on your system, you may have to reduce this number until `ker` can successfully complete.

The possible values for the model parameterization flag on the ninth argument and also the meanings of the following arguments are summarized in Table 4.5. For the half-space example, the model parameterization flag is set to 2, which selects the isotropic model parameterized using the P- and S-velocity and density (Table 4.5). I have set the 10th, 11th and 12th arguments to 1, 2 and 3, respectively, which indicates that `ker` will compute and output $K_{dsn}^{\ln\alpha}$, $K_{dsn}^{\ln\beta}$ and $\check{K}_{dsn}^{\ln\rho}$. If we would like to obtain $K_{dsn}^{\ln\alpha}$ and $K_{dsn}^{\ln\beta}$ only, we can remove the 12th argument. If we would

Table 4.5 The model parameterization flag on the ninth argument to `setup_ker.py`, the corresponding model parameterization and the meaning of the arguments following

Flag	Parameterization	Arguments following
0	Isotropic model parameterized using the bulk and shear moduli and density (Sect. 4.3.2)	1: $K_{dsn}^{\ln\kappa}$, 2: $K_{dsn}^{\ln\mu}$, 3: $K_{dsn}^{\ln\rho}$
1	Isotropic model parameterized using the two Lamé parameters and density (Sect. 4.3.3)	1: $K_{dsn}^{\ln\lambda}$, 2: $\check{K}_{dsn}^{\ln\mu}$, 3: $K_{dsn}^{\ln\rho}$
2	Isotropic model parameterized using the P- and S-velocity and density (Sect. 4.3.4)	1: $K_{dsn}^{\ln\alpha}$, 2: $K_{dsn}^{\ln\beta}$, 3: $\check{K}_{dsn}^{\ln\rho}$
3	Transversely isotropic model with a vertical symmetry axis parameterized using the five Love's parameters (Sect. 4.3.5)	1: K_{dsn}^A , 2: K_{dsn}^C , 3: K_{dsn}^N , 4: K_{dsn}^L , 5: K_{dsn}^F
4	Fully anisotropic model parameterized using the 21 asymptotic parameters (Sect. 4.3.6)	1: K_{dsn}^A , 2: K_{dsn}^C , 3: K_{dsn}^N , 4: K_{dsn}^L , 5: K_{dsn}^F , 6: K_{dsn}^Jc , 7: K_{dsn}^Js , 8: K_{dsn}^{Kc} , 9: K_{dsn}^{Ks} , 10: K_{dsn}^{Mc} , 11: K_{dsn}^{Ms} , 12: K_{dsn}^{Gc} , 13: K_{dsn}^{Gs} , 14: K_{dsn}^{Bc} , 15: K_{dsn}^{Bs} , 16: K_{dsn}^{Hc} , 17: K_{dsn}^{Hs} , 18: K_{dsn}^{Dc} , 19: K_{dsn}^{Ds} , 20: K_{dsn}^{Ec} , 21: K_{dsn}^{Es} , 22: $K_{dsn}^{\ln\rho}$

like to obtain $K_{dsn}^{\ln\beta}$ only, we can set the tenth argument to 2 and remove the following arguments. The same usage is also applicable to other values of the model parameterization flag.

Note that there is not a flag value corresponding to the kernels of anelastic attenuation quality factors $K_{dsn}^{Q\kappa^{-1}}$ and $K_{dsn}^{Q\mu^{-1}}$ (Sect. 4.3.7) in Table 4.5. To compute $K_{dsn}^{Q\kappa^{-1}}$ and $K_{dsn}^{Q\mu^{-1}}$, we only need to set the model parameterization flag to 0 and change the WPKs according to Sect. 4.3.7. I have included a MATLAB script `wpk_bbxcora_Q.m` in F3DWI to show how to change the WPK for the broadband cross-correlation amplitude anomaly (Sect. 4.2.6) for the P, PP and PS waves on the x_1 -component synthetic velocity seismogram from 100001 to IN.RC01 in the half-space example. The calculations in `wpk_bbxcora_Q.m` follow Eq. 4.296 in Sect. 4.3.7. The modified WPKs for the three waveforms are shown in Fig. 4.23 and the data sensitivity kernels $K_{dsn}^{Q\kappa^{-1}}$ and $K_{dsn}^{Q\mu^{-1}}$ for the three waveforms are shown in Fig. 4.24.

The script `setup_ker.py` will generate the correct directory structure and also the input files for `ker`. It will create the directory `KER` under the current project work directory (key `WRKDIR` in the parameter file, Line 3, Listing 2.5), if it does not exist yet, and also a sub-directory named as `sourceID_receiverID.componentID` inside `KER`. For the half-space example, the `sourceID` is 100001, the `receiverID` is IN.RC01 and the

```

1 2      ! model parameterization flag
2 1 1 1 ! kernel selection flag
3 /home/pochen/wrk/F3DWI/HS/VM/XYZGRD ! full path of mesh file
4 /home/pochen/wrk/F3DWI/HS/VM/HSGRD ! path to material properties
5 /home/pochen/wrk/F3DWI/HS/FWF/100001/ ! path to forward wavefield
6 /home/pochen/wrk/F3DWI/HS/FWF/IN.RC01 ! path to RGT
7 /home/pochen/wrk/F3DWI/HS/KER/100001_IN.RC01.1/wpk ! WPK file
8 /home/pochen/wrk/F3DWI/HS/KER/100001_IN.RC01.1/ker ! output
9 201 40 ! horizontal grid location of source
10 40 40 ! horizontal grid location of receiver
11 4 4 4 ! grid decimation rates in x, y, z
12 1      ! component ID: 1-x, 2-y, 3-z, 4-radial, 5-transverse
13 0      ! visco
14 1e4    ! kthresh
15 6900 484 ! gbuff, tbuff

```

Listing 4.2 The kernel input file for the half-space example.

componentID is 1 (i.e., the fourth argument to `setup_ker.py`) and this sub-directory is named as `100001_IN.RC01.1`. The `sourceID` and `receiverID` are identical to those given on the second column of the source-receiver list file (Listing 2.6). Inside this sub-directory, `setup_ker.py` will create two symbolic links: `ker`, which points to the parallel binary executable `ker` in the `bin` directory, and `wpk`, which points to the WPK file specified in the fifth argument to `setup_ker.py` (Table 4.4). And `setup_ker.py` will also create an ASCII text file `ker.in`, which is the master input file for `ker`.

The content of the kernel input file `ker.in` created by `setup_ker.py` for the half-space example is shown in Listing 4.2. Line 1 is the model parameterization flag specified in the 9th argument to `setup_ker.py`. Line 2 consists of a set of flags indicating which kernel should be computed. For the model parameterization flag of 2, three types of kernels can be computed, $K_{dsn}^{\ln\alpha}$, $K_{dsn}^{\ln\beta}$ and $\check{K}_{dsn}^{\ln\rho}$ (Table 4.5). If we require only $K_{dsn}^{\ln\alpha}$, the first number on line 2 should be set to 1 and the other two numbers should be set to 0. If we require only $K_{dsn}^{\ln\beta}$, the second number on line 2 should be set to 1 and the other two numbers should be set to 0. For the example shown in Listing 4.2, we require all three types of kernels, therefore all three numbers are set to 1. These flags in `ker.in` have been set automatically by `setup_ker.py` when I specified the 9th-12th arguments. Line 3 is the full path of the mesh file `XYZGRD` (line 5, Listing 2.5). Line 4 is the full path and the prefix of the material property files (line 7, Listing 2.5). Line 5 specifies the path to the forward wavefield from the source. Line 6 specifies the path to the RGT from the receiver. Line 7 is the full path to the WPK file, which is now the full path of the symbolic link `wpk` created by `setup_ker.py`. Line 8 specifies the path and the prefix of the kernel output files. Line 9 and 10 are the horizontal grid indices of the source and the receiver, respectively, which should be identical to the corresponding values inside the `srlst.grd` file (Listing 2.11, Sect. 2.2.4.2). Line 11 consists of the values assigned to `NXSKP`, `NYSKP` and `NZSKP` (line 21–23,

Listing 2.5). Line 12 is the component number specified in the fourth argument to `setup_ker.py` (Table 4.4). Line 13 is the value assigned to the key `VISCO` in the parameter file (line 17, Listing 2.5). Line 14 is the `kthresh` value specified in the sixth argument to `setup_ker.py` (Table 4.4). Line 15 consists of the two buffer sizes, `gbuff` and `tbuff`, specified in the eighth and ninth arguments to `setup_ker.py`, respectively (Table 4.4).

We can run the parallel binary executable `ker` from inside the sub-directory `KER/100001_IN.RC01.1` by typing the following command in the terminal,

```
mpiexec -n 8 -f ~/tools/mpi/share/myhosts ./ker ker.in
```

The total number of processors used in the command above is eight, which is the same as the total number of processors used in the wave-propagation simulations (Sect. 2.2.6) and also the same as the number of sub-meshes. The number of processors used in the kernel calculation can be fewer than the number of sub-meshes. In such a case, some or all of the processors will have to compute the kernels for more than one sub-mesh. The kernel calculation involves essentially a temporal convolution and a temporal integration at each spatial grid point. This calculation is easily parallelized, as the calculations on different grid points have no inter-dependence, therefore incur zero inter-processor communication overhead on distributed-memory computer clusters, as well as different types of co-processors or accelerators (e.g., GPU, Intel Xeon Phi).

4.4.3 Kernel Output and Visualization

After `ker` successfully completes its calculations, a set of binary files named as `kerprocID` are generated under the sub-directory `KER/100001_IN.RC01.1`. Here the prefix `ker` was specified on line 8 of `ker.in` (Listing 4.2) and the `procID` is the nine-digit number for identifying the sub-meshes. This set of files store the kernel values at grid points \mathbf{x} where the absolute value of the kernel $|K(\mathbf{x})| \geq K_{\max}/k_{\text{thresh}}$ (Table 4.4) and K_{\max} is the maximum absolute value of the kernel throughout the whole mesh. If a sub-mesh has no kernel values above the threshold $K_{\max}/k_{\text{thresh}}$, then no output file will be written for that sub-mesh. It is therefore possible that the number of kernel output files is less than the total number of sub-meshes.

Inside the kernel output file for a sub-mesh, the kernel values are stored inside a `nnz-by-3` matrix, where `nnz` is the total number of non-zero (i.e., above the threshold) kernel values inside the sub-mesh. The format of this matrix resembles the

coordinate list (COO) format⁷ for storing sparse matrices. The first column of this matrix stores the row number, `rowt`, the second column stores the column number, `icol`, and the third column stores the actual kernel value. The row number `rowt` corresponds to the column number of the WPK inside the WPK file that was used as the input to `ker` (line 6, Listing 4.2). For the half-space example, we have six WPKs (i.e., broadband cross-correlation delay-time and amplitude anomaly for the P, PP and PS waves) and three different kernel types (i.e., $K_{dsn}^{\ln\alpha}$, $K_{dsn}^{\ln\beta}$ and $\check{K}_{dsn}^{\ln\rho}$), therefore we have 18 different kernels in total. The row number `rowt` will range from 1 to 6 with one corresponding to the WPK for the broadband cross-correlation delay-time of the P wave and six corresponding to the WPK for the amplitude anomaly of the PS wave (Table 4.3). The column number `icol` is computed using the following formula

$$\text{icol} = \text{ikx} + \text{nkx} * (\text{iky} - 1) + \text{nkx} * \text{nky} * (\text{ikz} - 1) + \text{nkx} * \text{nky} * \text{nkz} * (\text{iktype} - 1),$$

where $(\text{ikx}, \text{iky}, \text{ikz})$ are the global grid indices in the kernel recording mesh (Sect. 2.2.5), nkx , nky and nkz are the total number of recording grid points in the x_1 , x_2 and x_3 directions, respectively, and iktype is an index for the different kernel types (Table 4.5). For the half-space example, the total number of grid points NX , NY and NZ are 240, 76, 180 (Listing 2.3) and the grid decimation rates NXSKP , NYSKP and NZSKP are set to 4 (line 21–23, Listing 2.5), therefore nkx , nky and nkz are 60, 19 and 45, respectively. Since the model parameterization flag is 2 and I have selected all three kernel types (line 1–2, Listing 4.2), iktype ranges from 1 to 3 with 1 corresponding to $K_{dsn}^{\ln\alpha}$ and 3 corresponding to $\check{K}_{dsn}^{\ln\rho}$. If I had selected two kernel types by changing line 2 in Listing 4.2 to, say, 0 1 1, then iktype would range from 1 to 2 with 1 corresponding to $K_{dsn}^{\ln\beta}$ and 2 corresponding to $\check{K}_{dsn}^{\ln\rho}$. This particular format for storing the kernel output is chosen to minimize the I/O overhead when loading all kernels from disk to memory for solving the Gauss-Newton normal equation using our highly optimized parallel LSQR code.

The kernel output files are written in binary format. If we would like to convert them into an ASCII text file, we can use the sequential binary executable `rdker` in the `bin` directory. Our parallel LSQR code reads the binary kernel output files directly. The conversion to an ASCII text file is merely for visualization purposes. To use the `rdker` utility, we can execute the following two commands under the sub-directory `KER/100001_IN.RC01.1` in the terminal.

```
cat ker?????????? > ker_all
rdker ker_all 60 19 45 3 > bbxcor.ker2
```

⁷ The COO format stores a list of (row, column, value) tuples.

```

1  2  1  1  4.23408e-13  1  3
2  2  1  1  1.99032e-14  1  4
3  3  1  1 -5.20901e-13  1  3
4  3  1  1  9.6122e-14  1  4
5  3  1  1 -6.22554e-15  1  5
6  4  1  1 -3.06489e-13  1  3
7  4  1  1 -3.43248e-14  1  4
8  4  1  1 -1.30518e-14  1  5
9  5  1  1  1.34905e-13  1  3
10 5  1  1 -2.08818e-14  1  4

```

Listing 4.3 The first 10 lines inside the ASCII kernel output file. The non-zero kernel values are not ordered in any specific way.

The first command is for concatenating the binary kernel output files for different sub-meshes into a single binary file, named `ker_all` here, which is then used as the first command-line argument to `rdker` in the second command. The command `cat` is a utility of the Linux/Unix operating system. Our parallel LSQR code can read the concatenated binary file `ker_all` as well. Once `ker_all` is generated, the individual kernel files for each sub-mesh `ker????????` can be deleted from the disk to reduce the total number of files. The binary executable `rdker` in `bin` takes five command-line arguments. The first argument is the binary kernel file `ker_all`, the second, third and fourth arguments are `nkx`, `nky` and `nkz`, respectively. The fifth argument is the total number of kernel types stored in the binary kernel file. For the half-space example, `nkx`, `nky` and `nkz` are 60, 19 and 45, respectively, as computed above. And we have 3 types of kernels stored in `ker_all`. The output of `rdker` is re-directed into the ASCII output file, named `bbxcor.ker2` here. This ASCII text file has six columns. The first three columns are for `ikx`, `iky` and `ikz`, i.e., the grid indices of the non-zero kernel value in the kernel recording mesh. The fourth column is for the kernel value at that grid point. The fifth column is for `iktype`, i.e., the type of the kernel. The sixth column is for `irowt`, which is the index for the WPK. A snippet of `bbxcor.ker2` for our half-space example is shown in Listing 4.3. I have included a MATLAB function `read_ker.m` in `F3DWI` for reading the ASCII kernel output file generated by `rdker` into a 6-dimensional MATLAB matrix, which can then be plotted using the MATLAB driver script `ker_bbxcorr_plot.m` also included in `F3DWI`. The definition of the function `read_ker.m` is in the following

```
function K=read_ker(kerfnam,nkx,nky,nkz,nktype,nwpk)
```

The first input is the full path of the ASCII output file generated by `rdker`. The second to fourth inputs are `nkx`, `nky` and `nkz`, respectively. The fifth input is the total number of kernel types. The sixth input is the total number of WPKs. The figure generated by the driver script `ker_bbxcorr_plot.m` should be similar to that shown in Fig. 4.8.

```

1 /home/pochen/wrk/HS/VM/
2 11
3 -118 33.50
4 -118 33.73
5 50000
6 35500
7 50
8 16 8 16

```

Listing 4.4 Input file to the mesh generation code for the larger mesh.

4.4.4 More Kernel Examples

The examples we have been discussing so far only use eight CPU cores and can be completed on a single desktop (or laptop) computer. The `F3DWI` code is capable of scaling to more than tens of thousands of cores. I will expand the volume of the half-space structural model to about 50 km in the x_1 direction and to about 25 km in the x_2 direction and keep the dimension in the x_3 direction unchanged. I will reduce the grid spacing for wave-propagation simulations from 200 to 50 m. The resulting mesh has over 371 million grid points and I will use 2048 cores on the Blue Gene/Q system Mira at ALCF (Sect. 2.2).

The input file to the binary executable `FD_GRID_XYZ_rect` (Sect. 2.2.1) for this larger mesh is shown in Listing 4.4. Snippets in the corresponding mesh file `XYZGRD` is shown in Listing 4.5. To satisfy the CFL stability condition in the wave-propagation simulation, I have reduced the time-sampling interval from 0.015 s to 0.00375 s. The spatial decimation rates for the uniform recording grid points `NXSKP`, `NYSKP` and `NZSKP` are all set to 4 (the value here can be increase to 16 without significant loss of kernel resolution) and the temporal decimation rate `NTSKP` is set to 2. The parameter file is shown in Listing 4.6.

4.4.4.1 S-Wave From a Double-Couple Source

I have added a double-couple source 100002 to the source-receiver list. The modified source-receiver list `srlst` is shown in Listing 4.7. The source-time function is kept the same as shown in Fig. 2.7 and Listing 2.7. The double-couple source 100002 is located at the same location as the explosive source 100001, which I have been using in previous examples. The source-type flag for 100002 on column 7, line 4 is 4, which indicates that column 8–13 in the following specifies the moment tensor in the box coordinate (Table 2.3, Sect. 2.2.4.1). This moment tensor was generated by a double-couple source with 45° dip, 45° rake and a strike in the direction of the x_2 -axis of the box coordinate. Every time we change the source-receiver list file, we need to re-run the following two scripts in the terminal under the project `JOBS` directory.

```

1  1008  512  720      ! Global grid numbers NX, NY, NZ
2    16    8   16      ! Chunks (#of proc.) in X, Y, Z
3    63   64   45      ! Grid numbers per chunk NXT, NYT, NZT
4  407110.70348136901      3706961.4617379787
5  407357.58726096048      3732510.2644716040
6  457705.22787301656      3732023.7439901000
7  457458.34409342508      3706474.9412564747
8                720  -0.0000000000000000
9                719  -50.0000000000000000
10               718  -100.0000000000000000
11               .
12               .
13               .
14               3  -35850.0000000000000000
15               2  -35900.0000000000000000
16               1  -35950.0000000000000000

```

Listing 4.5 The first ten lines and the last three lines in the mesh file generated for the larger mesh.

```

1  BINDIR = /home/pochen/F3DWI/bin
2  SCRIPTDIR = /home/pochen/F3DWI/py
3  WRKDIR = /home/pochen/wrk/HS
4  JOBDIR = /home/pochen/wrk/HS/JOBS
5  XYZGRD = /home/pochen/wrk/HS/VM/XYZGRD
6  UTMZONE = 11
7  MEDIA = /home/pochen/wrk/HS/VM/HSGRD
8  MINIMUM VELOCITY = 3500.0
9  MAXIMUM VELOCITY = 6500.0
10 SOURCE RECEIVER LIST = /home/pochen/wrk/HS/JOBS/srinfo/sr1st
11 TMAX = 15.0
12 DT = 0.00375
13 CFL = 0.9792
14 USEPML = 1
15 PMLCOEF = 3.5
16 BOUNDARY NODE NUMBER = 10
17 VISCO = 0
18 SOURCE STENCIL SIZE IN X = 1
19 SOURCE STENCIL SIZE IN Y = 1
20 SOURCE STENCIL SIZE IN Z = 1
21 NXSKP = 4
22 NYSKP = 4
23 NZSKP = 4
24 NTSKP = 2

```

Listing 4.6 The parameter file used for the larger mesh.

```

proc_sr1st.py param_HS.dat 1 7
set_sr_grd.py param_HS.dat

```

```

1  0 IN.RC01   1 160 257 256
2  0 SF.RC02   1 160 257 720
3  0 SF.RC03   1 804 257 720
4  1 100002    1 804 257 256   4 -0.707106781186547 0
      0.707106781186547 0.5 0 0.5 /home/pochen/wrk/HS/JOBS/ srinfo/
      src_hs.dat
5  2 IN.RC01.2 1 160 257 256 -2 0 1 0 /home/pochen/wrk/HS/JOBS/
      srinfo/imp_hs.dat
6  2 SF.RC02.3 1 160 257 720 -2 0 0 1 /home/pochen/wrk/HS/JOBS/
      srinfo/imp_hs.dat
7  2 SF.RC03.3 1 804 257 720 -2 0 0 1 /home/pochen/wrk/HS/JOBS/
      srinfo/imp_hs.dat

```

Listing 4.7 The modified source-receiver list for the half-space example. A double-couple point source is added on line 4. The body-force sources associated with two surface receivers are added on line 6–7. The receiver lines associated with the two surface receivers are added on line 2–3.

I will be computing data sensitivity kernels for the S-wave from 100002 to the x_2 -component of the receiver IN.RC01. Therefore I will need the forward wavefield from 100002 and the RGT for IN.RC01.2. To set up the directories for the forward wavefield calculation for 100002 and the RGT calculations for IN.RC01.2, I will need to execute the following command in the terminal under the project JOBS directory.

```
setup_awp.py param_HS.dat FWF 4 5
```

This command generates two sub-directories under the project FWF directory: 100002 and IN.RC01.2. Afterwards, I will need to run the parallel binary executable awp to generate the wavefields following the instructions given in Sect. 2.2.6. On Mira, parallel jobs must be submitted to the job scheduling system (Footnote 27) and the parallel job launcher is different from the one installed on my Linux box. Such supercomputers usually have detailed user’s guide, which can be consulted with when running parallel jobs on them.

After the wave-propagation simulations are completed, the synthetic seismograms from 100002 to IN.RC01 can be extracted by executing the following command in the terminal under the JOBS directory (Sect. 2.2.7)

```
fwf2syn.py param_HS.dat 4 4 1 1
```

and the synthetic seismograms are stored inside the file IN.RC01.rot under the sub-directory SYN/100002/. The synthetic velocity seismograms on the x_1 , x_2 and x_3 components of IN.RC01 are shown in Fig. 4.27.

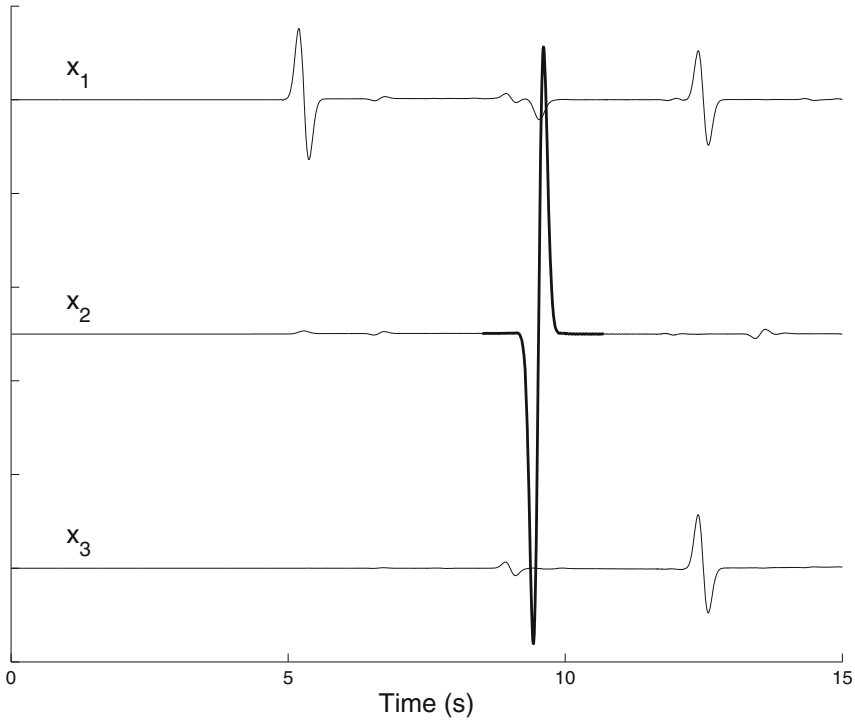


Fig. 4.27 Three-component synthetic velocity seismograms from the double-couple source 100002 to IN.RC01 in our half-space example with the larger mesh. *Top trace*: x_1 component; *center trace*: x_2 component; *bottom trace*: x_3 component. The S-wave on the x_2 component (*black thick line on center trace*) is used for computing the WPK.

```

1  2 3998
2  1.134000e+03 1.134000e+03
3  1.427000e+03 1.427000e+03
4  0.000000e+00 -0.000000e+00
5  ...

```

Listing 4.8 The first four lines inside the WPK file for the double-couple S-wave example.

I will use the WPKs for the broadband cross-correlation delay time and amplitude anomaly (Sect. 4.2.6) for computing the data sensitivity kernels in this example. The WPKs can be generated using the MATLAB function `wpk_bbxcorr.m` as shown in Sect. 4.4.1. The WPKs for the delay time and amplitude anomaly derived from the S-wave shown in Fig. 4.27 are plotted in Fig. 4.28 and a snippet of the generated WPK file, named `bbxcorr_S.wpk`, is shown in Listing 4.8.

To set up the environment for the kernel calculations, I can execute the following command in the terminal under the JOBS directory

```

1 2
2 1 1 1
3 /home/pochen/wrk/HS/VM/XYZGRD
4 /home/pochen/wrk/HS/VM/HSGRD
5 /home/pochen/wrk/HS/FWF/100002/
6 /home/pochen/wrk/HS/FWF/IN.RC01
7 /home/pochen/wrk/HS/KER/100002_IN.RC01.2/wpk
8 /home/pochen/wrk/HS/KER/100002_IN.RC01.2/ker
9 804 257
10 160 257
11 4 4 4
12 2
13 0
14 1e7
15 3072 1427

```

Listing 4.9 The kernel input file for our double-couple S-wave example.

```

setup_ker.py param_HS.dat 4 1 2 /home/pochen/wrk/HS/
JOBS/srinfo/bbxcor_S.wpk 1e7 3072 1427 2 1 2 3

```

where the values for `gbuff` and `tbuff` were found using the method discussed in Sect. 4.4.2. The model parameterization flag on the 9th argument is set to 2, which means that I am computing the data sensitivity kernels with respect to the relative perturbations in P- and S-velocity and density (Table 4.5). The sub-directory `KER/100002_IN.RC01.2` is then generated by `setup_ker.py`, together with the kernel input file. The kernel input file `ker.in` generated by `setup_ker.py` is shown in Listing 4.9. The data sensitivity kernels can then be obtained by running `ker`. The kernel output files can be concatenated into a single file using the following command inside the sub-directory `KER/100002_IN.RC01.2`

```

cat ker?????????? > ker_all

```

and the individual kernel output files for each sub-mesh can then be deleted. The ASCII kernel output file can be generated by executing the following command inside the sub-directory `KER/100002_IN.RC01.2`

```

rdker ker_all 252 128 180 3 > bbxcor_S.ker2

```

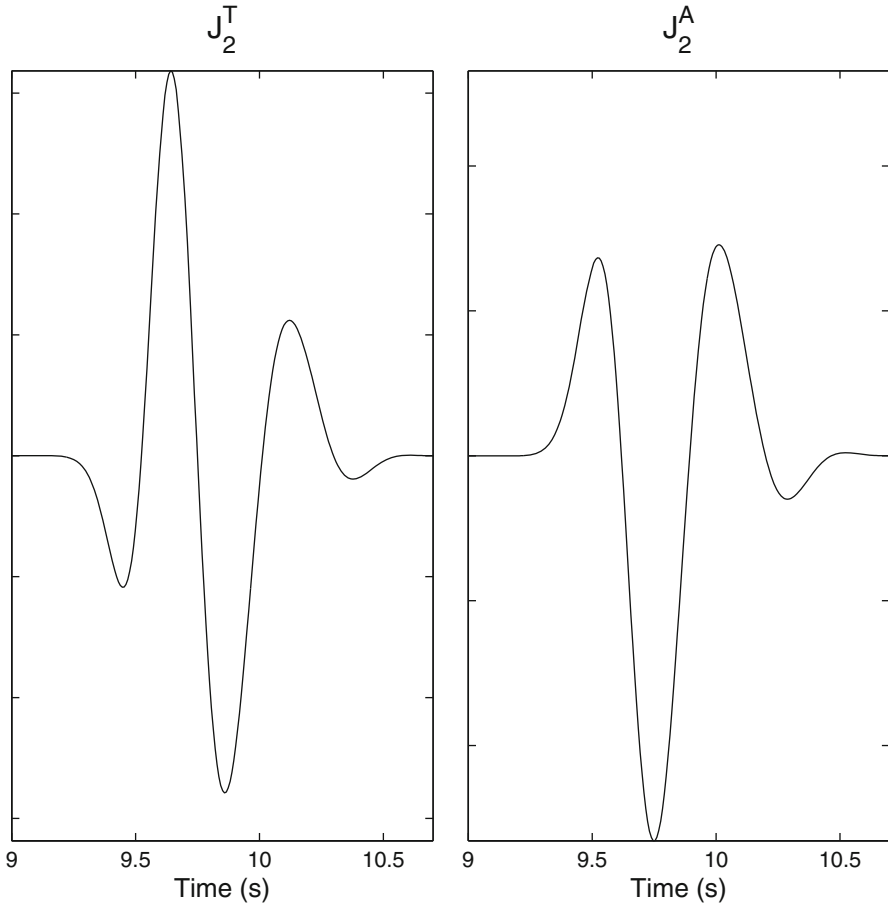


Fig. 4.28 The WPKs for the broadband cross-correlation delay time (*left*) and amplitude anomaly (*right*) computed from the selected S-wave shown in Fig. 4.27.

and this ASCII file `bbxcor_s.ker2` can be read into MATLAB using the function `read_ker.m` and plotted.

The source-receiver plane cross-section views of the two data sensitivity kernels are shown in Fig. 4.29. Note that the amplitude kernel in Fig. 4.29 differs from that shown in (Zhao et al. 2005) by a minus sign, which is due to the differences in the definition of the amplitude misfit (Sect. 4.2.6). Compared with the data sensitivity kernels for the P-wave shown in Fig. 4.8, the S-wave kernel for delay-time in Fig. 4.29 also exhibits the counter-intuitive “banana-doughnut” phenomena. The Fresnel zone of the S-wave kernels in Fig. 4.29 is narrower than the Fresnel zone of the P-wave kernels in Fig. 4.8 because of the shorter wavelength of the S-wave (Eq. 4.135, Sect. 4.2.9). Unlike the P-wave kernels for the explosive source shown in Fig. 4.8), the kernels shown in Fig. 4.29 are not symmetric with respect

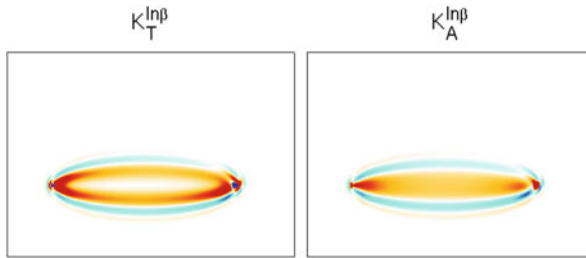


Fig. 4.29 Cross-section views of the data sensitivity kernel for the broadband cross-correlation delay time (*left*) and amplitude anomaly (*right*) with respect to relative perturbations in S-velocity for the S-wave on the x_2 -component synthetic velocity seismogram from 100002 to IN.RC01 in our half-space example. The cross-section lies in the source-receiver plane perpendicular to the x_2 direction.

to the source-receiver path, which is due to the S-wave radiation pattern from the double-couple source used in the calculation.

4.4.4.2 Rayleigh Wave From a Vertical Body-Force

I will solve the Lamb's problem in the half-space structural model. Two receivers SF.RC02 and SF.RC03 are placed at the surface (Fig. 3.3) and an upward vertical body-force source is applied at SF.RC03 and the synthetic velocity seismograms are recorded at SF.RC02. The data sensitivity kernels will be computed for the Rayleigh wave on the vertical component of SF.RC02 and we therefore need the RGT for the x_3 component of SF.RC02. The source-receiver list file for this example is shown in Listing 4.7. The vertical body force applied on SF.RC03 is on line 7 and the body-force source for the RGT of SF.RC02 is on line 6. The receiver line for SF.RC02 is line 2. Since the source-receiver list file is kept unchanged from that used in Sect. 4.4.4.1, I do not need to re-execute the scripts `proc_srlst.py` and `set_sr_grd.py`. For this example, I have changed the spatial grid decimation rate NXSKP, NYSKP and NZSKP in the parameter file to 8 (line 21–23, Listing 2.5) and re-executed the script `set_ker_grd.py` (Sect. 2.2.5). To set up the directories for the forward wave-propagation simulations for SF.RC03 and the RGT calculations for SF.RC02, I can execute the following command in the terminal under the project JOBS directory.

```
setup_awp.py param_HS.dat FWF 6 7
```

This command generates two sub-directories under the project FWF directory: SF.RC02.3 and SF.RC03.3 and I need to run the parallel executable `awp` to

generate the forward wavefield and the RGT. After the wave-propagation simulations are completed, I can extract the synthetic velocity seismograms by executing the following command in the JOBS directory.

```
fwf2syn.py param_HS.dat 7 7 2 2
```

The synthetic velocity seismograms on the x_1 , x_2 and x_3 components of `SF.RC02` are stored in the file `SF.RC02.rot` under the sub-directory `SYN/SF.RC03.3` and are plotted in Fig. 4.30. The dominant waveform on the vertical-component synthetic velocity seismogram is a Rayleigh wave.

For this example, I will use the WPKs for the frequency-dependent phase and amplitude misfits (Sect. 4.2.7) at 0.5 and 2.0 Hz. A MATLAB function `wpk_dtau.m` is included to generate the WPKs for the frequency-dependent phase-delay time, group-delay time and amplitude-reduction time. The definition of this function is the following

```
function [sconv, Jp, Jg, Jq]=wpk_dtau(syn,stf_grn,dt,
ntskp,w,F),
```

and it is written according to Eqs. 4.92, 4.93 and 4.94 in Sect. 4.2.7. It takes six inputs. The first five inputs have the same meaning as in the `wpk_bbxcorr.m` function discussed in Sect. 4.4.1. The last input `F` is a vector containing a list of the frequencies at which the frequency-dependent misfit measurements are made. In this example, `F` contains two numbers: 0.5 and 2.0, for the two frequencies at which the kernels will be computed. The first output `sconv` has the same meaning as that in `wpk_bbxcorr.m` (Sect. 4.4.1). The second to fourth outputs are the WPKs for the frequency-dependent phase-delay, group-delay and amplitude-reduction time, respectively, in the same format as the second and third outputs of `wpk_bbxcorr.m` (Sect. 4.4.1). The WPKs for the Rayleigh wave on the vertical-component synthetic velocity seismogram from `SF.RC03` to `SF.RC02` (Fig. 4.30) are shown in Fig. 4.31. The WPKs can be written into an ASCII text file in the format specified in Sect. 4.4.1. A snippet of the WPK file, which is stored in the file `JOBS/srinfo/dtau_R.wpk` is shown in Listing 4.10. The ordering of the 6 columns inside the WPK file is such that the first three columns correspond to the WPKs of the phase-delay, group-delay and amplitude-reduction time at 0.5 Hz, respectively, and the last three columns correspond to the WPKs of the phase-delay, group-delay and amplitude-reduction time at 2.0 Hz, respectively.

I will compute the data sensitivity kernels of the frequency-dependent phase-delay time, group-delay time and amplitude-reduction time with respect to perturbations in the 21 asymptotic parameters (Sect. 4.3.6). To set up the directory for kernel calculations, I can execute the following command under the JOBS directory

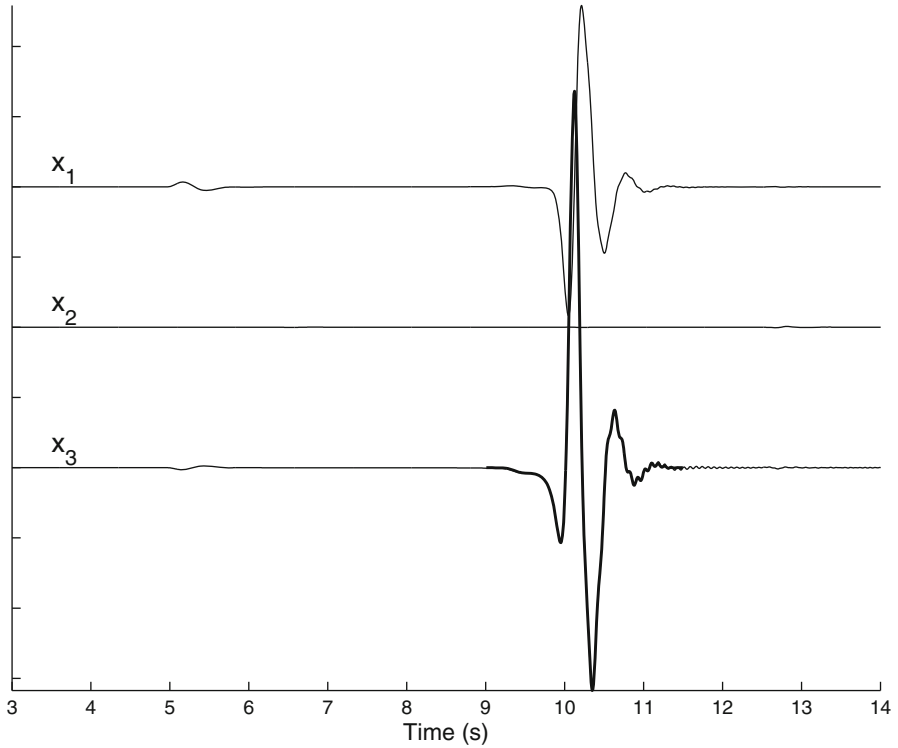


Fig. 4.30 Three-component synthetic velocity seismograms from the vertical body-force source applied at SF.RC03 to the receiver SF.RC02 in the half-space model. *Top trace:* x_1 component; *center trace:* x_2 component; *bottom trace:* x_3 component. The thick black solid line on the bottom trace shows the waveform used for computing the WPKs.

```

1 6 3998
2 1.134000e+03 1.134000e+03 1.134000e+03 1.134000e+03 1.134000e+03
  1.134000e+03
3 1.667000e+03 1.667000e+03 1.667000e+03 1.667000e+03 1.667000e+03
  1.667000e+03
4 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
  -0.000000e+00
5 ...

```

Listing 4.10 The first 4 lines inside the WPK file generated for the frequency-dependent phase and amplitude data functionals on the vertical-component Rayleigh wave.

```

setup_ker.py param_HS.dat 7 2 3 /home/pochen/wrk/HS/
JOBS/srinfo/dtau_R.wpk 1e7 384 1667 4 1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16 17 18 19 20 21 22

```

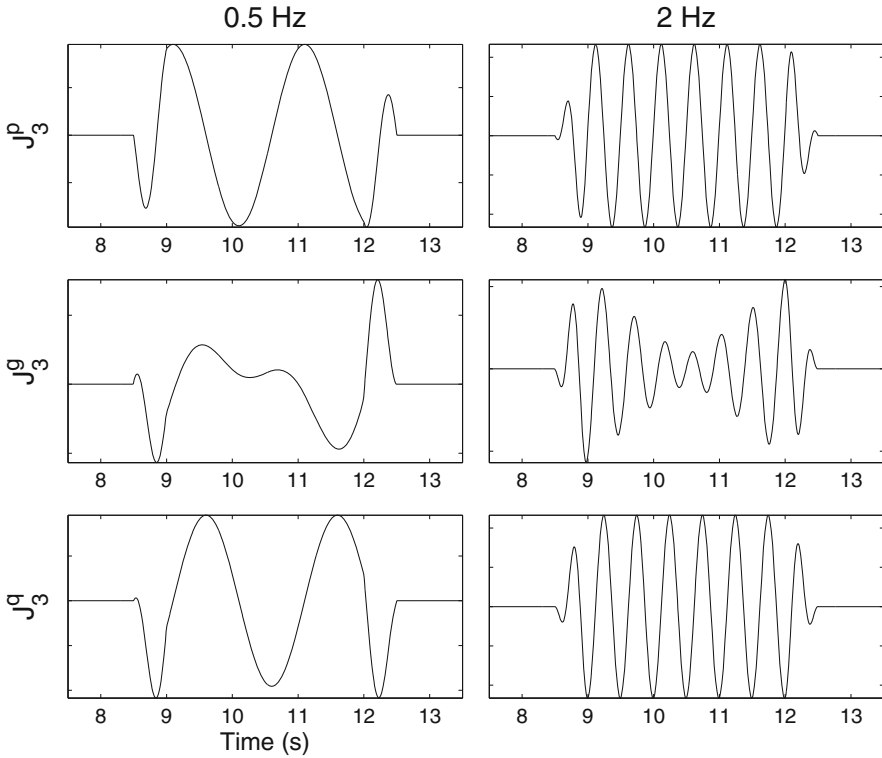


Fig. 4.31 The WPKs for frequency-dependent phase-delay time (*top row*), group-delay time (*center row*) and amplitude-reduction time (*bottom row*) at 0.5 Hz (*left column*) and 2 Hz (*right column*). The synthetic velocity waveform used for computing these WPKs is shown in thick *black solid line* in Fig. 4.30.

Here I have set the model parameterization flag to 4 and the kernels for all 21 asymptotic parameters will be computed (Table 4.5). The temporal buffer size `tbuffer` is set to 1667 here, which is obtained from the third line in the WPK file (Listing 4.10) and the spatial buffer size `gbuffer` is set to 384 here, which is obtained using the command shown in Sect. 4.4.2. The sub-directory `KER/SF.RC03.3_SF.RC02.3` is then generated, which also contains the kernel input file `ker.in` (Listing 4.11), a soft link to the WPK file (Listing 4.10) and a soft link to the parallel binary executable `ker`. The kernel output files can be concatenated into a single binary file, converted into an ASCII file using `rdker` and plotted using MATLAB. Map-views and cross-section views of the data sensitivity kernels for the three different data functionals at two different frequencies with respect to the asymptotic parameter L are shown in Fig. 4.32. Map-views of the data sensitivity kernels for the phase-delay time at 2.0 Hz with respect to the perturbations in the 21 asymptotic parameters are shown in Fig. 4.33. The source-receiver

```

1 4
2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3 /home/pochen/wrk/HS/VM/XYZGRD
4 /home/pochen/wrk/HS/VM/HSGRD
5 /home/pochen/wrk/HS/FWF/SF.RC03.3/
6 /home/pochen/wrk/HS/FWF/SF.RC02
7 /home/pochen/wrk/HS/KER/SF.RC03.3_SF.RC02.3/wpk
8 /home/pochen/wrk/HS/KER/SF.RC03.3_SF.RC02.3/ker
9 804 257
10 160 257
11 8 8 8
12 3
13 0
14 1e6
15 384 1667

```

Listing 4.11 The kernel input file for our Rayleigh wave example.

geometry used in this example can be useful when computing data sensitivity kernels for misfit measurements derived from the ambient-noise Green’s function (e.g., Chen et al. 2010a; Xu et al. 2013; Lee et al. 2014b).

Note that the data sensitivity kernels with respect to the anisotropic parameters shown in this example were computed in an isotropic reference model. Both the forward wavefield and the RGT were computed in the isotropic half-space structural model using `awp`. This type of data sensitivity kernels can be used for deriving anisotropic perturbations to an isotropic reference model. The forward wave-propagation simulation code `awp` does not account for anisotropy at the current stage. If data sensitivity kernels with respect to anisotropic parameters for an anisotropic reference model are needed, we will need to use a forward wave-propagation simulation code that can account for anisotropy to calculate the forward wavefield from the source and the RGT for the receiver. The formulations given in Sects. 4.3.5 and 4.3.6 are still applicable under such circumstances.

4.4.4.3 Head Waves

I will compute the data sensitivity kernels for head waves refracted along an internal velocity interface. The source-receiver geometry and the problem set-up are summarized in Fig. 4.34 and are identical to those in (Zhang et al. 2007). The modeling volume is discretized into a uniform mesh with 400-m grid spacing. The input file to the mesh generation code `FD_GRID_XYZ_rect` is shown in Listing 4.12 and a snippet of the output file `XYZGRD` is shown in Listing 4.13. For this example, I have about 11 million grid points and I will be using 256 processors on Mira. The velocity model for this layer-over-half-space structural model can be obtained by modifying the codes shown in Listing 2.4 to add a logical construct for assigning the velocity and density values based on the value of `izg`.

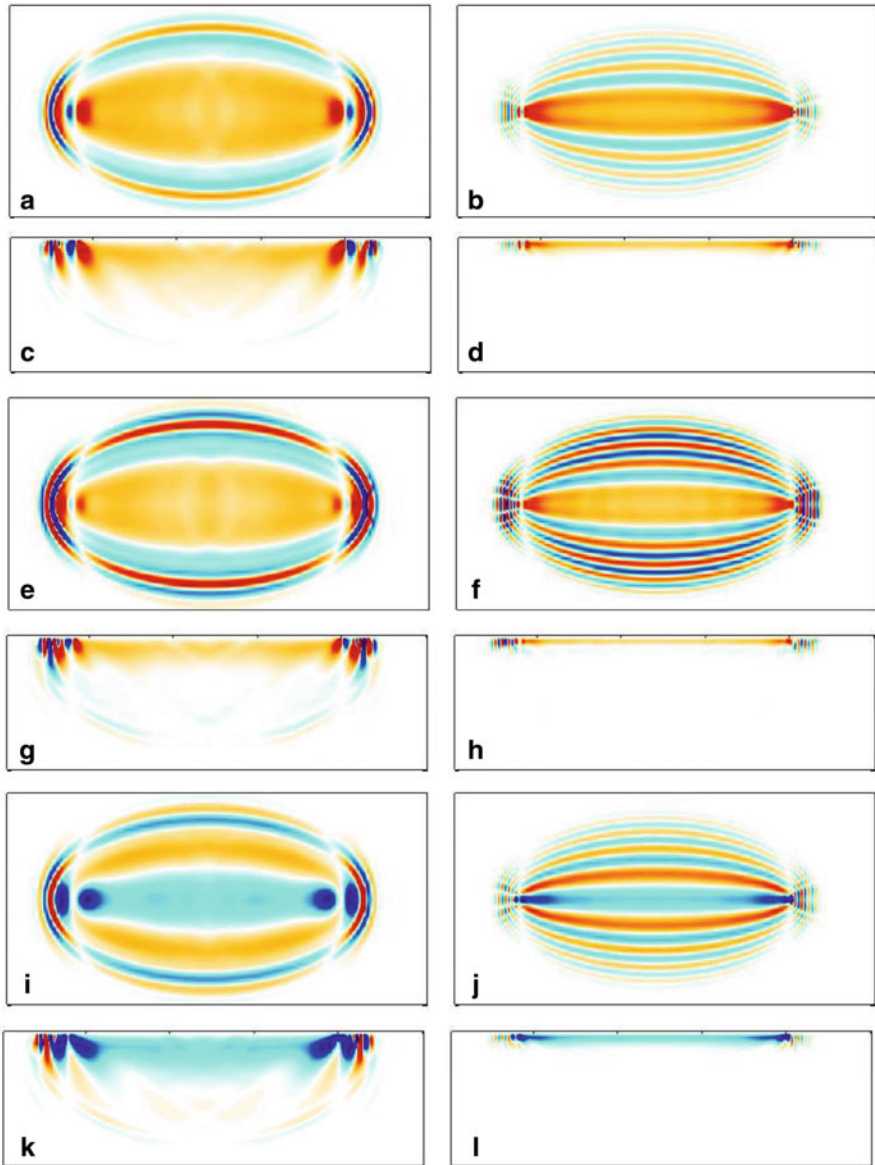


Fig. 4.32 Data sensitivity kernels of the frequency-dependent phase and amplitude misfits with respect to the asymptotic parameter L for the vertical-component Rayleigh wave shown in Fig. 4.30. *Left column*: kernels at 0.5 Hz; *right column*: kernels at 2.0 Hz. **a–d** kernels for frequency-dependent phase-delay time $\delta\tau_p$; **e–h** kernels for frequency-dependent group-delay time $\delta\tau_g$; **i–l** kernels for frequency-dependent amplitude-reduction time $\delta\tau_q$. **a, b, e, f, i, j** map-views of the kernels at the depths of maximum sensitivity, which is about 1.6 km at 0.5 Hz and about 400 m at 2.0 Hz; **c, d, g, h, k, l** cross-section views perpendicular to the source-receiver plane.

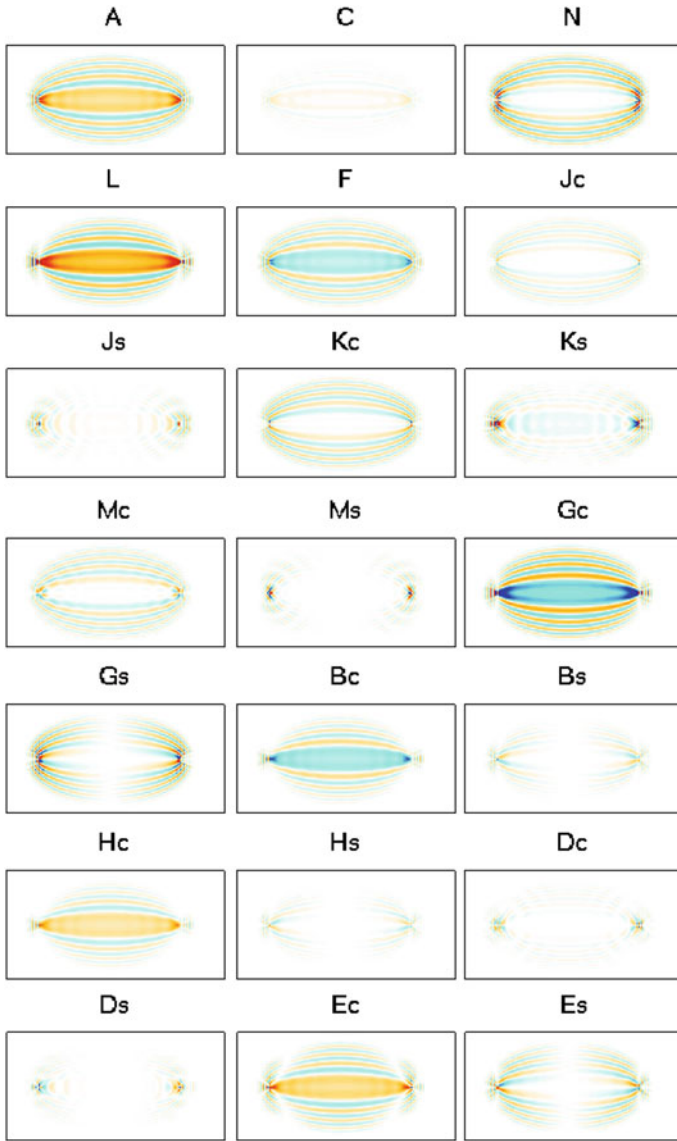


Fig. 4.33 Map-views of the data sensitivity kernels for the 2-Hz phase-delay time with respect to the perturbations in the 21 asymptotic parameters. The map-views are at the depths of the kernel's maximum amplitudes.

The source-receiver file has 3 lines (Listing 4.14). The source 100003 is set to be an explosive source with the source-time function given in the file `src_hw.dat`, which has the same format as explained in Sect. 2.2.4.1. This source-time function was generated using Eq. 2.122 with $c_1 = 20$ and $c_2 = 2$. I

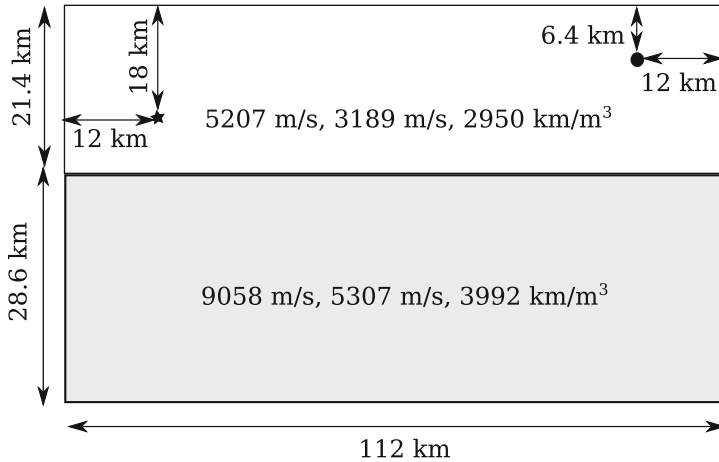


Fig. 4.34 The layer-over-half-space seismic velocity model and the source-receiver geometry used for computing the data sensitivity kernels for head waves. The spatial dimension is 112 km in x_1 , 120 km in x_2 and 50 km in x_3 . This figure shows the cross-section view in the source-receiver plane perpendicular to the x_2 direction. The velocity interface is at 21.4-km depth. The P- and S-velocity and density are 5207 m/s, 3189 m/s and 2950 kg/m³, respectively, for the *top* layer, and 9058 m/s, 5307 m/s and 3992 kg/m³, respectively, for the *bottom* layer. The source 100003 (*black star*) is located at 18-km depth and the receiver IN.RC04 (*black circle*) is located at 6.4-km depth. Both the source and the receiver are located at $x_2=60$ km. The source-receiver horizontal distance is 88 km.

```

1 /home/pochen/wrk/HW/VM/
2 11
3 -118.00000 31.0000
4 -118.00000 32.1000
5 111000
6 49000
7 400
8 8 8 4

```

Listing 4.12 The input file to the mesh generation code for the head wave example.

will be computing the data sensitivity kernels for the vertical-component synthetic velocity seismogram from 100003 to IN.RC04, therefore I will need the RGT for the x_3 component of IN.RC04. The body-force source for this RGT is on line 3 of Listing 4.14. The source-time function for this body-force source is identical to that used in the previous half-space examples (Fig. 2.11). The second line in Listing 4.14 is the receiver line for IN.RC04.

The parameter file for the head-wave example is shown in Listing 4.15. The time sampling interval is set to 0.02 s on line 12. The total duration of the synthetic seismograms is set to 20 s on line 11. The spatial grid decimation rate is set to 2

```

1   280  312  128      ! Global grid numbers NX, NY, NZ
2     8   8   4      ! Chunks (#of proc.) in X, Y, Z
3   35  39  32      ! Grid numbers per chunk NXT, NYT, NZT
4  404529.66513544734      3429832.8747296482
5  405665.76867925981      3554227.4743449953
6  517260.92396119499      3553208.2689150157
7  516124.82041738252      3428813.6692996686
8           128  -0.0000000000000000
9           127  -400.00000000000000
10          126  -800.00000000000000
11  ...

```

Listing 4.13 The first ten lines in the mesh output file for the head-wave example.

```

1  1 100003  1 30 156 83 4 1 1 1 0 0 0 /home/pochen/wrk/HW/JOBS
   /srinfo/src_hw.dat
2  0 IN.RC04  1 250 156 112
3  2 IN.RC04.3 1 250 156 112 -2 0 0 1 /home/pochen/wrk/HW/JOBS/
   srinfo/imp_hw.dat

```

Listing 4.14 The source-receiver list for the head-wave example.

in each direction on line 21–23 and the temporal decimation rate is set to 2 on line 24.

After the input files are set up correctly, I can execute the following sequence of scripts in the project JOBS directory

```

proc_srlst.py param_HW.dat 1 3
set_sr_grd.py param_HW.dat
set_ker_grd.py param_HW.dat
setup_awp.py param_HW.dat FWF 1 1
setup_awp.py param_HW.dat FWF 3 3

```

which will then create two sub-directories FWF/100003 and FWF/IN.RC04.3 under the project FWF directory for the wave-propagation simulations using `awp`. After the wave-propagation simulations are completed for the forward wavefield from 100003 and the RGT for IN.RC04, I can extract the synthetic velocity seismograms by executing the following command in the JOBS directory

```

fwf2syn.py param_HW.dat 1 1 2 2

```

```

1 BINDIR = /home/pochen/F3DWI/bin
2 SCRIPTDIR = /home/pochen/F3DWI/py
3 WRKDIR = /home/pochen/wrk/HW
4 JOBDIR = /home/pochen/wrk/HW/JOBS
5 XYZGRD = /home/pochen/wrk/HW/VM/XYZGRD
6 UTMZONE = 11
7 MEDIA = /home/pochen/wrk/HW/VM/HWGRD
8 MINIMUM VELOCITY = 3189.0
9 MAXIMUM VELOCITY = 9650.0
10 SOURCE RECEIVER LIST = /home/pochen/wrk/HW/JOBS/srinfo/srlist
11 TMAX = 20.0
12 DT = 0.02
13 CFL = 0.9792
14 USEPML = 1
15 PMLCOEF = 3.5
16 BOUNDARY NODE NUMBER = 10
17 VISCO = 0
18 SOURCE STENCIL SIZE IN X = 1
19 SOURCE STENCIL SIZE IN Y = 1
20 SOURCE STENCIL SIZE IN Z = 1
21 NXSKP = 2
22 NYSKP = 2
23 NZSKP = 2
24 NTSKP = 2

```

Listing 4.15 The parameter file for our head-wave example.

and the synthetic seismograms are stored in the file `SYN/100003/IN.RC04.rot`. The head wave on the vertical-component synthetic velocity seismogram is shown in Fig. 4.35.

The WPKs for the broadband cross-correlation delay time and amplitude anomaly (Sect. 4.2.6) can then be computed from the synthetic waveform shown in Fig. 4.35 using the MATLAB function `wpk_bbxcorr.m` (Sect. 4.4.1). And the WPKs are shown in Fig. 4.36. The WPKs are stored inside the file `JOBS/srinfo/bbxcor_HW_loh.wpk`. The directory for the kernel calculations can then be generated using the following command in the `JOBS` directory.

```

setup_ker.py param_HW.dat 1 2 3 /home/pochen/wrk/HW/
JOBS/srinfo/bbxcor_HW_loh.wpk 1e7 5760 376 2 1 2 3

```

Here I am computing the data sensitivity kernels with respect to relative perturbations in the P- and S-velocity and density. This command will generate the sub-directory `KER/100003_IN.RC04.3`, the kernel input file (Listing 4.16) and the two soft links pointing to the WPK file and the parallel binary executable `ker`. After executing the kernel calculation code `ker`, the binary kernel output files can be concatenated into a single binary file, converted into ASCII format and plotted.

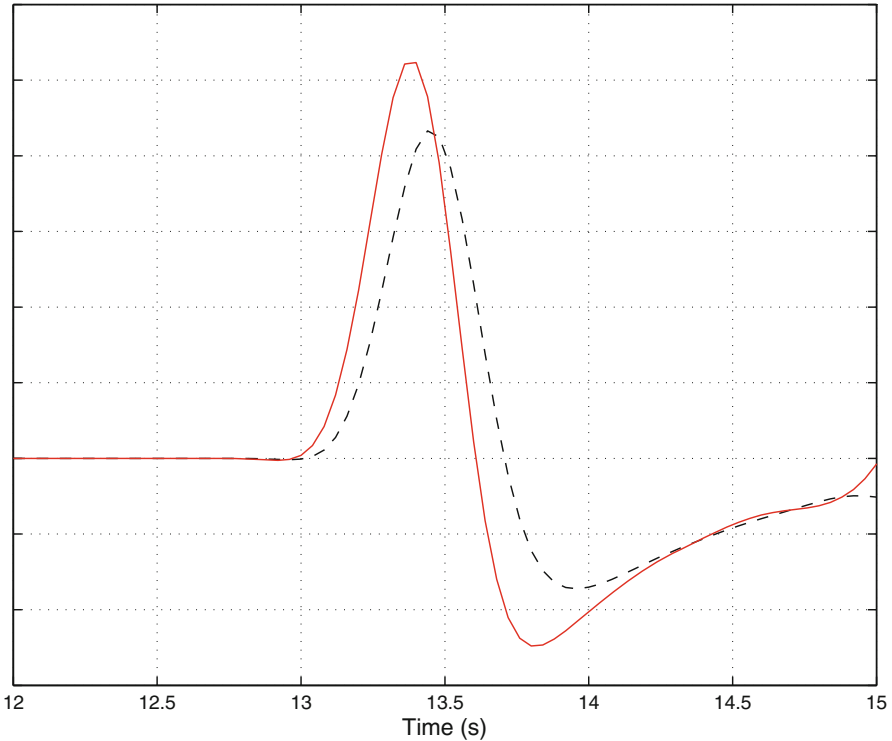


Fig. 4.35 Head waves on the vertical-component synthetic velocity seismograms from 100003 to IN.RC04. *Black dash line*: synthetic seismogram generated using the layer-over-half-space model shown in Fig. 4.34; *red solid line*: synthetic seismogram generated using the structural model whose P- and S-velocity in the *bottom* layer have a positive gradient of 0.02 km/s/km with depth.

The data sensitivity kernels of the broadband cross-correlation phase delay time and amplitude anomaly with respect to P-velocity are shown in Fig. 4.37. As pointed out in (Zhang et al. 2007), the sensitivity is located mostly on only one side of the refracted ray path below the velocity interface and can be characterized as a “split banana-doughnut” kernel.

I now modify the layer-over-half-space velocity model shown in Fig. 4.34 so that the P- and S-velocity inside the bottom layer increase linearly with depth from 9058 m/s and 5307 m/s at the velocity interface with a gradient of 0.02 km/s/km. At the bottom of our modeling volume the P- and S-velocity reach 9650 m/s and 5899 m/s, respectively. This velocity model can also be generated by modifying the codes shown in Listing 2.4. The modified velocity model, as a function of depth, is shown in Fig. 4.38. The forward wavefield from 100003 and the RGT for IN.RC04 can be re-computed using this “layer-over-gradient” velocity model and the synthetic seismogram is plotted in Fig. 4.35. This velocity gradient in the bottom layer can cause dispersion in the head wave. The data sensitivity kernels of

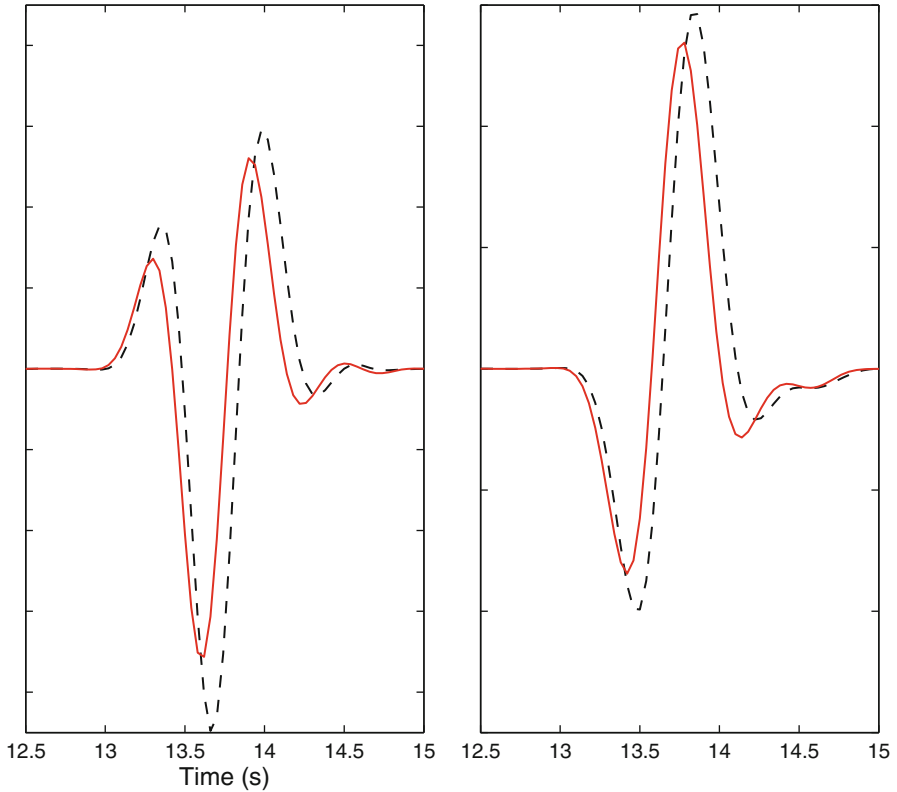


Fig. 4.36 The WPKs of the broadband cross-correlation delay time (*left*) and amplitude anomaly (*right*) for the head wave generated using the layer-over-half-space model (*black dash lines*) and the head wave generated using the model with the bottom layer having a positive gradient of 0.02 km/s/km with depth.

the broadband cross-correlation misfits computed using the “layer-over-gradient” velocity model (Fig. 4.39) are highly similar to those computed using the layer-over-half-space velocity model (Fig. 4.37), except that the sensitivity has extended slightly to larger depths for kernels computed using the “layer-over-gradient” model (Fig. 4.39). For this dispersive headwave we can compute the data sensitivity kernels of the frequency-dependent phase and amplitude misfits (Sect. 4.2.7). I have computed the WPKs at 1.0 and 2.0 Hz using `wpk_dtau.m` (Sect. 4.4.4.2) and re-computed the data sensitivity kernels. Figure 4.40 shows the data sensitivity kernels of the frequency-dependent phase-delay time with respect to P-velocity at 1.0 and 2.0 Hz. At higher frequencies the width of the first Fresnel zone is narrower and sensitivity is compressed more towards the velocity interface. The shape of the frequency-dependent phase-delay time kernels still shows the “split banana-doughnut” character with the sensitivity mostly located below the refracted ray path.

```
1 2
2 1 1 1
3 /home/pochen/wrk/HW/VM/XYZGRD
4 /home/pochen/wrk/HW/VM/HWGRD
5 /home/pochen/wrk/HW/FWF/100003/
6 /home/pochen/wrk/HW/FWF/IN.RC04
7 /home/pochen/wrk/HW/KER/100003_IN.RC04.3/wpk
8 /home/pochen/wrk/HW/KER/100003_IN.RC04.3/ker
9 30 156
10 250 156
11 2 2 2
12 3
13 0
14 1e7
15 5760 387
```

Listing 4.16 The kernel input file for our head-wave example.

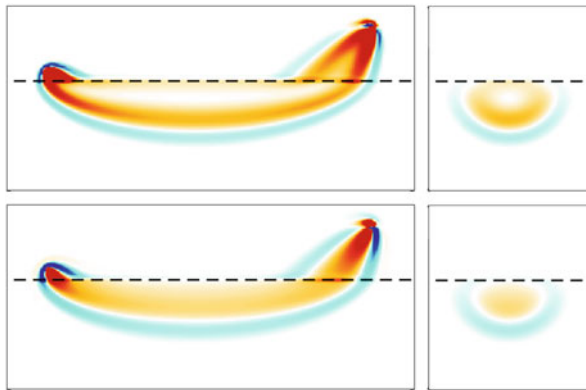


Fig. 4.37 Cross-section views of the data sensitivity kernels of the broadband cross-correlation delay time (*upper*) and amplitude anomaly (*lower*) for the head wave shown as *black dash lines* in Fig. 4.35 with respect to the relative perturbation in P-velocity. *Left column*: cross-sections in the source-receiver plane perpendicular to the x_2 axis; *right column*: cross-sections mid-way between source and receiver ($x_1 = 56$ km) perpendicular to the x_1 axis. The *black dash lines* indicate the velocity interface at 21.4 km depth (Fig. 4.34).

The fringe patterns shown in Fig. 4.40 are due to the narrow-band nature of the frequency-dependent phase-delay misfit.

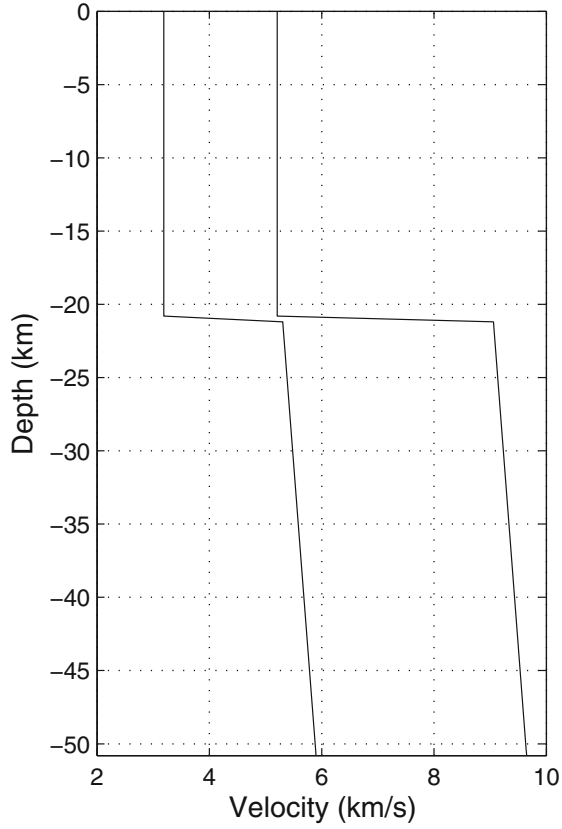


Fig. 4.38 The P- and S-velocity as functions of depth in the “layer-over-gradient” model. The top layer has constant P- and S-velocity with values identical to those shown in Fig. 4.34 and the bottom layer has a constant positive gradient of 0.02 km/s/km with depth. The velocities in the bottom layer increase from 9058 m/s and 5307 m/s at the interface to 9650 m/s and 5899 m/s at the bottom of our model.

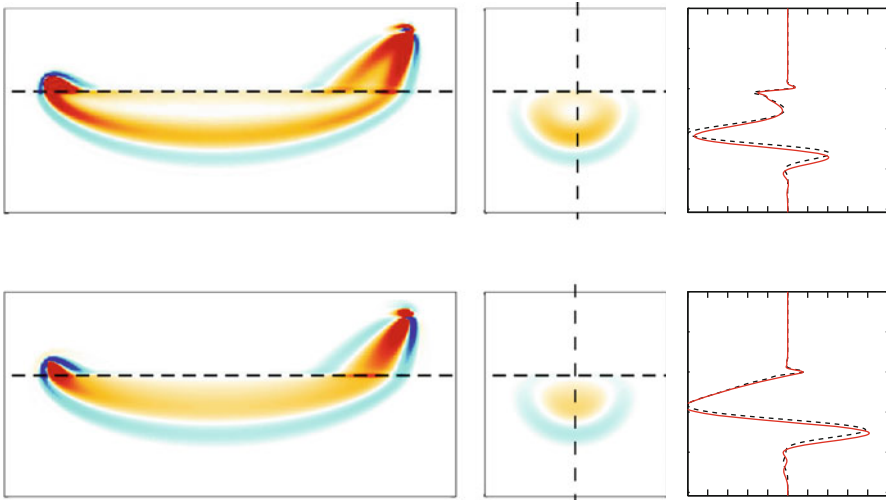


Fig. 4.39 Cross-sections of the data sensitivity kernels for the broadband cross-correlation delay time (*top row*) and amplitude anomaly (*bottom row*) with respect to P-velocity. *Left column*: cross-sections on the source-receiver plane perpendicular to the x_2 axis; *center column*: cross-sections mid-way between source and receiver and perpendicular to the x_1 axis; *right column*: sensitivity along the vertical dash lines shown on the center column for the “layer-over-half-space” structural model (*black dash lines*) and the “layer-over-gradient” structural model (*red solid lines*), the vertical axis is depth and the horizontal axis is the value of the data sensitivity kernel. The horizontal black dash lines on the left and center columns show the location of the velocity interface at 21.4 km depth.

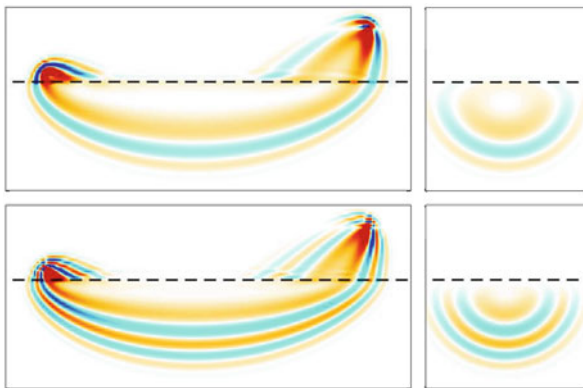


Fig. 4.40 Cross-sections of the data sensitivity kernels for the frequency-dependent phase-delay time with respect to P-velocity at 1.0 Hz (*top row*) and 2.0 Hz (*bottom row*). *Left column*: cross-section at the source-receiver plane perpendicular to the x_2 axis; *right column*: cross-section at the plane mid-way between source and receiver and perpendicular to the x_1 axis.

References

- Aki, K., & Richards, P. G. (2002). *Quantitative seismology* (second). Sausalito, California: University Science Books.
- Babuska, V., & Cara, M. (1991). *Seismic anisotropy in the earth*. European Consortium for Mathematics in Industry. The Netherlands: Kluwer Academic Publishers.
- Backus, G. (1962). Long-wave elastic anisotropy produced by horizontal layering. *Journal of Geophysical Research*, *67*(11), 4427–4440.
- Backus, G. E. (1977a). Interpreting the seismic glut moments of total degree two or less. *Geophysical Journal International*, *51*(1), 1–25.
- Backus, G. E. (1977b). Seismic sources with observable glut moments of spatial degree two. *Geophysical Journal International*, *51*(1), 27–45.
- Backus, G., & Gilbert, F. (1968). The resolving power of gross earth data. *Geophysical Journal International*, *16*(2), 169–205.
- Ben-Menahem, A. (1961). Radiation of seismic surface-waves from finite moving sources. *Bulletin of the Seismological Society of America*, *51*(3), 401–435.
- Beydoun, W., & Tarantola, A. (1988). First Born and Rytov approximations: Modeling and inversion conditions in a canonical example. *The Journal of the Acoustical Society of America*, *83*(3), 1045–1055.
- Bozdağ, E., Trampert, J., & Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, *185*(2), 845–870.
- Brenders, A., & Pratt, R. (2007). Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model. *Geophysical Journal International*, *168*(1), 133–151.
- Chen, P. (2011). Full-wave seismic data assimilation: Theoretical background and recent advances. *Pure and Applied Geophysics*, *168*(10), 1527–1552.
- Chen, M., & Tromp, J. (2007). Theoretical and numerical investigations of global and regional seismic wave propagation in weakly anisotropic earth models. *Geophysical Journal International*, *168*(3), 1130–1152.
- Chen, P., Jordan, T., & Zhao, L. (2005a). Finite-moment tensor of the 3 september 2002 Yorba Linda earthquake. *Bulletin of the Seismological Society of America*, *95*(3), 1170–1180.
- Chen, Q., Zhou, C., Luo, J., & Ming, D. (2005b). Fast segmentation of high-resolution satellite images using watershed transform combined with an efficient region merging approach. In *Combinatorial image analysis*, pp. 621–630. Berlin Heidelberg: Springer.
- Chen, P., Jordan, T. H., & Zhao, L. (2007a). Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, *170*(1), 175–181.
- Chen, P., Zhao, L., & Jordan, T. H. (2007b). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, *97*(4), 1094–1120.
- Chen, P., Jordan, T. H., & Lee, E.-J. (2010a). Perturbation kernels for generalized seismological data functionals (GSDF). *Geophysical Journal International*, *183*(2), 869–883.
- Chen, P., Jordan, T. H., & Zhao, L. (2010b). Resolving fault plane ambiguity for small earthquakes. *Geophysical Journal International*, *181*(1), 493–501.
- Chernov, L. (1960). *Wave propagation in a random medium*. Dover Publications.
- Chevrot, S. (2006). Finite-frequency vectorial tomography: A new method for high-resolution imaging of upper mantle anisotropy. *Geophysical Journal International*, *165*(2), 641–657.
- Cochard, A., Igel, H., Schuberth, B., Suryanto, W., Velikoseltsev, A., Schreiber, U., Wassermann, J., Scherbaum, F., & Vollmer, D. (2006). Rotational motions in seismology: Theory, observation, simulation. In *Earthquake source asymmetry, structural media and rotation effects*, pp. 391–411. Berlin Heidelberg: Springer.
- Coupric, M., & Bertrand, G. (1997). Topological gray-scale watershed transformation. In *Optical science, engineering and instrumentation '97*, pp. 136–146. International Society for Optics and Photonics.

- Coupric, M., Najman, L., & Bertrand, G. (2005). Quasi-linear algorithms for the topological watershed. *Journal of Mathematical Imaging and Vision*, 22(2–3), 231–249.
- Dahlen, F. A., & Baig, A. M. (2002). Fréchet kernels for body-wave amplitudes. *Geophysical Journal International*, 150, 440–466.
- Dahlen, F. A., & Tromp, J. (1998). *Theoretical global seismology*. Princeton, New Jersey: Princeton University Press.
- Dahlen, F. A., & Zhou, Y. (2006). Surface-wave group-delay and attenuation kernels. *Geophysical Journal International*, 165(2), 545–554.
- Dahlen, F., Hung, S., & Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes-I. Theory. *Geophysical Journal International*, 141, 157–174.
- Diersen, S., Lee, E., Spears, D., Chen, P., & Wang, L. (2011). Classification of seismic windows using artificial neural networks. In *Procedia Computer Science proceedings* Tsukuba, Japan.
- Dziewonski, A. M., & Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, 25(4), 297–356.
- Ekström, G., Tromp, J., & Larson, E. (1997). Measurements and global models of surface wave propagation. *Journal of Geophysical Research*, 102(B4), 8137–8157.
- Fichtner, A. (2010). *Full seismic waveform modelling and inversion* (Advances in geophysical and environmental mechanics and mathematics). Springer.
- Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, 175(2), 665–685.
- Gallier, J. (2000). *Curves and surfaces in geometric modeling: Theory and algorithms* (Computer graphics and geometric modeling series). Morgan Kaufmann Publishers.
- Gauthier, O., Virieux, J., & Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*, 51(7), 1387–1403.
- Gee, L. S., & Jordan, T. H. (1992). Generalized seismological data functionals. *Geophysical Journal International*, 111(2), 363–390.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1), 51–83.
- Heidbreder, G. (1967). Multiple scattering and the method of Rytov. *Journal of the Optical Society of America*, 57, 1477.
- Hellweg, M., & Boatwright, J. (1999). Mapping the rupture process of moderate earthquakes by inverting accelerograms. *Journal of Geophysical Research*, 104(B4), 7319–7328.
- Holschneider, M., Diallo, M. S., Kulesh, M., Ohrnberger, M., Luck, E., & Scherbaum, F. (2005). Characterization of dispersive surface waves using continuous wavelet transforms. *Geophysical Journal International*, 163(2), 463–478.
- Hung, S.-H., Dahlen, F., & Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes II. examples. *Geophysical Journal International*, 141(1), 175–203.
- Ishimaru, A. (1978). *Wave propagation and scattering in random media* (Vol. 2). New York: Academic Press.
- Kanamori, H., & Anderson, D. L. (1977). Importance of physical dispersion in surface wave and free oscillation problems: Review. *Reviews of Geophysics*, 15(1), 105–112.
- Keller, J. (1969). Accuracy and validity of the Born and Rytov approximations. *Journal of the Optical Society of America*, 59(8), 1003–1004.
- Kulesh, M., Holschneider, M., Diallo, M., Xie, Q., & Scherbaum, F. (2005). Modeling of wave dispersion using continuous wavelet transforms. *Pure and Applied Geophysics*, 162(5), 843–855.
- Larson, E. W., Tromp, J., & Ekström, G. (1998). Effects of slight anisotropy on surface waves. *Geophysical Journal International*, 132(3), 654–666.
- Lee, E.-J., & Chen, P. (2013). Automating seismic waveform analysis for full 3-D waveform inversions. *Geophysical Journal International*, 194(1), 572–589.
- Lee, W. H., Igel, H., & Trifunac, M. D. (2009a). Recent advances in rotational seismology. *Seismological Research Letters*, 80(3), 479–490.

- Lee, W. K., Celebi, M., Todorovska, M., & Igel, H. (2009b). Introduction to the special issue on rotational seismology and engineering applications. *Bulletin of the Seismological Society of America*, 99(2B), 945–957.
- Lee, E.-J., Chen, P., Jordan, T. H., & Wang, L. (2011). Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. *Geophysical Journal International*, 186(1), 311–330.
- Lee, E.-J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research*, 119(8), 6421–6451.
- Liu, Q., & Gu, Y. (2012). Seismic imaging: From classical to adjoint tomography. *Tectonophysics*, 566, 31–66.
- Liu, H.-P., Anderson, D. L., & Kanamori, H. (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition. *Geophysical Journal of the Royal Astronomical Society*, 47(1), 41–58.
- Love, A. (2013). *A treatise on the mathematical theory of elasticity*. Cambridge: Cambridge University Press.
- Luo, Y., & Shuster, G. (1991). Wave-equation traveltimes inversion. *Geophysics*, 56(5), 645–653.
- Maggi, A., Tape, C., Chen, M., Chao, D., & Tromp, J. (2009). An automated time-window selection algorithm for seismic tomography. *Geophysical Journal International*, 178, 257–281.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2), 431–441.
- Marquering, H., Dahlen, F., & Nolet, G. (1999). Three-dimensional sensitivity kernels for finite-frequency traveltimes: The banana-doughnut paradox. *Geophysical Journal International*, 137(3), 805–815.
- Maupin, V. (1985). Partial derivatives of surface wave phase velocities for flat anisotropic models. *Geophysical Journal International*, 83(2), 379–398.
- McGuire, J. J. (2004). Estimating finite source properties of small earthquake ruptures. *Bulletin of the Seismological Society of America*, 94(2), 377–393.
- McGuire, J., Zhao, L., & Jordan, T. (2001). Teleseismic inversion for the second-degree moments of earthquake space-time distributions. *Geophysical Journal International*, 145(3), 661–678.
- McGuire, J., Zhao, L., & Jordan, T. (2002). Predominance of unilateral rupture for a global catalog of large earthquakes. *Bulletin of the Seismological Society of America*, 92(8), 3309–3317.
- Montagner, J.-P., & Nataf, H.-C. (1986). A simple method for inverting the azimuthal anisotropy of surface waves. *Journal of Geophysical Research*, 91(B1), 511–520.
- Mori, J. (1996). Rupture directivity and slip distribution of the m 4.3 foreshock to the 1992 Joshua tree earthquake, Southern California. *Bulletin of the Seismological Society of America*, 86(3), 805–810.
- Mori, J., & Hartzell, S. (1990). Source inversion of the 1988 Upland, California, earthquake: Determination of a fault plane for a small event. *Bulletin of the Seismological Society of America*, 80(3), 507–518.
- Nocedal, J., & Wright, S. (2006). *Numerical optimization* (Springer series in operations research and financial engineering). New York: Springer.
- Okada, T., Umino, N., Ito, Y., Matsuzawa, T., Hasegawa, A., & Kamiyama, M. (2001). Source processes of 15 September 1998 m 5.0 Sendai, Northeastern Japan, earthquake and its m 3.8 foreshock by waveform inversion. *Bulletin of the Seismological Society of America*, 91(6), 1607–1618.
- Panning, M. P., & Nolet, G. (2008). Surface wave tomography for azimuthal anisotropy in a strongly reduced parameter space. *Geophysical Journal International*, 174(2), 629–648.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3), 888–901.
- Pratt, R. G., & Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data. *Geophysics*, 64(3), 902–914.

- Pratt, R. G., Shin, C., & Hick, G. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2), 341–362.
- Press, W. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge: Cambridge University Press.
- Quackenbush, L. J. (2004). A review of techniques for extracting linear features from imagery. *Photogrammetric Engineering & Remote Sensing*, 70(12), 1383–1392.
- Ritsema, J., Rivera, L., Komatitsch, D., Tromp, J., & van Heijst, H. (2002). Effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios. *Geophysical Research Letters*, 29(10), 72–1.
- Romanowicz, B., & Snieder, R. (1988). A new formalism for the effect of lateral heterogeneity on normal modes and surface waves-II. general anisotropic perturbation. *Geophysical Journal International*, 93(1), 91–99.
- Sieminski, A., Liu, Q., Trampert, J., & Tromp, J. (2007). Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods. *Geophysical Journal International*, 168(3), 1153–1174.
- Sieminski, A., Trampert, J., & Tromp, J. (2009). Principal component analysis of anisotropic finite-frequency sensitivity kernels. *Geophysical Journal International*, 179(2), 1186–1198.
- Sipkin, S. A., & Jordan, T. H. (1980). Multiple ScS travel times in the Western Pacific: Implications for mantle heterogeneity. *Journal of Geophysical Research*, 85(B2), 853–861.
- Sirgue, L., & Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1), 231–248.
- Smith, M. L., & Dahlen, F. (1973). The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research*, 78(17), 3321–3333.
- Snieder, R., & Lomax, A. (1996). Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes. *Geophysical Journal International*, 125(3), 796–812.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the Southern California crust. *Science*, 325(5943), 988–992.
- Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128, 365–399.
- Tatarskii, V., Oceanic, U. S. N., Administration, A., & (U.S.), N. S. F. (1971). *The effects of the turbulent atmosphere on wave propagation*. Israel Program for Scientific Translations.
- Thomsen, L. (1986). Weak elastic anisotropy. *Geophysics*, 51(10), 1954–1966.
- Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1), 195–216.
- Voigt, W. (1928). *Lehrbuch der kristallphysik: (mit ausschluss der kristalloptik)*. B.G. Teubners Sammlung von Lehrbüchern auf dem Gebiete der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. B. G. Teubner.
- Woodward, M. (1992a). A qualitative comparison of the first-order Born and Rytov approximations. *SEP-60: Stanford exploration project*, (pp. 203–214).
- Woodward, M. J. (1992b). Wave-equation tomography. *Geophysics*, 57(1), 15–26.
- Woodward, R., & Masters, G. (1991). Global upper mantle structure from long-period differential travel times. *Journal of Geophysical Research*, 96(B4), 6351–6377.
- Wu, R.-S. (1989). The perturbation method in elastic wave scattering. *Pure and Applied Geophysics*, 131(4), 605–637.
- Wu, R.-S. (2003). Wave propagation, scattering and imaging using dual-domain one-way and one-return propagators. *Pure and Applied Geophysics*, 160(3–4), 509–539.
- Wu, R.-S., & Aki, K. (1988). Introduction: Seismic wave scattering in three-dimensionally heterogeneous earth. *Pure and Applied Geophysics*, 128(1), 1–6.
- Xu, Z., Chen, P., & Chen, Y. (2013). Sensitivity kernel for the weighted norm of the frequency-dependent phase correlation. *Pure and Applied Geophysics*, 170(3), 353–371.
- Zhang, Z., Shen, Y., & Zhao, L. (2007). Finite-frequency sensitivity kernels for head waves. *Geophysical Journal International*, 171(2), 847–856.

- Zhao, L., Jordan, T. H., & Chapman, C. H. (2000). Three-dimensional Fréchet differential kernels for seismic delay times. *Geophysical Journal International*, *141*(3), 558–576.
- Zhao, L., Jordan, T. H., Olsen, K. B., & Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, *95*(6), 2066–2080.
- Zhao, L., Chen, P., & Jordan, T. H. (2006). Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. *Bulletin of the Seismological Society of America*, *96*(5), 1753–1763.

Chapter 5

Optimization Algorithms

Solutions to seismic inverse problems are often implemented using optimization algorithms, in which we search for an optimal model that can minimize the discrepancies between model predictions and actual observations and also satisfy some constraints simultaneously. The theoretical framework developed in Chaps. 3 and 4 allows us to discuss and compare existing optimization algorithms for full-3D waveform inversion, to analyze their advantages and disadvantages, and may also inspire the development of new algorithms in the future. As we will see in this chapter, this theoretical framework is quite general. It encompasses gradient-based optimization algorithms, in which the gradient of the objective function can be computed using the adjoint-wavefield (AW) method, and also Hessian-based algorithms, in which an approximate Hessian can be constructed through the scattering-integral (SI) method. It also allows us to use objective functions defined using any types of waveform misfit measure, as well as any model parameterization that is suitable for the problem at hand.

When discussing optimization algorithms, we need to distinguish between local optimization and global optimization. As mentioned in Sect. 4.2.8, for fully three-dimensional tomographic inversions it is often mandatory to use local optimization algorithms because the model space is large and the computational cost for solving the three-dimensional (visco)elastic wave equation is high. At a local optimal model, the objective function is smaller (suppose we are solving a minimization problem) than at all other feasible models *nearby* in model space. Local optimization algorithms cannot guarantee to find the global optimal model, which is the model with the smallest objective function value among *all* feasible models. For convex optimization problems, and particularly for linear optimization, the local optimal model is also the global optimal model. And as discussed briefly in Sect. 4.2.8, it is possible to make our full-3D tomography problem more convex through the appropriate choice of the data functionals used for quantifying waveform discrepancies. For general nonlinear problems, local solutions are not always global solutions. In this chapter, I will only discuss local optimization algorithms. I note, however, that many global optimization algorithms require solutions of many local optimization problems and in such situations our discussion on local optimization algorithms here is still relevant. The model space can be reduced, sometimes significantly, through a

judicious choice of regionalization (e.g., Käüfl et al. 2013). The computational cost for solving three-dimensional wave equations are reducing rapidly, especially with the advent of energy-efficient many-core co-processors (Sect. 1.4). It is possible that global optimization algorithms can be adopted for solving full-3D tomography problems in the not-too-distant future.

Most classical deterministic local optimization algorithms are iterative. They start with an initial guess and proceed to a sequence of gradually improving estimates following a certain updating strategy that utilizes information about the objective function at the current iteration or accumulated through several iterations. And those local optimization algorithms can be classified based on their updating strategies. Throughout the iterative process, the definition of the objective function (e.g., the observations used for defining the objective function) is fixed. In realistic full-3D waveform inversions, the amount (and/or the time-frequency-domain extent, the types of the data functionals) of observed waveforms used in the inversion often increases with iteration as the model improves and the model-predicted (i.e., synthetic) seismograms are capable of fitting more of all available observed waveforms (e.g., Tape et al. 2010; Lee et al. 2014b). This means that not only the value, but also the definition of the objective function can change from iteration to iteration in realistic full-3D waveform inversions. We therefore need to be cautious when adopting optimization algorithms that use information accumulated through more than one iteration, as the objective function from a previous iteration may have a quite different definition from the objective function at the current iteration.

In the following I will discuss two classes of local optimization algorithms that have been successfully applied to solve full-3D tomography problems using real seismic waveform data.

5.1 The Scattering-Integral (SI) Method

The scattering-integral (SI) method is based on the Gauss–Newton algorithm (e.g., Pratt et al. 1998) for solving least-squares problems. It explicitly constructs the approximate Hessian matrix using the data sensitivity kernel for every misfit measurement. This method is named after the integrals used in constructing the data sensitivity kernels (Chen et al. 2007a), which are referred to as the “scattering-integral”. Similar integrals, but different types of Green’s functions and forward wavefields, have been used for constructing Born kernels and Rytov kernels in an extensive literature related to diffraction tomography (e.g., Devaney 1981; Wu and Toksöz 1987; Woodward 1992a, b) and inverse scattering (e.g., Beylkin 1985; Bleistein et al. 2001). The scattering-integral has also been used in computing data sensitivity kernels in global seismology, in which the Green’s functions and forward wavefields are often computed in one-dimensional reference structural models through normal mode summation or using asymptotic ray theory (e.g., Dahlen et al. 2000; Zhao et al. 2000; Zhao and Jordan 2006). The extension of this technique

to fully three-dimensional isotropic elastic reference models and broadband cross-correlation measurements was implemented in (Zhao et al. 2005; Zhao and Jordan 2006) based on the use of reciprocity and RGTs. In Chap. 3, I have shown that the requirement on reciprocity can be removed by replacing the RGTs with the time-reversed adjoint Green's tensors. In Chap. 4, I have extended the SI method to arbitrary data functionals and also anisotropic and anelastic kernels. The SI method was successfully applied to image the structure of the Los Angeles Basin using P- and S-waves from local small earthquakes at frequencies up to 1.2 Hz in (Chen et al. 2007) and to image the crustal structure of Southern California using body- and surface-waves from local earthquakes and surface-waves from ambient-noise Green's functions at frequencies up to 0.2 Hz in (Lee et al. 2014b).

5.1.1 The Gauss–Newton Algorithm

The least-squares problem is so pervasive among many disciplines and a huge literature has been developed for its solution. The main approach is the Gauss–Newton algorithm. In least-squares problems, we consider an objective function that is the sum of squares, i.e.,

$$f(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^m r_j^2(\mathbf{x}), \quad (5.1)$$

where $\mathbf{x} \in \mathbb{R}^n$ are the parameters we are trying to optimize. Each $r_j(\mathbf{x})$ is often referred to as the “residual”, which is an individual measurement of the discrepancy between a prediction made using a model and the corresponding actual observation. This special form of $f(\mathbf{x})$ in Eq. 5.1 often makes least-squares problems easier to solve than general optimization problems. In order to see why this is the case, we first introduce a residual vector, i.e.,

$$\mathbf{r}(\mathbf{x}) = [r_1(\mathbf{x}), r_2(\mathbf{x}), \dots, r_m(\mathbf{x})]^T, \quad (5.2)$$

which is a column vector with m rows and maps \mathbb{R}^n to \mathbb{R}^m . The objective function in Eq. 5.1 can be re-written as

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|_2^2, \quad (5.3)$$

where $\|\cdot\|_2$ is the L_2 -norm. The Jacobian matrix $\mathbf{A}(\mathbf{x})$ is the $m \times n$ matrix of the first-order partial derivatives of the residuals, i.e.,

$$\mathbf{A}(\mathbf{x}) = \left[\frac{\partial r_j}{\partial x_i} \right]_{j=1,2,\dots,m; i=1,2,\dots,n} = \begin{bmatrix} \nabla r_1(\mathbf{x})^T \\ \nabla r_2(\mathbf{x})^T \\ \vdots \\ \nabla r_m(\mathbf{x})^T \end{bmatrix}, \quad (5.4)$$

where $\nabla r_j(\mathbf{x}) = \left[\frac{\partial r_j}{\partial x_1}, \frac{\partial r_j}{\partial x_2}, \dots, \frac{\partial r_j}{\partial x_n} \right]^T$ is a column-vector with n rows and $\nabla r_j(\mathbf{x})^T$ is a row-vector with n columns. The gradient and Hessian of the objective function $f(\mathbf{x})$ can then be expressed in terms of the Jacobian matrix as

$$\nabla f(\mathbf{x}) = \sum_{j=1}^m r_j(\mathbf{x}) \nabla r_j(\mathbf{x}) = \mathbf{A}(\mathbf{x})^T \mathbf{r}(\mathbf{x}), \quad (5.5)$$

$$\begin{aligned} \mathbf{H}(\mathbf{x}) &= \nabla \nabla f(\mathbf{x}) = \sum_{j=1}^m \nabla r_j(\mathbf{x}) \nabla r_j(\mathbf{x})^T + \sum_{j=1}^m r_j(\mathbf{x}) \nabla \nabla r_j(\mathbf{x}) \\ &= \mathbf{A}(\mathbf{x})^T \mathbf{A}(\mathbf{x}) + \sum_{j=1}^m r_j(\mathbf{x}) \nabla \nabla r_j(\mathbf{x}). \end{aligned} \quad (5.6)$$

If the first-order derivatives of the residuals can be evaluated, the Jacobian matrix is available, which gives not only the gradient of the objective function (Eq. 5.5), but also the first term $\mathbf{A}(\mathbf{x})^T \mathbf{A}(\mathbf{x})$ in the Hessian (Eq. 5.6), without the need to calculate any second-order derivatives of $r_j(\mathbf{x})$. This availability of part of the Hessian from only first-order derivatives of the residuals is a special feature that is unique to least-squares problems. In practice, the term $\mathbf{A}(\mathbf{x})^T \mathbf{A}(\mathbf{x})$ is often the dominant term in the Hessian, either because the residuals are small (i.e., $r_j(\mathbf{x})$ is small), or because the residuals are close to affine (e.g., linear with respect to \mathbf{x}) near the solution (i.e., $\nabla \nabla r_j(\mathbf{x})$ is small), or both. In the standard Newton's method for optimization, the descent direction at the k th iteration, $\delta \mathbf{x}_k$, is obtained by solving the normal equation

$$\nabla \nabla f(\mathbf{x}_k) \delta \mathbf{x}_k = -\nabla f(\mathbf{x}_k). \quad (5.7)$$

For least-squares problems we can approximate the Hessian $\nabla \nabla f(\mathbf{x}_k)$ using its first term $\mathbf{A}(\mathbf{x}_k)^T \mathbf{A}(\mathbf{x}_k)$ and arrive at the Gauss–Newton normal equation

$$\mathbf{A}(\mathbf{x}_k)^T \mathbf{A}(\mathbf{x}_k) \delta \mathbf{x}_k = -\mathbf{A}(\mathbf{x}_k)^T \mathbf{r}(\mathbf{x}_k). \quad (5.8)$$

It can be verified that as long as the Jacobian $\mathbf{A}(\mathbf{x}_k)$ has full rank and the gradient is nonzero, the direction $\delta \mathbf{x}_k$ obtained by solving Eq. 5.8 is always descending, therefore suitable for a line search¹. Once the Gauss–Newton normal Eq. 5.8 is solved, we can update the solution using

$$\mathbf{x}_{k+1} = \mathbf{x}_k + c \delta \mathbf{x}_k, \quad (5.9)$$

where the constant scalar c can be 1 or the optimal step length determined through the line search. The procedure can be iterated until a pre-specified convergence criterion is met.

¹ Once a descent direction is given, the line search determines how far we should move along the descent direction, i.e., the optimal step length along the descent direction. The other widely used technique is known as the “trust region” method.

Extensions of the Gauss–Newton algorithm can be understood in terms of the different approximations made on the second term of the Hessian (Eq. 5.6). The very successful Levenberg–Marquardt algorithm (Levenberg 1944; Marquardt 1963) is based on replacing the second term in the Hessian (Eq. 5.6) with a scaled identity matrix, i.e.,

$$\mathbf{H}(\mathbf{x}) \approx \mathbf{A}(\mathbf{x})^T \mathbf{A}(\mathbf{x}) + c\mathbf{I}, \quad (5.10)$$

where c is a scalar constant. This approach provides a remedy to the Gauss–Newton algorithm when the Jacobian is rank-deficient, or nearly so. When c is very large, the normal Eq. 5.7 approaches

$$c\mathbf{I}\delta\mathbf{x}_k = -\nabla f(\mathbf{x}_k), \quad (5.11)$$

and we recover the steepest-descent algorithm. When c is very small, we recover the Gauss–Newton algorithm. Marquardt also gave another approximation. Instead of using the scaled identity matrix, the matrix

$$\mathbf{D} = [D_{ij}] = \begin{cases} (\mathbf{A}^T \mathbf{A})_{ii} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (5.12)$$

is used. This choice avoids some problems when the parameters are of very different scales. It is also possible to use a combination of both matrices, i.e.,

$$\mathbf{H}(\mathbf{x}) \approx \mathbf{A}(\mathbf{x})^T \mathbf{A}(\mathbf{x}) + c\mathbf{I} + \mathbf{D}, \quad (5.13)$$

where c is a small constant. An approximation to the second term can also be constructed based on the secant equation (e.g., Dennis et al. 1981). Consider the second-order derivative of the residual $\nabla\nabla r_j$, at two consecutive iterations k and $k + 1$, we have

$$\nabla\nabla r_j(\mathbf{x}_{k+1})(\mathbf{x}_{k+1} - \mathbf{x}_k) = \nabla r_j(\mathbf{x}_{k+1}) - \nabla r_j(\mathbf{x}_k). \quad (5.14)$$

Since we do not have $\nabla\nabla r_j(\mathbf{x}_{k+1})$, we can construct an approximation, $(S_j)_{k+1}$, which mimics the behavior of $\nabla\nabla r_j(\mathbf{x}_{k+1})$ over the step from k to $k + 1$. We require that this approximation should also satisfy Eq. 5.14, i.e.,

$$(S_j)_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) = \nabla r_j(\mathbf{x}_{k+1}) - \nabla r_j(\mathbf{x}_k). \quad (5.15)$$

The RHS of Eq. 5.15 is merely a subtraction of the j th row of the Jacobian matrices between the two consecutive steps. Denote the approximation to the second term in the Hessian $\sum_{j=1}^m r_j(\mathbf{x})\nabla\nabla r_j(\mathbf{x})$ for the $(k + 1)$ th step as \mathbf{B}_{k+1} , then it must satisfy

$$\begin{aligned} \mathbf{B}_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) &= \sum_{j=1}^m r_j(\mathbf{x}_{k+1})(S_j)_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) \\ &= \sum_{j=1}^m r_j(\mathbf{x}_{k+1})[\nabla r_j(\mathbf{x}_{k+1}) - \nabla r_j(\mathbf{x}_k)] \\ &= [\mathbf{A}(\mathbf{x}_{k+1})^T - \mathbf{A}(\mathbf{x}_k)^T] \mathbf{r}(\mathbf{x}_{k+1}). \end{aligned} \quad (5.16)$$

Equation 5.16 is not enough to constrain \mathbf{B}_{k+1} uniquely and additional conditions (e.g., \mathbf{B}_{k+1} must be symmetric) can be imposed to obtain an explicit formula for \mathbf{B}_{k+1} (Dennis et al. 1981).

For problems with large residuals or large $\nabla \nabla r_j$, the various extensions of the Gauss–Newton algorithm may provide better convergence rate than the Gauss–Newton algorithm before reaching a neighborhood of the optimal solution (i.e., during earlier iterations). When the second term in the Hessian is negligible (i.e., small residual or small $\nabla \nabla r_j$ or both), the Gauss–Newton algorithm has quadratic convergence rate, which is faster than the superlinear or linear convergence rate of the conjugate-gradient, the steepest-descent and other quasi-Newton algorithms. It is also possible to combine the Gauss–Newton algorithm with other types of algorithms (e.g., BFGS) to construct hybrid algorithms, in which the Gauss–Newton update is taken if the reduction in the objective function is sufficiently large and a BFGS-type update is taken otherwise (e.g., Fletcher 2013).

5.1.2 F3DT-SI

Full-3D tomography based on the scattering-integral method (F3DT-SI) uses the Gauss–Newton algorithm to minimize a quadratic objective function defined in terms of the waveform misfit measurements and model regularization, i.e.,

$$\chi^2(\mathbf{m}, \mathbf{m}_k) = \Delta \mathbf{d}^T(\mathbf{m}, \mathbf{m}_k) \mathbf{C}_d^{-1} \Delta \mathbf{d}(\mathbf{m}, \mathbf{m}_k) + (\mathbf{m} - \mathbf{m}_k)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_k), \quad (5.17)$$

and the spatially discretized structural model \mathbf{m} is updated through a finite series of perturbations, i.e.,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta \mathbf{m}_k, \quad k = 0, 1, 2, \dots, K. \quad (5.18)$$

In Eq. 5.17, $\Delta \mathbf{d}$ is a column vector composed of the data functional differences Δd^{sn} (Eq. 4.5), \mathbf{C}_d is a positive-definite data covariance matrix and \mathbf{C}_m is a positive-definite model covariance matrix. The k th model update $\Delta \mathbf{m}_k$ is constructed by minimizing this quadratic objective function at the k th reference model \mathbf{m}_k .

There are statistical motivations for choosing such an objective function. It might be reasonable to assume that the individual misfits Δd^{sn} are independent and identically distributed with a certain Gaussian probability density function. Under such an assumption, the likelihood of a given collection of misfits can be expressed in terms of the product of the Gaussian distribution for every misfit in this collection. The exponent of the likelihood has the form $-\Delta \mathbf{d}^T \mathbf{C}_d^{-1} \Delta \mathbf{d}$, which is the negative of the first term on the RHS of Eq. 5.17. If we use the *maximum likelihood* estimate, the likelihood is maximized when $\Delta \mathbf{d}^T \mathbf{C}_d^{-1} \Delta \mathbf{d}$ is minimized. If there are correlations among the individual misfits, the likelihood can be expressed in terms of a general multivariate Gaussian distribution with a non-diagonal data covariance

matrix \mathbf{C}_d . A priori constraints can also be incorporated into the analysis using the *Gaussian-Bayesian* framework and can give rise to the second term on the RHS of Eq. 5.17. Detailed derivations based on statistical inference can be found in e.g., (Tarantola 2005).

The Jacobian matrix at the k th iteration is then given by

$$\begin{aligned} \mathbf{A}(\mathbf{m}_k) &= \left[\frac{\partial \Delta d^{sn}(\mathbf{m}, \mathbf{m}_k)}{\partial \mathbf{m}} \right] \Big|_{\mathbf{m}=\mathbf{m}_k} \\ &= - \left[\frac{\partial d^{sn}(\mathbf{m}, \mathbf{m}_k)}{\partial \mathbf{m}} \right] \Big|_{\mathbf{m}=\mathbf{m}_k} = -\mathbf{K}_k, \end{aligned} \quad (5.19)$$

where \mathbf{K}_k is the kernel matrix whose rows are the spatially discretized data sensitivity kernels of the data functionals d^{sn} evaluated at the reference model \mathbf{m}_k . The gradient of the objective function (i.e., the Fréchet derivative of χ^2 with respect to the structural model \mathbf{m}) at the k th iteration can then be expressed using the kernel matrix as

$$\nabla_{\mathbf{m}} \chi^2(\mathbf{m}, \mathbf{m}_k) \Big|_{\mathbf{m}=\mathbf{m}_k} = -\mathbf{K}_k^T \mathbf{C}_d^{-1} \Delta \mathbf{d}(\mathbf{m}_k, \mathbf{m}_k). \quad (5.20)$$

Considering Eq. 4.3, we have

$$\Delta d^{sn}(\mathbf{m}_k, \mathbf{m}_k) = \bar{d}^{sn} - d^{sn}(\mathbf{m}_k, \mathbf{m}_k) = \bar{d}^{sn}, \quad (5.21)$$

where \bar{d}^{sn} is the misfit between the observed wavefield and the synthetic wavefield computed using the reference structural model at the k th iteration \mathbf{m}_k . Equation 5.20 indicates that the gradient of the objective function is in fact a data-weighted superposition of individual data sensitivity kernels and the misfit data is given by \bar{d}^{sn} . Equation 5.20 has implications for the adjoint-wavefield method, as we will see in Sect. 5.2. The Gauss–Newton approximation of the Hessian can be expressed using the kernel matrix as

$$\mathbf{H}_k \approx \mathbf{K}_k^T \mathbf{C}_d^{-1} \mathbf{K}_k + \mathbf{C}_m^{-1}. \quad (5.22)$$

The Gauss–Newton normal equation for the k th iteration is then given by

$$\left(\mathbf{K}_k^T \mathbf{C}_d^{-1} \mathbf{K}_k + \mathbf{C}_m^{-1} \right) \Delta \mathbf{m}_k = \mathbf{K}_k^T \mathbf{C}_d^{-1} \mathbf{d}_k, \quad (5.23)$$

where the column vector \mathbf{d}_k is composed of \bar{d}^{sn} and $\Delta \mathbf{m}_k$ is the k th model update (Eq. 5.18). It can be verified that the Gauss–Newton normal Eq. 5.23 is identical to the normal equation of the following linear least-squares problem

$$\min_{\Delta \mathbf{m}_k} \left\| \begin{bmatrix} \mathbf{C}_d^{-1/2} \mathbf{K}_k \\ \mathbf{C}_m^{-1/2} \end{bmatrix} \Delta \mathbf{m}_k - \begin{bmatrix} \mathbf{C}_d^{-1/2} \mathbf{d}_k \\ 0 \end{bmatrix} \right\|_2^2. \quad (5.24)$$

Therefore the nonlinear least-squares optimization problem defined in Eq. 5.17 can be solved by repeated solution of the linear least-squares problem defined in Eq. 5.24.

5.1.3 Scalable Parallel LSQR (SPLSQR)

The Gauss–Newton normal Eq. 5.23 or the linear least-squares problem in Eq. 5.24 can be solved using a variety of linear system solvers (e.g., Cholesky factorization, QR factorization, singular-value decomposition, conjugate-gradient). When the linear system is very large, iterative methods that solve the linear least-squares problem in Eq. 5.24 directly can be highly efficient. In such iterative methods, we only need the ability to perform successive matrix-vector multiplications with the Jacobian and its transpose without having to form the Gauss–Newton approximate Hessian explicitly. A highly successful iterative method is the Least Squares with QR factorization (LSQR) algorithm as described in (Paige and Saunders 1982). The LSQR algorithm is a member of the conjugate-gradient family of Krylov subspace iterative methods. It uses an iterative Lanczos process to construct orthonormal vectors in the model and data spaces. It is one of the most popular linear system solvers used in seismic tomography due to its efficiency and stability in solving large, sparse and ill-conditioned problems (e.g., Nolet 1985; Iyer and Hirahara 1993).

Implementations of the LSQR algorithm on distributed-memory computer clusters have been discussed previously in e.g., (Baur and Austen 2004; Liu et al. 2006; Balay et al. 1997, 2014a, b). These traditional implementations can have significant inter-processor communication overhead depending upon the structure of the linear system. Such inter-processor communication overhead can significantly limit the parallel scalability of the code. To address the scalability problem in traditional implementations, we have developed a new scalable parallel implementation optimized for the seismic tomography problem in (Huang et al. 2012, 2013; Lee et al. 2013). We call this new implementation “Scalable Parallel LSQR” (SPLSQR). SPLSQR has a scalable inter-processor communication volume distributed among a fixed or a modest number of neighboring processors. The largest linear system we have solved using SPLSQR has over 38 million columns, 261 million rows and 144 billion non-zeros. We were able to solve this linear system for 300 LSQR iterations on 12,000 cores on the IBM iDataPlex system (code named “Yellowstone”) at the NCAR-Wyoming Supercomputing Center (NWSC) in less than 5 min of wall-time.

Technical details about SPLSQR have been documented in (Huang et al. 2012, 2013; Lee et al. 2013). In (Lee et al. 2013), we gave a thorough analysis of the entire algorithm, as well as realistic performance evaluations obtained from our F3DT-SI application in Southern California (Lee et al. 2014b, Chap. 6). There are a number of innovative features in SPLSQR and among them three are critical for scalability. First, on distributed-memory computer clusters, each processor stores only a portion of the complete matrix and vectors in its own memory. Therefore how to partition the matrix and vectors among the processors affects both the computational load on each processor and the inter-processor communication overhead. We partition the kernel matrix $\mathbf{C}_d^{-1/2} \mathbf{K}$ in Eq. 5.24 (here we have dropped the iteration index k to reduce clutter) according to columns (i.e., different processor owns a range of columns of $\mathbf{C}_d^{-1/2} \mathbf{K}$) and partition the damping matrix $\mathbf{C}_m^{-1/2}$ according to rows. This data partitioning scheme can significantly reduce inter-processor

communication overhead for highly under-determined linear systems (Lee et al. 2013). Second, unlike the kernel matrix $\mathbf{C}_d^{-1/2}\mathbf{K}$, whose non-zero pattern can be highly irregular depending upon the distribution of seismic sources and receivers, the model discretization and some other factors, the damping matrix $\mathbf{C}_m^{-1/2}$ is usually highly sparse and has specific structures. The inter-processor communication overhead depends upon this internal structure of the damping matrix. In (Lee et al. 2013) we have shown that in order to minimize the inter-processor communication overhead it is important to re-order $\mathbf{C}_m^{-1/2}$ such that its bandwidth is as small as possible (i.e., as close to a diagonal matrix as possible). Three, the overall performance of the code depends upon the amount of data stored on each processor (i.e., memory balance), the number of calculations performed by each processor (i.e., computation balance) and also the number of inter-processor communications for each processor (i.e., communication balance). Our load-balancing scheme takes into account all three factors. When partitioning the kernel matrix according to columns, we ensure that the number of nonzeros on each processor is about the same (i.e., memory and computation balance). The re-ordered damping matrix is partitioned according to rows at the same time as we partition the kernel matrix to give a rough estimate of the inter-processor communication overhead. The resulting partition provides a good compromise among all three factors.

We have compared the performance of SPLSQR with the parallel LSQR implementation in the Portable, Extensible Toolkit for Scientific Computation (PETSc) library (Balay et al. 1997, 2014a, b), a widely used, highly optimized parallel linear algebra library. SPLSQR shows significant speed-up compared to PETSc for our F3DT-SI application in Southern California. On Yellowstone we were able to obtain about 17 times of speed-up using 2400 cores and about 74 times of speed-up using 9600 cores (Lee et al. 2013). When comparing the total wall-time spent only on inter-processor communication, the speed-up was about 62 times on 2400 cores and about 201 times on 9600 cores. SPLSQR has also been ported to distributed-memory computer clusters equipped with GPU accelerators and obtained significant speed-up compared with its purely CPU implementation (Huang et al. 2012). The optimization and implementation of SPLSQR are tailored to the specific needs of our seismic tomography problem and may not be applicable to more general linear systems encountered in other fields. The LSQR code in PETSc is suitable for a wider class of problems.

5.1.4 The Roughing Operator

When solving the linear least-squares problem in Eq. 5.24 using iterative linear system solvers, such as the LSQR algorithm, we need to specify specific forms for the data-weighting matrix $\mathbf{C}_d^{-1/2}$ and the damping (or regularization) matrix $\mathbf{C}_m^{-1/2}$. The data-weighting matrix can often be determined from a certain noise model for the misfit measurements used in the inversion. Examples about how to construct

such noise models for frequency-dependent phase and amplitude misfits measured on earthquake seismograms were shown in e.g., (Chen et al. 2007; Lee et al. 2014b) and examples for misfit measurements made on ambient-noise Green's functions are shown in (Lee et al. 2014b). In the absence of a robust noise model, we can assume that $\mathbf{C}_d^{-1/2}$ is the identity matrix, which will give all misfits equal weights.

The damping matrix $\mathbf{C}_m^{-1/2}$ is sometimes called the *roughing operator* (e.g., Tarantola 2005), as its inverse $\mathbf{C}_m^{1/2}$ is often a smoothing operator. By using a roughing operator for $\mathbf{C}_m^{-1/2}$ in Eq. 5.24 we are imposing our prior information on the inversion, which is that, in the absence of other information, the model perturbation $\Delta \mathbf{m}_k$ should be smooth in space. By minimizing Eq. 5.24, we are penalizing spatially rough model perturbations among all possible solutions. One possible choice for such a roughing operator is the Laplacian $\nabla^2 = \nabla \cdot \nabla$ (e.g., deGroot-Hedlin and Constable 1990; Sambridge 1990; Tarantola 2005). If we would also like to penalize model perturbations that are too large, we can include the identity operator \mathbf{I} into $\mathbf{C}_m^{-1/2}$. In practice, we can use a linear combination of the identity operator and the Laplacian, i.e.,

$$\mathbf{C}_m^{-1/2} = \theta_1 \mathbf{I} + \theta_2 \nabla^2, \quad (5.25)$$

where the regularization parameters $\theta_{1,2}$ can be chosen based on the trade-off between model smoothness/norm and the fit to the data. If the model perturbation $\Delta \mathbf{m}_k$ is discretized on a uniform grid, the Laplacian can be approximated numerically using spatial finite-differencing operators. For other types of model discretizations, it is also possible to explicitly construct numerical approximations of Eq. 5.25.

The smoothing operator $\mathbf{C}_m^{1/2}$ corresponding to the roughing operator in Eq. 5.25 has connections with the exponential model covariance function. To see this, we consider the one-dimensional stationary exponential covariance function

$$f(|x - y|) = e^{-2\pi\theta|x-y|}, \quad (5.26)$$

where θ is a parameter proportional to the inverse of the correlation length. We would like to find the function, $g(|x - z|)$, such that for any $h(y)$ we have

$$\int_{-\infty}^{\infty} dx g(|x - z|) \int_{-\infty}^{\infty} dy f(|x - y|) h(y) = h(z). \quad (5.27)$$

Equation 5.27 implies that

$$\int_{-\infty}^{\infty} dx g(|x - z|) f(|x - y|) = \delta(y - z), \quad (5.28)$$

The Fourier transform of 5.26 over x can be evaluated explicitly as

$$\hat{f}(\xi) = \frac{1}{\pi} \frac{\theta}{\theta^2 + \xi^2} e^{-i\xi y}, \quad (5.29)$$

whose amplitude is known as the Lorentzian function. Using the inverse Fourier transform, we can express $f(|x - y|)$ as

$$f(|x - y|) = f(|y - x|) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\xi e^{-i\xi x} \left(\frac{1}{\pi} \frac{\theta}{\theta^2 + \xi^2} e^{i\xi y} \right). \quad (5.30)$$

Bring Eq. 5.30 into Eq. 5.28, we obtain

$$\int_{-\infty}^{\infty} d\xi \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-i\xi x} g(|x - z|) \right] \left(\frac{1}{\pi} \frac{\theta}{\theta^2 + \xi^2} e^{i\xi y} \right) = \delta(y - z). \quad (5.31)$$

Considering Eq. 3.94, the Fourier transform of $g(|x - z|)$ over x must be

$$\hat{g}(\xi) = \sqrt{\frac{\pi}{2}} \frac{\theta^2 + \xi^2}{\theta} e^{-i\xi z} \quad (5.32)$$

in order to satisfy Eq. 5.28. The inverse Fourier transform of Eq. 5.32 can then be expressed as

$$g(|x - z|) = \pi\theta\delta(x - z) - \frac{\pi}{\theta}\delta''(x - z), \quad (5.33)$$

where $\delta''(x)$ is the second-order derivative of the Dirac delta function with respect to x . The first term on the RHS of Eq. 5.33 is the identity operator scaled by $\pi\theta$, which is proportional to the inverse of the correlation length. The second term on the RHS of Eq. 5.33 is the one-dimensional Laplacian (integration-by-parts twice) scaled by $-\pi/\theta$, which is proportional to the correlation length. The function $g(|x - z|)$ is in fact the one-dimensional roughing operator corresponding to the exponential covariance function in Eq. 5.26.

5.1.5 Annihilator Matrix

In inversions that involve multiple types of model parameters we may wish to maximize the sensitivities to some types of model parameters while minimizing the sensitivities to other types of model parameters. For instance, when inverting for structural parameters, the misfit measurements are also affected by errors in the assumed source parameters. The partial derivatives of the misfits with respect to the source parameters can be computed and we may formulate a joint inversion that includes perturbations to both structural parameters and source parameters as the unknown $\Delta\mathbf{m}_k$. The effects of the unknown perturbations to the source parameters can be projected away by using the annihilator matrix. The same approach can be applied to remove the effects of any types of unwanted model parameters, as long as the Fréchet kernels of the misfits with respect to those model parameters can be computed. These unwanted model parameters are sometimes called *nuisance parameters* and the procedure for projecting away their effects on the misfits is called *denuisancing*, which has been used to remove path effects in relative

earthquake source relocation techniques (e.g., Jordan and Sverdrup 1981) and FMT inversions (Chen et al. 2005). Similar techniques have also been well studied in the context of ordinary least squares (OLS) for estimating the unknown parameters in linear regression models (e.g., Montgomery et al. 2012).

To simplify notation, we can re-write Eq. 5.24 in the following form

$$\mathbf{K}\Delta\mathbf{m} = \begin{bmatrix} \mathbf{K}_1 & \mathbf{K}_2 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{m}_1 \\ \Delta\mathbf{m}_2 \end{bmatrix} = \mathbf{K}_1\Delta\mathbf{m}_1 + \mathbf{K}_2\Delta\mathbf{m}_2 = \mathbf{d}, \quad (5.34)$$

where I have dropped the iteration index k to reduce clutter, $\Delta\mathbf{m}_{1,2}$ are two types of unknown model parameters and $\mathbf{K}_{1,2}$ are their corresponding kernel sub-matrices. Suppose we would like to annihilate the effects of $\Delta\mathbf{m}_2$ from Eq. 5.34, we first construct the projection matrix

$$\mathbf{P} = \mathbf{K}_2 \left(\mathbf{K}_2^T \mathbf{K}_2 \right)^{-1} \mathbf{K}_2^T. \quad (5.35)$$

The annihilator matrix is then given by

$$\mathbf{Q} = \mathbf{I} - \mathbf{P}. \quad (5.36)$$

Left-multiplying \mathbf{Q} on both sides of Eq. 5.34, we obtain

$$\begin{aligned} \mathbf{Q}\mathbf{K}_1\Delta\mathbf{m}_1 + (\mathbf{I} - \mathbf{P})\mathbf{K}_2\Delta\mathbf{m}_2 &= \mathbf{Q}\mathbf{d} \\ \Rightarrow \mathbf{Q}\mathbf{K}_1\Delta\mathbf{m}_1 + \left[\mathbf{K}_2 - \mathbf{K}_2 \left(\mathbf{K}_2^T \mathbf{K}_2 \right)^{-1} \mathbf{K}_2^T \mathbf{K}_2 \right] \Delta\mathbf{m}_2 &= \mathbf{Q}\mathbf{d} \\ \Rightarrow \mathbf{Q}\mathbf{K}_1\Delta\mathbf{m}_1 &= \mathbf{Q}\mathbf{d}. \end{aligned} \quad (5.37)$$

Equation 5.37 does not have $\Delta\mathbf{m}_2$. In Eq. 5.35, the projection matrix is defined in terms of one particular form (i.e., the left inverse) of the Moore-Penrose pseudoinverse (e.g., Campbell and Meyer 2009), which we denote as \mathbf{K}_2^+ in the following, i.e.,

$$\mathbf{K}_2^+ = \left(\mathbf{K}_2^T \mathbf{K}_2 \right)^{-1} \mathbf{K}_2^T. \quad (5.38)$$

The inverse $\left(\mathbf{K}_2^T \mathbf{K}_2 \right)^{-1}$ in Eq. 5.38 exists when \mathbf{K}_2 has full column rank. If we left-multiply \mathbf{K}_2^+ on \mathbf{K}_2 , we obtain

$$\mathbf{K}_2^+ \mathbf{K}_2 = \left(\mathbf{K}_2^T \mathbf{K}_2 \right)^{-1} \mathbf{K}_2^T \mathbf{K}_2 = \mathbf{I}. \quad (5.39)$$

If \mathbf{K}_2 has full row rank, then $\left(\mathbf{K}_2 \mathbf{K}_2^T \right)^{-1}$ exists and the pseudoinverse can be computed as

$$\mathbf{K}_2^+ = \mathbf{K}_2^T \left(\mathbf{K}_2 \mathbf{K}_2^T \right)^{-1}. \quad (5.40)$$

The pseudoinverse given in Eq. 5.40 is called the right inverse. If we right-multiply \mathbf{K}_2^+ on \mathbf{K}_2 , we obtain

$$\mathbf{K}_2 \mathbf{K}_2^+ = \mathbf{K}_2 \mathbf{K}_2^T \left(\mathbf{K}_2 \mathbf{K}_2^T \right)^{-1} = \mathbf{I}. \quad (5.41)$$

It can be shown that the Moore-Penrose pseudoinverse is defined and unique for all matrices and can always be computed through singular-value decomposition (e.g., Campbell and Meyer 2009). The projection matrix in Eq. 5.35 can then be written in terms of the pseudoinverse as

$$\mathbf{P} = \mathbf{K}_2 \mathbf{K}_2^+, \quad (5.42)$$

and it can be shown that the identities

$$\mathbf{P} \mathbf{K}_2 = \mathbf{K}_2 \mathbf{K}_2^+ \mathbf{K}_2 = \mathbf{K}_2, \quad (5.43)$$

$$\mathbf{K}_2^+ \mathbf{P} = \mathbf{K}_2^+ \mathbf{K}_2 \mathbf{K}_2^+ = \mathbf{K}_2^+ \quad (5.44)$$

always hold. The annihilator matrix \mathbf{Q} defined in Eq. 5.36 always exists and is the orthogonal projector onto the null-space of \mathbf{K}_2^+ . The analysis above can also be extended to complex matrices, in which case the transpose needs to be replaced with the Hermitian (conjugate) transpose.

Once the annihilator matrix \mathbf{Q} is constructed, the application of Eq. 5.37 involves left-multiplying \mathbf{Q} with \mathbf{K}_1 and \mathbf{d} , which produces linear combinations of the data sensitivity kernels in \mathbf{K}_1 on the LHS and also linear combinations of misfit measurements in \mathbf{d} on the RHS.

The double-difference seismic source re-location technique (e.g., Waldhauser and Ellsworth 2000, 2002) can be obtained by treating perturbations to structural parameters as the nuisance parameters and project away their effects using the annihilator matrix constructed from the Fréchet kernels of the misfits with respect to the structural parameters. In the conventional double-difference technique, two events separated by a large distance needs to be linked together through a series of intermediate pairs (e.g., Got et al. 1994). Using the formulation given here, the annihilator matrix can automatically account for such situations. If we concatenate the linear system (Eq. 5.37) obtained by projecting away the path effects with the original linear system (Eq. 5.24) we can solve for source locations and structural parameters jointly and we have obtained a full-wave version of the double-difference ray tomography method given in e.g., (Zhang and Thurber 2003). As mentioned at the beginning of this section, the same technique can also be used to project away effects of unwanted source parameters and/or structural parameters in tomography studies.

5.2 The Adjoint-Wavefield (AW) Method

The adjoint-state method, which has been driving rapid advances in other disciplines (e.g., Bengtsson et al. 1981; Bennett 1992; Daley 1993; Malanotte-Rizzoli 1996; Wunsch 1996; Kalnay 2003), was adopted to solve seismic imaging problems in some early work of e.g., (Bamberger et al. 1977, 1982) and formalized in the space-time domain in e.g., (Tarantola 1984, 1988a) and in the space-frequency domain in e.g., (Pratt et al. 1998; Pratt 1999). It has been successfully applied in seismic

exploration industry for more than a decade. It was adapted to large-scale passive-source tomography studies in (Tromp et al. 2005), in which extensions to different types of data functionals and connections with finite-frequency kernels were made. In seismic inverse problems, the state variable is the wavefield, therefore we use the term “adjoint-wavefield” method (AWM) instead of the more general term “adjoint-state” method.

The adjoint-wavefield method can be derived using different approaches. A detailed derivation of AWM based on the Lagrange multiplier method, which is commonly employed to solve optimal control problems with strong constraints imposed by partial-differential equations (e.g., Akcelik et al. 2002, 2003; Biegler 2003), has been documented in (Liu 2006). In the Lagrange multiplier method the objective function defined in terms of waveform misfit measurements is minimized and the seismic wave equation, as well as its initial/boundary conditions, is included into the objective function using Lagrange multipliers. Variations with respect to the Lagrange multipliers and the application of the Karush–Kuhn–Tucker (KKT) conditions lead to the forward and adjoint wave equations and also a set of equations for updating model parameters.

A more general formulation based on the variational data assimilation approach was given in (Chen 2011). This formulation provides a weakly-constrained generalized inverse, in which the dynamic model (i.e., the wave equation and its associated initial/boundary conditions) is allowed to contain errors and the solution of the inverse problem does not have to satisfy the dynamic model exactly. The strongly-constrained inverse, which is identical to the conventional AWM, can be obtained as a limiting case, in which the errors in the dynamic model approach zero. In meteorology and oceanography it has been shown that one can obtain non-physical model parameters as a result of neglecting errors in the dynamic model (e.g., Zupanski and Zupanski 2006). The effects of neglecting errors in the dynamic model in seismic inverse problems have not been fully investigated in the literature. However, it is clear that our deterministic dynamic model is not perfect and it is difficult, if possible at all, to fully eliminate all its deficiencies. The impact of the errors in the dynamic model, in particular the impact on the estimation of model parameters, needs to be evaluated, especially when the procedure for solving seismic inverse problems is becoming more and more precise. The weakly-constrained variational data assimilation approach formulated in (Chen 2011) provides a systematic means to accommodate errors in the dynamic model when solving seismic inverse problems.

The derivation given in the following is not based on the Lagrange multiplier method or the variational data assimilation approach. Instead, I will use the adjoint representation theorem (Eq. 3.335) formulated in Sect. 3.3.3.3. As we will see in the following, the application of the adjoint representation theorem can significantly simplify the derivation of the adjoint method for computing the gradient of the objective function (Sect. 5.2.1) and also for computing the Hessian-vector product (Sect. 5.2.3).

5.2.1 The Gradient

In Sect. 5.1.2 we have seen that the gradient of the objective function $\nabla_{\mathbf{m}}\chi^2(\mathbf{m}, \mathbf{m}_k)$ evaluated at the reference model \mathbf{m}_k can be expressed as the data-weighted superposition of the data sensitivity kernels for every misfit in the data set (i.e., Eq. 5.20). Since at the reference model we have $\Delta d^{sn} = \tilde{d}^{sn}$ (Eq. 5.21), we can re-write Eq. 5.20 as

$$\nabla_{\mathbf{m}}\chi^2(\mathbf{m}, \mathbf{m}_k) \Big|_{\mathbf{m}=\mathbf{m}_k} = -\mathbf{K}_k^T \mathbf{C}_d^{-1} \mathbf{d}_k, \quad (5.45)$$

where the column vector \mathbf{d}_k is composed of \tilde{d}^{sn} for the k th iteration. In the following, I will drop the iteration index k to reduce clutter. If we consider the model parameters composed of density and elastic moduli,

$$\mathbf{m}(\mathbf{x}) = \begin{bmatrix} \rho(\mathbf{x}) \\ c_{ijklm}(\mathbf{x}) \end{bmatrix}, \quad (5.46)$$

the corresponding data sensitivity kernels were given in Eqs. 4.32 and 4.35. If we denote

$$\mathbf{C}_d^{-1} \mathbf{d} = \tilde{\mathbf{d}} = [\tilde{d}^{sn}], \quad (5.47)$$

the RHS of Eq. 5.45 can be expressed as

$$-\mathbf{K}^T \mathbf{C}_d^{-1} \mathbf{d} = - \begin{bmatrix} \sum_{sn} K_{d^{sn}}^\rho(\mathbf{x}) \tilde{d}^{sn} \\ \sum_{sn} K_{d^{sn}}^{c_{ijklm}}(\mathbf{x}) \tilde{d}^{sn} \end{bmatrix}. \quad (5.48)$$

Bring Eqs. 4.32 and 4.35 into the RHS of Eq. 5.48, we obtain

$$\begin{aligned} - \sum_{sn} K_{d^{sn}}^\rho(\mathbf{x}) \tilde{d}^{sn} &= \sum_s \sum_{j=1}^3 \int_0^T d\tau \frac{\partial^2 u_j^s(\mathbf{x}, \tau)}{\partial \tau^2} \\ &\left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \sum_n \tilde{d}^{sn} J_p^{sn}(\mathbf{x}', t) G_{jp}(\mathbf{x}, t - \tau; \mathbf{x}', 0) \right], \quad (5.49) \\ - \sum_{sn} K_{d^{sn}}^{c_{ijklm}}(\mathbf{x}) \tilde{d}^{sn} &= \sum_s \int_0^T d\tau \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \\ \frac{\partial}{\partial x_k} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \sum_n \tilde{d}^{sn} J_p^{sn}(\mathbf{x}', t) G_{jp}(\mathbf{x}, t - \tau; \mathbf{x}', 0) \right], \quad (5.50) \end{aligned}$$

where I have replaced the temporal convolution operator “*” with time integrals and used Eqs. 3.407–3.409 in Eq. 5.49. Equations 5.49 and 5.50 are based on the reciprocity relation, which was used when deriving the data sensitivity kernels in

Chap. 3 and 4. As mentioned in Sects. 3.4.1, 3.5.2 and at the beginning of Chap. 4, more general formulas that do not rely on reciprocity can be obtain by replacing $G_{jp}(\mathbf{x}, t - \tau; \mathbf{x}', 0)$ with the time-reversed adjoint Green's tensor $G_{jp}^\dagger(\mathbf{x}, t - \tau; \mathbf{x}', 0)$ (Eq. 3.345) and Eqs. 5.49 and 5.50 can then be written as

$$-\sum_{sn} K_{dsn}^\rho(\mathbf{x}) \tilde{d}^{sn} = \sum_s \sum_{j=1}^3 \int_0^T d\tau \frac{\partial^2 u_j^s(\mathbf{x}, \tau)}{\partial \tau^2} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \sum_n \tilde{d}^{sn} J_p^{sn}(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau - t; \mathbf{x}', 0) \right], \quad (5.51)$$

$$-\sum_{sn} K_{dsn}^{c_{jklm}}(\mathbf{x}) \tilde{d}^{sn} = \sum_s \int_0^T d\tau \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \frac{\partial}{\partial x_k} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \sum_n \tilde{d}^{sn} J_p^{sn}(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau - t; \mathbf{x}', 0) \right]. \quad (5.52)$$

If we define the adjoint source field as

$$[f_p^s]^\dagger(\mathbf{x}', t) = \sum_n \tilde{d}^{sn} J_p^{sn}(\mathbf{x}', t), \quad (5.53)$$

then Eqs. 5.51 and 5.52 can be expressed as

$$-\sum_{sn} K_{dsn}^\rho(\mathbf{x}) \tilde{d}^{sn} = \sum_s \sum_{j=1}^3 \int_0^T d\tau \frac{\partial^2 u_j^s(\mathbf{x}, \tau)}{\partial \tau^2} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 [f_p^s]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) \right], \quad (5.54)$$

$$-\sum_{sn} K_{dsn}^{c_{jklm}}(\mathbf{x}) \tilde{d}^{sn} = \sum_s \int_0^T d\tau \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \frac{\partial}{\partial x_k} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 [f_p^s]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) \right]. \quad (5.55)$$

The terms inside the square brackets on the RHS of Eqs. 5.54 and 5.55 are in the form of the first term on the RHS of the adjoint representation theorem, Eq. 3.335, which means that if we use the adjoint source field in Eq. 5.53 as the source and solve the adjoint (visco)elastic wave equation coupled with the homogeneous

adjoint initial/boundary conditions backwards in time, the resulting adjoint wavefield $\left[u_j^s \right]^\dagger(\mathbf{x}, \tau)$ will be identical to the terms in the square brackets in Eqs. 5.54 and 5.55. Therefore as a result of Eq. 3.335 we have

$$\begin{aligned} \left[u_j^s \right]^\dagger(\mathbf{x}, \tau) &= \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \left[f_p^s \right]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) \\ &= \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \left[f_p^s \right]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau - t; \mathbf{x}', 0) \\ &= \int_V dV(\mathbf{x}') \sum_{p=1}^3 \left[f_p^s \right]^\dagger(\mathbf{x}', \tau) * G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', 0) \end{aligned} \quad (5.56)$$

and Eqs. 5.54 and 5.55 can be expressed in terms of $\left[u_j^s \right]^\dagger(\mathbf{x}, \tau)$ as

$$\begin{aligned} - \sum_{sn} K_{dsn}^\rho(\mathbf{x}) \tilde{d}^{sn} &= \sum_s \sum_{j=1}^3 \int_0^T d\tau \frac{\partial^2 u_j^s(\mathbf{x}, \tau)}{\partial \tau^2} \left[u_j^s \right]^\dagger(\mathbf{x}, \tau), \\ &= - \sum_s \left[\sum_{j=1}^3 \frac{\partial \left[u_j^s \right]^\dagger(\mathbf{x}, t)}{\partial t} \otimes \frac{\partial u_j^s(\mathbf{x}, t)}{\partial t} \Bigg|_{t=0} \right], \end{aligned} \quad (5.57)$$

$$\begin{aligned} - \sum_{sn} K_{dsn}^{c_{jklm}}(\mathbf{x}) \tilde{d}^{sn} &= \sum_s \int_0^T d\tau \frac{\partial \left[u_j^s \right]^\dagger(\mathbf{x}, \tau)}{\partial x_k} \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \\ &= \sum_s \left[\frac{\partial \left[u_j^s \right]^\dagger(\mathbf{x}, t)}{\partial x_k} \otimes \frac{\partial u_l^s(\mathbf{x}, t)}{\partial x_m} \Bigg|_{t=0} \right], \end{aligned} \quad (5.58)$$

where I have applied integration-by-parts at the second equality in Eq. 5.57 and the integrals over τ are expressed as correlations “ \otimes ” at zero-lag (e.g., Tarantola 1988b). The terms inside the square brackets on the RHS of Eqs. 5.57 and 5.58 are sometimes called the *event kernel* for the s -th source in the literature. The gradient of the objective function is obtained through the summation of all event kernels (i.e., the summation over the source index s).

As shown in Eq. 5.53, the adjoint source field used for generating the adjoint wavefield of the s -th source is a data-weighted superposition of the WPKs for the s -th source. As shown in Eqs. 5.57 and 5.58, the gradient of the objective function can be constructed through two wave-propagation simulations per source: one forward simulation and one adjoint simulation. The gradient with respect to density is obtained by summing the zero-lag temporal correlation between the forward

particle-velocity field and the adjoint particle-velocity field for each source. The gradient with respect to elastic moduli is obtained by summing the zero-lag temporal correlation between the forward strain field and the adjoint strain field for each source. Derivations of the gradient with respect to other types of model parameters, including anelastic attenuation and anisotropy, based on the adjoint method are straightforward based on formulations given in Sect. 4.3 and have been documented in e.g., (Tarantola 1988b; Tromp et al. 2005; Liu and Tromp 2006; Sieminski et al. 2007).

When solving for the adjoint wavefield, if we would like to step forward in time, rather than backward in time, we need to reverse the time axis of the adjoint source field by making the change $t \rightarrow T - t$, i.e.,

$$\left[f_p^s \right]^\dagger(\mathbf{x}', t) = \sum_n \tilde{a}^{sn} J_p^{sn}(\mathbf{x}', T - t), \quad (5.59)$$

and after the simulation completes, we then need to reverse the time axis of the resulting wavefield by making the change $t \rightarrow T - t$ again. To understand why the two time-reversal operations are necessary to obtain the adjoint wavefield $\left[u_j^s \right]^\dagger(\mathbf{x}, t)$, consider the temporal convolution in Eq. 5.56 between the adjoint source field and the adjoint Green's tensor, which I abbreviate as $f^\dagger(\tau) * G^\dagger(\tau)$. In the frequency domain, the convolution can be expressed as multiplication $\hat{f}^\dagger(\omega)\hat{G}^\dagger(\omega)$. By stepping forward in time, we are multiplying with the conjugate of the adjoint Green's tensor $\left[\hat{G}^\dagger(\omega) \right]^*$. To get back the correct result, we can conjugate the adjoint source and then conjugate the product, i.e., $\left\{ \left[\hat{f}^\dagger(\omega) \right]^* \left[\hat{G}^\dagger(\omega) \right]^* \right\}^* = \hat{f}^\dagger(\omega)\hat{G}^\dagger(\omega)$. The conjugation on the adjoint source corresponds to the time reversal of the adjoint source field in the time domain. The conjugation on the product corresponds to the time reversal of the simulation results in the time domain.

At the current stage of computational technology, full-3D (visco)elastic seismic wave-propagation simulations are still computationally expensive. An important advantage of the adjoint method for constructing the gradient of the objective function is that the number of wave-propagation simulations are independent of the number of receivers and the number of misfit measurements on each seismogram. The summation over receiver and misfit indices, which are included in the index n in the formulation above is accounted for in the adjoint source field. For tomography studies involving a large number of receivers and misfit measurements, but a limited number of sources, the adjoint method is highly efficient for constructing the gradient of the objective function.

In principle, the adjoint method can also be used to construct the Jacobian matrix, which is composed of the data sensitivity kernel for each misfit measurement (Eq. 5.19), and also the approximate Hessian (Eq. 5.22). The calculation will involve two simulations per misfit measurement. In realistic tomography studies

involving a large number of misfit measurements, the computational cost for constructing the Jacobian using the adjoint method can become prohibitive and the scattering-integral method is more preferable. A more in-depth analysis of the computational cost associated with both the adjoint method and the scattering-integral method was presented in (Chen et al. 2007a).

5.2.2 F3DT-AW

The gradient of the objective function computed using the adjoint method can be conditioned before being used in a conjugate-gradient, quasi-Newton or other types of gradient-based optimization algorithms (e.g., Tarantola 1988b; Fichtner et al. 2009; Tape et al. 2010; Chen 2011). Such a conditioning step can be highly important for improving the convergence rate of the optimization algorithm. The theoretical justification for this conditioning step can be made based on the model covariance operator \mathbf{C}_m (e.g., Tarantola 1988b; Chen 2011), which describes our *a priori* assumptions about the model perturbation. If we denote the gradient, which is a function of space, as $\mathbf{K}_{\chi^2}^m(\mathbf{x})$, the conditioned gradient can be written as

$$\tilde{\mathbf{K}}_{\chi^2}^m(\mathbf{x}) = \int_V dV(\mathbf{x}') \mathbf{K}_{\chi^2}^m(\mathbf{x}') \mathbf{C}_m(\mathbf{x}, \mathbf{x}'). \quad (5.60)$$

The model covariance is usually a smoothing operator (Sect. 5.1.4) and Eq. 5.60 can often be implemented efficiently through a spatial convolution (filtering). One issue in conditioning the gradient is how to deal with the excessively large sensitivities in close proximity of the sources and the receivers. Some techniques have been developed to address this issue by taking into account the geometric spreading effect in the gradient (e.g., Fichtner et al. 2009).

In practice, the conditioned gradient is discretized in space and the gradient with respect to different model parameters can be concatenated into a single column vector, which we denote in the following as $\tilde{\mathbf{K}}_k$ for the k th iteration. In a typical conjugate-gradient algorithm, the descent direction at the k th iteration, denoted as \mathbf{P}_k , is constructed from the descent direction at the $(k - 1)$ -th iteration, \mathbf{P}_{k-1} , and the gradient of the k th iteration $\tilde{\mathbf{K}}_k$ using the formula

$$\mathbf{P}_k = -\tilde{\mathbf{K}}_k + b_k \mathbf{P}_{k-1}, \quad (5.61)$$

where b_k is a scalar computed from the gradients of the previous and the current iterations. For the first iteration, the descent direction is set to the negative gradient, therefore the first iteration is always a steepest-descent iteration. There are at least five widely-used, different choices for computing the scalar b_k in Eq. 5.61. The original Fletcher-Reeves formula (Fletcher and Reeves 1964),

$$b_k = \frac{\tilde{\mathbf{K}}_k^T \tilde{\mathbf{K}}_k}{\tilde{\mathbf{K}}_{k-1}^T \tilde{\mathbf{K}}_{k-1}}, \quad (5.62)$$

the one due to Polak and Ribière (1969),

$$b_k = \frac{\tilde{\mathbf{K}}_k^T (\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1})}{\tilde{\mathbf{K}}_{k-1}^T \tilde{\mathbf{K}}_{k-1}}, \quad (5.63)$$

the one due to Hestenes and Stiefel (1952) and also known as the Beale-Sorenson formula (Sorenson 1969; Beale 1972),

$$b_k = -\frac{\tilde{\mathbf{K}}_k^T (\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1})}{(\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1})^T \mathbf{P}_{k-1}}, \quad (5.64)$$

the one due to Dai and Yuan (1999),

$$b_k = -\frac{\tilde{\mathbf{K}}_k^T \tilde{\mathbf{K}}_k}{(\tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1})^T \mathbf{P}_{k-1}}, \quad (5.65)$$

and the one due to Hager and Zhang (2005),

$$b_k = \left(\mathbf{Y}_{k-1} - 2\mathbf{P}_{k-1} \frac{\tilde{\mathbf{Y}}_{k-1}^T \tilde{\mathbf{Y}}_{k-1}}{\tilde{\mathbf{Y}}_{k-1}^T \mathbf{P}_{k-1}} \right)^T \frac{\tilde{\mathbf{K}}_k}{\tilde{\mathbf{Y}}_{k-1}^T \mathbf{P}_{k-1}},$$

where $\mathbf{Y}_{k-1} = \tilde{\mathbf{K}}_k - \tilde{\mathbf{K}}_{k-1}$.

(5.66)

Once the descent direction is obtained, a line search can be performed to find the optimal step length a_k along the descent direction and the model vector can be updated to the next iteration,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + a_k \mathbf{P}_k. \quad (5.67)$$

In practice, a quadratic (or cubic) interpolation for the line search is often sufficient for finding the optimal step length a_k (e.g., Tromp et al. 2005). A modification of the nonlinear conjugate-gradient algorithm is to restart the iterations by setting $b_k = 0$ after certain criteria are met (Powell 1977). The restart procedure refreshes the search by erasing old information from previous iterations and taking a steepest descent step. Some restart criteria have been provided in the literature (e.g., Powell 1977; Dai and Yuan 2001).

The Hager–Zhang formula (Eq. 5.66) seems to give the best numerical results for a quite large class of problems (Hager and Zhang 2005). There are numerical evidences that the Fletcher-Reeves formula (Eq. 5.62) can become slower than the steepest-descent algorithm after the iteration enters a region in which the objective function is quadratic (e.g., Powell 1976). Under such circumstances the Polak–Ribière formula (Eq. 5.63) is preferred. Using the Polak–Ribière formula, if a very

small step is taken at a certain iteration, the next descent direction will be close to the negative gradient (i.e., a steepest descent iteration).

Despite the simple structure of the conjugate-gradient algorithm, its performance is often very good, especially for large-scale problems. Its implementation is quite straightforward and highly amenable to automation using scientific workflow tools (e.g., Zhang et al. 2013).

5.2.3 Hessian-Vector Product

As mentioned in the end of Sect. 5.2.1, for tomography studies involving a large number of misfit measurements, it may not be practical to explicitly evaluate the Jacobian (therefore the Hessian) using the adjoint method. However, as shown in e.g., (Akcelik et al. 2002, 2003), the adjoint method can be used to evaluate the Hessian-vector product $\mathbf{p} = \mathbf{H}\mathbf{q}$ very efficiently, where $\mathbf{H} = \mathbf{K}^T\mathbf{K}$ is the Gauss–Newton approximate Hessian in Eq. 5.22. In this section, I will use the representation theorem (Eq. 3.326) and the adjoint representation theorem (Eq. 3.335) to show how this calculation can be done through one forward simulation and one adjoint simulation per source.

The Hessian-vector product can be written as

$$\mathbf{p} = (\mathbf{K}^T\mathbf{K})\mathbf{q} = \mathbf{K}^T(\mathbf{K}\mathbf{q}) \quad (5.68)$$

and we first consider the term inside the parenthesis $\mathbf{K}\mathbf{q}$. Suppose the vector \mathbf{q} is made up by a fourth-order tensor corresponding to the elastic moduli and we denote it as $q_{jklm}(\mathbf{x})$. The matrix-vector multiplication $\mathbf{K}\mathbf{q}$ involves the spatial integral over $V(\mathbf{x})$ and the summation over j, k, l, m . Bring in the data sensitivity kernel for elastic moduli (Eq. 4.35), we have

$$\mathbf{K}\mathbf{q} = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \left[\int_V dV(\mathbf{x}) \int_0^T d\tau \sum_{jk} \frac{\partial G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t)}{\partial x_k} \sum_{lm} \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} q_{jklm}(\mathbf{x}) \right], \quad (5.69)$$

where I have replaced the RGT with the time-reversed adjoint Green's tensor and used Eq. 3.345 to replace the time-reversed adjoint Green's tensor with the adjoint Green's tensor and I have replaced the temporal convolution operator “*” with the time integral over τ . Apply Eq. 3.297, we have

$$G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) = G_{pj}(\mathbf{x}', t; \mathbf{x}, \tau). \quad (5.70)$$

We can introduce the moment source

$$m_{jk}(\mathbf{x}, \tau) = \sum_{lm} \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} q_{jklm}(\mathbf{x}). \quad (5.71)$$

Bring Eqs. 5.70 and 5.71 into Eq. 5.69, we obtain

$$\mathbf{K}\mathbf{q} = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) \left[\int_V dV(\mathbf{x}) \int_0^T d\tau \sum_{jk} \frac{\partial G_{pj}(\mathbf{x}', t; \mathbf{x}, \tau)}{\partial x_k} m_{jk}(\mathbf{x}, \tau) \right], \quad (5.72)$$

where the term inside the square bracket is in the form of the second term in the representation theorem (Eq. 3.326). If we denote the term inside the square bracket as $w_p(\mathbf{x}', t)$, i.e.,

$$w_p(\mathbf{x}', t) = \int_V dV(\mathbf{x}) \int_0^T d\tau \sum_{jk} \frac{\partial G_{pj}(\mathbf{x}', t; \mathbf{x}, \tau)}{\partial x_k} m_{jk}(\mathbf{x}, \tau), \quad (5.73)$$

then $w_p(\mathbf{x}', t)$ can be obtained by solving the forward elastic wave equation coupled with the homogeneous initial/boundary conditions using the moment source given by Eq. 5.71 as the source. The term $\mathbf{K}\mathbf{q}$ can then be expressed as

$$\mathbf{K}\mathbf{q} = - \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 J_p^{sn}(\mathbf{x}', t) w_p(\mathbf{x}', t) \equiv \mathbf{e} = [e^{sn}], \quad (5.74)$$

which is a vector, denoted as \mathbf{e} , indexed by s, n .

We now consider the term $\mathbf{K}^T \mathbf{e}$, whose calculation is similar to the gradient calculation shown in Sect. 5.2.1. This matrix-vector multiplication involves the summation over s, n . Bring in the expression for the data sensitivity kernels (Eq. 4.35) and replace the RGT with the time-reversed adjoint Green's tensor, we obtain

$$\mathbf{K}^T \mathbf{e} = \sum_{sn} K_{dsn}^{cijklm}(\mathbf{x}) e^{sn} = \sum_s \int_0^T dt \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \frac{\partial}{\partial x_k} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 \sum_n e^{sn} J_p^{sn}(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) \right]. \quad (5.75)$$

If we introduce the adjoint source field

$$[f_p^s]^\dagger(\mathbf{x}', t) = \sum_n e^{sn} J_p^{sn}(\mathbf{x}', t), \quad (5.76)$$

then Eq. 5.75 can be re-written as

$$\mathbf{K}^T \mathbf{e} = \sum_s \int_0^T dt \frac{\partial u_l^s(\mathbf{x}, \tau)}{\partial x_m} \frac{\partial}{\partial x_k} \left[\int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 [f_p^s]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t) \right], \quad (5.77)$$

and the term inside the square bracket is in the form of the first term of the adjoint representation theorem (Eq. 3.335). If we denote the term inside the square bracket as $[w_j^s]^\dagger$, i.e.,

$$[w_j^s]^\dagger(\mathbf{x}, \tau) = \int_V dV(\mathbf{x}') \int_0^T dt \sum_{p=1}^3 [f_p^s]^\dagger(\mathbf{x}', t) G_{jp}^\dagger(\mathbf{x}, \tau; \mathbf{x}', t), \quad (5.78)$$

then $[w_j^s]^\dagger$ can be obtained by solving the adjoint elastic wave equation coupled with the homogeneous adjoint initial/boundary conditions backwards in time. The matrix-vector multiplication $\mathbf{K}^T \mathbf{e}$ can then be expressed as the zero-lag temporal correlation

$$\mathbf{K}^T \mathbf{e} = \sum_s \int_0^T d\tau \frac{\partial u_i^s(\mathbf{x}, \tau)}{\partial x_m} \frac{\partial [w_j^s]^\dagger(\mathbf{x}, \tau)}{\partial x_k} \equiv p_{jklm}(\mathbf{x}) = \mathbf{p}, \quad (5.79)$$

where the vector \mathbf{p} is also made of a fourth-order tensor similar to the elastic moduli and can be written in index form as $p_{jklm}(\mathbf{x})$.

Note that the derivation above can also be extended to density-like \mathbf{q} , in which case $\mathbf{K}\mathbf{q}$ will involve only the spatial integral over $V(\mathbf{x})$ and $w_p(\mathbf{x}', t)$ will be computed using a body-force source.

In (Akcelik et al. 2002, 2003) the Hessian-vector product is used as a matrix-free solution to the Gauss–Newton normal equation. If the Gauss–Newton normal equation is solved using a conjugate-gradient-type iterative algorithm, then each iteration involves the Hessian-vector product, which can be carried out using the forward and adjoint wave-propagation simulations as derived above, instead of explicitly forming the Jacobian matrix. The computational cost for this approach can be quite high, as it replaces cheaper matrix-vector multiplications with more expensive wave-propagation simulations. The benefit is that it does not require large amounts memory to store the complete Jacobian matrix, which can be very large in realistic tomography studies (Lee et al. 2013). The computational cost of this approach is also analyzed and compared with F3DT-SI and F3DT-AW in (Chen et al. 2007a). The Hessian-vector product can also be used in resolution analysis (e.g., Fichtner and Trampert 2011).

5.2.4 Software

The calculation of the gradient of the objective function using the adjoint method involves three basic steps: (1) constructing the adjoint source field for the s -th source (Eq. 5.59); (2) running `awp` using the adjoint source field constructed in step (1) as the source; and (3) running the parallel binary executable `awm`. These three steps need to be repeated for every source in the data set and the results generated by `awm`

need to be conditioned and summed to give the gradient of the objective function at the current iteration. In the following, I will demonstrate how to use F3DWI for the adjoint calculations using the half-space example.

5.2.4.1 Adjoint Source Field

The adjoint source field for the s -th source is also specified as a source line in the source-receiver list file `srlst`. I'll use the explosive source 100001 in our half-space example to demonstrate the specifications.

Suppose we have made broadband cross-correlation delay-time measurements on the P-waves arriving at the x_1 -components of the two receivers: `IN.RC01` and `SF.RC02` (Fig. 3.3). The adjoint source field for 100001 therefore involves the time-reversed WPKs for the broadband cross-correlation delay time measured on the x_1 -component velocity seismograms at the two receiver locations. Here we have assumed that the synthetic seismograms from 100001 to `IN.RC01` and `SF.RC02` have already been computed. The source line for this adjoint source field in `srlst` is shown on line 6 of Listing 2.6 and descriptions about the format were included in Sect. 2.2.4.1. The first column on this line is set to 2, indicating that the body-force sources are applied at receiver locations. The second column is a unique source ID, which can be arbitrary and I set at `100001.adj` to indicate that it is the adjoint source field for 100001. The grid flag on the third column is set to 2, indicating that it is a finite source involving 2 point sources. The fourth column specifies the full-path of the text file `100001.adj` describing the two point sources making up the adjoint source field for 100001. The content of this text file is shown in Listing 2.10 and the format is identical to that in `srlst`.

The source-time function files `IN.RC01.1.adj` and `SF.RC02.1.adj` in the last column of Listing 2.10 contain the time-reversed WPKs for the two broadband cross-correlation delay-time measurements. The generation of the WPK files were discussed in Sect. 4.4.1. If we use the MATLAB function `wpk_bbxcorr.m`, in addition to the five inputs discussed in Sect. 4.4.1 we can specify a sixth input, which is the misfit measurement ΔT^{sn} in seconds (Sect. 4.2.6). The output `wpkt` will then contain the time-reversed WPK for broadband cross-correlation delay time multiplied with ΔT^{sn} in the source-time function file format (Listing 2.7). If we specify a seventh input, which is the broadband cross-correlation amplitude anomaly measurement ΔU^{sn} (Sect. 4.2.6), the output `wpka` will contain the time-reversed WPK for the broadband cross-correlation amplitude anomaly multiplied with ΔU^{sn} in the source-time function file format (Listing 2.7). Since we do not have real misfit measurements, we will assume that $\Delta T^{sn} = 1.0$ at both receivers and specify the sixth input to `wpk_bbxcorr.m` as `1.0`. The output `wpkt` can be directly saved into the text files shown in the last column of Listing 2.10.

If the misfit measurement is made on the radial (transverse) component, the eighth column in Listing 2.10 should be set to the cosine of the angle between the radial (transverse) direction and the x_1 direction of the box coordinate and the ninth

column in Listing 2.10 should be set to the cosine of the angle between the radial (transverse) direction and the x_2 direction of the box coordinate.

5.2.4.2 Adjoint Wavefield

After the adjoint source field is constructed, I need to execute the script `proc_srlst.py` to convert the adjoint source field into the source file that can be read by `awp`. Under the project `JOBS` directory, using the source-receiver list `srlst` as shown in Listing 2.6, I can execute the following command in the terminal

```
proc_srlst.py param_HS.dat 1 8
```

where lines 7 and 8 not shown in Listing 2.6 are the receiver line for `SF.RC02` and a source line for a body-force in the x_1 direction at `SF.RC02`. Even though only line 6 is used in this example, I still included all the other lines in `srlst` when executing `proc_srlst.py` to ensure that the output files generated by `proc_srlst.py` are still consistent with other examples shown previously. After executing this command, the source file for the adjoint source `100001.adj` will be stored inside the `JOBS/srinfo` and named `100001.adj.src`. Snippets of this source file are shown in Listing 2.13.

After the source file is generated, I need to set up the correct directory structure for running `awp`. The script `setup_awp.py`, which has been used before when computing forward wavefields from sources and RGTs from receivers, can also be used for adjoint wavefield simulations. I can execute the following command under project `JOBS` directory in the terminal

```
setup_awp.py param_HS.dat AWF 6 6
```

where the second command-line input has been changed from `FWF` in previous examples to `AWF`. The directory `AWF` will then be generated under the project directory and a sub-directory named `100001.adj` will be created under `AWF`. Then I can run `awp` from inside the sub-directory `100001.adj` in the same way as in previous examples. The resulting wavefield is the adjoint wavefield prior to the time-reversal operation. The time-reversal will be carried out by `awm`, which will be discussed in Sect. 5.2.4.3.

5.2.4.3 Event Kernel

The zero-lag correlation between the forward and the adjoint wavefields for one source is sometimes called the “event kernel”. The summation of all event kernels

gives the gradient of the objective function (Eqs. 5.57 and 5.58). The event kernels for each source can be computed using the parallel binary executable `awm` inside the `bin` directory. To set up the correct environment for running `awm`, a python script `setup_awm.py` is included in the `py` directory.

The syntax for using `setup_awm.py` is very similar to that for `setup_ker.py`, which I discussed in Sect. 4.4.2. To set up the environment for computing the event kernel of 100001 with respect to relative perturbations in P- and S-velocity and density, I can execute the following command under the `JOBS` directory in the terminal

```
setup_awm.py param_HS.dat 2 6 2 1 2 3
```

where the first command-line argument is the parameter file, the second is the line number for the source of the forward wavefield in `srlst` (Listing 2.6), the third is the line number for the adjoint source in `srlst`, the fourth to the last command-line arguments have identical format as the ninth to the last arguments in `setup_ker.py` (Table 4.5, Sect. 4.4.2). In this particular example, the model parameterization flag on the fourth argument is set to 2, which means that I am choosing the isotropic model parameterized using the P- and S-velocity and density. The fifth–seventh argument indicate that the event kernels with respect to relative perturbations in P- and S-velocity and density will be computed. This command will create the `AWM` directory under the project directory and a sub-directory `100001`, which is the ID of the source for the forward wavefield in `srlst`, under the `AWM` directory. Inside the sub-directory `AWM/100001`, it will create a symbolic link `awm`, which points to the binary executable `awm` in the `bin` directory, and a text input file `awm.in`, which will be read by `awm`. The content of `awm.in` for this example is shown in Listing 5.1. The first two lines inside this file have identical format as the first two lines in the kernel input file (Listing 4.2). The third line is the full-path and the prefix for the kernel recording grid generated by the script `set_ker_grd.py` (Sect. 2.2.5). The fourth line is the full-path to the reference structural model (i.e., the value assigned to the key `MEDIA` in the parameter file, line 7, Listing 2.5). The fifth and sixth lines are the full-path and prefix for the forward wavefield generated by 100001 and the adjoint wavefield (before time-reversal) generated by the adjoint source `100001.adj`. The seventh line is the full-path and prefix for the output event kernel files that will be generated by `awm`. The eighth line shows the number of sub-meshes in each dimension (Fig. 2.9). The ninth line shows the number of grid points in each dimension inside each sub-mesh. These numbers are the grid points used in the wave-propagation simulations, not the kernel recording grid points. The tenth line shows the total number of time steps in the forward and adjoint wavefields and the time-sampling interval after temporal decimation by the factor `NTSKP` (line 24, Listing 2.5). The 11th line shows the spatial decimation rate `NXSKP`, `NYSKP` and `NZSKP` (line 21–23, Listing 2.5).

```

1 2
2 1 1 1
3 /home/pochen/wrk/F3DWI/HS/VM/HSGRDREC
4 /home/pochen/wrk/F3DWI/HS/VM/HSGRD
5 /home/pochen/wrk/F3DWI/HS/FWF/100001/output_vlm/SX96PV
6 /home/pochen/wrk/F3DWI/HS/AWF/100001.adj/output_vlm/SX96PV
7 /home/pochen/wrk/F3DWI/HS/AWM/100001/ker
8 2 2 2
9 120 38 90
10 500 0.03
11 2 4 2

```

Listing 5.1 The input file for the event kernel calculation used in the half-space example.

After the correct environment has been set up, I can execute the following command in the terminal from inside `AWM/100001`

```
mpiexec -n 8 -f ~/tools/mpi/share/myhosts awm awm.in
```

and a set of files named as `SX96PVprocID.rt`, where `procID` is the 9-digit sub-mesh (or processor) ID, are created inside the full-path specified on line 6 in Listing 5.1 and a set of files named as `kerprocID` are created inside the full-path specified on line 7 in Listing 5.1. Both sets of files are in binary format. The set of `SX96PVprocID.rt` files contain the adjoint wavefield (after the time-reversal) generated by the adjoint source `100001.adj`. The set of `kerprocID` files contain the event kernels saved on the kernel recording grid points.

If we would like to examine the event kernels, an optional step is to run the sequential binary executable `ker_collect` in the `bin` directory to combine the `kerprocID` files for all the sub-meshes into a binary file that can be loaded into MATLAB and plotted. From inside `AWM/100001`, I can execute the following command in the terminal

```
ker_collect awm.in
```

which uses the same input file `awm.in` (Listing 5.1) as the only command-line argument. For our example, this command will create 3 binary files `ker_001`, `ker_002` and `ker_003`, which are for the event kernels on the full kernel recording grid with respect to relative perturbations in P-velocity, S-velocity and density, respectively. These three binary files can be loaded into MATLAB using the function `read_awm.m`

```
function K=read_awm(kfnam,nkx,nky,nkz)
```

which takes 4 input arguments. The first input `kfnam` is the file name of the binary event kernel file generated by `ker_collect`, e.g., `ker_001`. The second to the fourth input arguments are the total number of kernel recording grid points in each dimension, which can be obtained by dividing the total number of grid points in each dimension used in the wave-propagation simulations (line 2, Listing 2.3) by `NXSKP`,

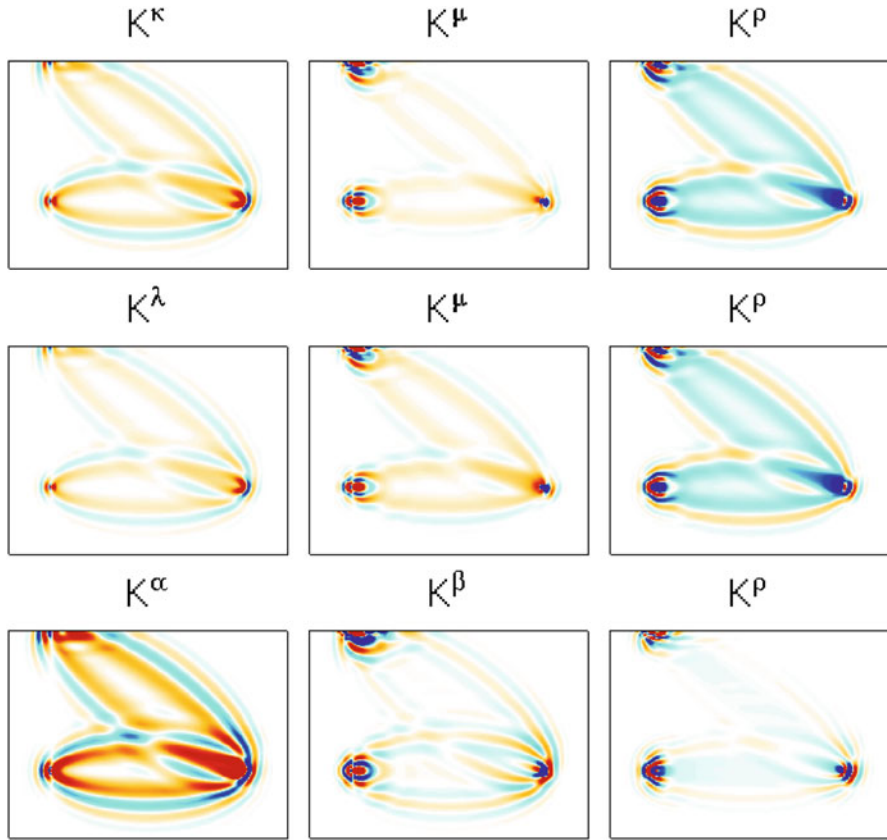


Fig. 5.1 Event kernels of broadband cross-correlation delay-time with respect to relative perturbations in κ , μ , ρ (top row), λ , μ , ρ (center row) and α , β , ρ (bottom row) for 100001 in the half-space example).

NYSKP and NZSKP, respectively. The event kernels for this example are shown in the bottom row of Fig. 5.1. I also calculated the event kernels with respect to the bulk, shear moduli and density (top row, Fig. 5.1) by setting the model parameterization flag to 0 when executing `setup_awm.py` and also the event kernels with respect to the two Lamé parameters and density (center row, Fig. 5.1) by setting the model parameterization flag to 1. As predicted by Eq. 5.45, the event kernel for 100001 shown in Fig. 5.1 is a data-weighted superposition of the data sensitivity kernels for all misfit measurements of 100001. The anisotropic event kernels with respect to the five Love's parameters (Sect. 4.3.5) and with respect to the 21 asymptotic parameters (Sect. 4.3.6) can also be computed by setting the model parameterization flag to 3 and 4, respectively. Figure 5.2 shows the anisotropic event kernels of the broadband cross-correlation delay time with respect to the 21 asymptotic parameters. The spatial distributions of the anisotropic sensitivities are different for the two different propagation paths involved in this example.

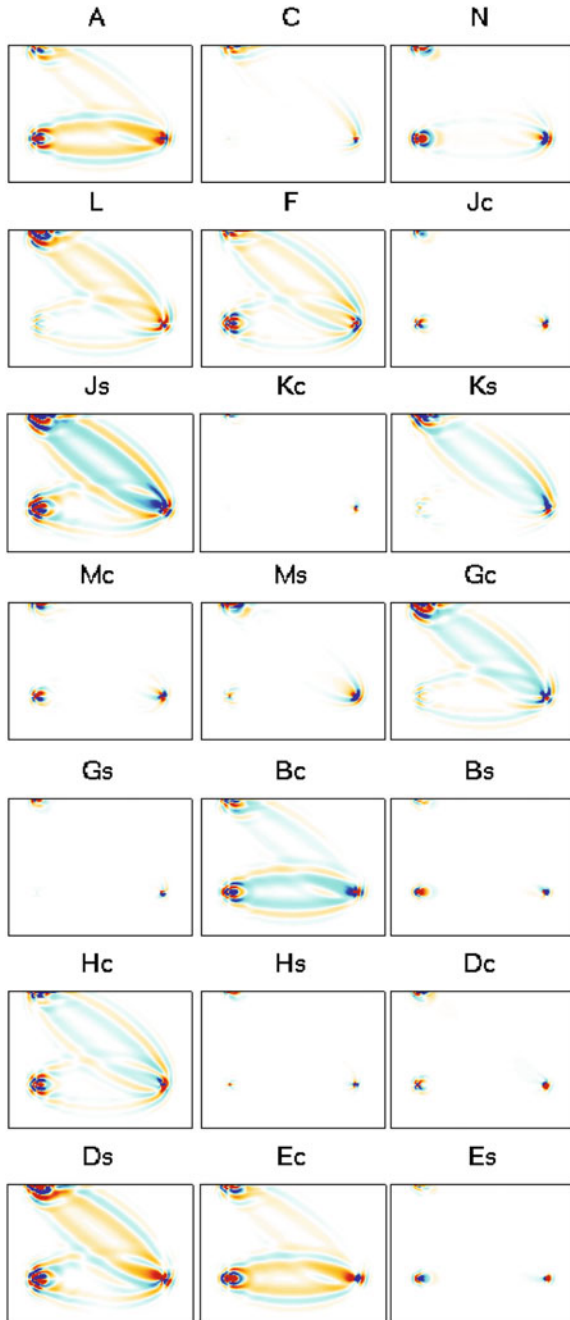


Fig. 5.2 Anisotropic event kernels of broadband cross-correlation delay time with respect to the 21 asymptotic parameters for 100001 in the half-space example.

References

- Akcelik, V., Biros, G., & Ghattas, O. (2002). Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation. *Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, (pp. 1–15).
- Akcelik, V., Bielak, J., Biros, G., Epanomeritakis, I., Fernandez, A., Ghattas, O., Kim, E. J., Lopez, J., O’Hallaron, D., Tu, T., et al. (2003). High resolution forward and inverse earthquake modeling on terascale computers. In *Supercomputing, 2003 ACM/IEEE Conference* (pp. 52–52). IEEE.
- Balay, S., Gropp, W. D., McInnes, L. C., & Smith, B. F. (1997). Efficient management of parallelism in object oriented numerical software libraries. In E. Arge, A. M. Bruaset, & H. P. Langtangen (Eds.), *Modern software tools in scientific computing* (pp. 163–202). New York: Springer Science & Business Media.
- Balay, S., Abhyankar, S., Adams, M., Brown, J., Brune, P., Buschelman, K., Eijkhout, V., Gropp, W. D., Kaushik, D., Knepley, M. G., McInnes, L. C., Rupp, K., Smith, B. F., & Zhang, H. (2014a). *PETSc Users Manual*. Technical Report ANL-95/11—Revision 3.5, Argonne National Laboratory.
- Balay, S., Abhyankar, S., Adams, M., Brown, J., Brune, P., Buschelman, K., Eijkhout, V., Gropp, W. D., Kaushik, D., Knepley, M. G., McInnes, L. C., Rupp, K., Smith, B. F., & Zhang, H. (2014b). PETSc Web page. <http://www.mcs.anl.gov/petsc>. accessed 27 June 2015.
- Bamberger, A., Chavent, G., & Lailly, P. (1977). Une application de la théorie du contrôle à un problème inverse de sismique. *Annales de géophysique*, 33, 183–199.
- Bamberger, A., Chavent, G., Hemon, C., & Lailly, P. (1982). Inversion of normal incidence seismograms. *Geophysics*, 47(5), 757–770.
- Baur, O., & Austen, G. (2004). A parallel iterative algorithm for large-scale problems of type potential field recovery from satellite data. *Proceedings of the Joint CHAMP/GRACE Science Meeting, 1, 2*.
- Beale, E. (1972). A derivation of conjugate gradients. In F. A. Lootsma (Ed.), *Numerical methods for nonlinear optimization*, pp. 39–43. London: Academic Press.
- Bengtsson, L., Ghil, M., & Källén, E. (1981). *Dynamic meteorology: Data assimilation methods*. Applied mathematical sciences. New York: Springer.
- Bennett, A. (1992). *Inverse methods in physical oceanography*. Arnold and Caroline Rose Monograph Series of the American So. Cambridge: Cambridge University Press.
- Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26(1), 99–108.
- Biegler, L. (2003). *Large-scale PDE-constrained optimization*. Lecture Notes in Computational Science and Engineering. Springer Berlin Heidelberg.
- Bleistein, N., Cohen, J., & Stockwell, J. (2001). *Mathematics of multidimensional seismic imaging, migration, and inversion*. Interdisciplinary Applied Mathematics. New York: Springer.
- Campbell, S., & Meyer, C. (2009). *Generalized inverses of linear transformations*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).
- Chen, P. (2011). Full-wave seismic data assimilation: Theoretical background and recent advances. *Pure and Applied Geophysics*, 168(10), 1527–1552.
- Chen, P., Jordan, T., & Zhao, L. (2005). Finite-moment tensor of the 3 september 2002 yorba linda earthquake. *Bulletin of the Seismological Society of America*, 95(3), 1170–1180.
- Chen, P., Jordan, T. H., & Zhao, L. (2007a). Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods. *Geophysical Journal International*, 170(1), 175–181.
- Chen, P., Zhao, L., & Jordan, T. H. (2007b). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, 97(4), 1094–1120.
- Dahlen, F., Hung, S., & Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes-I. Theory. *Geophysical Journal International*, 141, 157–174.

- Dai, Y.-H., & Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10(1), 177–182.
- Dai, Y., & Yuan, Y. (2001). An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research*, 103(1–4), 33–47.
- Daley, R. (1993). *Atmospheric data analysis*. Cambridge Atmospheric and Space Science Series. Cambridge: Cambridge University Press.
- deGroot-Hedlin, C., & Constable, S. (1990). Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data. *Geophysics*, 55(12), 1613–1624.
- Dennis, J. E., Jr., Gay, D. M., & Walsh, R. E. (1981). An adaptive nonlinear least-squares algorithm. *ACM Transaction on Mathematical Software*, 7(3), 348–368.
- Devaney, A. (1981). Inverse-scattering theory within the Rytov approximation. *Optics Letters*, 6(8), 374–376.
- Fichtner, A., & Trampert, J. (2011). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2), 775–798.
- Fichtner, A., Kennett, B., Igel, H., & Bunge, H. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3), 1703–1725.
- Fletcher, R. (2013). *Practical methods of optimization*. Chichester, England: Wiley.
- Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal*, 7(2), 149–154.
- Got, J.-L., Fréchet, J., & Klein, F. W. (1994). Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. *Journal of Geophysical Research*, 99(B8), 15375–15386.
- Hager, W. W., & Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization*, 16(1), 170–192.
- Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau Standards*, 49(6), 409–436.
- Huang, H., Wang, L., Lee, E.-J., & Chen, P. (2012). An MPI-CUDA implementation and optimization for parallel sparse equations and least squares (LSQR). *Procedia Computer Science*, 9, 76–85.
- Huang, H., Dennis, J. M., Wang, L., & Chen, P. (2013). A scalable parallel LSQR algorithm for solving large-scale linear system for tomographic problems: A case study in seismic tomography. *Procedia Computer Science*, 18, 581–590.
- Iyer, H., & Hirahara, K. (1993). *Seismic tomography: Theory and practice*. London: Chapman & Hall.
- Jordan, T. H., & Sverdrup, K. A. (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific. *Bulletin of the Seismological Society of America*, 71(4), 1105–1130.
- Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge: Cambridge University Press.
- Käuff, P., Fichtner, A., & Igel, H. (2013). Probabilistic full waveform inversion based on tectonic regionalization—Development and application to the Australian upper mantle. *Geophysical Journal International*, 193(1), 437–451.
- Lee, E.-J., Huang, H., Dennis, J. M., Chen, P., & Wang, L. (2013). An optimized parallel LSQR algorithm for seismic tomography. *Computers & Geosciences*, 61, 184–197.
- Lee, E.-J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research*, 119(8), 6421–6451.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *The Quarterly of Applied Mathematics*, 2(2), 164–168.
- Liu, Q. (2006). *Spectral-element simulations of 3-D seismic wave propagation and applications to source and structural inversions*. PhD thesis, California Institute of Technology.

- Liu, Q., & Tromp, J. (2006). Finite-frequency kernels based on adjoint methods. *Bulletin of the Seismological Society of America*, 96(6), 2383–2397.
- Liu, J.-S., Liu, F.-T., Liu, J., & Hao, T.-Y. (2006). Parallel LSQR algorithms used in seismic tomography. *Chinese Journal of Geophysics-CH*, 49(2), 483–488.
- Malanotte-Rizzoli, P. (1996). *Modern approaches to data assimilation in ocean modeling*. Elsevier Oceanography Series. Elsevier Science.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2), 431–441.
- Montgomery, D., Peck, E., & Vining, G. (2012). *Introduction to linear regression analysis*. Wiley Series in Probability and Statistics. Hoboken, New Jersey: Wiley.
- Nolet, G. (1985). Solving or resolving inadequate and noisy tomographic systems. *Journal of Computational Physics*, 61(3), 463–482.
- Paige, C. C., & Saunders, M. A. (1982). LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transaction on Mathematical Software*, 8(1), 43–71.
- Polak, E., & Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique Et Analyse Numérique*, 3(R1), 35–43.
- Powell, M. J. D. (1976). Some convergence properties of the conjugate gradient method. *Mathematical Programming*, 11(1), 42–49.
- Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12(1), 241–254.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3), 888–901.
- Pratt, R. G., Shin, C., & Hick, G. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2), 341–362.
- Sambridge, M. (1990). Non-linear arrival time inversion: Constraining velocity anomalies by seeking smooth models in 3-d. *Geophysical Journal International*, 102(3), 653–677.
- Sieminski, A., Liu, Q., Trampert, J., & Tromp, J. (2007). Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods. *Geophysical Journal International*, 168(3), 1153–1174.
- Sorenson, H. (1969). Comparison of some conjugate direction procedures for function minimization. *Journal of the Franklin Institute*, 288(6), 421–441.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180(1), 433–462.
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8), 1259–1266.
- Tarantola, A. (1988a). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128(1/2), 365–399.
- Tarantola, A. (1988b). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128, 365–399.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Philadelphia: Society for Industrial and Applied Mathematics.
- Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1), 195–216.
- Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, 90(6), 1353–1368.
- Waldhauser, F., & Ellsworth, W. L. (2002). Fault structure and mechanics of the Hayward Fault, California, from double-difference earthquake locations. *Journal of Geophysical Research*, 107(B3), ESE–3.
- Woodward, M. (1992a). A qualitative comparison of the first-order Born and Rytov approximations. SEP-60: Stanford Exploration Project, 203–214.

- Woodward, M. J. (1992b). Wave-equation tomography. *Geophysics*, 57(1), 15–26.
- Wu, R.-S., & Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52(1), 11–25.
- Wunsch, C. (1996). *The ocean circulation inverse problem*. Cambridge: Cambridge University Press.
- Zhang, H., & Thurber, C. H. (2003). Double-difference tomography: The method and its application to the Hayward fault, California. *Bulletin of the Seismological Society of America*, 93(5), 1875–1889.
- Zhang, X., Chen, P., & Pullammanappallil, S. (2013). Automating adjoint wave-equation travel-time tomography using scientific workflow. *Earthquake Science*, 26(5), 331–339.
- Zhao, L., & Jordan, T. H. (2006). Structural sensitivities of finite-frequency seismic waves: A full-wave approach. *Geophysical Journal International*, 165(3), 981–990.
- Zhao, L., Jordan, T. H., & Chapman, C. H. (2000). Three-dimensional Fréchet differential kernels for seismic delay times. *Geophysical Journal International*, 141(3), 558–576.
- Zhao, L., Jordan, T. H., Olsen, K. B., & Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, 95(6), 2066–2080.
- Zupanski, D., & Zupanski, M. (2006). Model error estimation employing an ensemble data assimilation approach. *Monthly Weather Review*, 134, 1337–1354.

Chapter 6

CVM-S4.26

Southern California is an ideal natural laboratory for testing the F3DWI methodology and software that I have discussed in previous chapters. In this chapter I will discuss the 3D seismic velocity model that En-Jui Lee and I have constructed for the Southern California crust, CVM-S4.26. Some background information about the different CVMs in Southern California and some model validation results have been documented in Sect. 2.3. The technical details about our entire inversion process has been documented in (Lee et al. 2014b). Some more extensive model validation tests have been documented in (Lee et al. 2014a, b). In this chapter, I will mainly focus on descriptions of the various features in CVM-S4.26. Some features have been imaged before in previous localized studies conducted by other research groups and those features are now rendered in full-3D at small scale lengths in CVM-S4.26. Some features, especially in the mid- and lower-crust, are new and their interpretations are under study.

6.1 Background

The Southern California continental margin has experienced a complex tectonic evolution. Through geologic times, it has undergone four different continental margin configurations (e.g., Ernst and Rubey 1981; Burchfiel et al. 1992): the Atlantic-type rifting margin in late Precambrian and early Paleozoic; the Japanese-type subduction margin with off-shore island arcs in late Paleozoic and early Mesozoic; the Andean-type classic subduction margin in late Mesozoic and early Cenozoic; the present-day California-type transform margin dominated by the dextral San Andreas Fault system. The long, intense tectonic activities have led to highly complex crustal geologic structures. The oldest rock found in this region is over 1.5 billion years old and has undergone numerous cycles of metamorphism. Some recent volcanic activities have produced highly fresh basalt, which just started to weather. The materials in sedimentary rocks in this region can have highly diverse origins. The conformance to known geologic features is therefore a critical test of any seismic velocity model for this region constructed through tomography.

Southern California is one of the best instrumented regions throughout the world. Instruments for recording local earthquakes, such as the Wood-Anderson torsion seismometer and the Benioff short-period seismometer, were installed at a few sites in this region in the 1920's and 1930's. The number of instrumented sites increased slowly from the 1930's through the 1970's. Several large earthquakes occurred in this region during this period of time, which caused damages and also highlighted the need for better instrumentation. The Southern California Seismic Network (SCSN), a cooperative seismic network now part of the California Integrated Seismic Network (CISN), was formed in the 1970's. The SCSN was composed of about 250 short-period instruments through much of the 1980's and 1990's. It was gradually supplemented by a network of digital stations with both broadband and strong-motion sensors under the TERRAscope program (e.g., Kanamori et al. 1991) and the TriNet program (e.g., Hauksson et al. 2001) in the 1990's and 2000's. When I started working on the full-3D tomography for Southern California crustal structure in 2007, the total number of broadband digital seismometers accessible through SCSN reached more than 200. The SCSN has been providing a wealth of high-fidelity waveform recordings of ground motion during earthquakes. More than 38,000 seismograms with high signal-to-noise ratios from local small to medium-sized earthquakes, which are common in Southern California, were used in deriving CVM-S4.26 (Lee et al. 2014b). The continuous-time seismic records provided by SCSN allow extraction of approximate Green's functions between two stations through cross-correlations of ambient seismic noises (e.g., Shapiro 2004; Shapiro et al. 2005; Ma et al. 2008). The ambient-noise Green's functions (ANGF) allow us to fully exploit the high density of SCSN and are highly complementary to the earthquake waveform recordings. More than 12,000 ANGFs were used in deriving CVM-S4.26 (Lee et al. 2014b).

6.1.1 *Convergence Rate*

The full-3D tomography that En-Jui Lee and I have carried out and documented in (Lee et al. 2014b) involves an iterated cycle of wave-propagation simulation, misfit measurement, kernel calculation and inversion. The number of waveforms, the time-frequency extent of those waveforms, the types of the misfit measurements and the inversion algorithm (i.e., F3DT-AW or F3DT-SI) can be different from iteration to iteration. The CMT solutions for the earthquake sources used in our tomography study were inverted at selected iterations (Lee et al. 2011). In (Lee et al. 2014b) we used the term "tomographic navigation" (Fig. 2 in Lee et al. 2014b) to represent this hands-on iterative process.

The first two iterations carried out by myself in 2007–2008 were based on the F3DT-SI method. The starting model was CVM-S4. The total number of observed waveforms that the synthetics were capable of fitting were quite limited and the size of the linear system in the Gauss-Newton normal equation was quite small. I used the LSQR code from the PETSc library to solve the Gauss-Newton normal equation

as I did in (Chen et al. 2007). In early 2009, En-Jui joined my research group at University of Wyoming and we started to work on the inversion together.

As the structural model improved, the total number of waveforms that the synthetics were capable of fitting increased and the linear system of the Gauss-Newton normal equation was becoming larger. The LSQR code from the PETSc library was not scalable at large core-counts and we started to develop a more scalable parallel LSQR code with our collaborators: Liqiang Wang and his students from the Computer Science Department at University of Wyoming, John Dennis from the National Center for Atmospheric Research (NCAR) and Galen Arnold from the National Center for Supercomputing Applications (NCSA). In the meanwhile, we switched from the F3DT-SI method to the F3DT-AW method in our inversion since the F3DT-AW method does not require solving a large linear system (Sect. 5.2.2). From iteration 3 to iteration 20 we used the AW method, which provided moderate and consistent misfit reduction (Fig. 2 in Lee et al. 2014b) and also improvements in our structural model for most of the early iterations. The disk storage cost and the I/O overhead were negligible during the F3DT-AW iterations.

The improvements in waveform fits we could obtain from the F3DT-AW iterations decreased quickly after iteration 16 (Fig. 2 in Lee et al. 2014b). The relative misfit reduction compared with the previous iteration was only about 0.8 % by iteration 20 and the small improvements in the structural model might not justify the computational cost for further iterations. One possibility for us at that point was to stop the iterations. In fact, the accumulative misfit reduction compared with the starting model CVM-S4 was already close to 70 % at iteration 20. However, our effort in developing a scalable parallel LSQR code (SPLSQR) was starting to bear fruits and we published two conference papers in computer science using test data sets from our Southern California inversion (Huang et al. 2012, 2013). So instead of stopping the iterations, we decided to switch back to the F3DT-SI method and try out our scalable parallel LSQR code SPLSQR using our full Southern California data set (Sect. 5.1.3).

Iteration 21 to 26 were carried out using the F3DT-SI method again (Fig. 2 in Lee et al. 2014b). The linear system of the Gauss-Newton normal equation was more than 450 times larger than the one I solved in (Chen et al. 2007). The performance of SPLSQR was highly impressive and we were able to achieve nearly two orders-of-magnitude performance improvement when comparing with the LSQR code in the PETSc library (Lee et al. 2013). The relative waveform misfit reduction increased from about 0.8 % in iteration 20 to about 14 % in iteration 21 (Fig. 2 in Lee et al. 2014b). By iteration 26, the relative misfit reduction compared with iteration 25 was about 3 % and the accumulative misfit reduction reached nearly 80 %. The 18 F3DT-AW iterations accounted for about 40 % accumulative misfit reduction and the 6 F3DT-SI iterations accounted for the other 40 %.

However, one cannot draw the conclusion that three F3DT-AW iterations are equivalent to one F3DT-SI iteration based on these numbers. The first two F3DT-SI iterations accounted for nearly 30 % accumulative misfit reduction. If I had carried out the first two iterations using the F3DT-AW method instead of the F3DT-SI method, the accumulative misfit reduction could have been comparable to 30 %.

The reason is that when the starting model is far from the optimal, the Gauss-Newton approximate Hessian might not be a satisfactory approximation to the exact Hessian and the convergence rate of the Gauss-Newton algorithm approaches that of the steepest descent algorithm, which is identical to the first iteration of the conjugate-gradient algorithm as discussed in Sect. 5.2.2.

The sharp increase in relative misfit reduction at iteration 21 (Fig. 2 in Lee et al. 2014b) could be due to the faster convergence rate of the Gauss-Newton algorithm. During those later iterations, we were primarily dealing with small eigenvalues of the Hessian, which might be associated with small-scale features in the structural model. The conjugate-gradient algorithm tends to get one eigenvalue (several if they are very closely clustered together) at each iteration. The contribution of one small eigenvalue to the overall misfit reduction could be quite small, which makes it difficult to decide if we should continue to iterate. The Gauss-Newton algorithm tends to get a group of eigenvalues at each iteration. The combined contribution from many small eigenvalues to the overall misfit reduction could be significant. The F3DT-SI method may have allowed us to extract the small-scale features in CVM-S4.26 more easily.

En-Jui defended his PhD thesis with flying colors and graduated in December 2013 and is now a post-doctoral scholar at the Southern California Earthquake Center (SCEC) working on further improving the crustal structure in central California. At the current stage my own effort is shifting towards updating the crustal structure model throughout the state-wide California. CVM-S4.26 is now incorporated into the starting model of the state-wide inversion. In future iterations, anisotropy, and possibly anelastic attenuation, will be included into the inversion.

6.1.2 Data Functionals

Another important aspect of our tomographic navigation process was the selection of waveforms and the type of data functionals at different iterations. All earthquake seismograms were band-pass filtered using a Butterworth filter with corners at 0.02 Hz and 0.2 Hz. The average spectrum of all ANGFs show two peaks at around 0.065 Hz and 0.14 Hz, corresponding to the primary and secondary microseisms. In the inversion documented in (Lee et al. 2014b), we treated the ANGFs as if they were generated by virtual vertical point body-force sources acting at receiver locations. To simplify the modeling of the source-time functions for the ANGFs, we separated the two microseism peaks through filtering. Each ANGF was filtered using two band-pass Butterworth filters, one with corners at 0.03 Hz and 0.1 Hz, the other with corners at 0.11 Hz and 0.18 Hz. The filtered ANGFs were called ANGF-low (i.e., filtered with the lower pass band) and ANGF-high (i.e., filtered with the higher pass band) in (Lee et al. 2014b). The ANGF-low and ANGF-high were treated as two independent data sets in our inversion.

In early iterations, we mainly used broadband cross-correlation delay time (Sect. 4.2.6) measured on well-recognized body-wave phases (e.g., P/Pnl and S

waves) from earthquakes and some well-isolated surface-wave arrivals from earthquakes and ANGFs. For some strongly dispersed surface waves, we also used the frequency-dependent group-delay time (Sect. 4.2.7). Both types of data functionals are relatively insensitive to errors in focal mechanisms of the earthquake sources (e.g., van Leeuwen and Mulder 2010; Xu et al. 2013) and usually do not require corrections for possible cycle-skipping errors. In early iterations (i.e., when the structural model was poor) and in some regions (e.g., the southern San Joaquin Basin), the kinematic shift between synthetic and observed waveforms could be quite large (up to several tens of seconds for propagation paths of 200 km and longer) and those two types of data functionals allowed us to invert those large kinematic shifts in a robust way efficiently.

As the structural and earthquake source models improved in later iterations, we switched to frequency-dependent phase delay time misfits (Sect. 4.2.7) measured on a much wider class of seismic phases. As discussed in Sect. 4.2.8, the better linearization provided by the Rytov approximation for the complex phase of a transmitted wave assumes that the phase misfits are measured one wave at a time. If the phase misfits are measured on a mixture of arrivals, the advantage of the Rytov approximation can disappear. Therefore to improve the linearity of the inversion, it is important to segment the seismograms and isolate individual waveforms before making the frequency-dependent phase-delay misfit measurements.

Many automated waveform selection algorithms have been developed in the past (e.g., Ritsema and van Heijst 2002; Sigloch and Nolet 2006; Lawrence and Shearer 2008). They usually pick time windows around certain seismic phases predicted by seismic ray theory in a one-dimensional structural model. For F3DT, the waveform selection process can no longer be limited to ray-theoretic predictions of conventional seismic phases and needs to be extended to accommodate waveforms generated through complex wave propagation processes in a highly heterogeneous 3D earth structural model. An automated waveform selection algorithm for F3DT based on the time-domain short-term average/long-term average ratio (STA/LTA) curve and a set of waveform selection criteria was presented in (Maggi et al. 2009). A purely time-domain waveform selection algorithm may have difficulty in separating waves arriving at overlapping time windows. En-Jui and I have developed a time-frequency domain seismogram segmentation and waveform selection algorithm based on the continuous wavelet transform and the topological watershed transform (Lee and Chen 2013). This algorithm allowed us to automatically separate waves with disjoint supports in either the time domain or the frequency domain or the time-frequency domain and helped us to improve the convexity of the objective functions through the iterations.

6.1.3 Model Evaluation

For nonlinear inverse problems, we do not have a general theory for the appraisal of the estimated model. If the model space is sufficiently small and the solution of the forward problem is sufficiently fast, we can adopt a stochastic sampling approach to investigate the resolution and the reliability of the estimated model (e.g., Tarantola 2005). But for F3DT in general, the model space is too large and the solution of the forward problem is too expensive at the current stage of the computing technology. The dimension of the model space can be reduced significantly if regionalization is adopted in the parameterization and Monte-Carlo-type stochastic searching algorithms can be adopted for both the estimation of the model and its appraisal (e.g., Käuffl et al. 2013). Such an approach might also be applicable in Southern California. As shown in Fig. 4 in (Lee et al. 2014b), the Southern California crust is composed of only a handful of blocks and within each block, the structure is approximately one-dimensional (i.e., depth-dependent), and the different blocks are bounded by either active faults or recognizable boundaries.

In (Lee et al. 2014b) we adopted a pragmatic approach for model assessment by comparing 2D cross-sections extracted from the 3D CVM-S4.26 with 2D velocity profiles obtained from previous active-source studies conducted by other research groups. Among the 14 cross-sections we have compared with, CVM-S4.26 shows better agreement with those active-source studies than CVM-S4 and CVM-H11.9 in general. Another important feature of CVM-S4.26 is its conformance with known fault structures, even though the faults were not built into our starting model CVM-S4 or our model parameterization. CVM-S4.26 also shows good correlation with local geology, different types of geophysical field measurements (e.g., gravity and magnetic surveys) and receiver-function studies. It was the consistency with so many independent evidences all combined that gave us the confidence in CVM-S4.26.

We note that model validations based on the fits to earthquake waveforms not included in the inversion, as done in e.g., (Tape et al. 2010), may not be as “independent” as we may expect. Even though those waveform data were not used in the inversion, the source-receiver paths (thereby the sampling pattern through the structure) might not be truly independent of the inversion data because small to medium-sized earthquakes tend to repeat in small clusters on faults with similar locations and focal mechanisms and the receivers are usually fixed in space. In the inversion for CVM-S4.26, we made extensive effort to improve the spatial coverage of our earthquake data by selecting earthquake seismograms with good signal-to-noise ratio and source-receiver paths that provide a fairly even coverage of the entire modeling volume, which leaves very little room for testing CVM-S4.26 using truly independent earthquake seismograms. With that being said, CVM-S4.26 does provide much better fits to waveforms not used in our inversion than CVM-S4 and CVM-H11.9. The waveform fitting tests documented in (Lee et al. 2014a) and Sect. 2.3.3 are based on the two recent earthquakes not used in deriving CVM-S4.26 or CVM-H11.9. The wave-train used for computing the RWM (Eq. 2.126) extends

from the first-arriving P-wave to the end of the surface wave, which also includes waveforms not used in the inversion. Recently, we have applied the same analysis to about 28,000 earthquake seismograms with relatively poor signal-to-noise ratio. The RWM values are generally worse for all three CVMs due to the poorer signal-to-noise ratio, but CVM-S4.26 still outperforms the other two CVMs.

Since the Gauss-Newton approximate Hessian (i.e., the Jacobian) is explicitly constructed in the F3DT-SI iterations, it might be tempting to use it for resolution analysis. However, we need to be highly cautious when following this path. First, the approximate Hessian is computed for a single iteration, therefore it is a local estimate, not a global estimate. The full inversion involving all 26 iterations is nonlinear and a local Hessian may not tell us much about the uncertainty of the full nonlinear inversion. Second, the objective function of the 26th iteration was defined in terms of phase misfits measured on selected waveforms (Sect. 4.2.7), not the waveform differences (Sects. 4.2.4 and 4.2.5). But the fact is that CVM-S4.26 can fit not only the phase, but also the waveforms of more than 50,000 earthquake seismograms and ANGFs from the first-arriving P wave to 30 s after the surface wave almost wiggle-by-wiggle (Lee et al. 2014a, b). The nonlinear constraints imposed by the waveforms, both used and not used in our inversion, are generally much stronger than the linear (or quasi-linear) constraints imposed by the phase of selected waveforms at a single iteration. The smallest checker size we were able to reliably recover in our checkerboard tests for the 26th iteration was around 10 km (Appendix F in Lee et al. 2014b). However, in our inversion some sharp velocity gradients with scale-lengths less than 10 km emerged through the 26 iterations and are consistent with known fault structures.

In (Lee et al. 2014b), the resolution analysis was limited to the checkerboard tests using the Gauss-Newton approximate Hessian computed for the broadband cross-correlation delay time and frequency-dependent phase-delay time for all the earthquake waveforms and ANGF waveforms used in the 26th iteration. The primary purpose for the checkerboard tests was to show that by including the ANGFs and using the frequency-dependent phase-delay time we were able to improve the resolution of the 26th iteration, perhaps significantly, comparing with inversions using the broadband cross-correlation delay time and earthquake waveforms only. However, the checkerboard tests for the 26th iteration may not provide a reliable estimate of the true resolution of the entire nonlinear inversion process involving all iterations.

6.2 Model Description

Map-view plots of the S-velocity, P-velocity and Poisson's ratio at 12 different depths evenly distributed between 2 km and 24 km for CVM-S4, CVM-S4.26 and CVM-H11.9 are shown in Figs. 6.1, 6.2 and 6.3. The map-view plots of S-velocity at 2-km, 10-km and 20-km depths are identical to those shown in Fig. 6.3a of (Lee et al. 2014b). The map-view plots of S-velocity at other depths and the map-view

plots for P-velocity and Poisson's ratio for CVM-S4.26 have not been published before. The entire modeling volume in our inversion for CVM-S4.26 was about 768-km long, 496-km wide and 50-km deep. It was discretized into a uniform mesh with 500-m grid spacing.

6.2.1 CVM-S4.26 and CVM-S4 Comparison

In general, the differences between CVM-S4.26 and our starting model CVM-S4 are large. The most visible difference lies in the Southern San Joaquin Basin, where CVM-S4 is relatively unconstrained. The maximum S-velocity perturbation with respect to CVM-S4 reaches 40 % inside the basin. Such a large perturbation was not obtained from a single iteration, but accumulated through many iterations, which suggests that our iterative tomographic navigation process was effective in accounting for the nonlinearity in structural inversions. At mid- and lower-crustal depths, CVM-S4.26 generally enhances the large-scale velocity gradient already existing in CVM-S4 and also adds many small-scale heterogeneities that are absent in CVM-S4. In regions with relatively poor data coverage, especially south of the US-Mexico border, south of the Imperial Valley, close to the offshore southwest boundary of our tomography box, at depths larger than about 26 km, the differences between CVM-S4.26 and the starting model CVM-S4 are generally small. As discussed in Sect. 5.1.4, in the absence of data constraints, we explicitly enforced the model perturbation to be smooth and small during our inversion through model regularization.

6.2.2 CVM-S4.26 and CVM-H11.9 Comparison

In general, the correlation between CVM-S4.26 and CVM-H11.9 is higher at shallower depths. In regions with similar data coverage (e.g., the Los Angeles Basin and surrounding areas), both models show similar structural patterns (e.g., at 2-km depth in Figs. 6.1, 6.2 and 6.3). However, CVM-S4.26 generally has stronger velocity variations (i.e., blue is bluer and red is redder in CVM-S4.26 than in CVM-H11.9) and more small-scale heterogeneities. Generally speaking, CVM-S4.26 is a rougher model with more structural details and CVM-H11.9 is a smoother model.

6.2.2.1 Starting Model Differences

Some of the differences between CVM-S4.26 and CVM-H11.9 can be attributed to the differences in the starting models. CVM-H11.9 was derived from an earlier version of the CVM-H series maintained at SCEC. The large-scale background

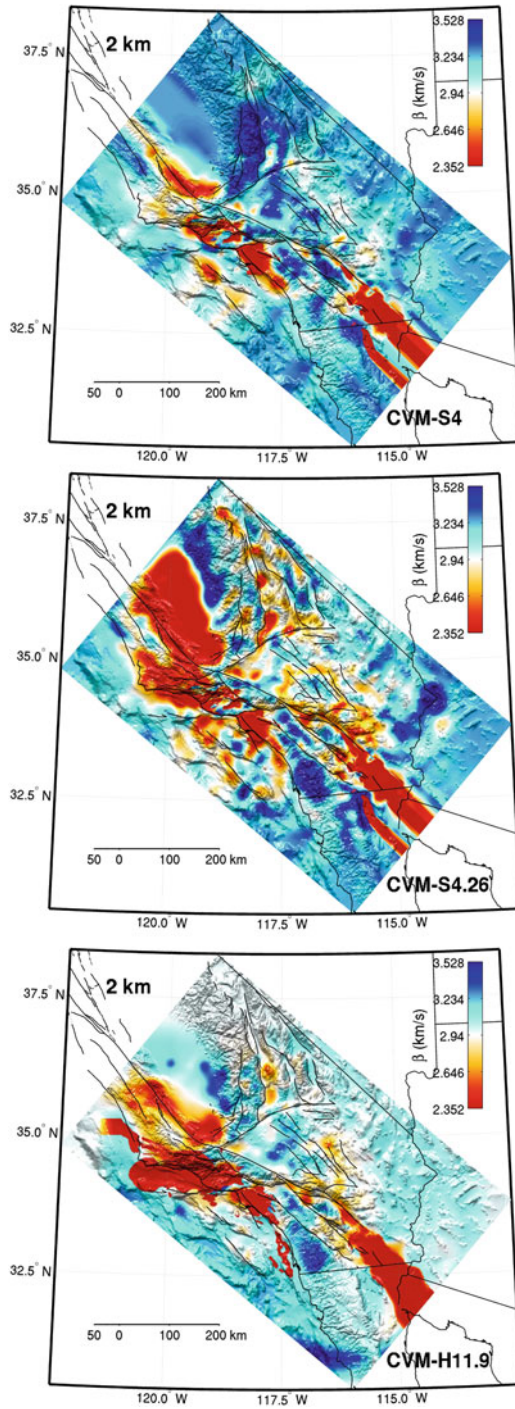


Fig. 6.1 Map-view of S-velocity at 2-km depth in CVM-S4 (top), CVM-S4.26 (center) and CVM-H11.9 (bottom). Warm colors indicate lower velocities and cold colors indicate higher velocities. Black solid lines: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages.

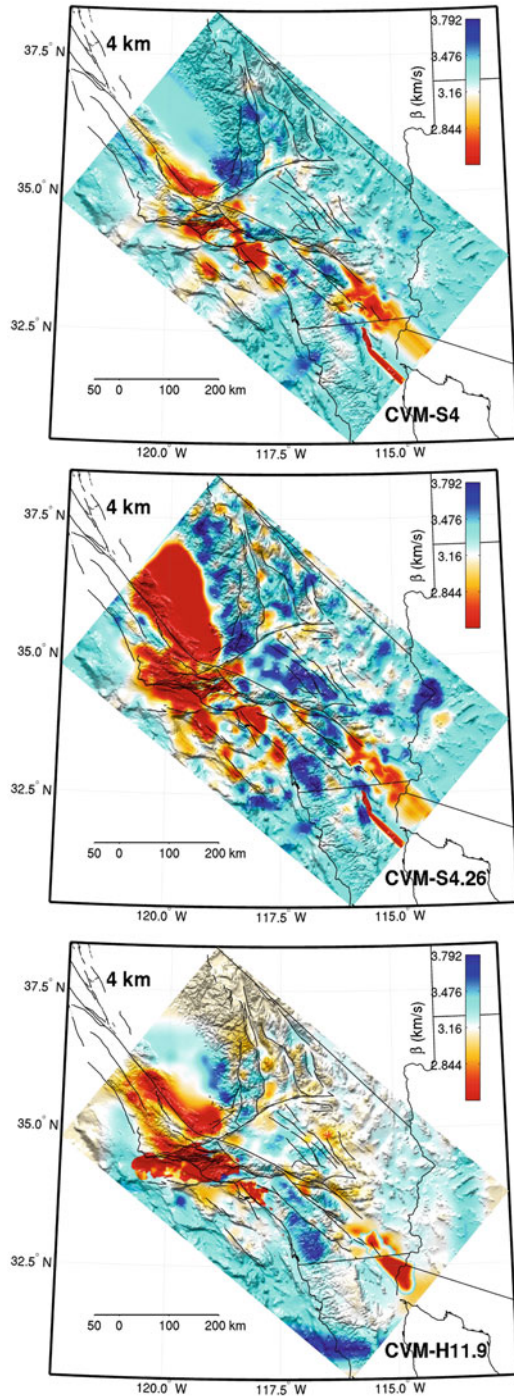


Fig. 6.1 (continued)

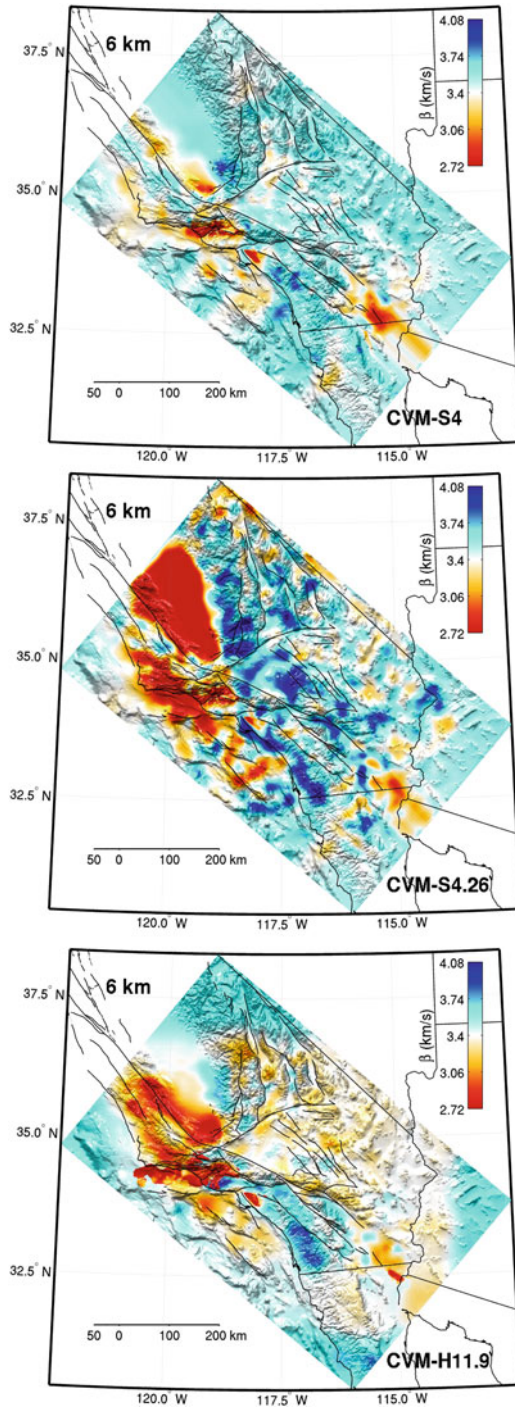


Fig. 6.1 (continued)

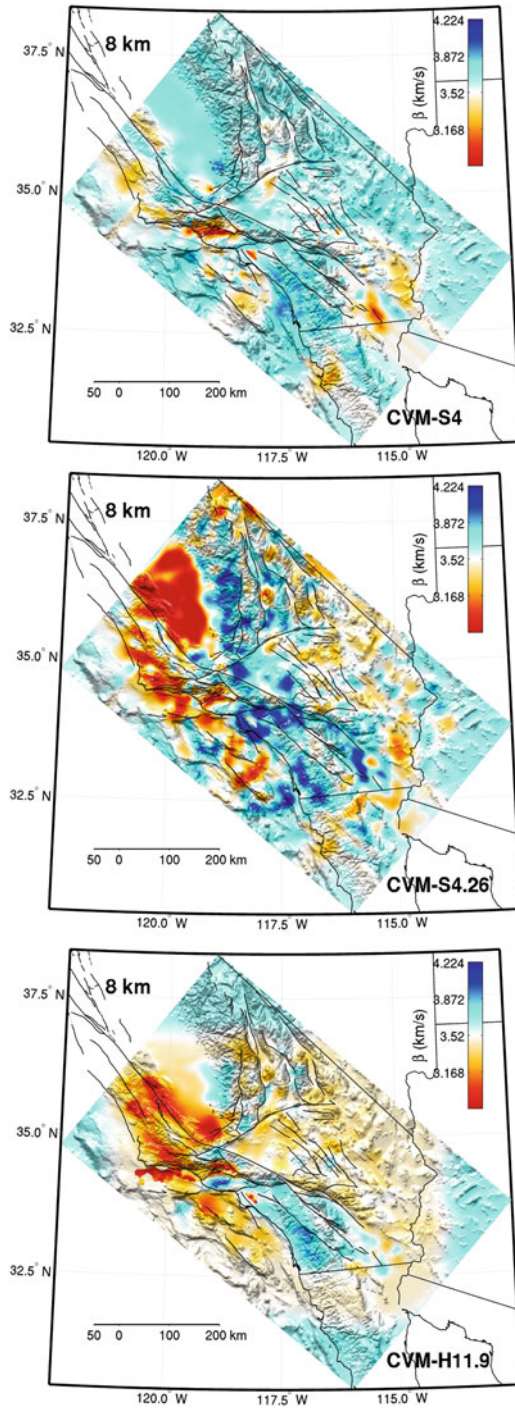


Fig. 6.1 (continued)

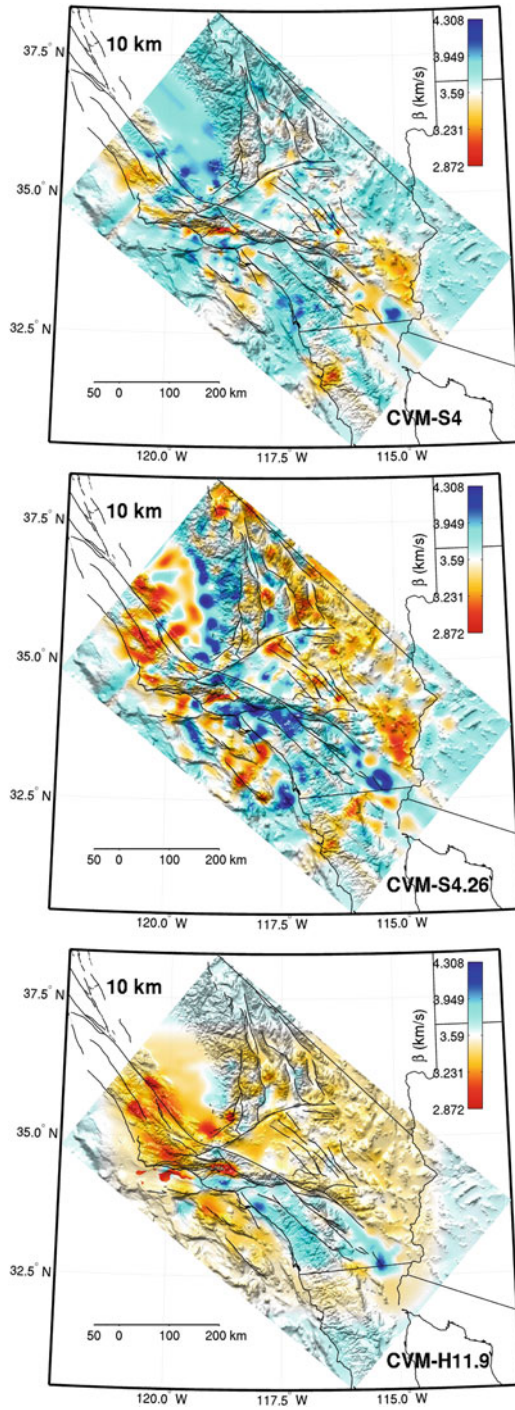


Fig. 6.1 (continued)

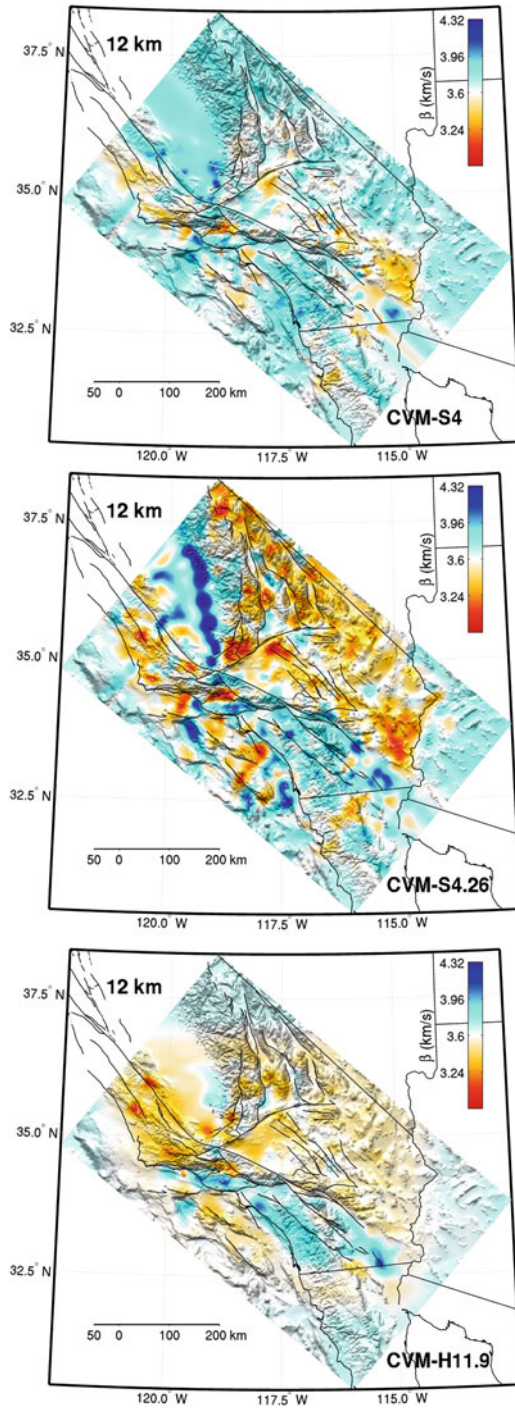


Fig. 6.1 (continued)

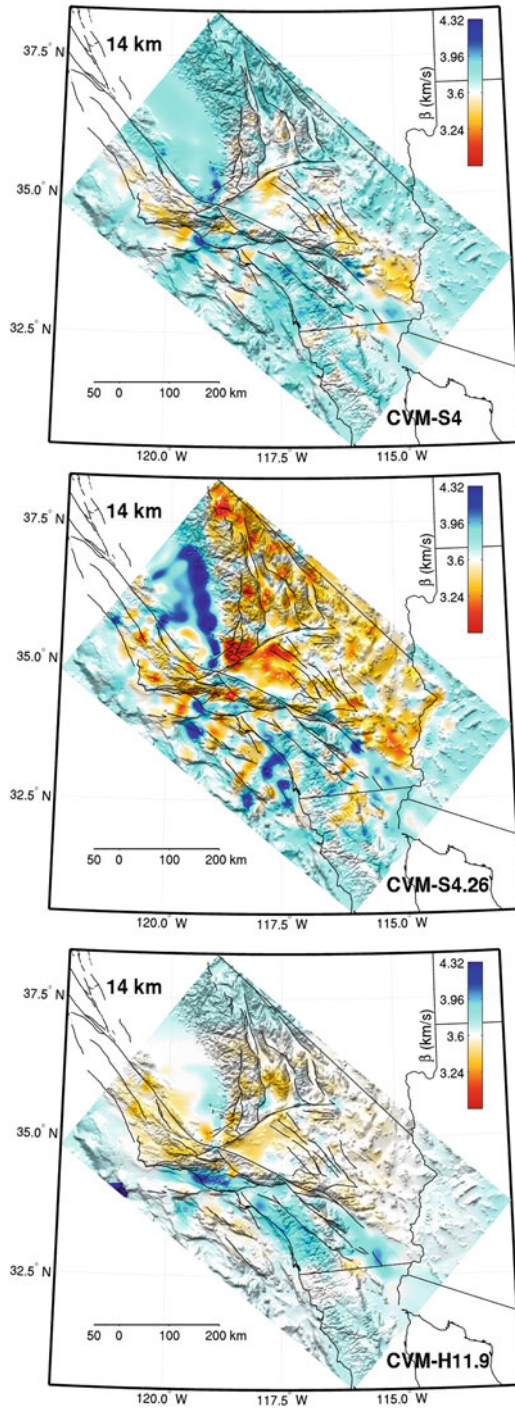


Fig. 6.1 (continued)

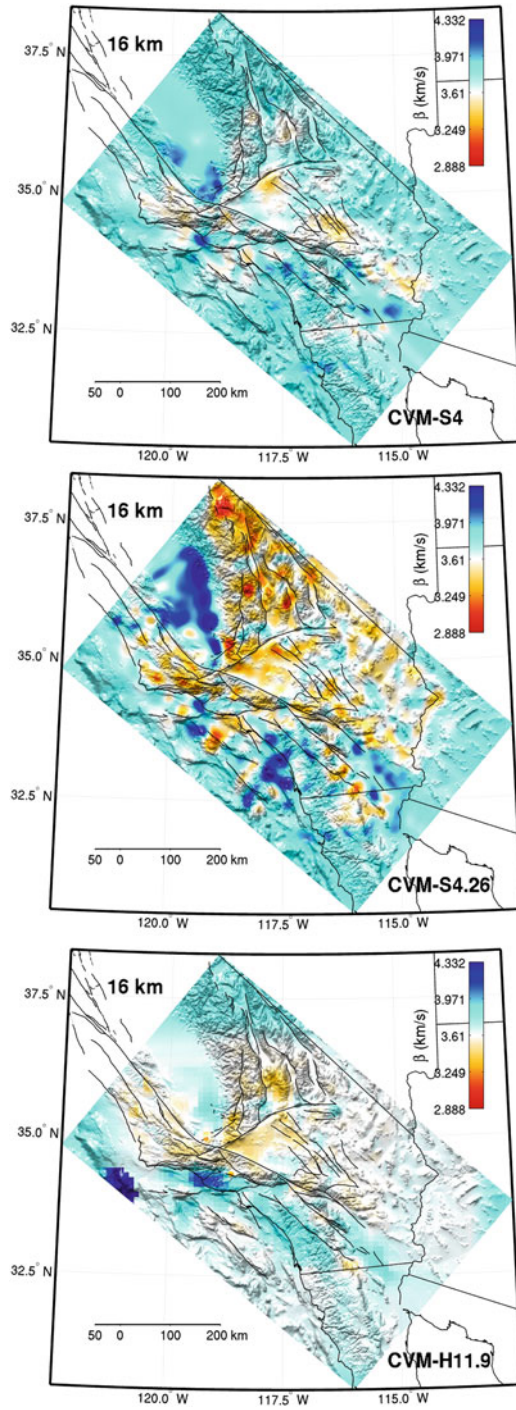


Fig. 6.1 (continued)

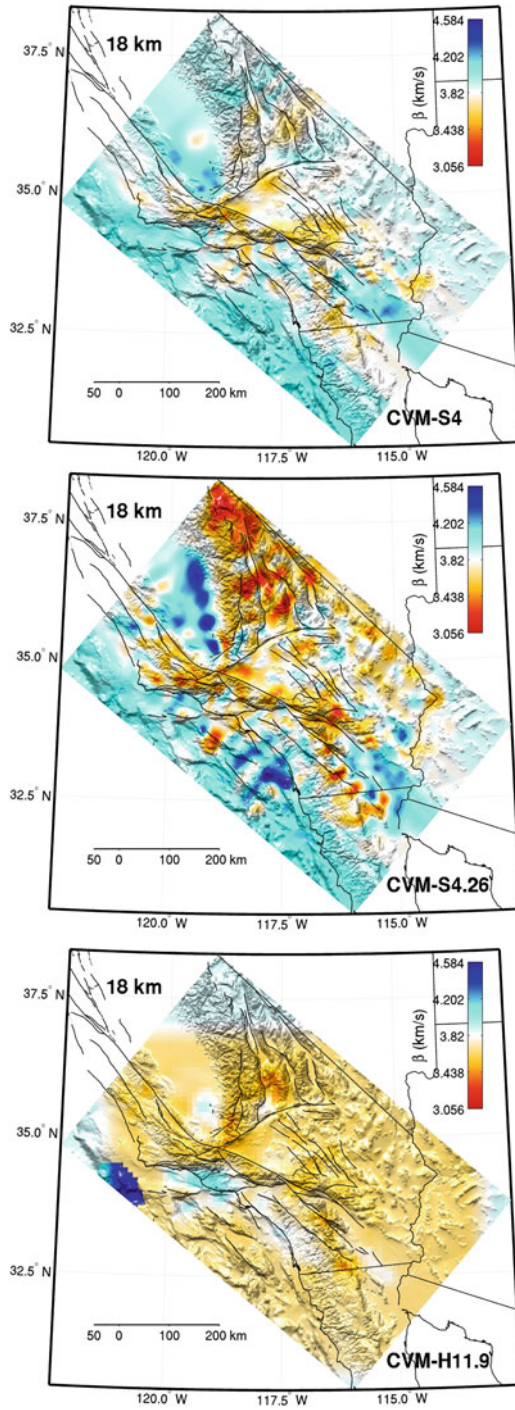


Fig. 6.1 (continued)

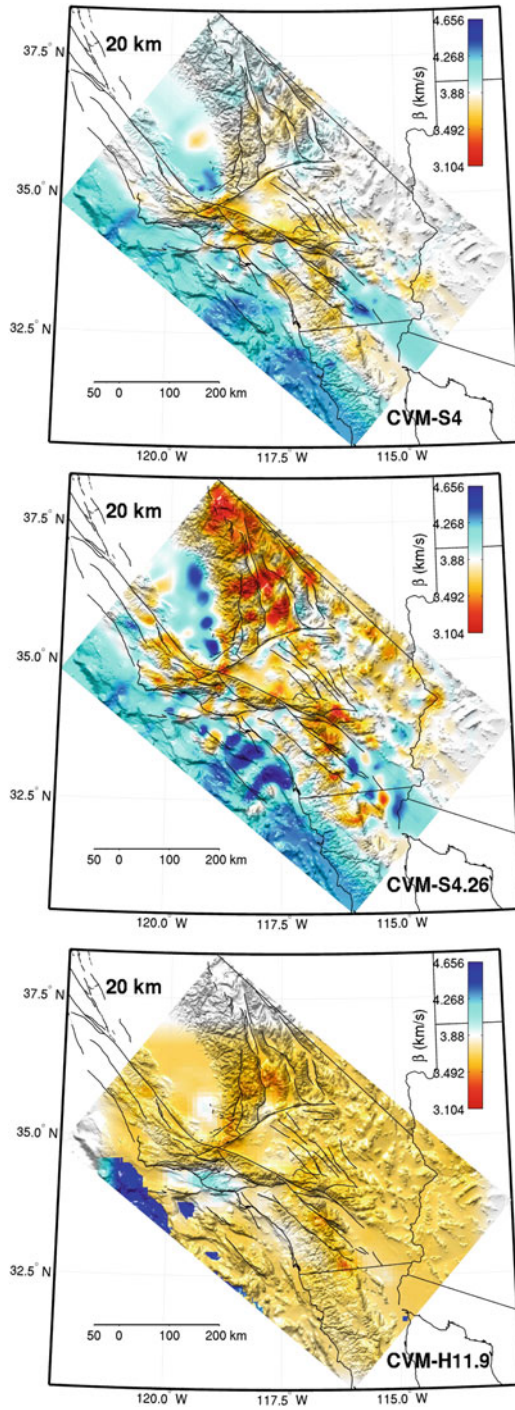


Fig. 6.1 (continued)

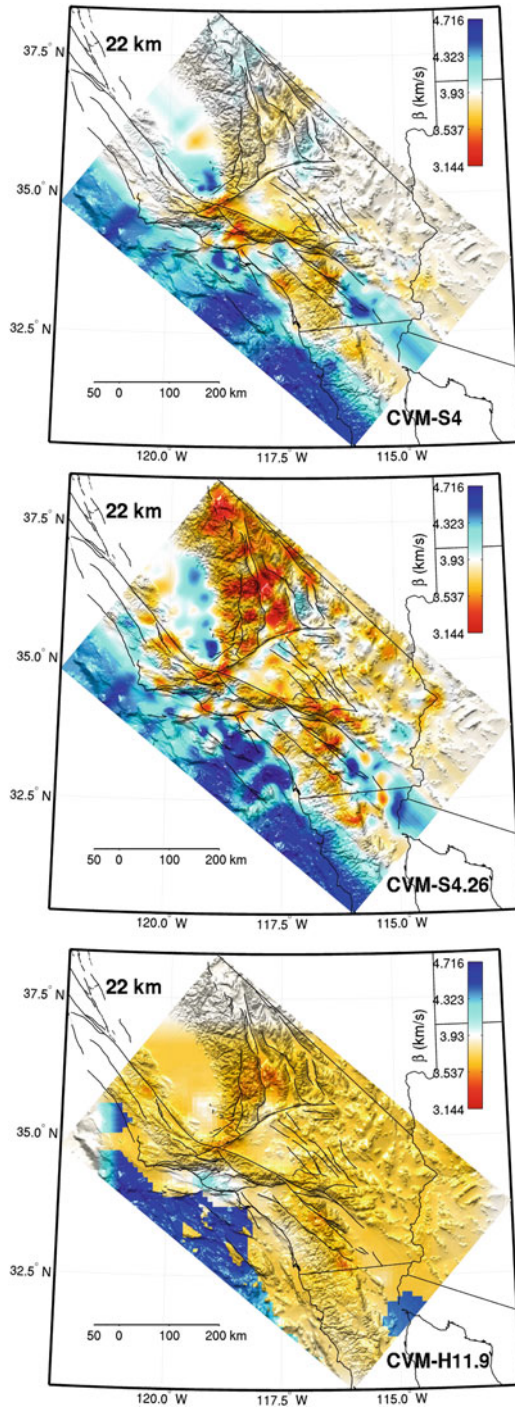


Fig. 6.1 (continued)

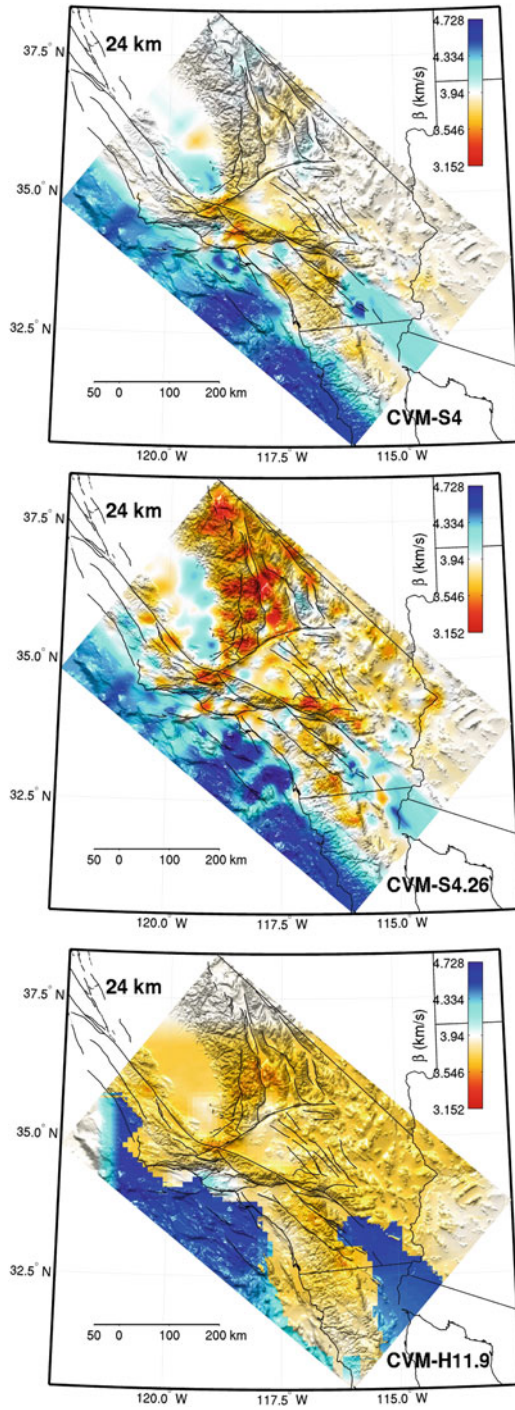


Fig. 6.1 (continued)

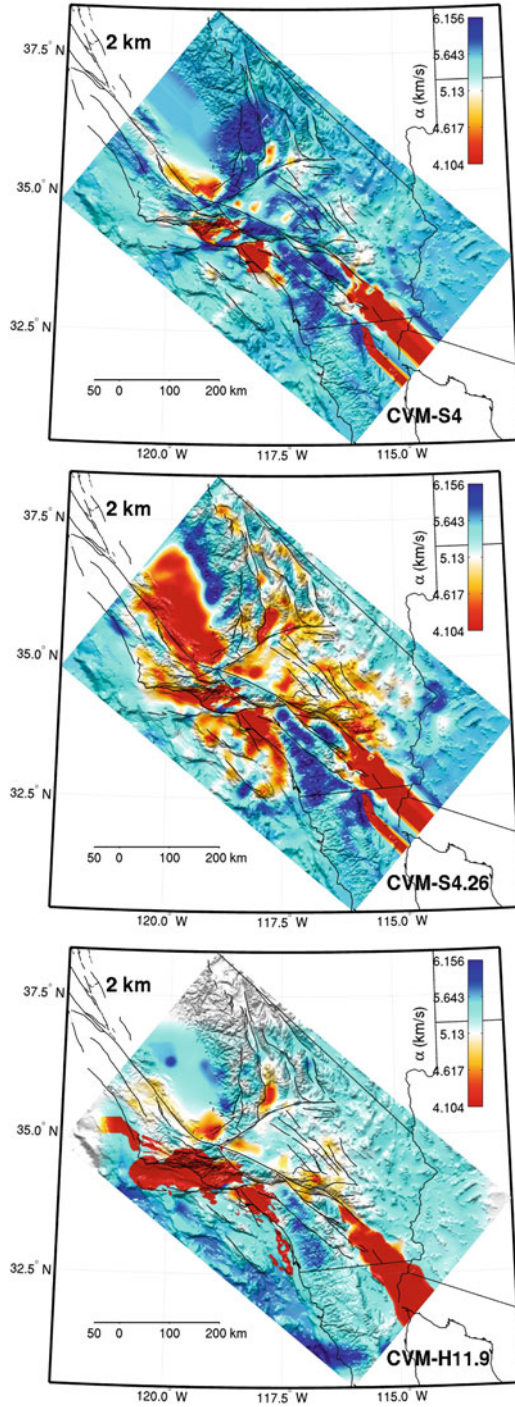


Fig. 6.2 Map-view of P-velocity at 2-km depth in CVM-S4 (top), CVM-S4.26 (center) and CVM-H11.9 (bottom). Warm colors indicate lower velocities and cold colors indicate higher velocities. Black solid lines: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages.

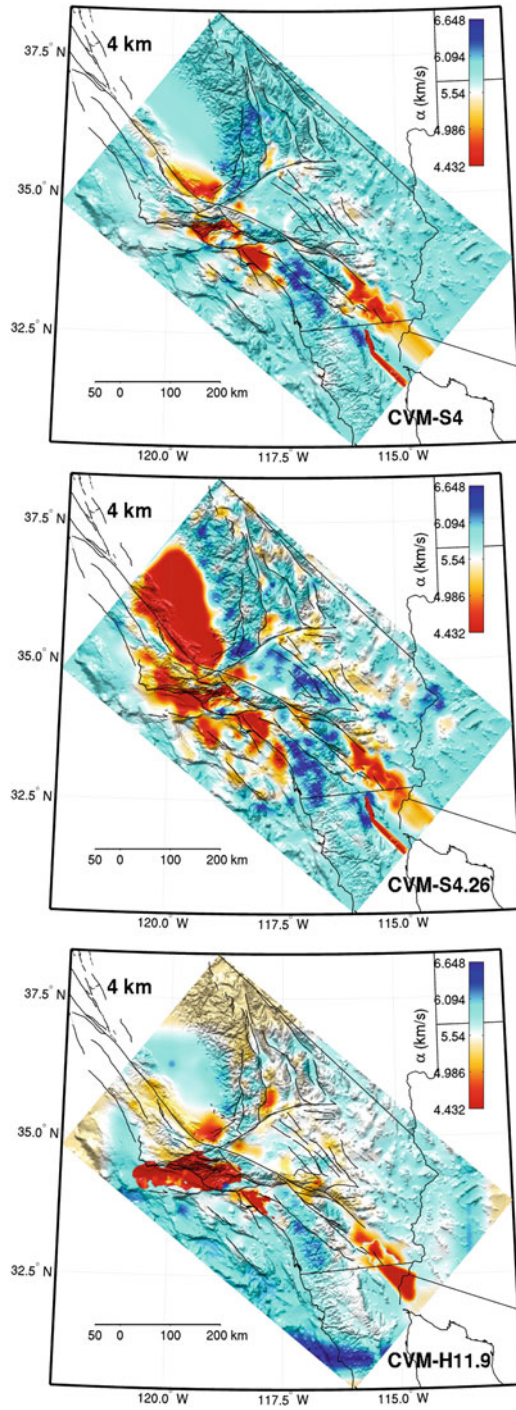


Fig. 6.2 (continued)

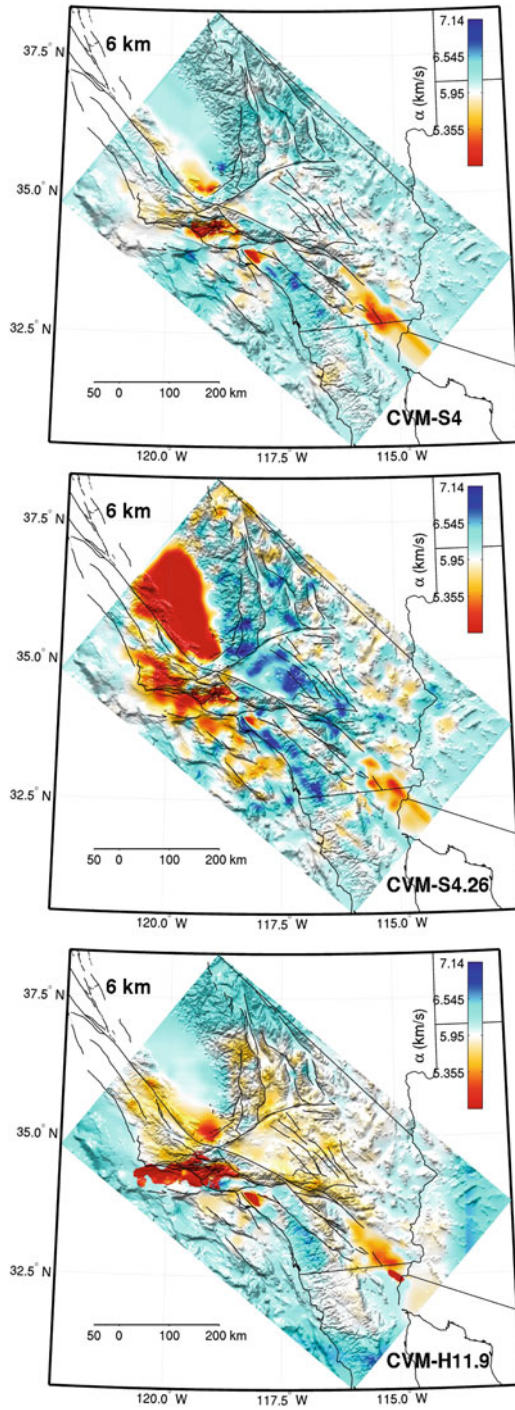


Fig. 6.2 (continued)

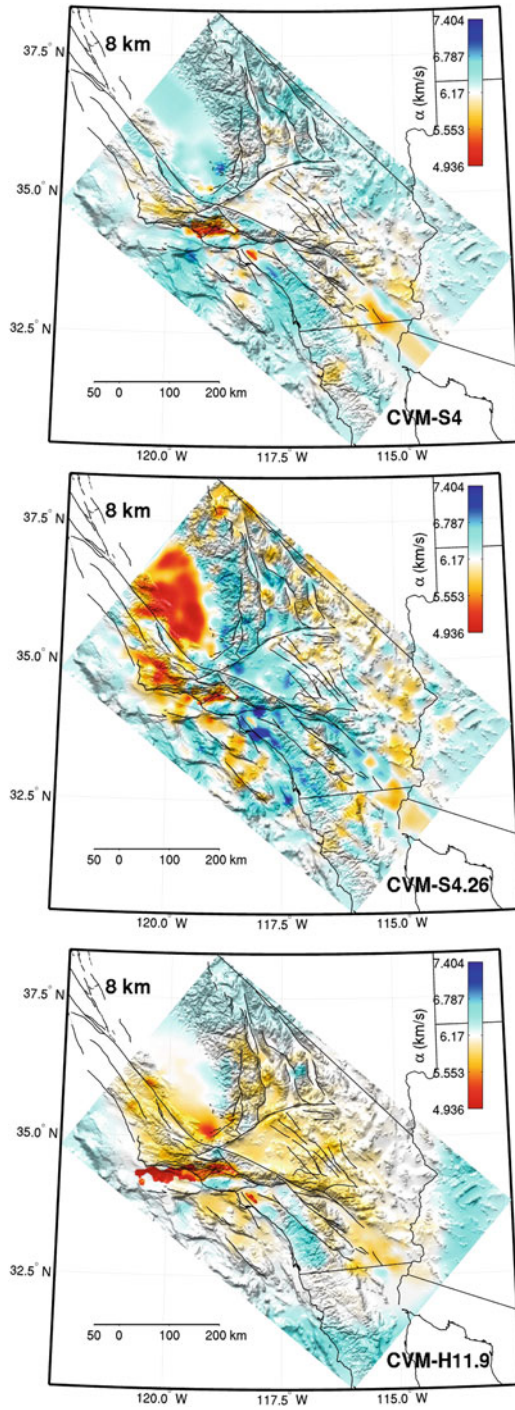


Fig. 6.2 (continued)

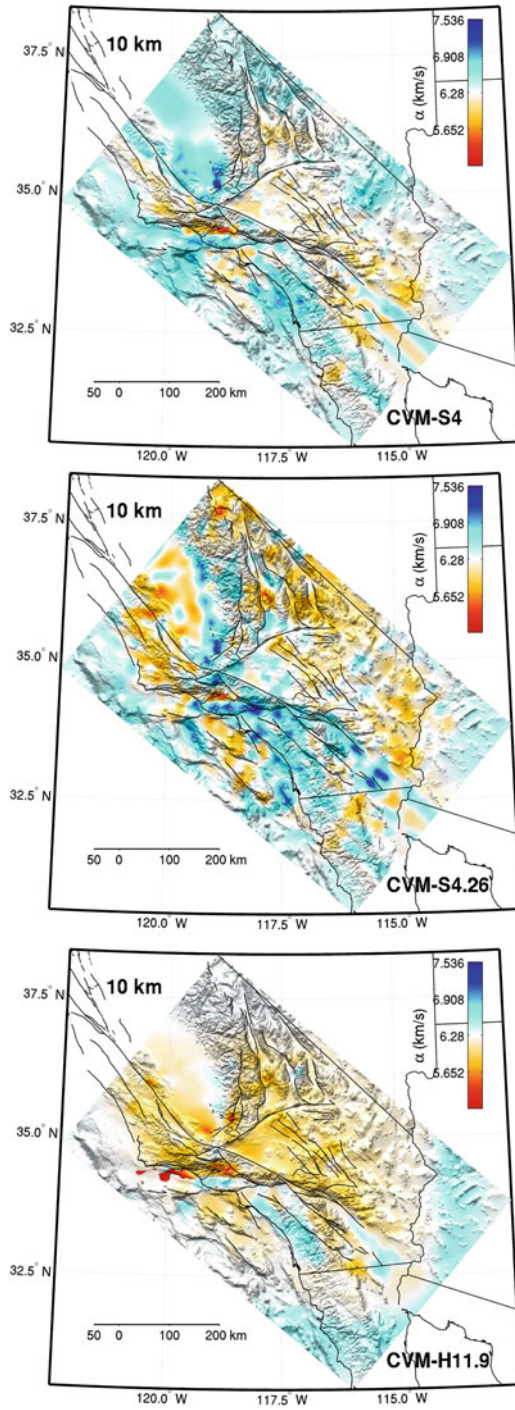


Fig. 6.2 (continued)

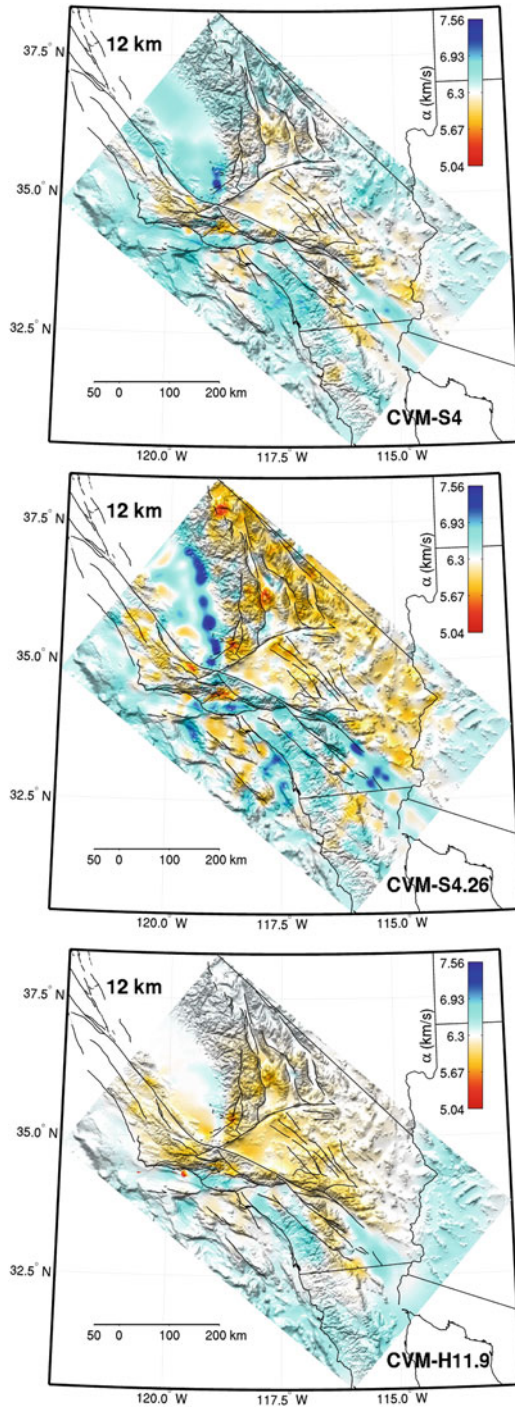


Fig. 6.2 (continued)

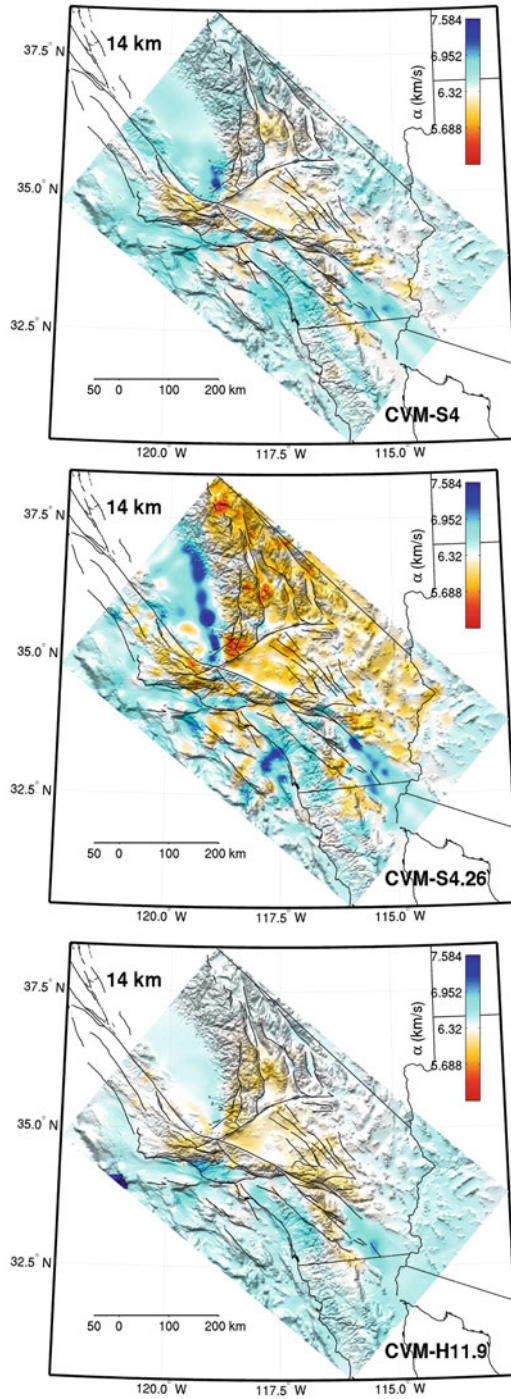


Fig. 6.2 (continued)

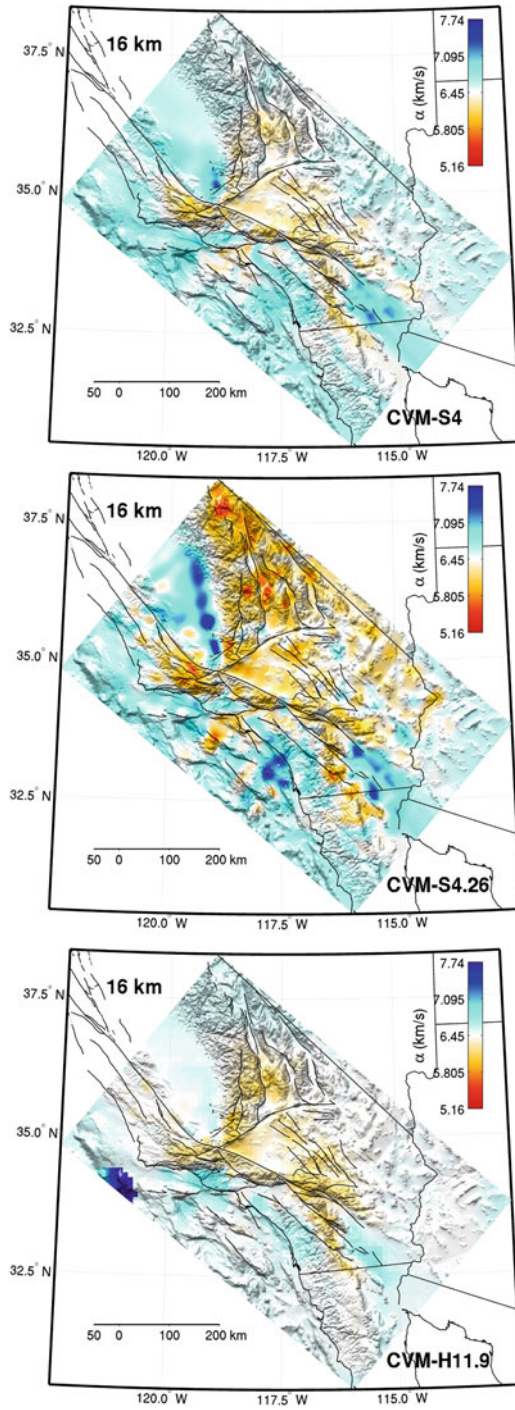


Fig. 6.2 (continued)

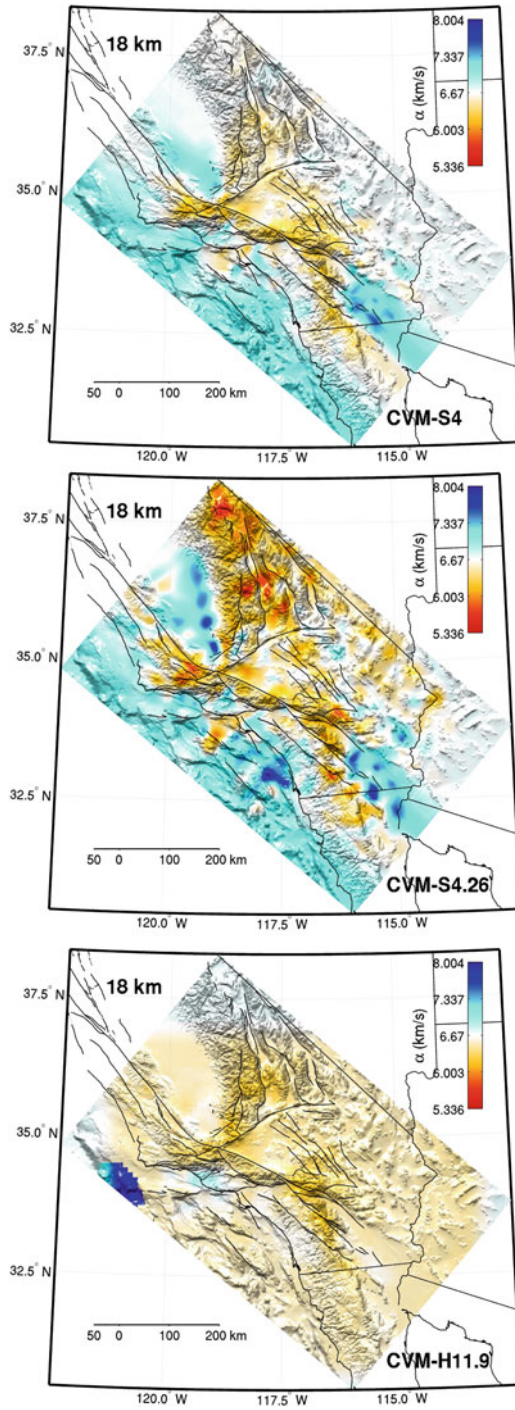


Fig. 6.2 (continued)

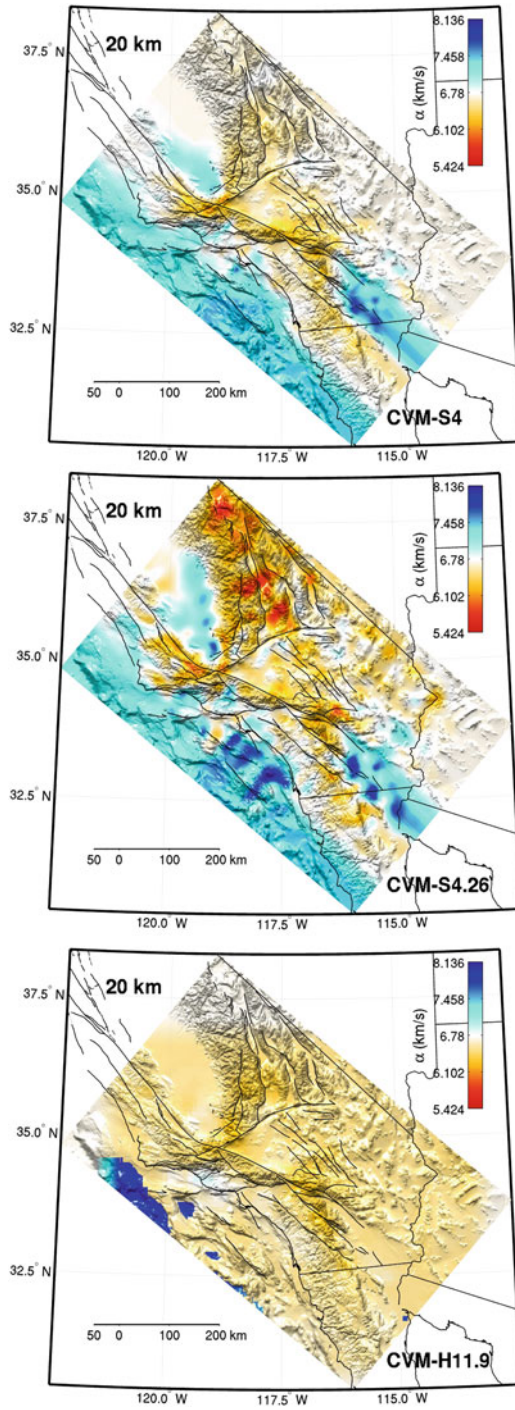


Fig. 6.2 (continued)

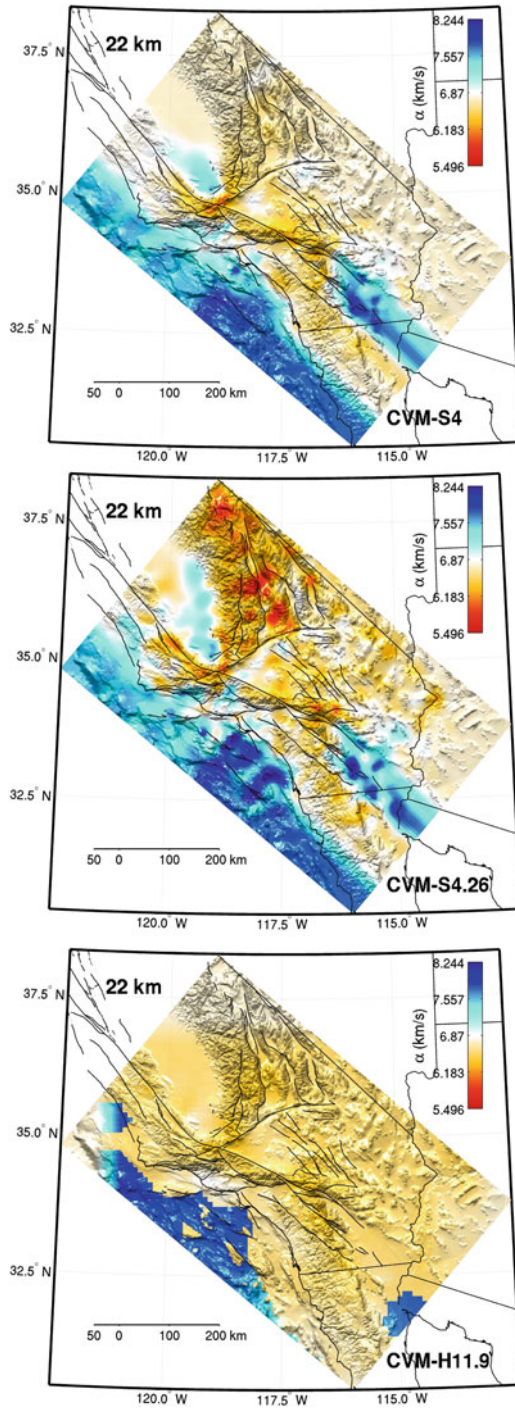


Fig. 6.2 (continued)

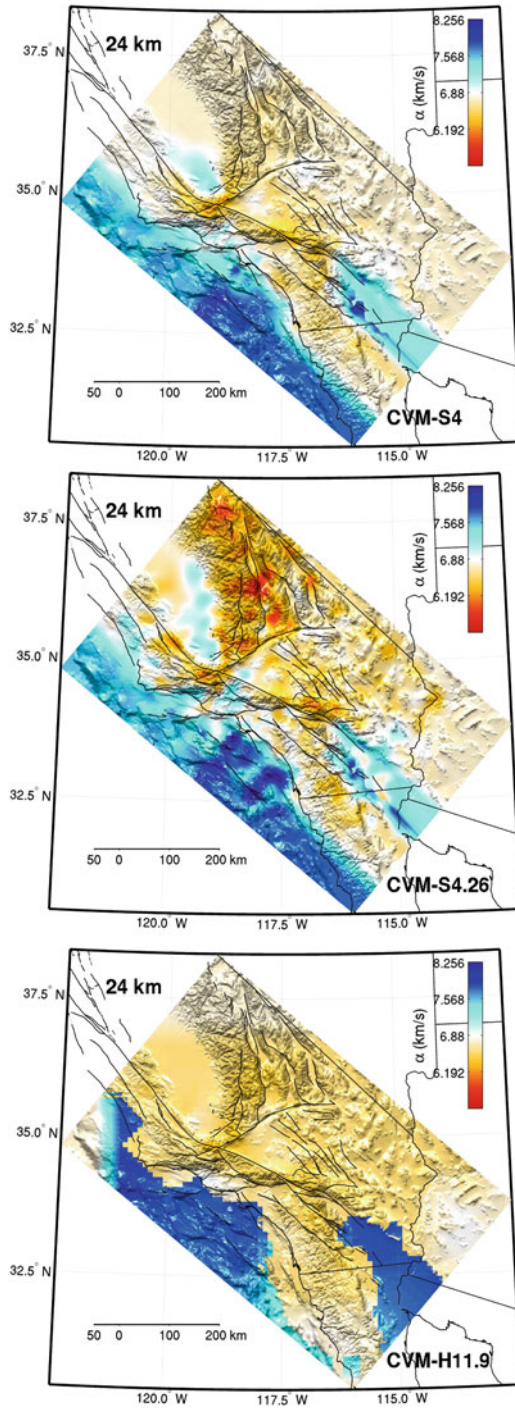


Fig. 6.2 (continued)

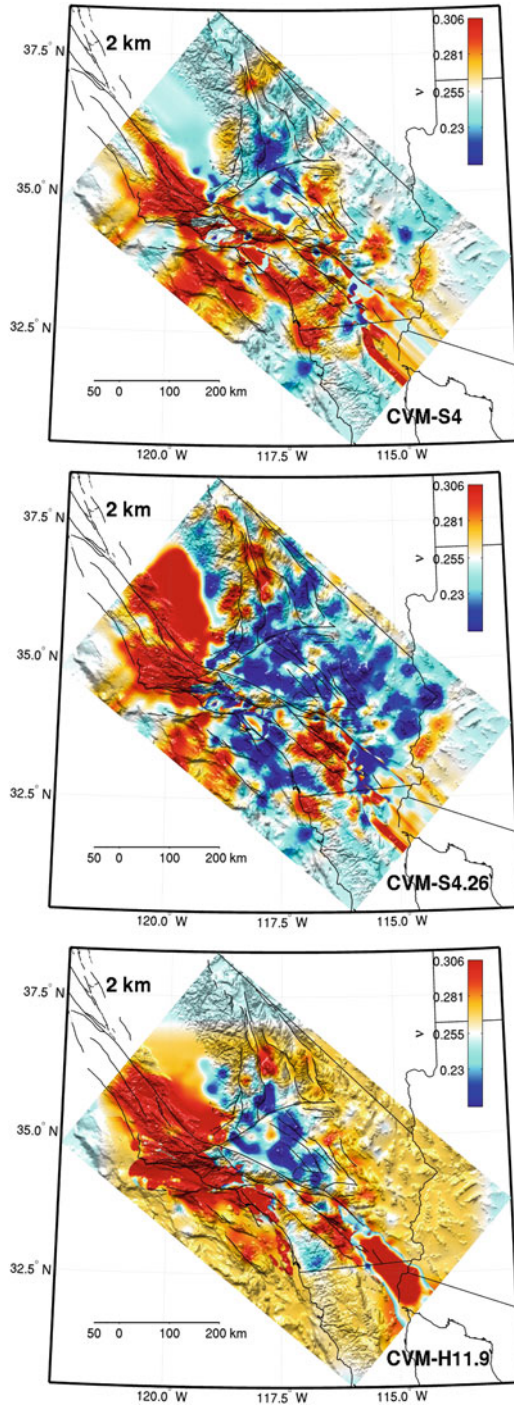


Fig. 6.3 Map-view of Poisson's ratio at 2 km depth in CVM-S4 (*top*), CVM-S4.26 (*center*) and CVM-H11.9 (*bottom*). Warm colors indicate higher Poisson's ratio and cold colors indicate lower Poisson's ratio. *Black solid lines*: faults, coast lines and political boundaries. Map-view plots at 4–24 km depths with 2-km interval are shown in the following pages.

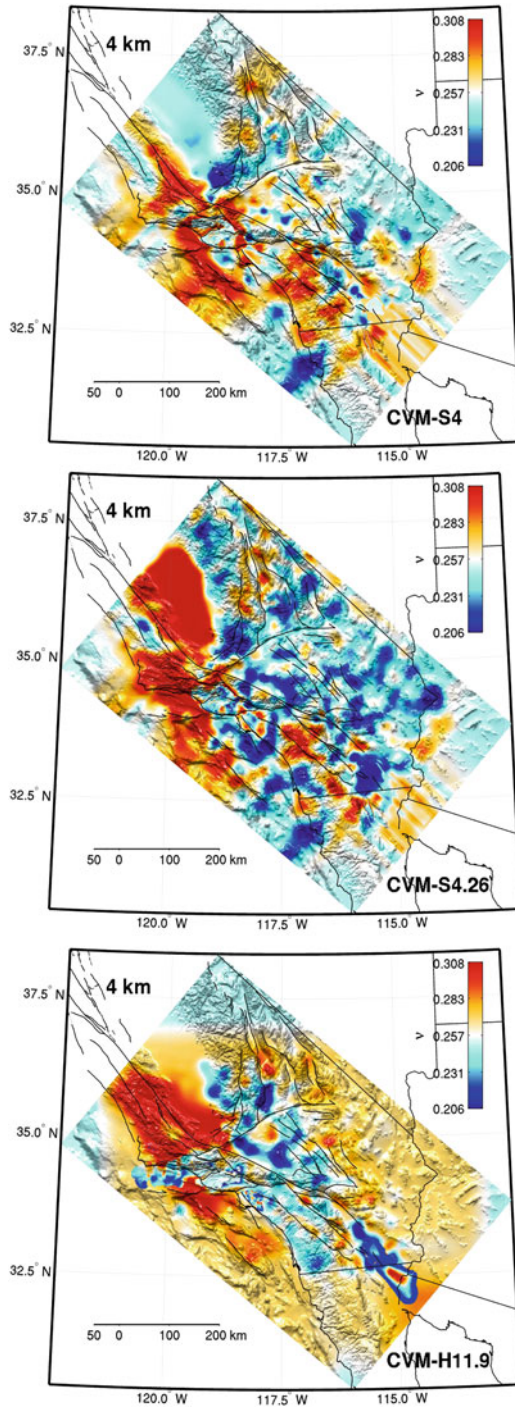


Fig. 6.3 (continued)

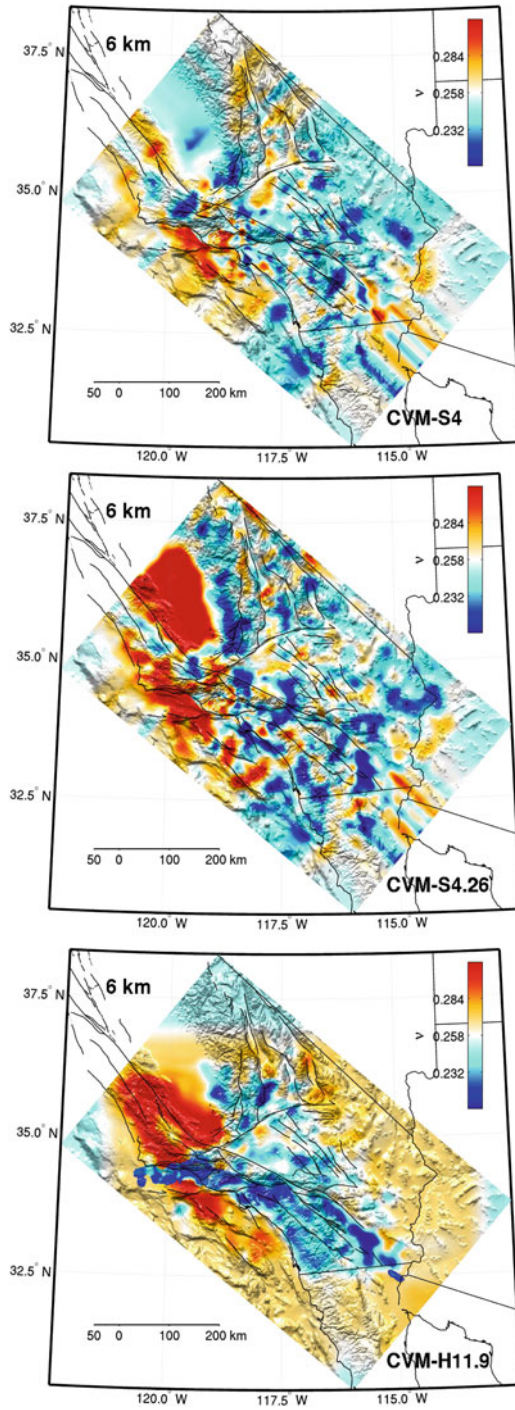


Fig. 6.3 (continued)

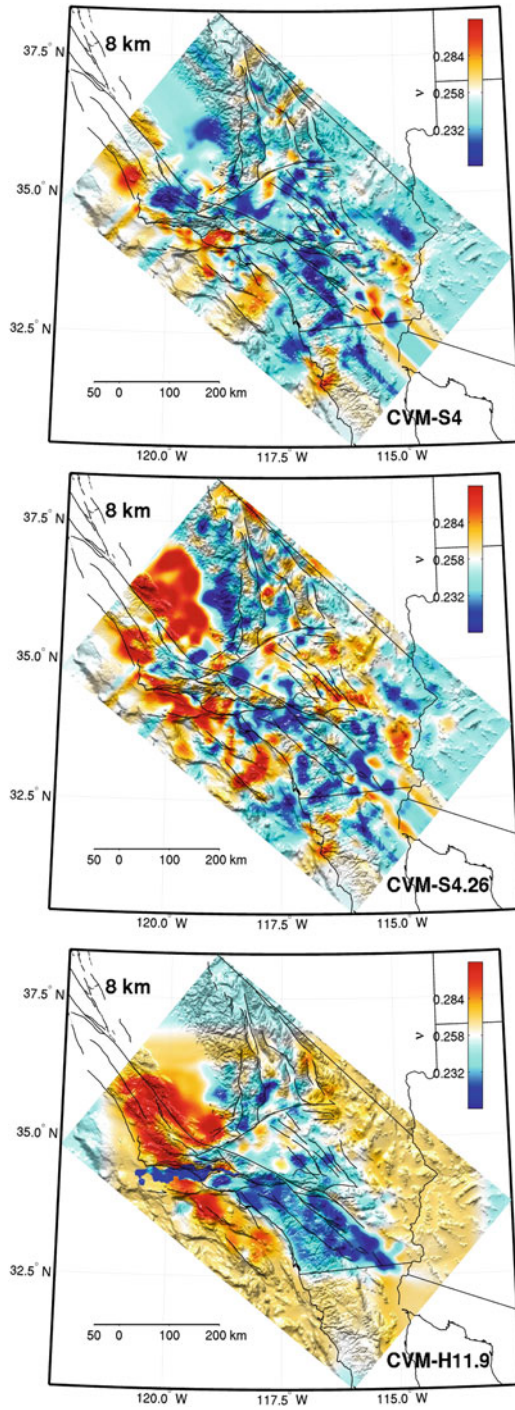


Fig. 6.3 (continued)

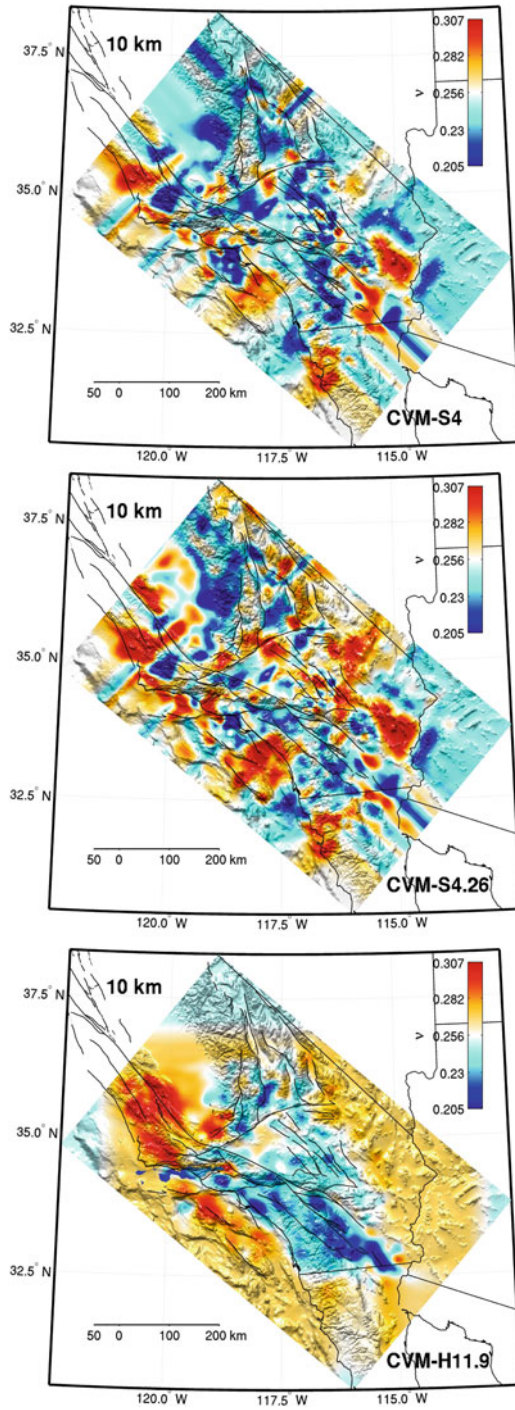


Fig. 6.3 (continued)

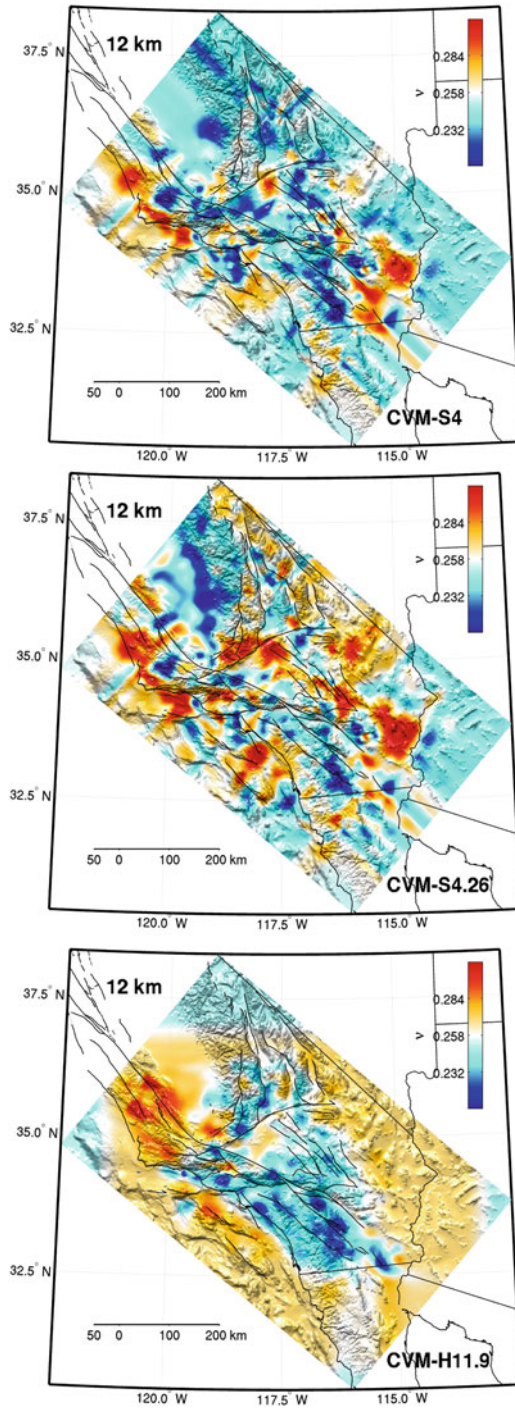


Fig. 6.3 (continued)

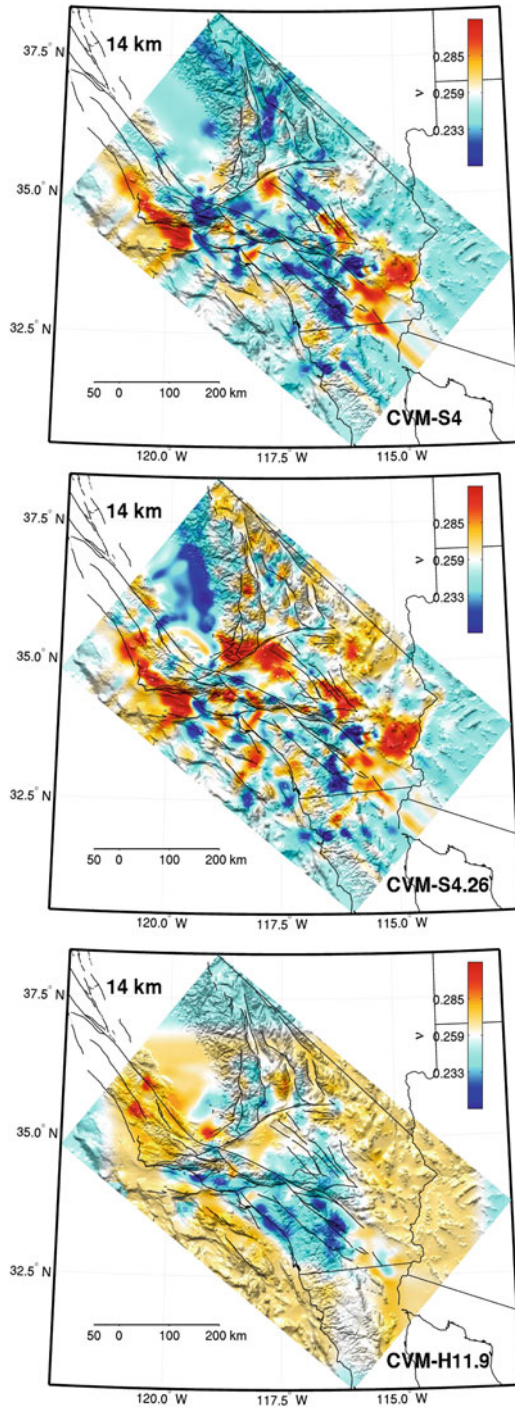


Fig. 6.3 (continued)

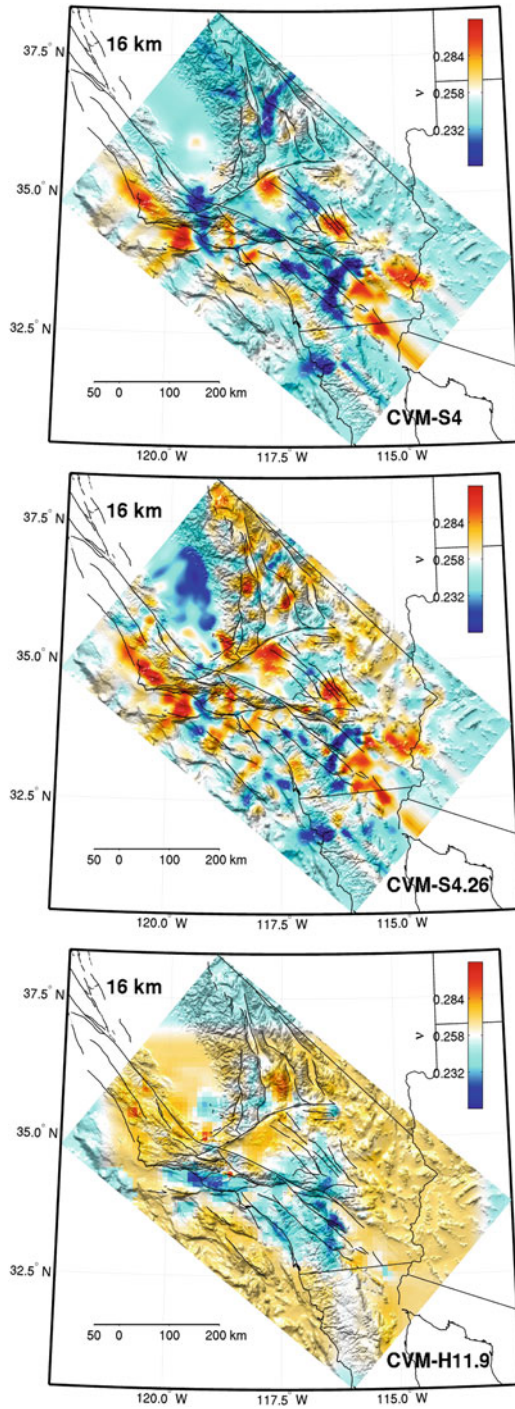


Fig. 6.3 (continued)

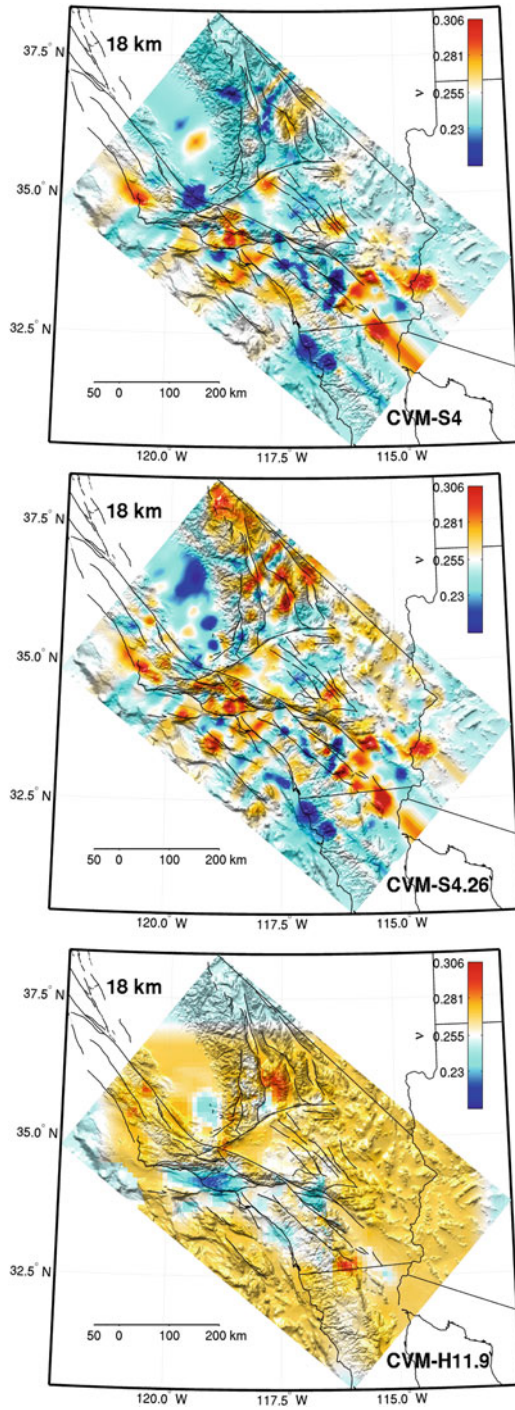


Fig. 6.3 (continued)

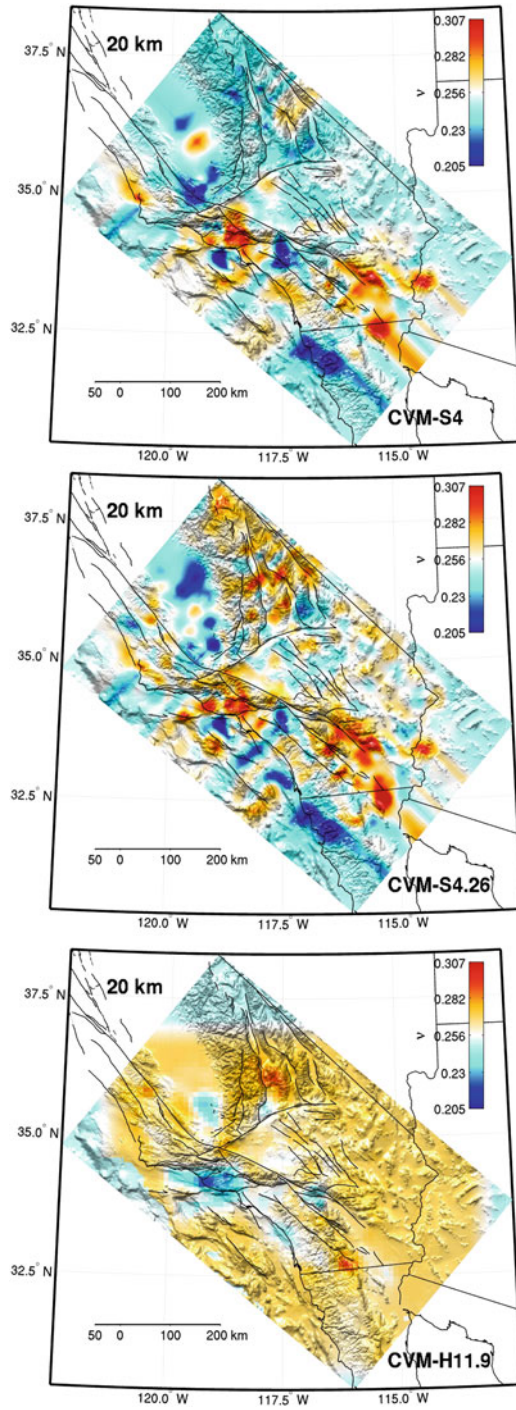


Fig. 6.3 (continued)

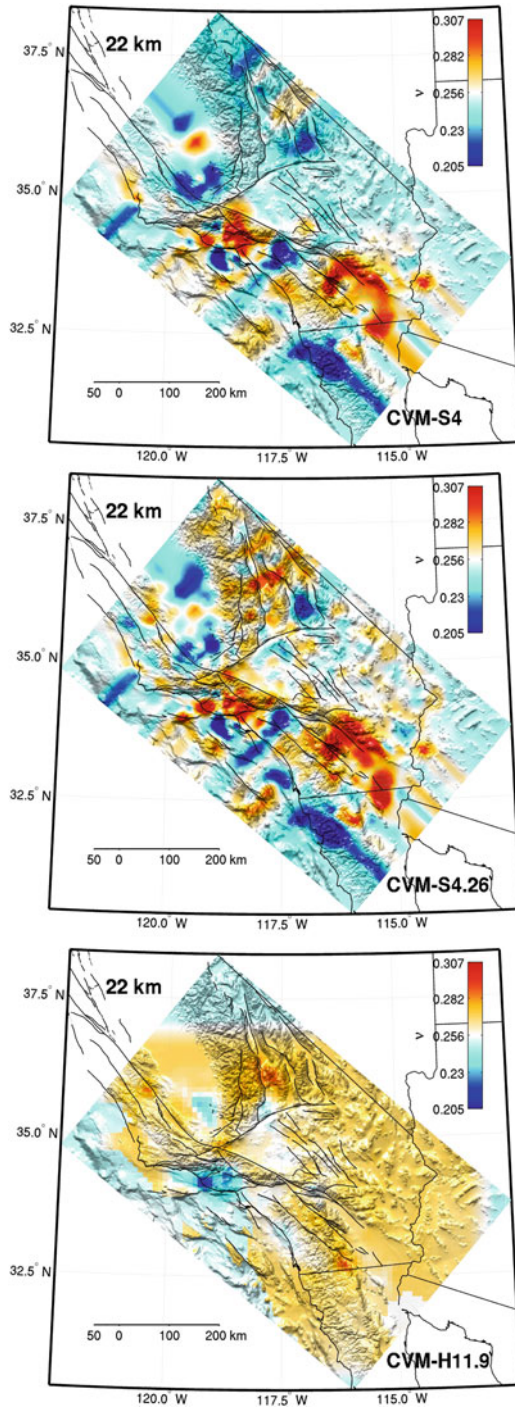


Fig. 6.3 (continued)

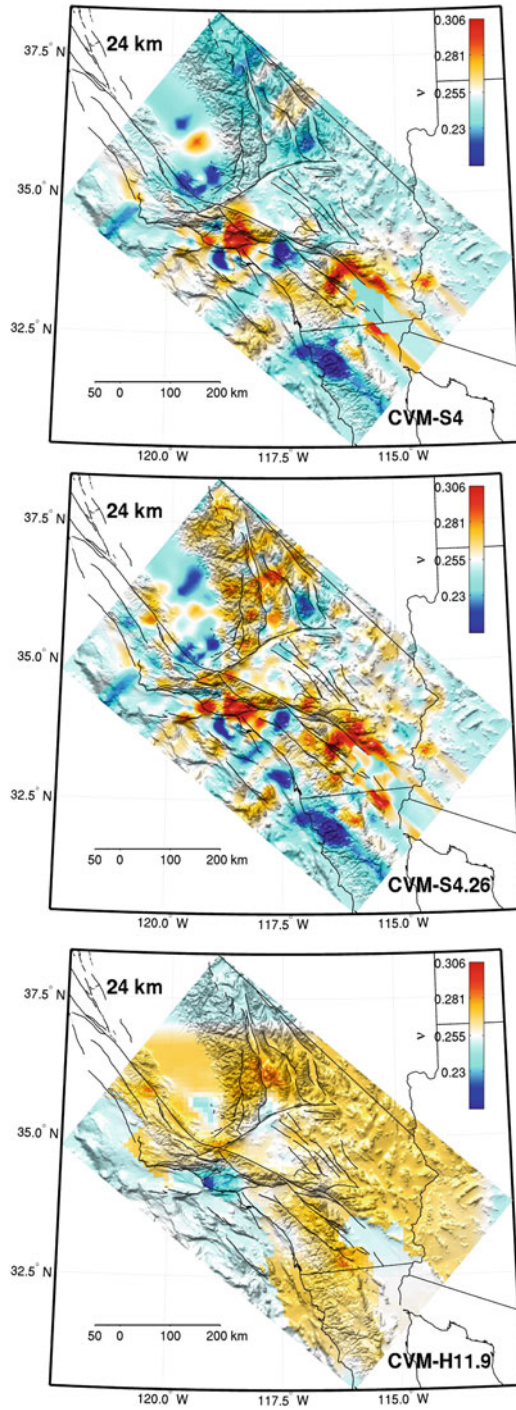


Fig. 6.3 (continued)

model used in both the CVM-S and the CVM-H series was based on the ray-travel-time tomography model of (Hauksson 2000). But the basin structures embedded in this common background regional model in the CVM-S and the CVM-H series were derived from different data sources. CVM-S mainly used rule-based basin structural models that relate seismic velocities to the age and depths of sediments and the rules were derived primarily from well-log data (Magistrale et al. 1996, 2000). The basin structures in CVM-H were determined from a large collection of sonic logs and seismic reflection profiles provided by the petroleum exploration industry (Süss and Shaw 2003). There are also other differences in the starting models, including differences in the Moho depth, the upper-mantle structure, the “geotechnical layer” very near the free surface and the density model. But these differences in the starting model probably do not have much effect on the differences between CVM-S4.26 and CVM-H11.9 shown in Figs. 6.1, 6.2 and 6.3.

6.2.2.2 Data Differences

In principle, F3DT-SI and F3DT-AW are physically equivalent, as we can derive one from the other mathematically. CVM-S4.26 was obtained through a combination of F3DT-SI and F3DT-AW and CVM-H11.9 was derived through F3DT-AW (Tape et al. 2010). Even though the starting models are different, if the data sets used in both inversions are identical, we may still expect that they should converge to a similar final model. The inverse problem is still non-unique of course, but the large-scale features should be well-constrained by the data sets. However, the data sets used in the two inversions were quite different. In the inversion for CVM-S4.26 (Lee et al. 2014b), about one quarter of the total misfit measurements used in the last (26th) iteration were derived from ambient-noise Green’s functions (Appendix A in Lee et al. 2014b), which were not used in the inversion for CVM-H11.9 (Tape et al. 2010).

The exact correspondence between data set differences and model differences is difficult to establish, as it may require inversions of the two different data sets separately from a common starting model and the entire iterative tomographic navigation process is nonlinear. Some first-order effects of including ANGFs into the inversion can be estimated by examining the spatial distributions of the earthquake sources and receivers used in both inversions (Fig. 1 in Lee et al. 2014b and Fig. 2 in Tape et al. 2010). The earthquakes used in both inversions were mainly clustered on and around major faults. Regions with fewer earthquakes include the southern San Joaquin Basin, the southern Sierra Nevada, the Walker Lane Belt east of southern Sierra Nevada, the western Mojave, the Basin and Range province to the southeast of the Mojave block and the western Peninsular ranges. The differences between CVM-S4.26 and CVM-H11.9 also tend to be larger in these regions. In the inversion for CVM-S4.26, ANGFs provided additional data coverage in conjunction with any existing earthquake data in these regions, while in the inversion for CVM-H11.9, the structures in these regions were constrained by the sparser earthquake data alone.

Figure A3 in (Lee et al. 2014b) shows the distribution of the frequency-dependent phase-delay misfits used in the last iteration of our inversion as a function of frequency and data type (i.e., earthquake or ANGF). The proportion of the misfits derived from ANGFs generally increases at lower frequencies. Small- to medium-sized earthquakes, such as those used in our inversion, usually cannot generate large amounts of low-frequency energy, which limits the signal-to-noise ratio, as well as the number of usable misfits, for earthquake data at lower frequencies. The average amplitude spectrum of the ANGFs (Fig. A1 in Lee et al. 2014b) shows significant energy at frequencies as low as 0.02 Hz, which highly complements the amplitude spectra of small- to medium-sized earthquakes used in our inversion. In CVM-S4.26, structures at larger depths were mainly constrained by a combination of earthquake body-wave data with relatively long propagation paths and low-frequency surface-wave data from ANGFs and medium-sized earthquakes. In the inversion for CVM-H11.9, only earthquake data from local small- to medium-sized earthquakes were used, which could limit the resolution of the model at larger depths. In general, the differences between CVM-S4.26 and CVM-H11.9 increase with depth, as shown in Figs. 6.1, 6.2 and 6.3.

There are also differences in the data functionals used in the two inversions, which can be quite important on the effective utilization of all available waveform data. As mentioned in Sect. 6.1.2, in early iterations, the observed and synthetic waveforms might be poorly aligned (i.e., large kinematic error) in regions where the starting model has large errors (e.g., the southern San Joaquin Basin). It might be tempting to reject those waveforms from the inversion all together based on certain waveform selection criteria (e.g., Maggi et al. 2009; Lee and Chen 2013). However, by using the broadband cross-correlation delay time and frequency-dependent group-delay time, which are mostly immune to cycle-skipping errors, we were able to include those poorly-aligned waveform pairs into our inversion, thereby improving the utilization of the available data. The linearization of the data functional can be highly important in such situations, as discussed in Sect. 4.2.8. When the kinematic error is large, the time- or frequency-domain waveform differences (Sects. 4.2.4 and 4.2.5) are not the suitable data functionals.

After the large kinematic errors were corrected, switching to the frequency-dependent phase-delay misfit with a very dense sampling over the entire frequency band was also important in fully utilizing all available waveform information in the data. In (Lee et al. 2014b), we made frequency-dependent phase-delay misfit measurements at 19 different frequencies evenly distributed between 0.02 Hz and 0.2 Hz. If the sampling over the frequency band is not dense enough, we may leave valuable waveform information unutilized. An extreme case is the broadband cross-correlation delay time, which gives only a single misfit measurement over the entire frequency band and cannot fully capture subtle differences in the waveforms. The checkerboard tests shown in Fig. F1 in (Lee et al. 2014b) (by comparing the last column with the first column) show the differences between inverting the broadband cross-correlation delay time and inverting the frequency-dependent phase-delay time measured at the 19 sampling frequencies using all waveforms in our last iteration. The misfits at different frequencies can be weighted differently in

the inversion according to their uncertainties given by the noise model (Appendix E in Lee et al. 2014b).

6.2.2.3 Algorithmic Differences

In the 18 AW iterations for deriving CVM-S4.26, our procedure was generally consistent with that given in (Tromp et al. 2005; Tape et al. 2007, 2010). There might be some minor differences in how the event kernels were conditioned in our procedure from that in (Tape et al. 2010). To reduce the excessively large sensitivities in the vicinity of the sources and receivers, we adopted the technique in (Fichtner et al. 2009), which conditions the event kernel by accounting for the geometric spreading effect. The gradient of the objective function, obtained by summing all conditioned event kernels, was smoothed by convolving with a three-dimensional spatial Gaussian function following (Tape et al. 2007, 2010). The horizontal and vertical widths of the Gaussian function were adjusted separately to maximize the misfit reduction for the current iteration. The widths of the Gaussian function were generally reduced with iteration.

By switching to the SI method at iteration 21, we were able to significantly accelerate the convergence rate of our inversion for CVM-S4.26. As discussed in Sect. 6.1.1 and in (Lee et al. 2014b), one iteration of the Gauss-Newton algorithm during the later stage of our inversion might account for many AW iterations based on the conjugate-gradient algorithm. In (Tape et al. 2010), a total number of 16 AW iterations were carried out, while in (Lee et al. 2014b) we have carried out 18 AW iterations and 6 SI iterations so far. Some of the differences between CVM-S4.26 and CVM-H11.9 might be just due to the fact that the total number of iterations carried out in (Lee et al. 2014b) was larger than that in (Tape et al. 2010). Those later iterations may account for some of the differences in small-scale heterogeneities between CVM-S4.26 and CVM-H11.9.

6.2.2.4 Earthquake Source Parameter Differences

In (Lee et al. 2014b), the procedure to account for errors in the earthquake source parameters can be roughly separated into two stages. In the first stage (iteration 1 to 14, Fig. 2 in Lee et al. 2014b), only ANGFs and seismograms from earthquakes with highly robust CMT solutions were used to correct large-scale errors in the structural model. In the second stage (iteration 15 to 26), CMT solutions for all earthquakes used in the inversion were inverted at some selected iterations using the updated structural model (Lee et al. 2011). The phase data of the ANGFs in Southern California are known to be quite robust and many ambient-noise tomography results have been obtained using these data (e.g., Shapiro et al. 2005). By inverting misfits obtained mainly from the ANGFs in the first stage, we were able to correct large-scale errors in the structural model without using seismograms from earthquake sources with highly uncertain CMT solutions. In the second stage, when

inverting for the CMT solutions, the synthetic Green's functions were computed from the updated structural model, therefore the inverted CMT solutions were not significantly biased by any large-scale errors in the structural model.

6.2.2.5 Summary

Among all the differences between the two inversions documented in (Lee et al. 2014b) and (Tape et al. 2010), respectively, the difference in data (Sect. 6.2.2.2) might be the most important factor for explaining the differences between CVM-S4.26 and CVM-H11.9. The difference in data includes the difference in the seismogram types (i.e., earthquake seismograms and ANGFs in Lee et al. 2014b and earthquake seismograms only in Tape et al. 2010) and also the difference in the data functionals used in both inversions. The difference in the inversion algorithm (Sect. 6.2.2.3) probably ranks the second in explaining the differences between CVM-S4.26 and CVM-H11.9. The difference in how the earthquake source parameters were handled might rank the third. The difference in the starting model might rank the fourth.

In addition to the differences discussed above, there are also other differences, such as those in the wave-equation solver (i.e., *AWP-ODC* in Lee et al. 2014b and *SPECFEM3D* in Tape et al. 2010). Details of *AWP-ODC* were given in Chapter 2 and the exact configuration used in our inversion for CVM-S4.26 was explained in Appendix B in (Lee et al. 2014b). Given the same velocity model, differences in synthetic seismograms generated using *AWP-ODC* and those generated using *SPECFEM3D* were likely to be very small at frequencies below 0.2 Hz. The effects of topography and the curvature of the Earth, which were not accounted for in *AWP-ODC* but included in *SPECFEM3D*, were also likely to be negligible at such low frequencies.

CVM-S4.26 reveals strong crustal heterogeneities throughout Southern California. Many features in CVM-S4.26 show good correlation with regional geology and tectonics. In the following, I will point out some interesting features in CVM-S4.26 that have not been well imaged in previous crustal-scale tomography studies but are consistent with geology and local studies.

6.2.3 *Southern Walker Lane Belt*

The Walker Lane belt was defined in (Stewart 1988) as the triangular region separating Sierra Nevada from the Great Basin bounded by the Garlock fault in the south. In our tomography box for CVM-S4.26, the southernmost Walker Lane belt is a 100-km-wide zone north of the Garlock fault bordering southeastern Sierra Nevada (Fig. 6.4). Previous researchers have suggested that the Miocene Basin and Range extension has been gradually encroaching into the Sierra Nevada and the present-day Sierra Nevada range front (Sierra Nevada Fault Zone in Fig. 6.4) is the western boundary of the Basin and Range province (e.g., Surpless et al. 2002). Recent GPS

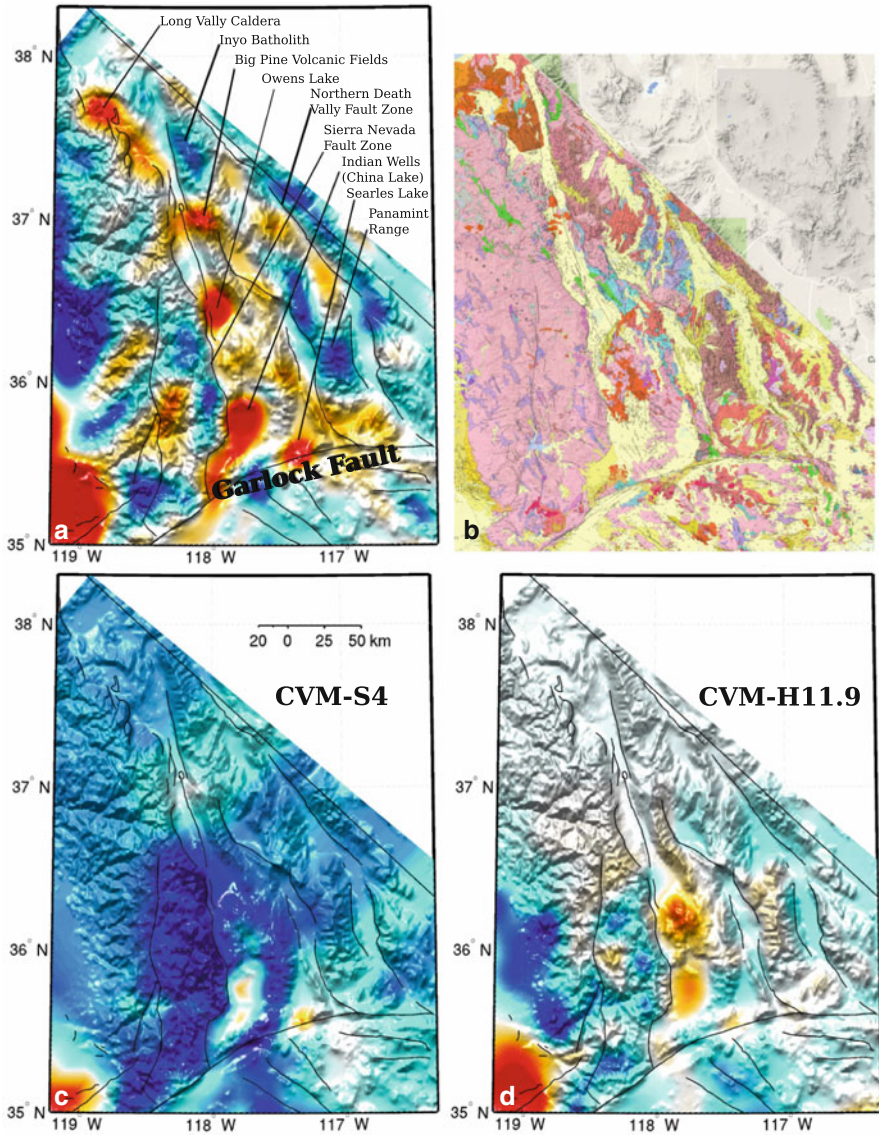


Fig. 6.4 S-velocity in CVM-S4.26 (*upper-left*), CVM-S4 (*lower-left*) and CVM-H11.9 (*lower-right*) compared with the geologic map (*upper-right*) around the Owens Valley and Death Valley region. The color range is identical to that in Fig. 6.1.

and earthquake data have shown that the Walker Lane belt is actively accommodating the relative motion between the Sierra Nevada block and the stable North American interior through transtensional deformation (e.g., McClusky et al. 2001; Unruh et al. 2003). GPS data have shown that the Walker Lane belt and the Eastern California Shear Zone south of the Garlock fault together accommodate about 20–25 % of the present-day relative plate motion between the Pacific and North American plates (e.g., Miller et al. 2001; Hammond and Thatcher 2007). Based on these observations, researchers have suggested that the Walker Lane belt is the northernmost extension of the Gulf of California rift. The eastern margin of the Sierra Nevada block will be the future plate boundary and the Sierra Nevada block will be transferred to the Pacific plate, as has Baja California (e.g., Faults and Henry 2008; Jayko and Bursik 2012).

In general, the structures at shallow depths in CVM-S4.26 are well correlated with surface geology. In Fig. 6.4, the S-velocities at 2-km depth in the three CVMs are compared with a geologic map¹ for this region. Several strong low-velocity anomalies imaged in CVM-S4.26 are correlated with (from north to south) the Long Valley Caldera, the Big Pine Volcanic Fields, the Indian Wells Valley (China Lake) and the Searles Lake. Several strong high-velocity anomalies imaged in CVM-S4.26 are correlated with the Inyo Batholith in the White Mountains, the Precambrian rocks in the Panamint Range and the Precambrian and Paleozoic rocks in the Grapevine Mountains to the northeast of the Northern Death Valley Fault Zone. In general, low-velocities in CVM-S4.26 are correlated with Quaternary deposits, Tertiary and Quaternary volcanic rocks in the geologic map, while high-velocities in CVM-S4.26 are correlated with Precambrian and Paleozoic rocks and Mesozoic plutonic and batholithic rocks. The Owens Valley is separated from the Sierra Nevada Batholith to the west by a sharp velocity gradient along the Sierra Nevada Fault Zone. The Northern Death Valley Fault Zone also lies along a strong velocity contrast in CVM-S4.26. The correlation between CVM-S4.26 and the surface geology in this region is quite remarkable, especially when considering that this region is already close to the northeastern boundary of our tomography box. In our inversion, the velocity structures in this region were mainly constrained by data from earthquakes located to the south of this region and ANGFs from both permanent and temporary broadband stations in and around this region (Fig. 1 in Lee et al. 2014b).

In Figs. 6.5, 6.6 and 6.7, I show cross-sections of the S-velocity, P-velocity and Poisson's ration in CVM-S4.26 along 13 profiles crossing the southernmost Walker Lane belt. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.8–6.13. On average, the velocities in CVM-S4.26 are lower than those in CVM-S4 and the Poisson's ratio is higher in CVM-S4.26 than that in CVM-S4 in this region. At mid- to lower-crustal depths, this region has below-average

¹ The geologic maps used in this chapter were obtained from the California Geological Survey (<http://www.quake.ca.gov/gmaps/GMC/stategeologicmap.html>). A detailed explanation about the different rock types in the geologic maps can be found at http://www.quake.ca.gov/gmaps/images/GMC_Explanation.pdf.

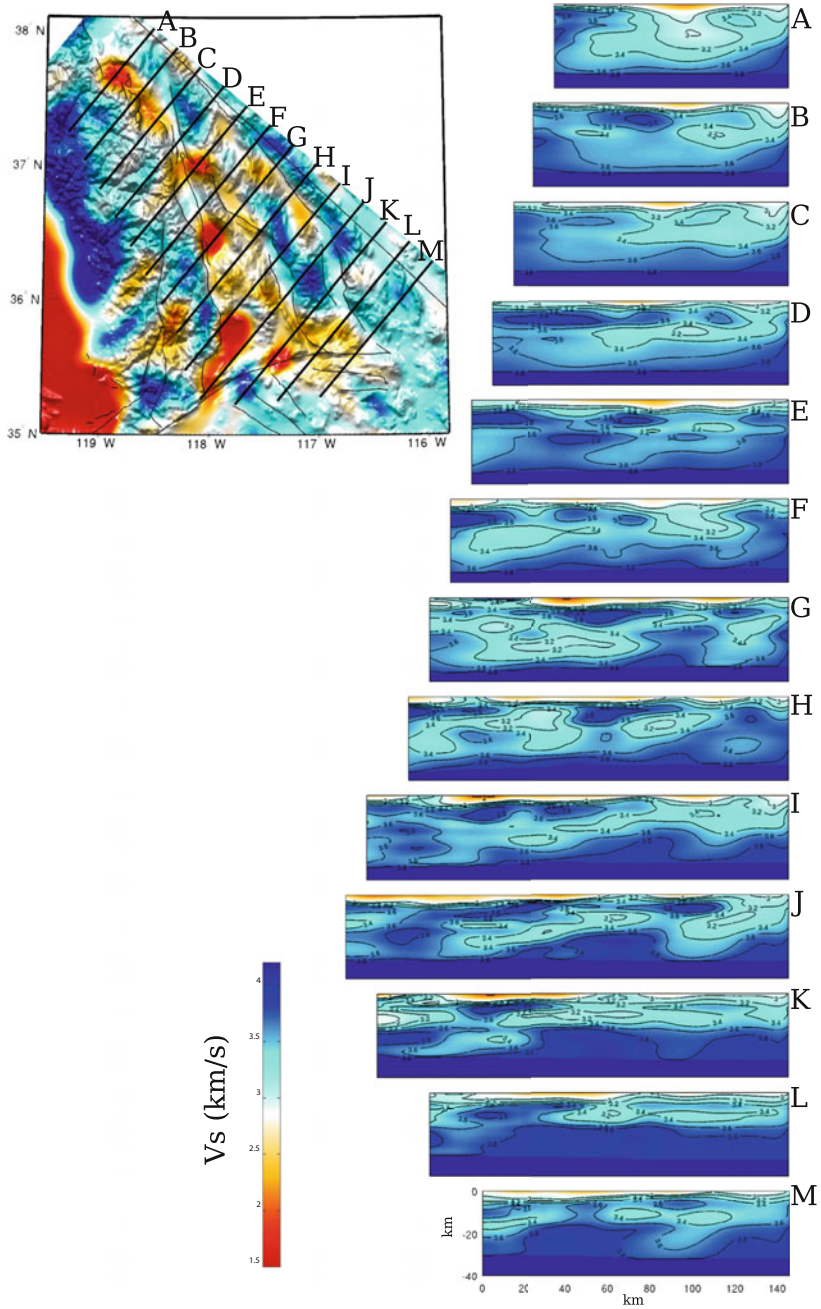


Fig. 6.5 Cross-sections of S-velocity in CVM-S4.26 along 13 profiles (A-M) across the Walker Lane belt. The map-view at *upper-left* corner shows S-velocity at 2-km depth and the color-scale is identical to the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown in the *lower-left* corner.

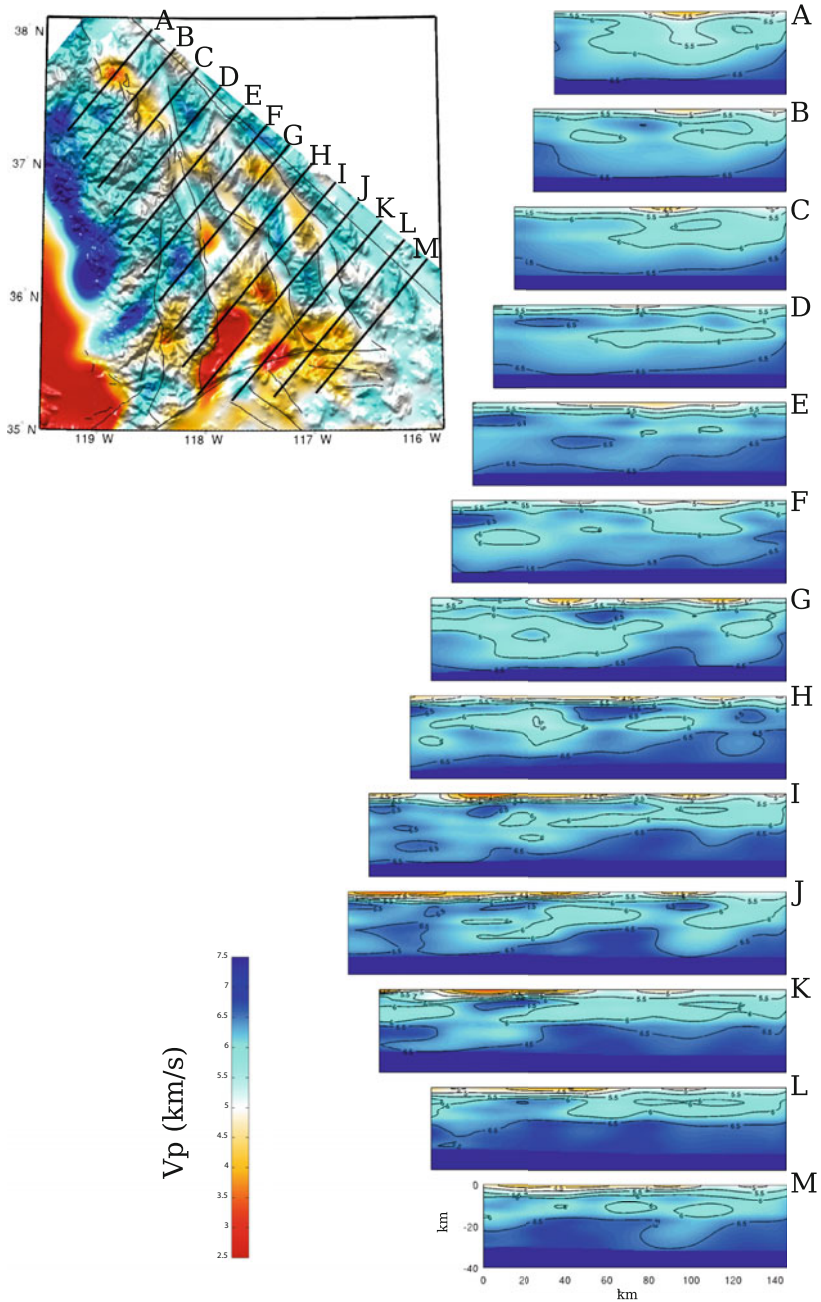


Fig. 6.6 Cross-sections of P-velocity in CVM-S4.26 along the 13 profiles (A-M) across the Walker Lane belt. The format is identical to that in 6.5.

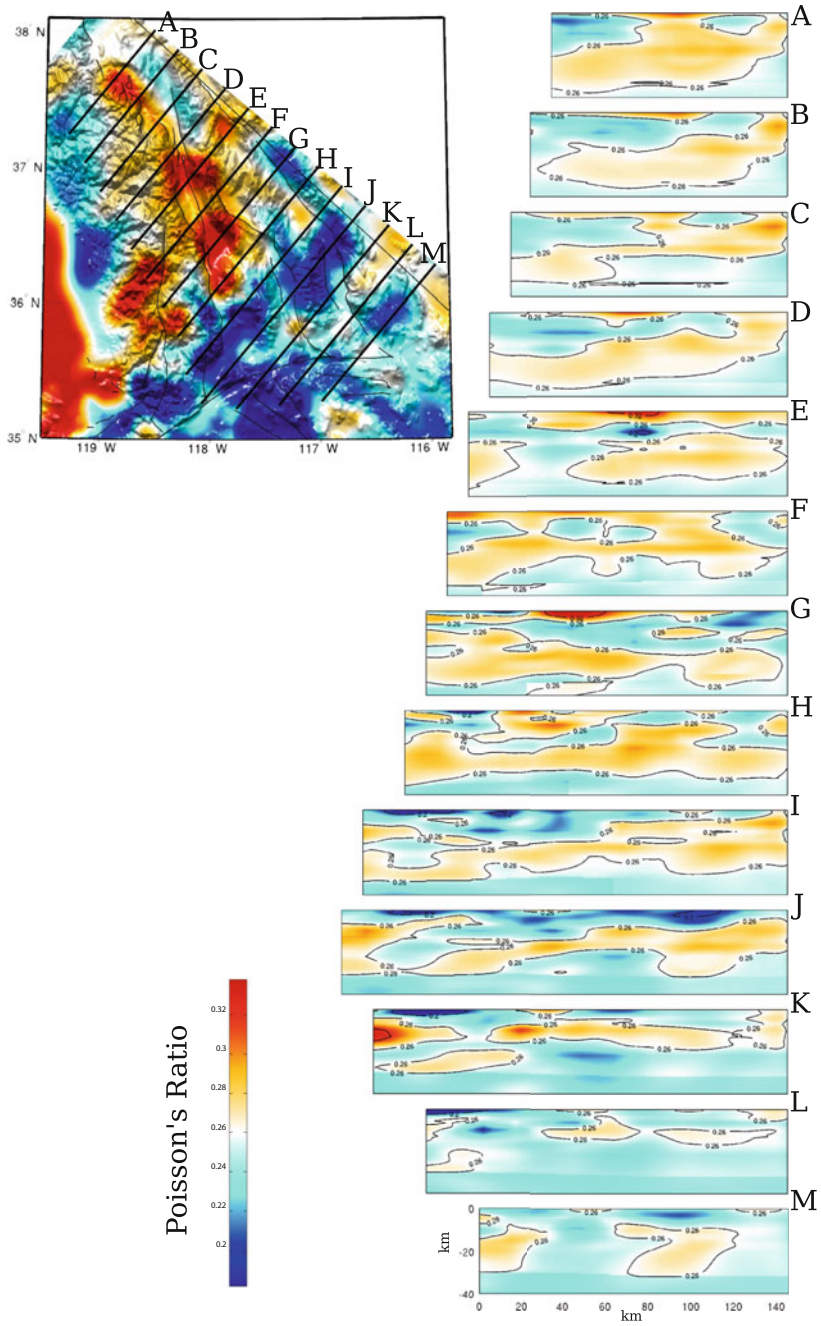


Fig. 6.7 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 13 profiles (A-M) across the Walker Lane belt. The format is identical to that in 6.5.

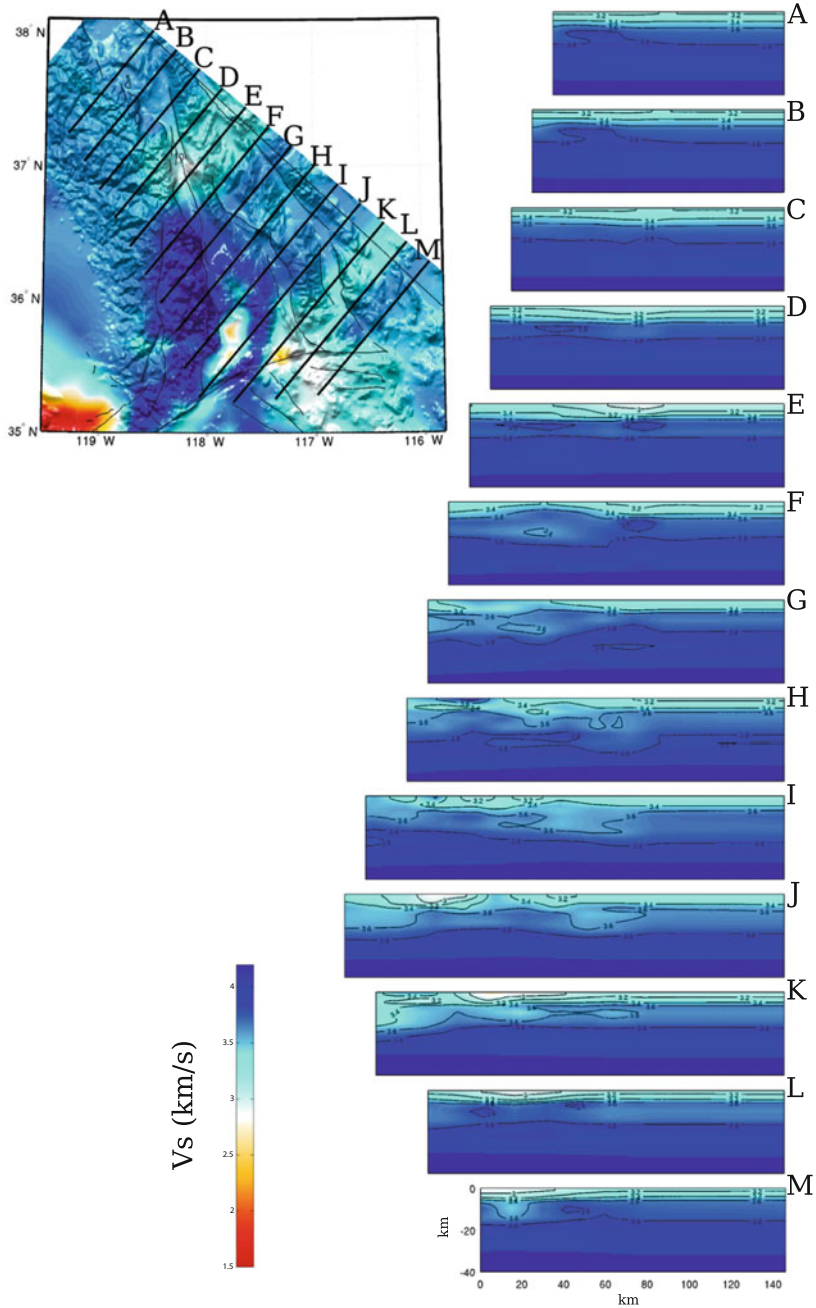


Fig. 6.8 Cross-sections of S-velocity in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.5.

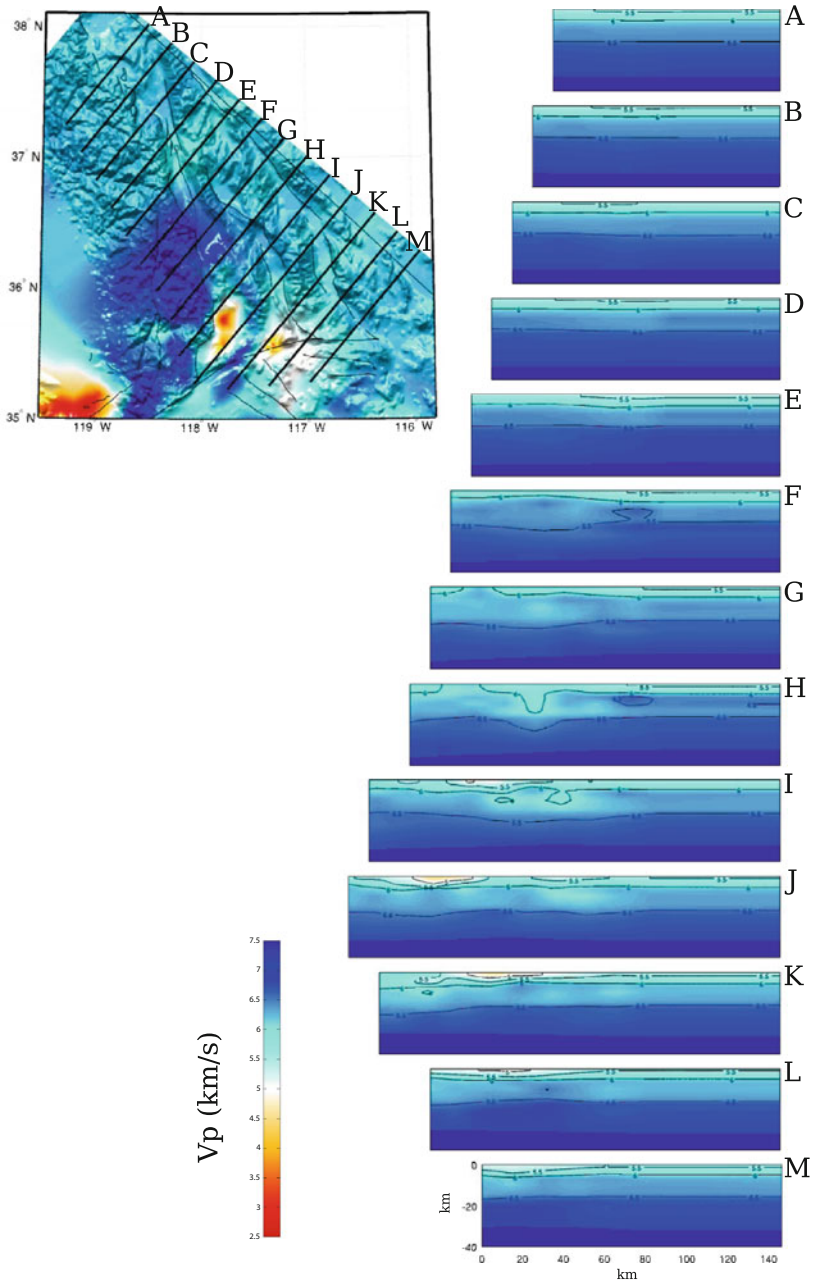


Fig. 6.9 Cross-sections of P-velocity in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.6.

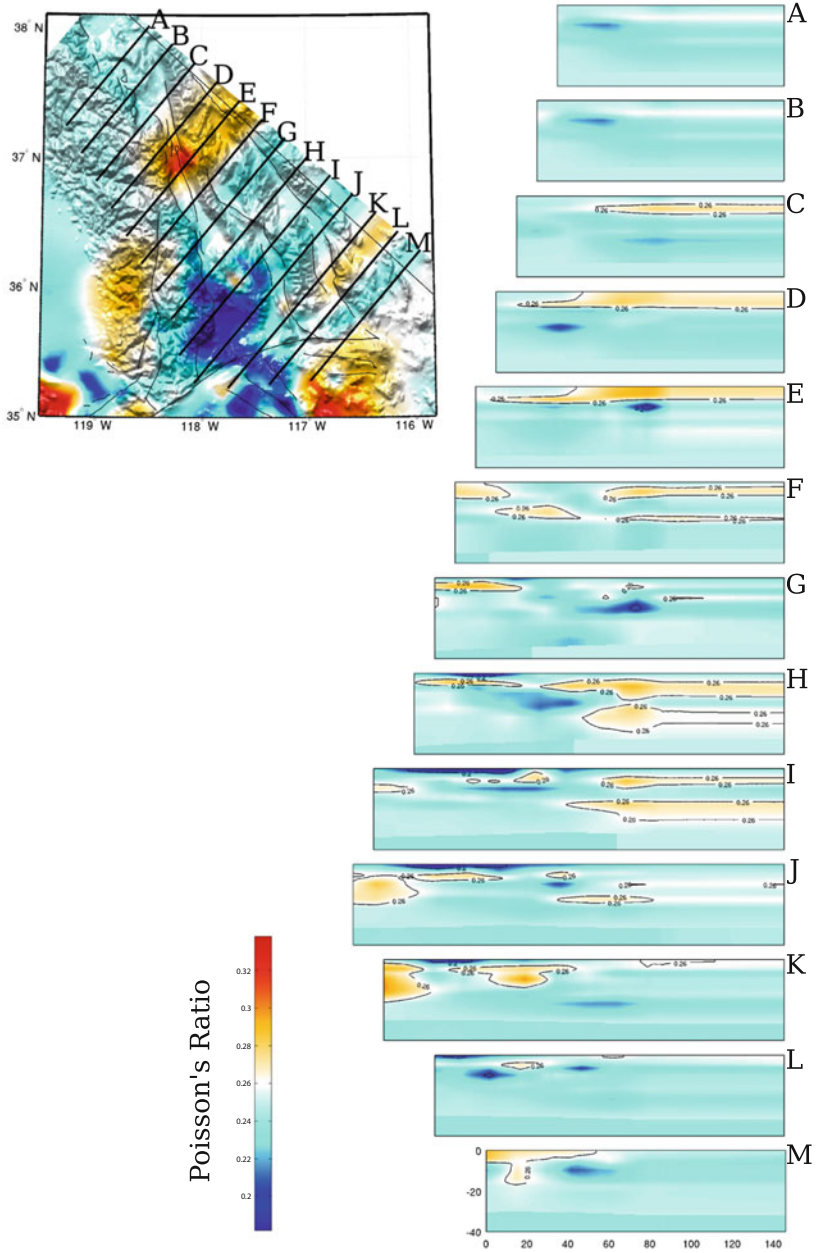


Fig. 6.10 Cross-sections of the Poisson's ratio in CVM-S4 along the 13 profiles. The format is identical to that in Fig. 6.7.

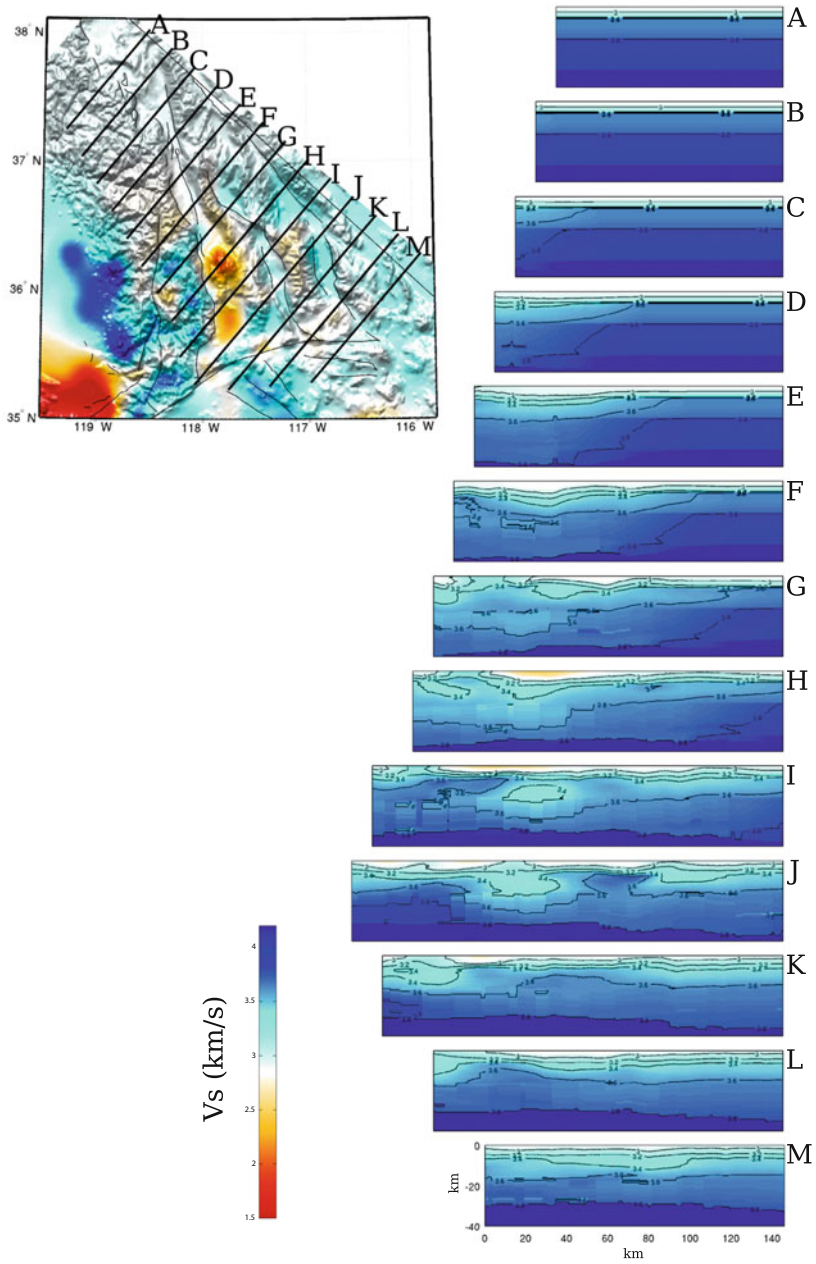


Fig. 6.11 Cross-sections of S-velocity in CVM-H11.9 along the 13 profiles. The format is identical to that in 6.5.

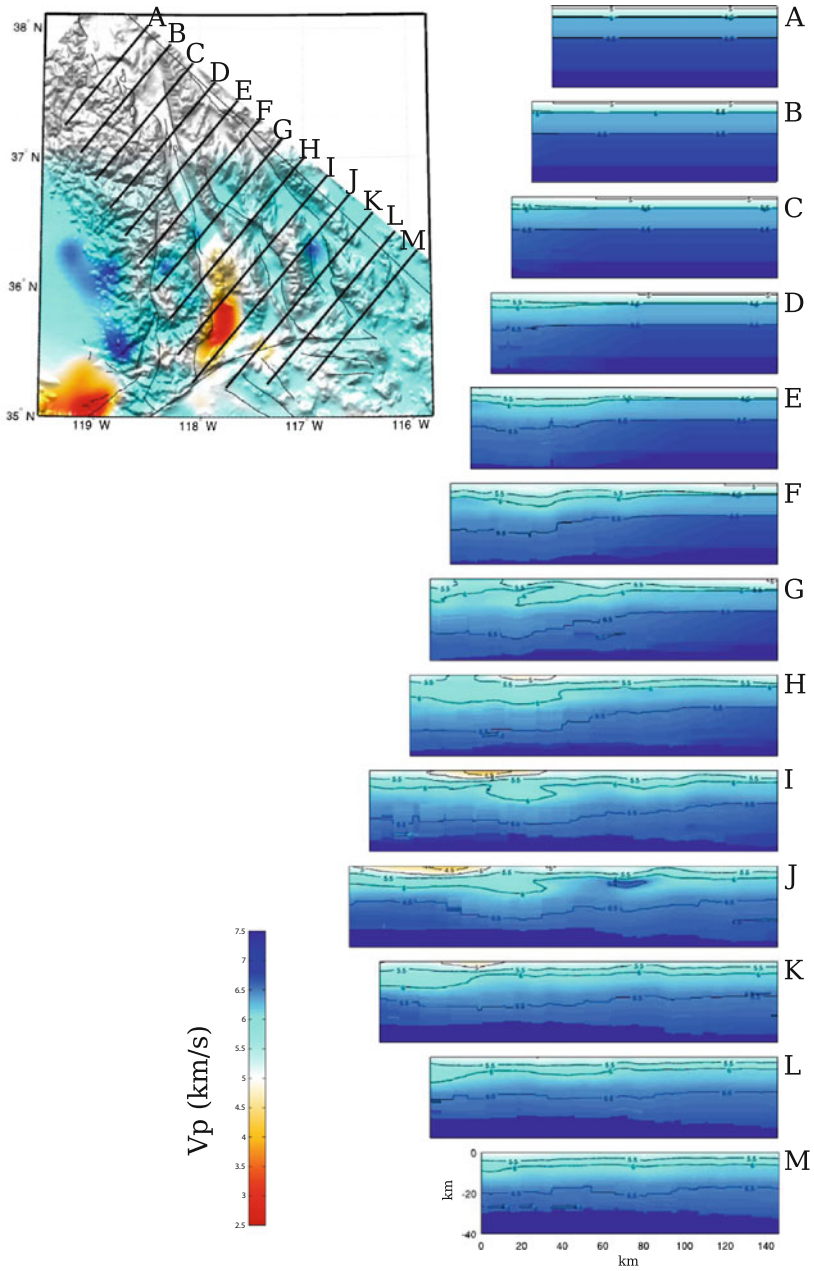


Fig. 6.12 Cross-sections of P-velocity in CVM-H11.9 along the 13 profiles. The format is identical to that in Fig. 6.6.

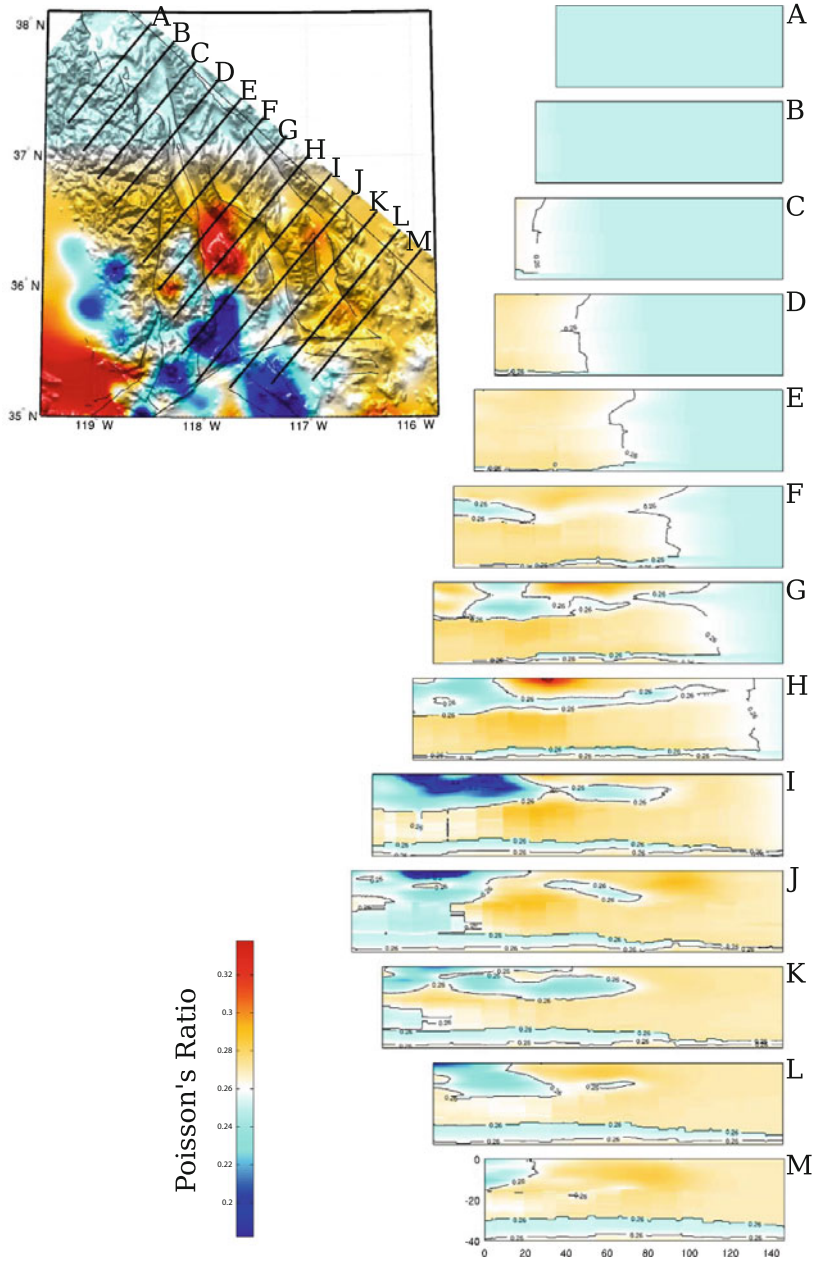


Fig. 6.13 Cross-sections of Poisson's ratio in CVM-H11.9 along the 13 profiles. The format is identical to that in Fig. 6.7.

velocities and above-average Poisson's ratio (Figs. 6.1, 6.2 and 6.3). As will be shown in Sects. 6.2.3.1 and 6.2.3.2, at some locations, CVM-S4.26 is in broad agreement with other localized tomography studies, which also imaged mid- and lower-crustal low-velocity anomalies. The interpretations offered in those localized studies, which include magmatic and hydrothermal activities, perhaps resulting from asthenospheric upwelling beneath this region (e.g., Gazel et al. 2012; Putirka et al. 2012), may also be applicable for explaining other mid- and lower-crustal low-velocity anomalies in this region. A more recent attenuation tomography of the western United States based on the ambient-noise Green's functions (Lawrence and Prieto 2011) has imaged a zone of high-attenuation extending from the Walker Lane south to the Salton Trough. In the lower crust, the high-attenuation zone in the Walker Lane belt was correlated with the low-viscosity compliant zone accommodating a large amount of strain (Hammond et al. 2009) and at shallow crustal depths, the high attenuation was interpreted to be from the increased scattering caused by extensive fractures in this region.

6.2.3.1 Long Valley Caldera

The low-velocity anomaly correlated with the Long Valley Caldera has a diameter of only about 25 km (about 1/40 of the length of our entire tomography region) (Fig. 6.14) and is located very close to the northeastern corner of our tomography box (Fig. 6.4) and we were still able to resolve it through our inversion.

At shallow depths the Long Valley Caldera is filled with the Bishop tuff (i.e., solidified volcanic ashes) from a massive, caldera-forming eruption occurred about 0.76 Ma (Hill 2006). Smaller eruptions fed by the residual magma chamber occurred in 500 Ka, 300 Ka and 100 Ka (Bailey 1976; Hildreth 2004). In 110-40 Ka, volcanic activities shifted to the west forming the Mammoth Mountain. In 40 Ka to about 300 years ago, volcanic activities shifted towards north, forming the Mono-Inyo Craters. Since 1978, the Long Valley Caldera has been experiencing intense unrest with strong earthquake swarms, thermal activities and swelling in the middle of the caldera (e.g., Langbein 2003; Prejean et al. 2003; Hill and Prejean 2005).

There have been geodetic and seismicity evidences supporting the hypothesis that the magma source is in the shallow crust (e.g., Battaglia et al. 1999; Prejean et al. 2002; Battaglia et al. 2003a, b; Langbein 2003; Tizzani et al. 2009). Ray-travel-time tomography using local earthquakes and receiver function analysis (Seccia et al. 2011) revealed a shallow high-velocity anomaly beneath the central-southern section of the resurgent dome, an elongated low-velocity anomaly in the upper crust and a broad low-velocity volume in the mid-crust. The data set used in our inversion did not allow us to resolve the shallow high-velocity anomaly, but the broad mid-crustal low-velocity anomaly is quite clear in CVM-S4.26 (Fig. 6.14) and is also in broad agreement with the teleseismic ray-travel-time tomography results in (Weiland et al. 1995). The cross-sections shown in Fig. 6.14 are at about the same locations as in (Seccia et al. 2011). The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Fig. 6.15.

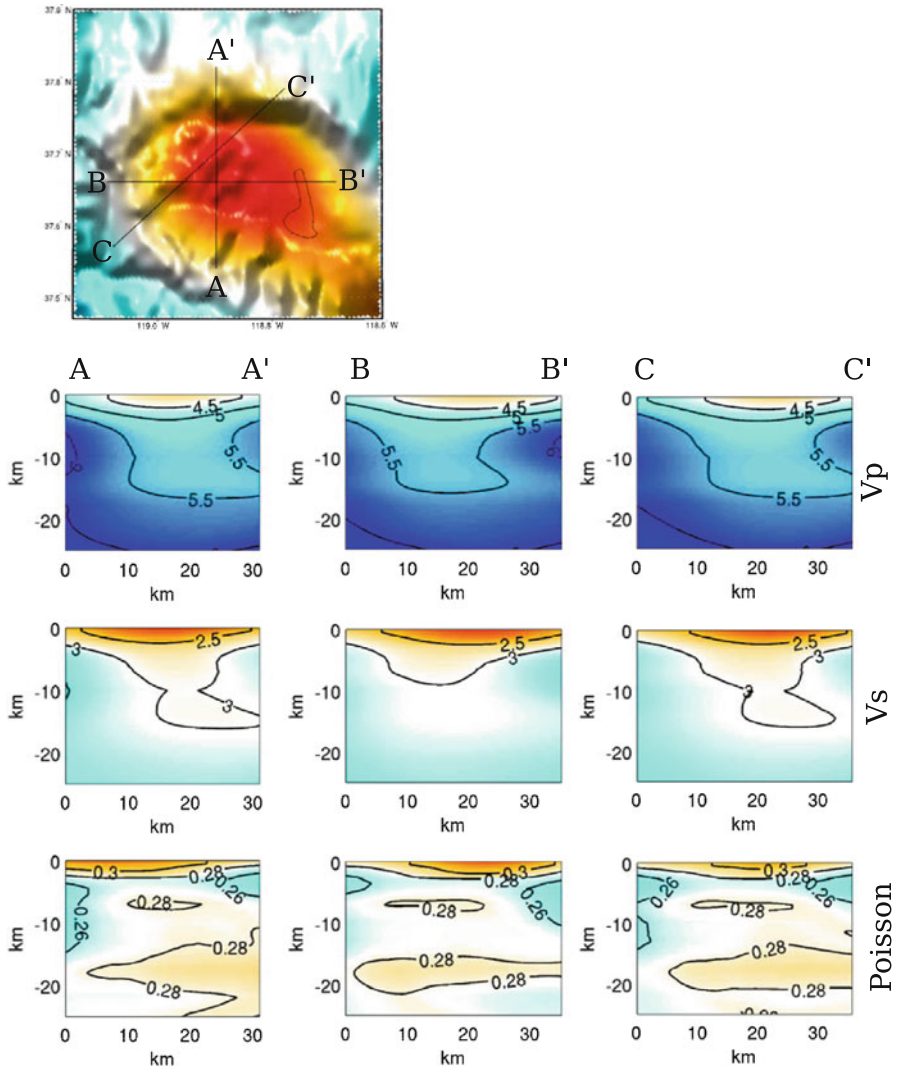


Fig. 6.14 The cross-section views show P- (top row) and S-velocity (center row) and Poisson's ratio (bottom row) in CVM-S4.26 along A-A' (left column), B-B' (center column) and C-C' (right column). The map-view plot at the upper-left corner shows the S-velocity in CVM-S4.26 at 2-km depth.

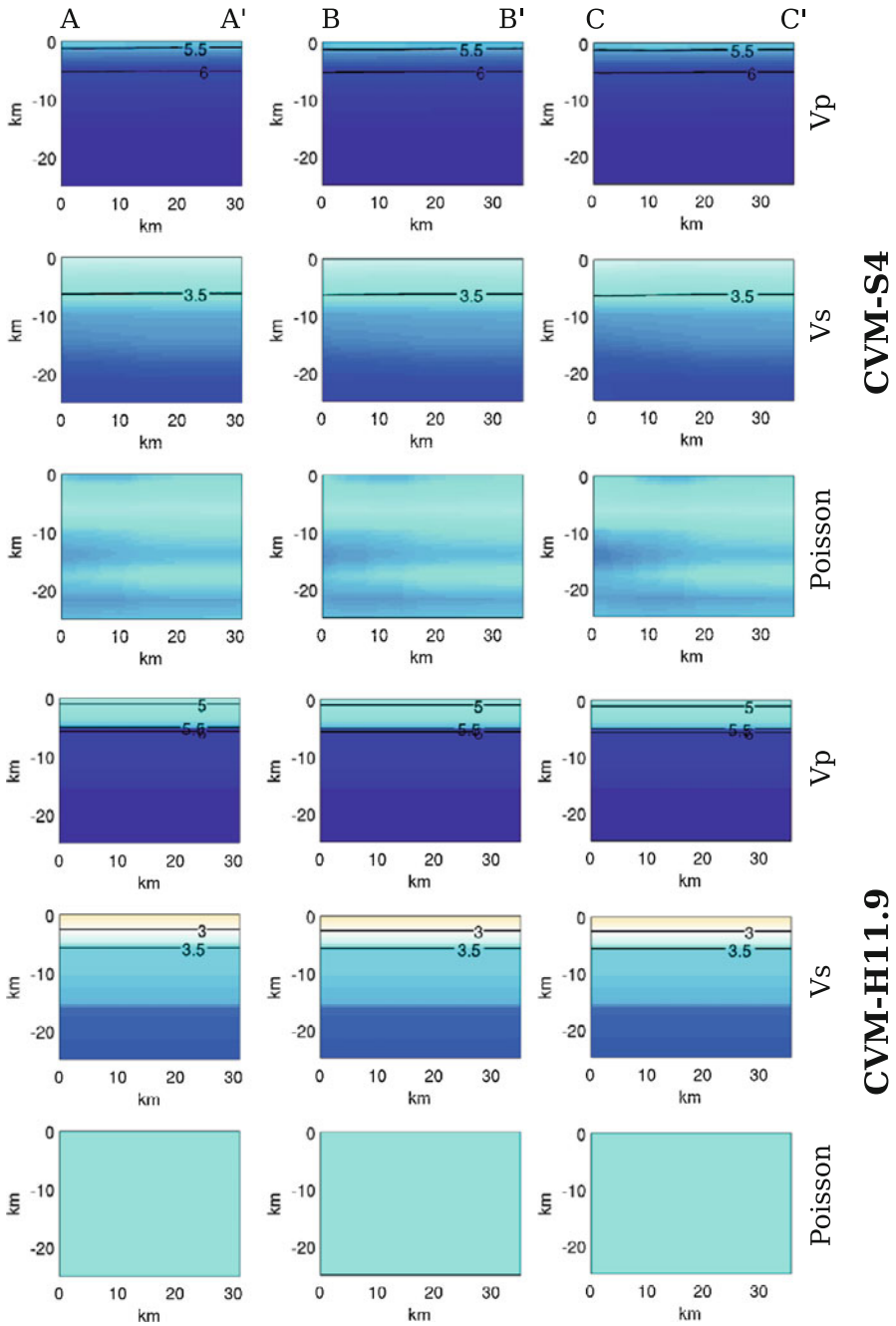


Fig. 6.15 The same cross-sections for CVM-S4 and CVM-H11.9. The format and color-scale are identical to those in Fig. 6.14.

6.2.3.2 Indian Wells Valley—Coso Geothermal Area

The Indian Well's Valley (Fig. 6.4) was also imaged as a prominent low-velocity anomaly on Profile P2 shown in Plate 1b of (Fliedner et al. 2000) and cross-section views of the P- and S-velocity along this profile for CVM-S4.26, CVM-S4 and CVM-H11.9 were shown in Fig. 7 in (Lee et al. 2014b). This low-velocity anomaly associated with the Indian Wells Valley is also consistent with the ray-travel-time tomography results obtained using local earthquakes (Hauksson and Unruh 2007). In (Hauksson and Unruh 2007), the low-P-velocity anomaly associated with the sediments in the Indian Wells Valley is most prominent in the upper 2-km and was imaged down to 4-km depth in some areas. This result is also consistent with the borehole data presented in (Monastero et al. 2002), which shows about 2-km of Tertiary sediments lying above Mesozoic crystalline basement, and also the gravity data analyzed in (Black et al. 2002), which shows that the sediments may reach 3-km depth in some areas.

The tomography result in (Hauksson and Unruh 2007) also shows low-P-velocity anomaly associated with late Cenozoic sediments in the Searles Lake down to at least 2-km depth, which is also consistent with our results shown in Fig. 6.4.

The Poisson's ratio in the Indian Wells Valley and the Searles Lake is unusually low in all three CVMs (2-km map-view in Fig. 6.3). Such a low Poisson's ratio is rare for sediments (e.g., Brocher 2005). However, it is consistent with the V_p/V_s tomography result in (Hauksson and Unruh 2007). The explanation given in (Hauksson and Unruh 2007) was unusual geological conditions, which may include high silica content and textural characteristics of the sediments, such as poor sorting.

In (Hauksson and Unruh 2007), a low-P-velocity anomaly was imaged in the depth range of 8 to 12 km beneath the Coso geothermal field. A set of map-view and cross-sections of P- and S-velocity, V_p/V_s ratio and Poisson's ratio associated with this low-velocity anomaly were shown in Figs. 9 and 10 in (Hauksson and Unruh 2007). In Figs. 6.16, 6.17 and 6.18, I show the map-views and cross-sections extracted from CVM-S4.26, CVM-H11.9 and CVM-S4, respectively, at about the same locations as in (Hauksson and Unruh 2007). The mid-crustal low-velocity body is also well imaged in CVM-S4.26 (Fig. 6.16). In CVM-H11.9 (Fig. 6.17), this low-velocity anomaly is shown as a vertical extrusion from the surface, rather than a more or less isolated low-velocity body as in Fig. 6.16 and in (Hauksson and Unruh 2007). The low-velocity body in the map-view plots in Fig. 6.16 appears shifted slightly to the north when compared with the corresponding map-view plots shown in (Hauksson and Unruh 2007). This might be due to the variable topography of the top surface of this low-velocity body in CVM-S4.26, as shown in the cross-section views in Fig. 6.16. The correlation in Poisson's ratio between CVM-S4.26 and CVM-H11.9 is quite high down to about 15-km depth and is also in broad agreement with that in (Hauksson and Unruh 2007). Below 15-km depth, the Poisson's ratio in CVM-S4.26 has better correlation with that in (Hauksson and Unruh 2007) than with the Poisson's ratio in CVM-H11.9. This mid-crustal low-velocity anomaly is not well imaged in CVM-S4 (Fig. 6.18).

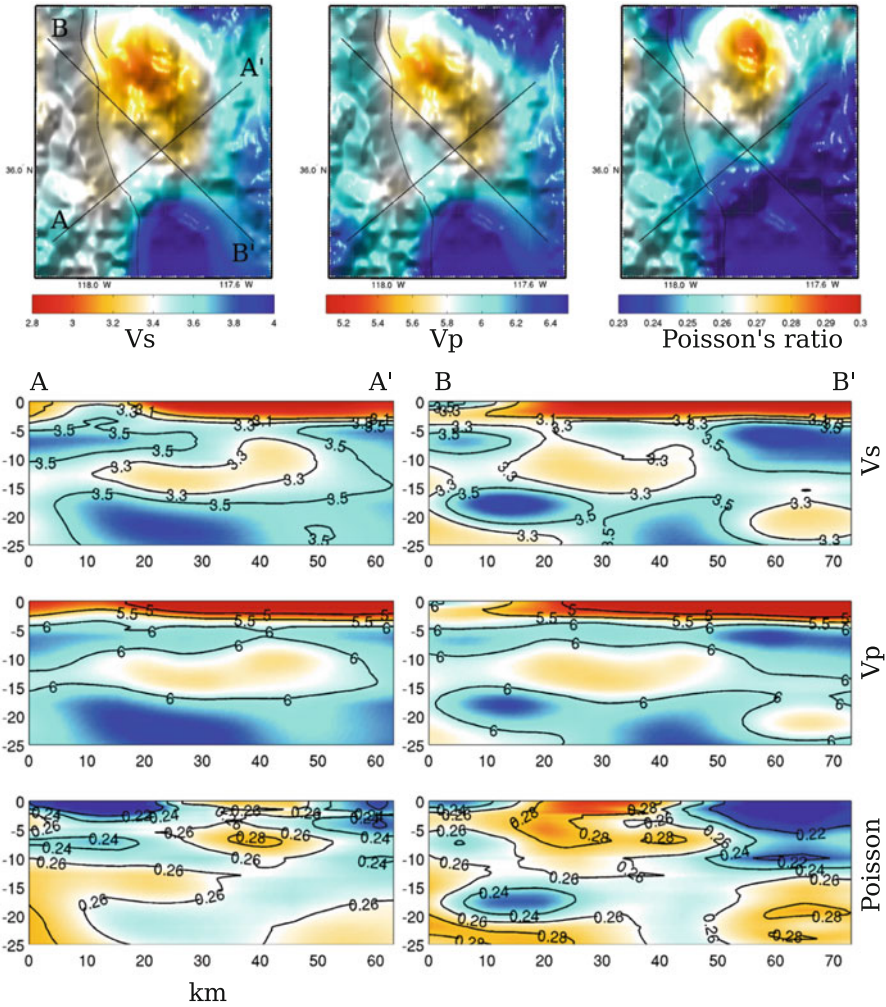


Fig. 6.16 The map-view plots (*top row*) show the S-velocity (*left*), P-velocity (*center*) and Poisson's ratio (*right*) at 10-km depth beneath the Coso geothermal area in CVM-S4.26. Locations of the two cross-sections, A-A' and B-B', are plotted on the S-velocity map-view. Cross-section plots show the S-velocity (*top row*), P-velocity (*center row*) and Poisson's ratio (*bottom row*) along A-A' (*left column*) and B-B' (*right column*).

Earlier results that suggest the presence of a mid- to lower-crustal low-velocity body beneath the Coso geothermal area were obtained using teleseismic data in (Reasenberget al. 1980), using receiver functions in (Wilson et al. 2003). One possible interpretation suggested in previous studies was the presence of partial melting associated with a magma chamber.

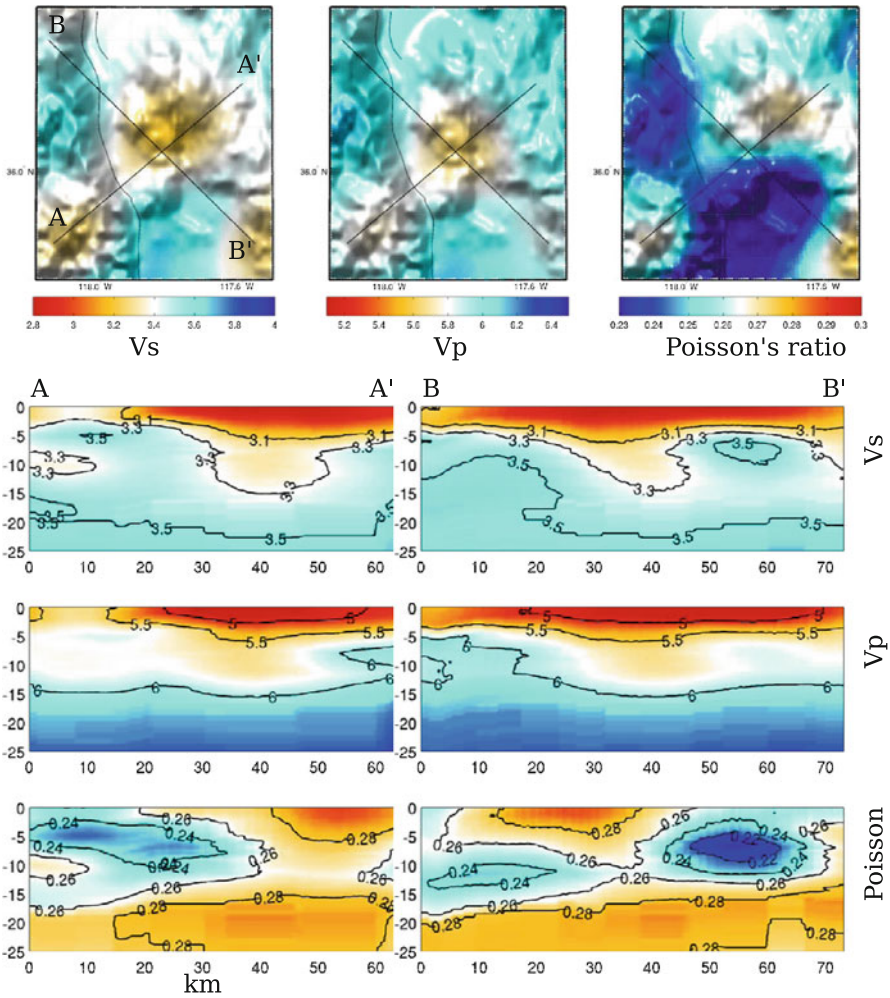


Fig. 6.17 Map-views and cross-sections extracted from CVM-H11.9. The format and color-scale are identical to those in Fig. 6.16.

6.2.3.3 Owens Lake

To the north of the Coso Mountains lies the Owens Lake, which is shown as a low-velocity anomaly in Fig. 6.4. Owens Lake was over 200 feet deep during the Pleistocene and is now a playa due to the loss of recharge water through climate change, irrigation and diversion of water from the Owens Valley to the Los Angeles Aqueduct (e.g., Von Huene et al. 1963). Gravity and seismic survey in the southern Owens Valley indicates that the depth to the bedrock ranges from about 1 km to

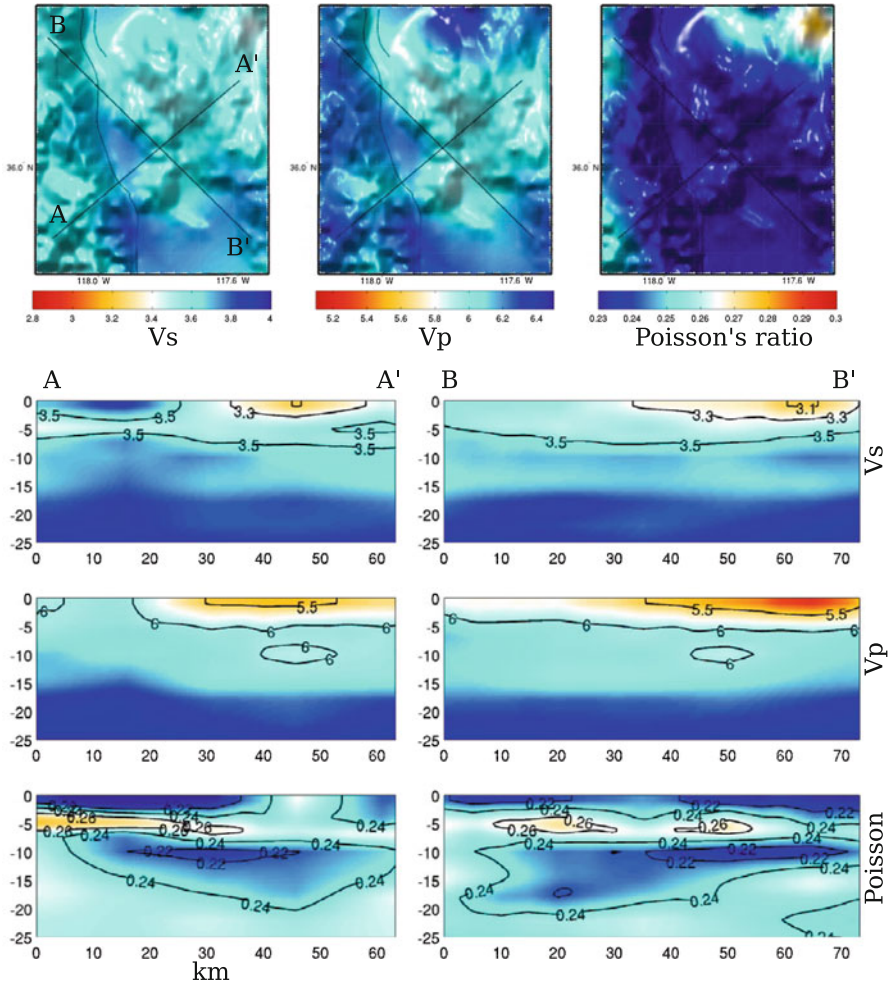


Fig. 6.18 Map-views and cross-sections extracted from CVM-S4. The format and color-scale are identical to those in Fig. 6.16.

about 3 km below the surface and the sediments are particularly thick east of the Alabama Hill near Lone Pine and beneath Owens Lake (Kane and Pakiser 1961).

The sediments filling the Owens Valley were eroded from the surrounding mountain ranges. The Owens Valley is a typical graben on the western edge of the Basin and Range province, formed by the transensional forces pulling the western part of California away from the stable North American interior (e.g., Faults and Henry 2008; Jayko and Bursik 2012). The surrounding mountain ranges have been undergoing significant uplifting (e.g., Ducea and Saleeby 1996; Fliedner et al. 1996; Wernicke et al. 1996; Ducea and Saleeby 1998; Zandt et al. 1998), exposing the

granitic plutons making up the core of the mountains to active weathering and erosion by ice, water and other forces. The resulting sediments have been carried down to the adjacent deep valleys.

The cross-section G in Fig. 6.5 shows a high-velocity anomaly beneath the Owens Lake basin. If this high-velocity anomaly also has higher density, as suggested in e.g., (Brocher 2005), the subsidence of the Owens Lake basin may have been driven, perhaps in part, by the isostatic response.

6.2.3.4 Panamint Range

The strongest high-velocity anomaly in CVM-S4.26 in this region is at the Panamint Range (Fig. 6.4), which is closely related to the formation of the Death Valley immediately to the east. The metamorphic rocks on the surface of the range have greenschist facies, representing a burial depth of around 10 km in the Jurassic (Labotka and Albee 1990).

Several hypotheses about the formation of the Death Valley and the Panamint Range have been discussed in the literature and they often involve low-angle detachment faulting (e.g., Stewart 1983; Wernicke et al. 1988; Snow and Wernicke 2000) and/or strike-slip pull-apart mechanisms (e.g., Burchfiel and Stewart 1966; Topping 1993; Serpa and Pavlis 1996; Miller and Prave 2002; Norton 2011). In the low-angle detachment faulting models, the Panamint Range was originally located on top or to the east of the Black Mountains, which is presently on the east of the Death Valley. The Panamint Range was part of the hanging wall and the Black Mountains were on the footwall. The detachment fault transported the Panamint Range westward for over 80 km to its current location. In the strike-slip pull-apart models, the detachment surfaces are normal faults linking the strike-slip faults.

The more recent interpretation in (Norton 2011) argues for a two-stage process. In the first stage (Miocene-Early Pliocene), the Panamint Range was formed as a large metamorphic core complex that exhumed mid- and lower-crustal rocks. This type of core complexes were proposed in studying extensional deformation in the Basin and Range province. In the second stage (last 3 Ma) the Death Valley was formed as a strike-slip pull-apart basin.

6.2.4 Southern Sierra Nevada

Our tomography box for CVM-S4.26 covers the southern portion of the Sierra Nevada Batholith (SNB) (Fig. 6.19). The SNB, together with the Southern California Batholith (SCB) (e.g., Luffi et al. 2009) and the Peninsular Ranges Batholith (PRB) (e.g., Schmidt et al. 2002) to the south, are part of the Cordilleran batholithic belt. The bulk of the batholithic belt was emplaced during the Mesozoic as a result of prolonged subduction of oceanic plates beneath the southwestern edge of the North American plate (e.g., Dickinson 1981).

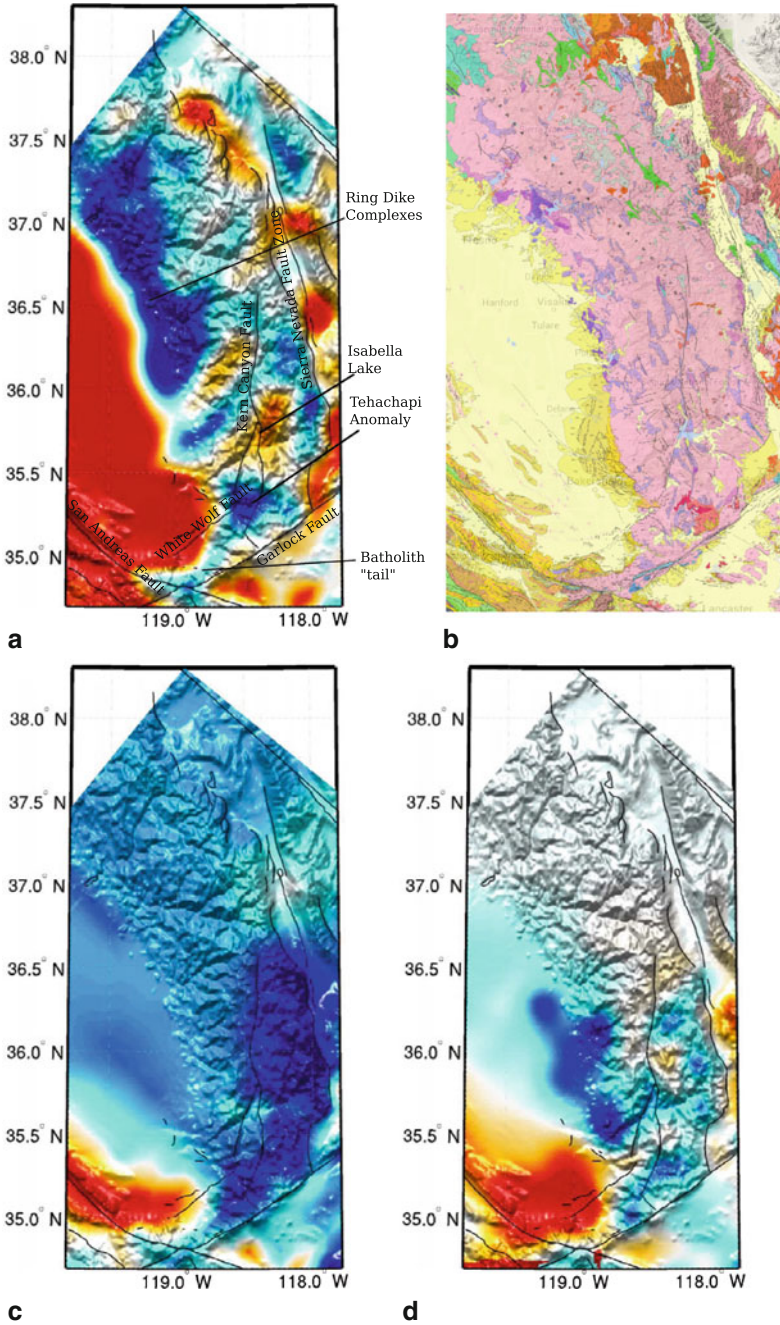


Fig. 6.19 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in and around the southern Sierra Nevada. The color scale is identical to that in the 2-km map-view in Fig. 6.1.

The S-velocity at 2-km depth in CVM-S4.26 shows good correlation with surface geology in this region (Fig. 6.19). In CVM-S4.26, the southern SNB shows above-average velocities and the eastern (i.e., the Sierra Nevada Fault zone) and western (i.e., the Sierra Nevada Foothills) boundaries of the high-velocity region associated with the southern SNB are well aligned with the boundaries of the SNB in the geologic map. At the southern end of the SNB, the surface distribution of the batholith extends southwestward along the Garlock fault and then bends northwestward along the San Andreas Fault in the geologic map (Fig. 6.19b). In CVM-S4.26, the narrow high-velocity anomaly associated with this batholith “tail” (Fig. 6.19a) follows almost exactly the same path as in the geologic map. The narrowest part of the batholith “tail” is less than 10-km wide.

Inside the high-velocity region associated with the SNB, there are a few low-velocity areas. One low-velocity anomaly is located around the Isabella Lake. Previous hydrologic studies have revealed extensive fracture networks associated with inter-connected faults in the crystalline bedrocks of the Kern Plateau (Howard et al. 1997; Ostdick 1997), especially in regions northeast of the Isabella Lake (Thyne et al. 1999). The lowest velocities inside this low-velocity anomaly in CVM-S4.26 is also located at about the same area (Fig. 6.19a).

6.2.4.1 SNB Mafic Cumulates

Geological and geochemical evidences suggest that the SNB is composed of a western domain and an eastern domain and the western domain has compositions that are more mafic than its eastern counterpart (e.g., Saleeby et al. 2003). One of the observations that have been used to support this interpretation in e.g., (Saleeby et al. 2003) is a well-exposed belt of intrusives lying between latitudes 36°N and 37°N (Fig. 6.19a), where a series of ring-dike complexes have been carefully studied in e.g. (Mack et al. 1979; Saleeby and Sharp 1980; Clemens-Knott and Saleeby 1999; Clemens-Knott et al. 2000). The ring-dike complexes, which are considered to be relics of Early Cretaceous basalt-andesite stratovolcanic centers, are composed of olivine-bearing hydrous mafic and ultramafic cumulates and crescent- to ring-shaped dikes of gabbro-norite and pyroxene quartz gabbrodiorite. In CVM-S4.26, a strong high-velocity anomaly lies at the same location as the exposed belt of intrusives (Fig. 6.19a). This high-velocity anomaly in CVM-S4.26 also correlates very well with gravity and magnetic anomalies in this region (e.g., Fig. 3b in Lee et al. 2014b).

In three-dimension, this belt of intrusives exposed on the surfaces appears to be connected with a much larger high-velocity body extending to at least 25-km depth. In Figs. 6.20–6.22, I show cross-section views of the S-velocity, P-velocity and Poisson’s ratio in CVM-S4.26 along 9 parallel profiles crossing the southern SNB. The same cross-sections for CVM-S4 and CVM-H11.9 are also shown in Figs. 6.23–6.28. The color scales used for plotting these cross-sections are identical to those used in Figs. 6.5–6.13 in Sect. 6.2.3. A high-velocity body with thickness over 20 km lies under the surface exposures of the intrusives and extends westwards underneath the southern San Joaquin Basin. In general, the Poisson’s ratio

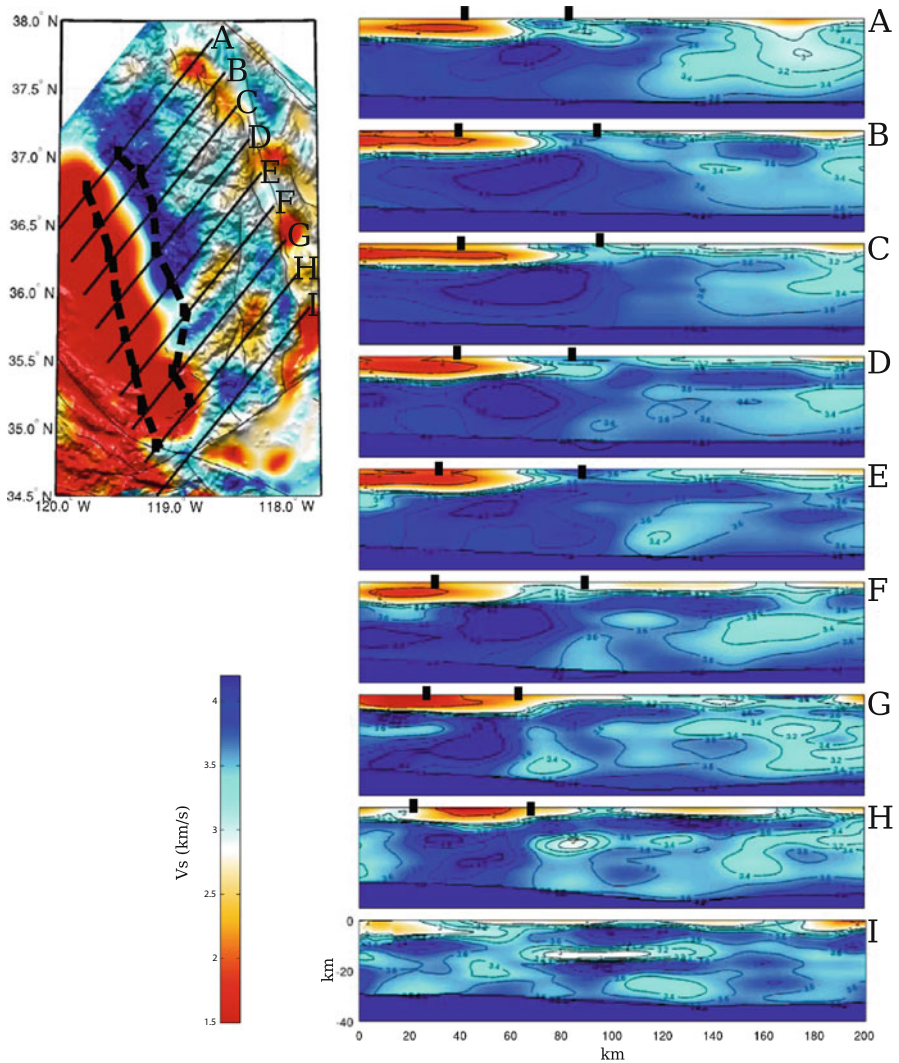


Fig. 6.20 Cross-section views of the S-velocity in CVM-S4.26 along 9 parallel profiles (A-I). The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the color scale identical to that in the 2-km map-view of Fig. 6.1. The color scale for the cross-sections is displayed in the *lower-left corner* and is identical to that used in Fig. 6.5. Black dots on the cross-sections indicate the approximate lateral extent of the high-velocity body. *Black dash lines* on the map-view are obtained by connecting the *black dots* on the cross-sections.

of this high-velocity body is below-average and much lower than those encountered under the southern Walker Lane Belt and eastern SNB. A rough estimate of the lateral extent of this high-velocity body is indicated using thick dash-lines on the map-views in Figs. 6.20–6.22 and also using black dots on the cross-sections in

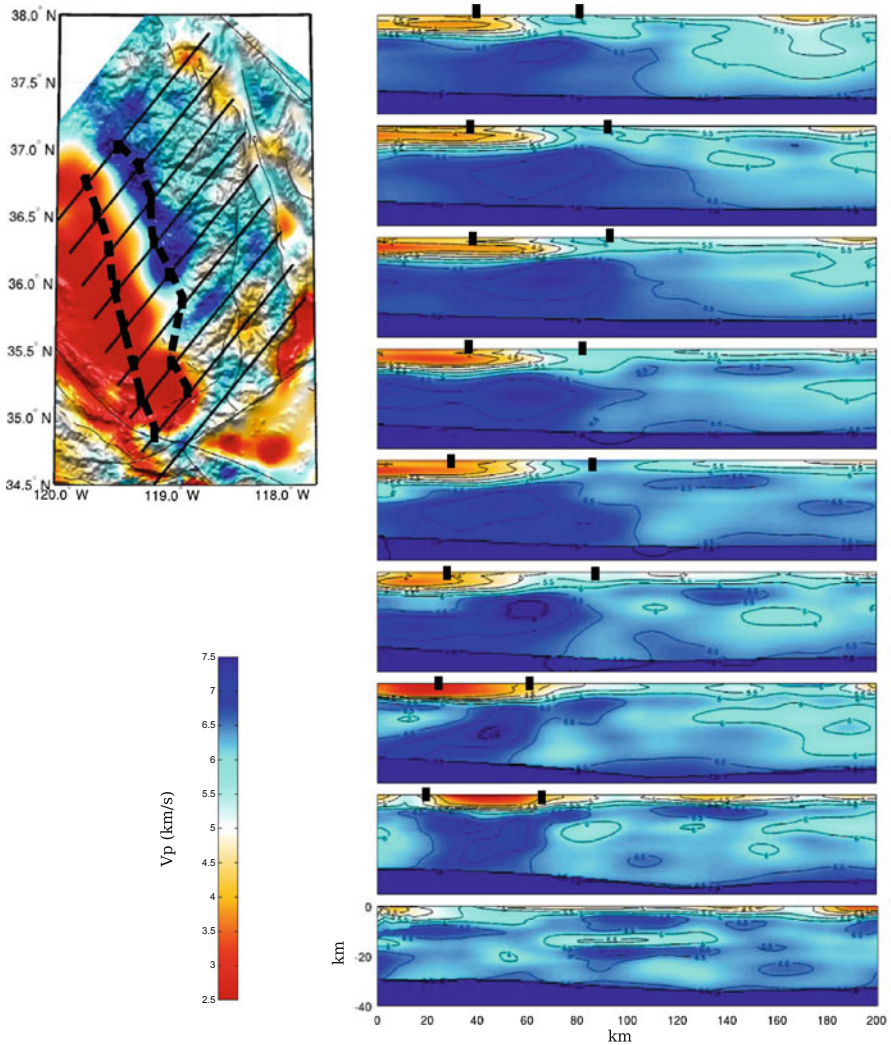


Fig. 6.21 Cross-sections of P-velocity in CVM-S4.26. The format is identical to that in Fig. 6.20.

Figs. 6.20–6.22. The lateral extent of this high-velocity body can also be examined on the 14- to 24-km map-views in Figs. 6.1 and 6.2.

Cross-sections passing through this thick high-velocity body have been obtained before through 2D active-source refraction tomography in (Fliedner et al. 2000) (see P5, P6 in Fliedner et al. 2000 and Fig. 7 in Lee et al. 2014b). This high-velocity body was not well-imaged along P1 in (Fliedner et al. 2000) due to poor ray-path coverage passing through this high-velocity body (gray area in P1 in Fliedner et al. 2000). In CVM-S4.26, this high-velocity body is well imaged in full-3D (Figs. 6.20–6.22 and P1, P5 and P6 in Fig. 7 in Lee et al. 2014b).

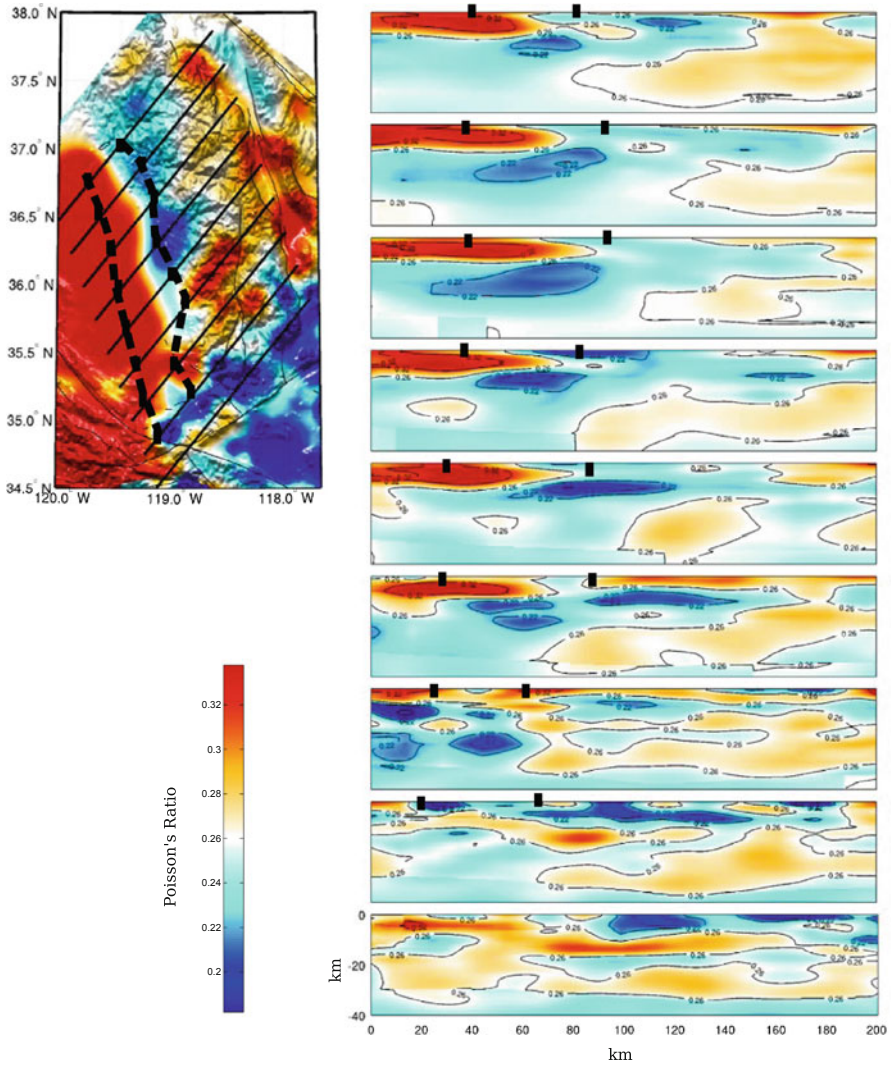


Fig. 6.22 Cross-sections of Poisson's ratio in CVM-S4.26. The format is identical to that in Fig. 6.20.

The compositional differences between the western and the eastern SNB can be attributed to the differences in the pre-batholithic metamorphic framework (e.g., Saleeby et al. 2003), which was composed of Paleozoic oceanic lithosphere to the west and the North American Proterozoic lithosphere to the east (Kistler and Peterman 1973; DePaolo 1981; Saleeby et al. 1987; Kistler 1990; Chen and Tilton 1991; Clemens-Knott et al. 1991; Coleman et al. 1992; Pickett and Saleeby 1994; Sisson et al. 1996). The Early Cretaceous batholith of the western domain was generated

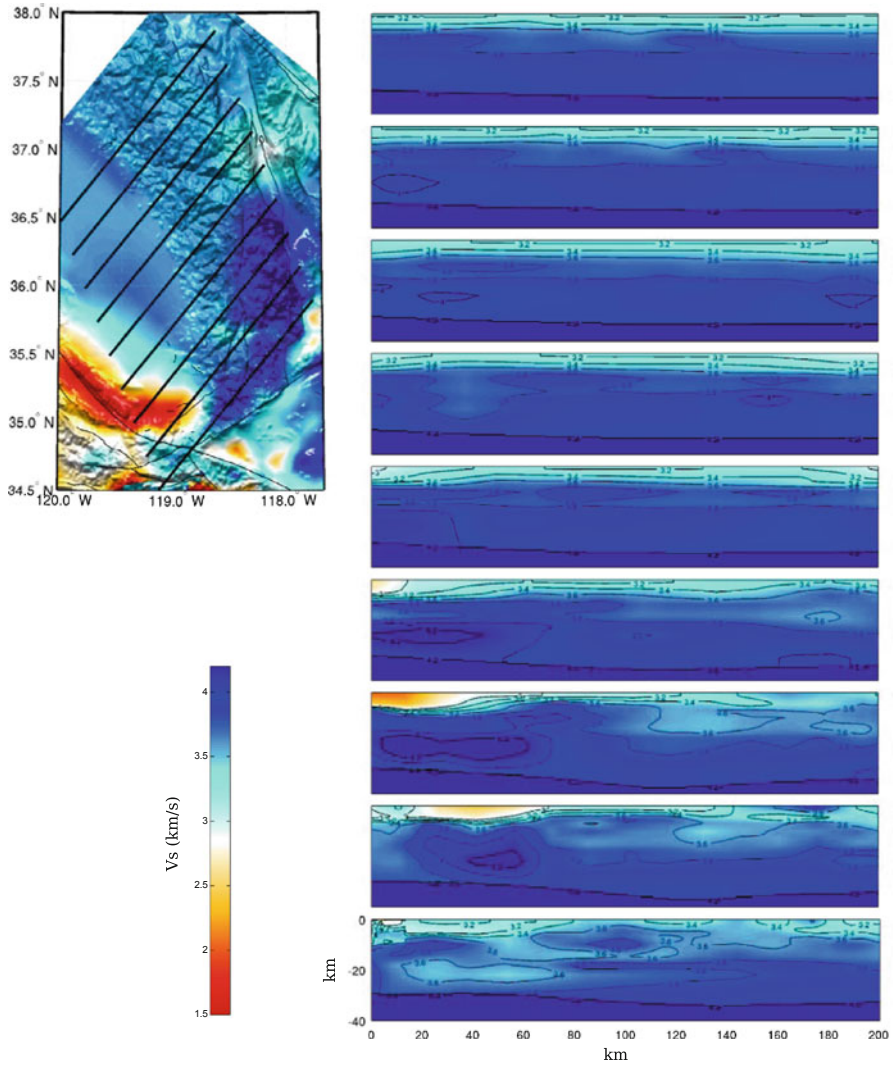


Fig. 6.23 Cross-sections of S-velocity in CVM-S4. The format is identical to that in Fig. 6.20.

primarily in a depleted mantle regime and ascended through an oceanic lithosphere that had been sutured with the North American lithosphere. Its composition was primarily mafic and similar to intro-oceanic arcs (Saleeby et al. 2003).

In CVM-S4.26, this thick high-velocity body is connected with a much thinner, slab-like high-velocity layer further to the west (Figs. 6.20–6.22). The western boundary of the thick high-velocity body (black dash lines and black dots in Figs. 6.20–6.22) was drawn at the estimated location of this connection. This thin

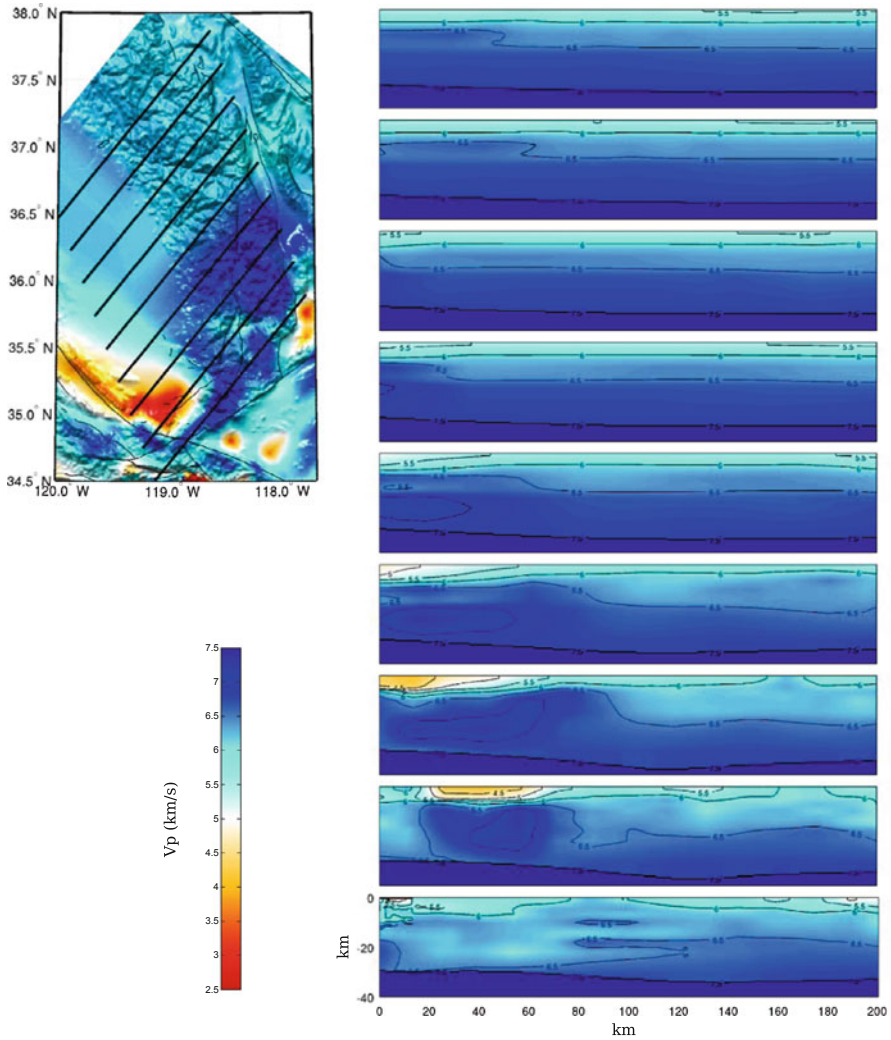


Fig. 6.24 Cross-sections of P-velocity in CVM-S4. The format is identical to that in Fig. 6.20.

high-velocity layer may have a different interpretation from the thick high-velocity body and will be discussed in Sect. 6.2.5.

To the east of the thick high-velocity body, the mid- to lower-crustal seismic velocities in CVM-S4.26 are generally below the laboratory measurements of felsic granulite (e.g., Christensen and Mooney 1995) and resemble those under the southern Walker Lane Belt (Figs. 6.5–6.7). The active-source studies in

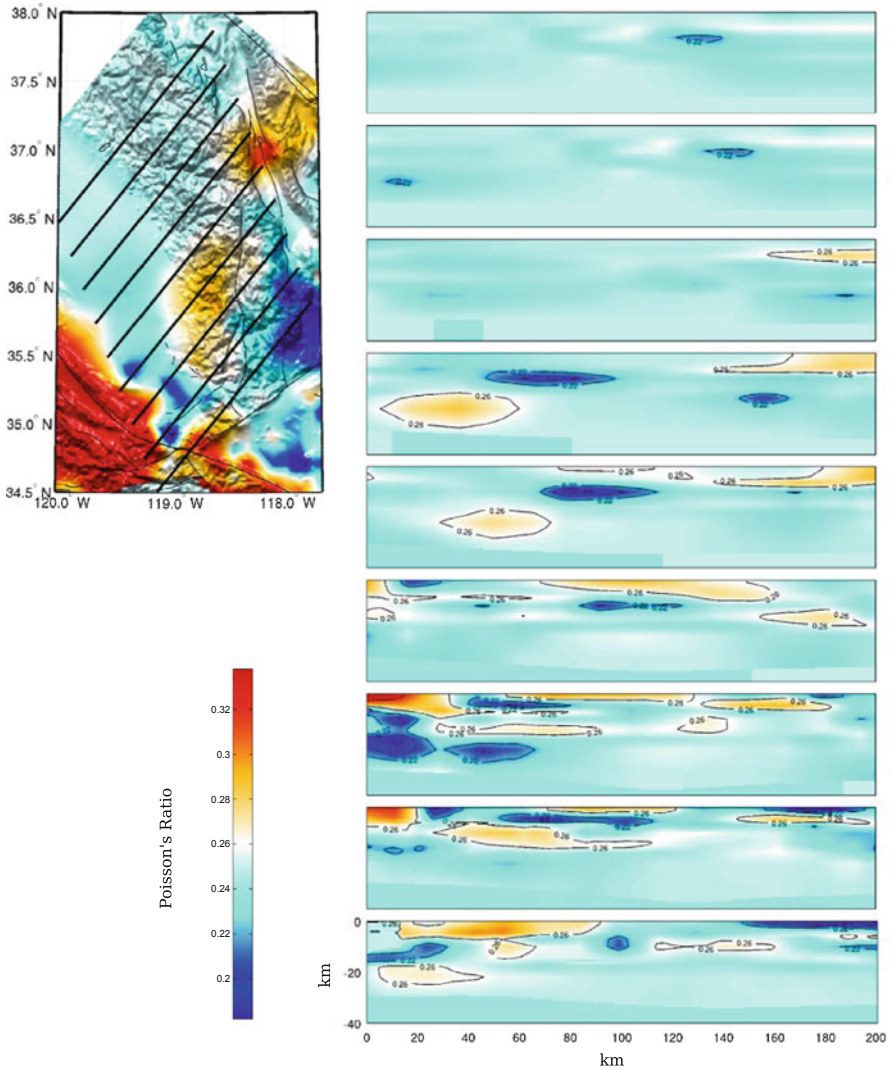


Fig. 6.25 Cross-sections of Poisson's ratio in CVM-S4. The format is identical to that in Fig. 6.20.

(Fliedner et al. 2000) also found below-average mid-crustal P-velocities under mostly the eastern SNB and one explanation suggested in (Fliedner et al. 2000) was increased temperature due to a heat pulse from the mantle. Considering our results in the southern Walker Lane Belt, such a heat pulse may have been generated by the very recent asthenospheric upwelling (e.g., Gazel et al. 2012; Putirka et al. 2012). However, different from the southern Walker Lane Belt, a persistent high-velocity layer lies at about 10-km depth underneath the eastern SNB (e.g., profile D-H in Fig. 6.20), perhaps because the lithosphere degradation process is only partly complete under eastern Sierra Nevada, as suggested in e.g., (Putirka et al. 2012).

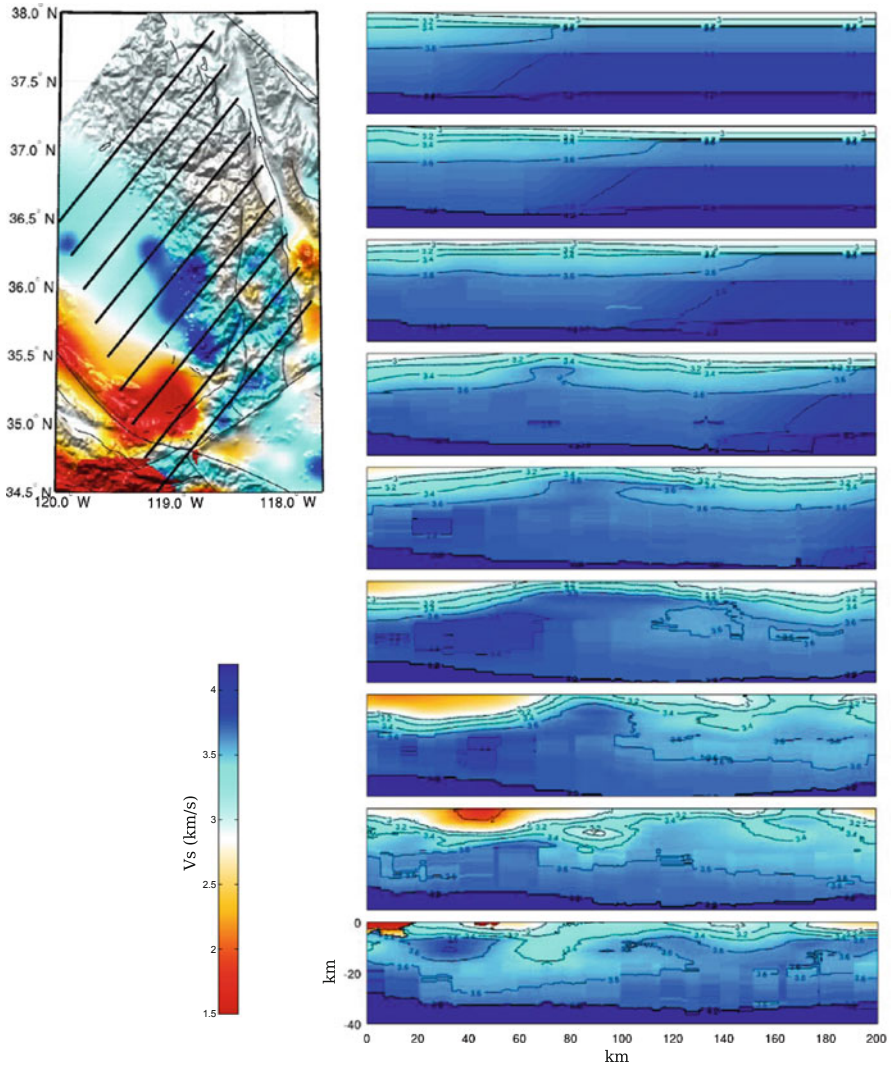


Fig. 6.26 Cross-sections of S-velocity in CVM-H11.9. The format is identical to that in Fig. 6.20.

6.2.4.2 Tehachapi Anomaly

Thermobarometric constraints show that the crustal section of the southernmost Sierra Nevada is tilted (e.g., Fig. 6 in Saleeby et al. 2003) with the level of exposure increasing from about 0.2-GPa pressure condition (about 6-km depth) at the southern region of Cretaceous-Paleogene orogenic plateau (e.g., House et al. 2001) to about 1-GPa pressure condition (about 30-km depth) at western Tehachapi mountains (e.g., Malin et al. 1995). This tilted crustal section has allowed direct

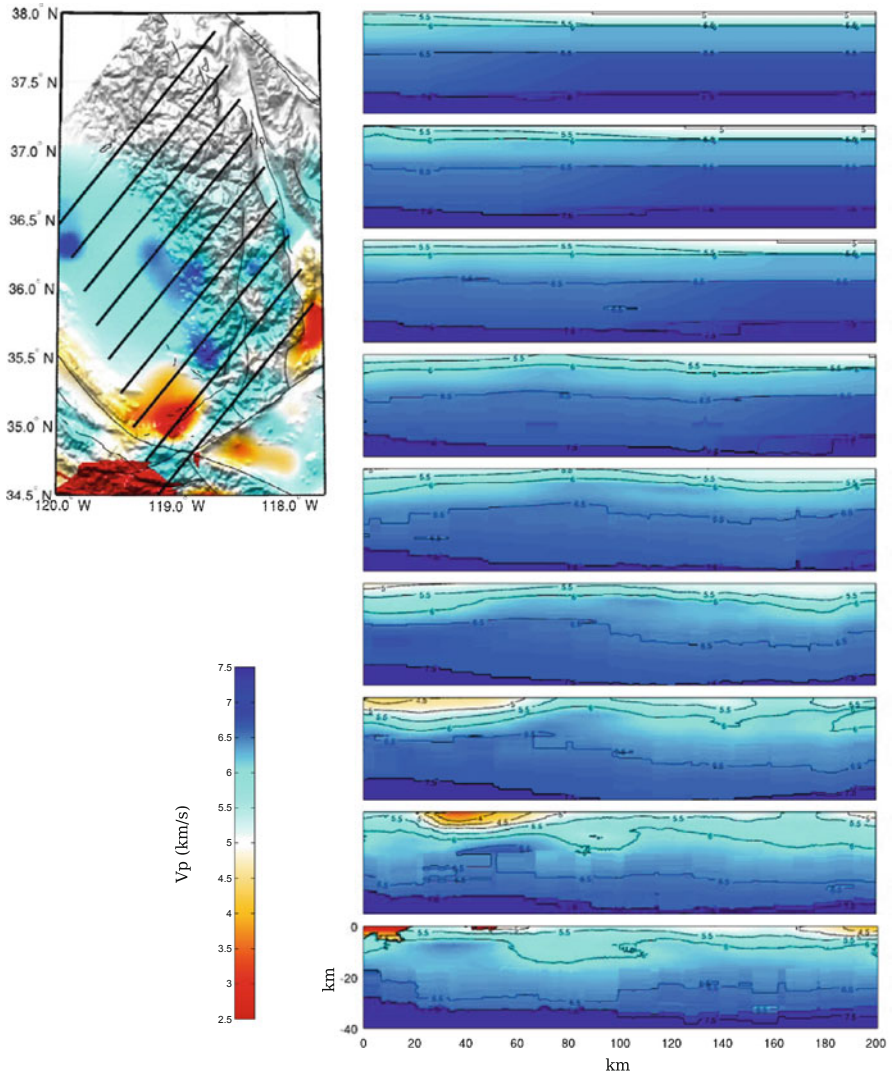


Fig. 6.27 Cross-sections of P-velocity in CVM-H11.9. The format is identical to that in Fig. 6.20.

observations of the mid- to lower-crustal batholithic rocks (Moore 1959; Saleeby 1981, 1990; Clemens-Knott 1992; Saleeby et al. 2003).

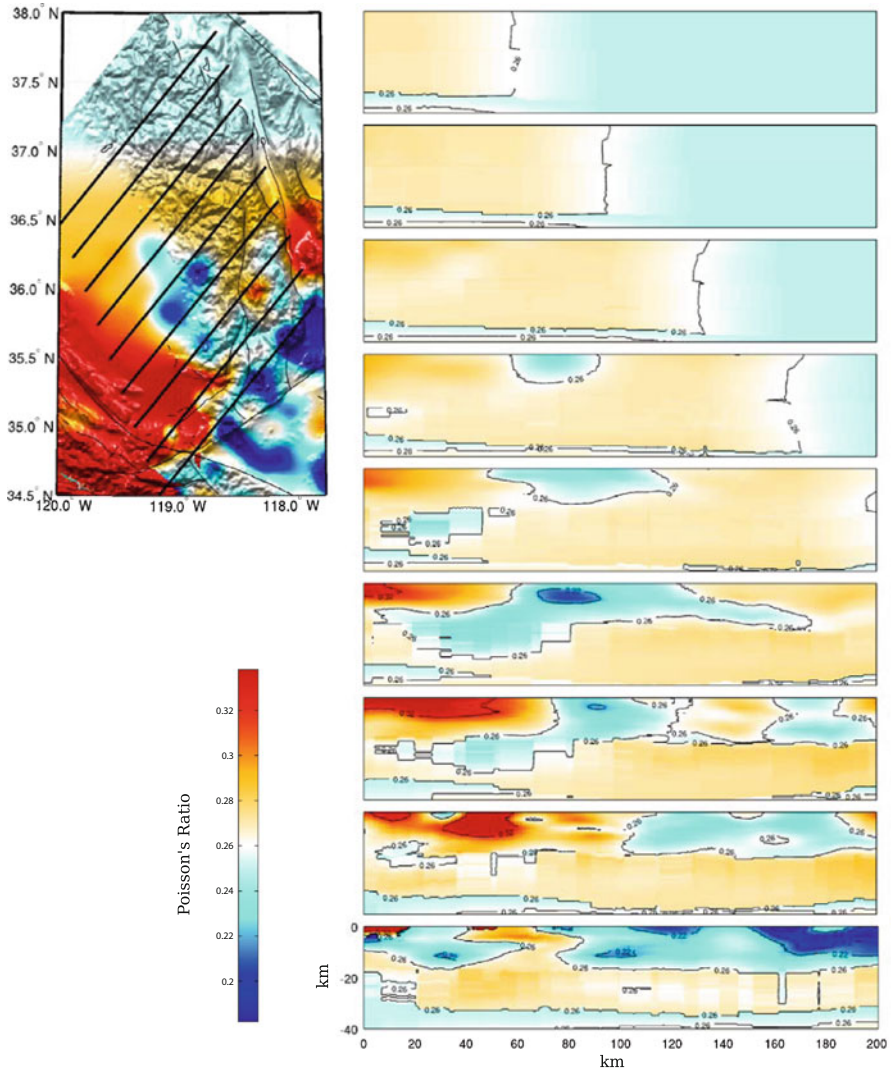


Fig. 6.28 Cross-sections of Poisson's ratio in CVM-H11.9. The format is identical to that in Fig. 6.20.

In Figs. 6.29–6.31, I show cross-sections of the S-velocity, P-velocity and Poisson's ratio in CVM-S4.26 along 7 profiles crossing the Tehachapi mountains and the surroundings. The same cross-sections in CVM-S4 and CVM-H11.9 are shown in Figs. 6.32–6.37. Along profiles B-D, the velocity contours bend upward between the White Wolf Fault and the Garlock Fault (Figs. 6.29–6.30), which is consistent with the findings in the active-source reflection/refraction studies along the CALCRUST profile in the same region (Malin et al. 1995). The CALCRUST study found

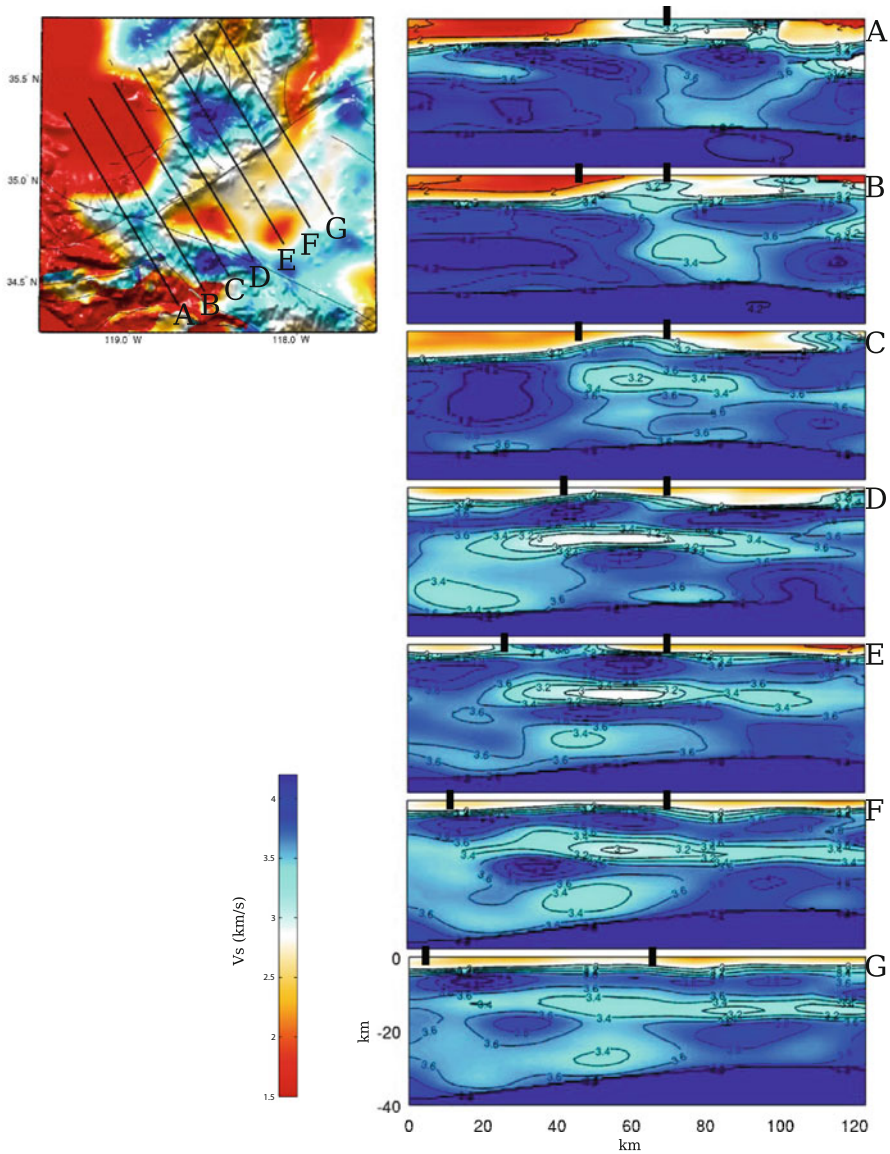


Fig. 6.29 Cross-sections of S-velocity in CVM-S4.26 along the 7 profiles (A-G). The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of surface traces of faults. The black dots at horizontal distance of about 65 km are for the Garlock Fault. The black dots at horizontal distance of about 45 km on profile B-D are for the White Wolf Fault. The black dots left of the Garlock fault on profile E-G are for the Kern Canyon Fault. Locations of the faults are shown on Fig. 6.19.

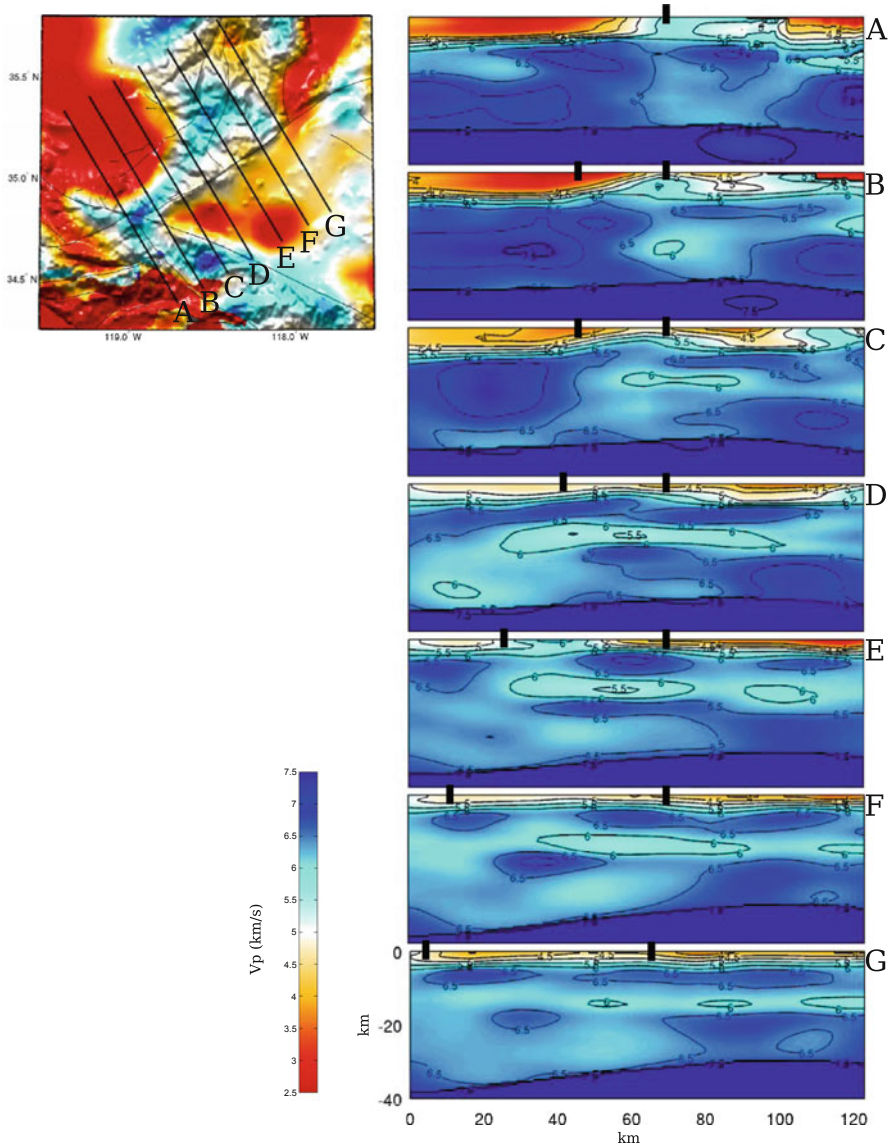


Fig. 6.30 Cross-sections of P-velocity in CVM-S4.26 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

that the batholithic rocks dip northward as a tilted slab underneath the southern San Joaquin Basin. The velocity contours also show some sharp discontinuities, which might be correlated with subsurface faults in this region. A low-velocity layer lies between the upper crust and the lower crust on profile C-G in Figs. 6.29–6.30. The minimum velocity inside this low-velocity layer is reached at about 12- to 14-km

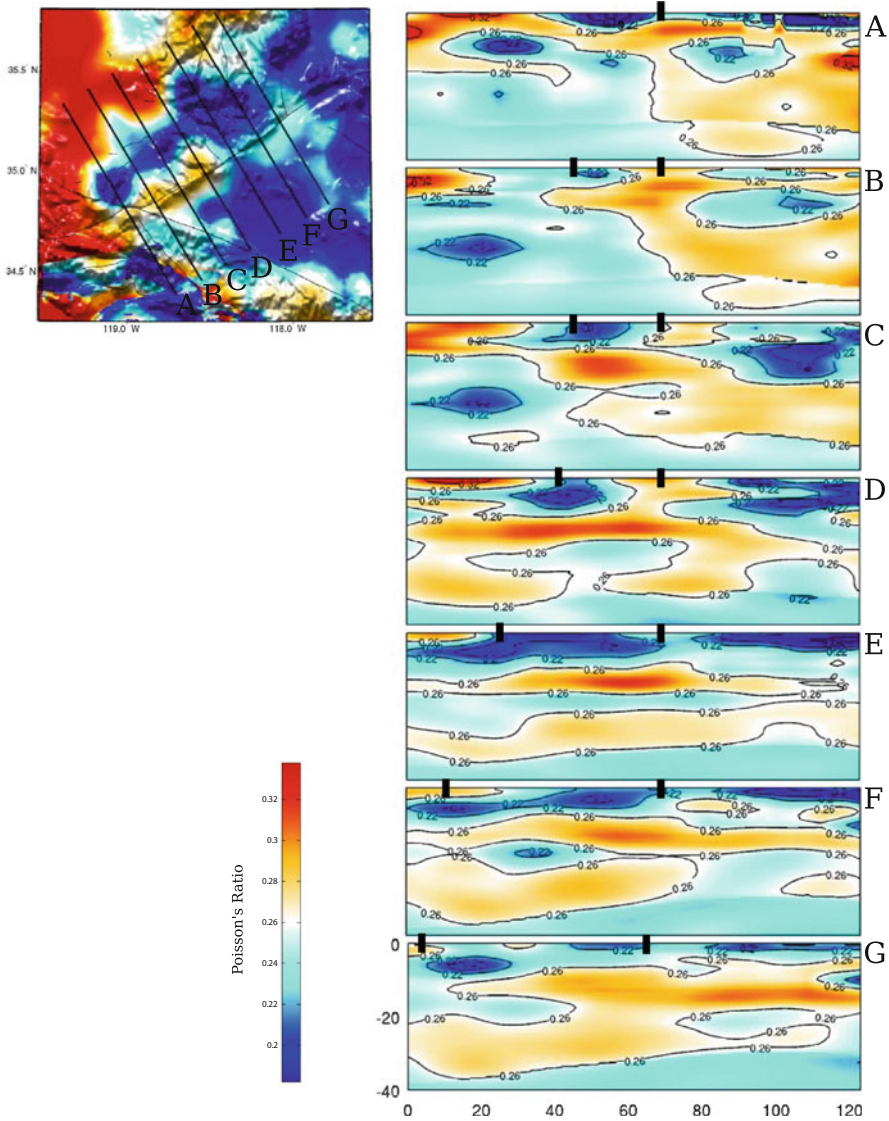


Fig. 6.31 Cross-sections of Poisson's ratio in CVM-S4.26 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

depth underneath the Tehachapi mountains, which can also be seen on the 12- and 14-km map-views in Figs. 6.1 and 6.2. This mid-crustal low-velocity layer is correlated with a zone of laterally discontinuous reflectors found in the CALCRUST study (Malin et al. 1995). The geologic map shows Tertiary volcanic rocks around

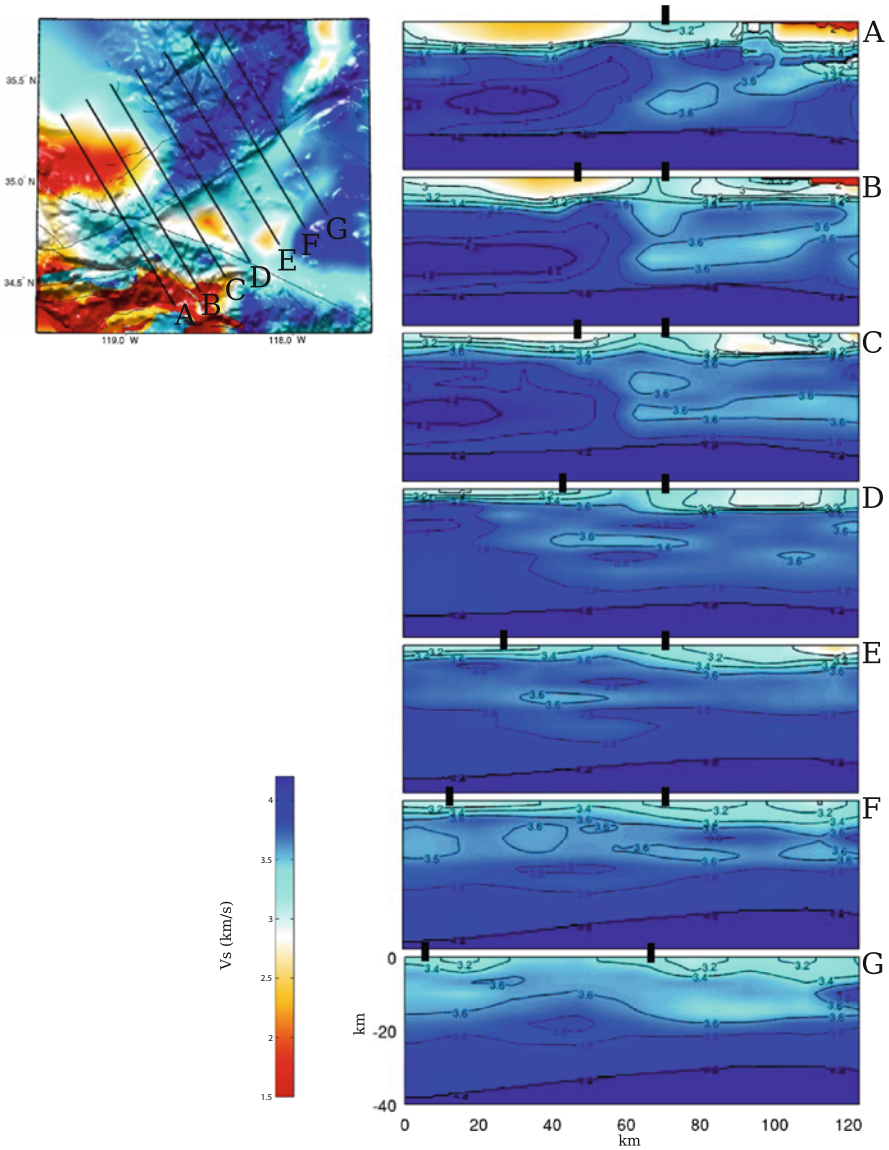


Fig. 6.32 Cross-sections of S-velocity in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

the Tehachapi mountains on the surface (Fig. 6.19b). The strong mid-crustal low-velocity anomaly underneath the Tehachapi mountains might be related to magmatic activities. This mid-crustal low-velocity layer extends south into the Mojave block on profile C-G in Figs. 6.29–6.30, which might be correlated with the reflector K in the COCORP study (Cheadle et al. 1986).

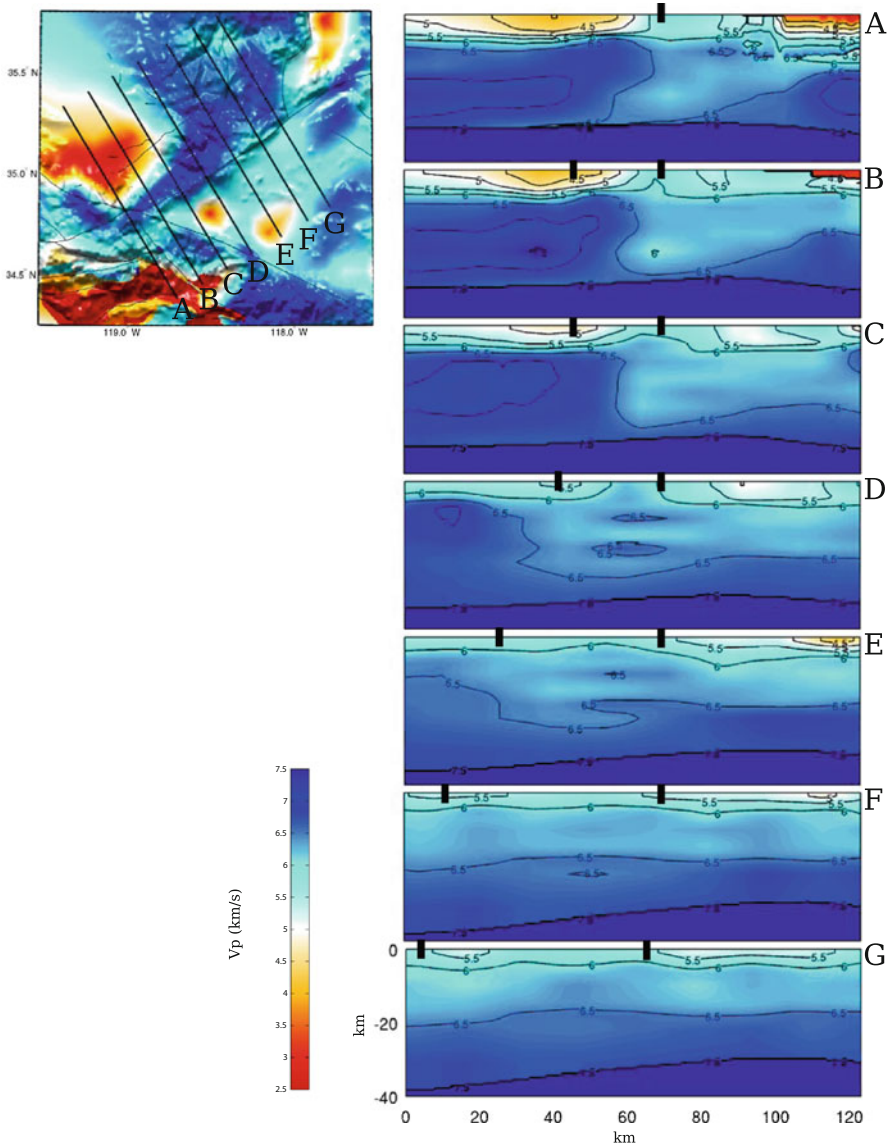


Fig. 6.33 Cross-sections of P-velocity in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

6.2.5 Southern San Joaquin Basin

The southern San Joaquin Basin (SSJB) is the southern tip of the Great Valley, which is a relic forearc basin formed in the Andean-type convergent continental margin during Mesozoic and Paleogene (e.g., Dickinson 1981; Ernst and Rubey 1981).

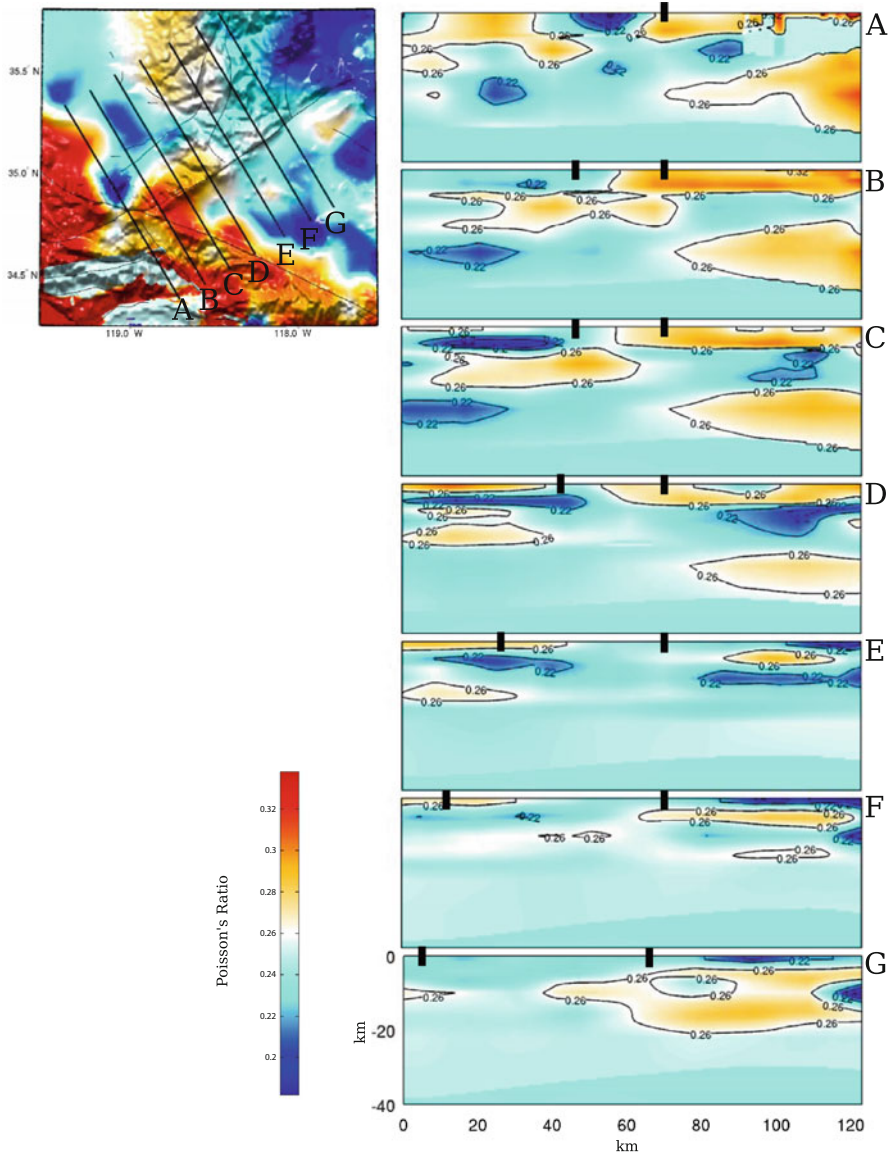


Fig. 6.34 Cross-sections of Poisson's ratio in CVM-S4 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

The Great Valley Sequence mainly consists of Mesozoic and Cenozoic mudstones, sandstones and conglomerates deposited in deep ocean, slope, shelf and subaerial environments (Bailey et al. 1970; Dickinson and Rich 1972; Ingersoll 1982). In

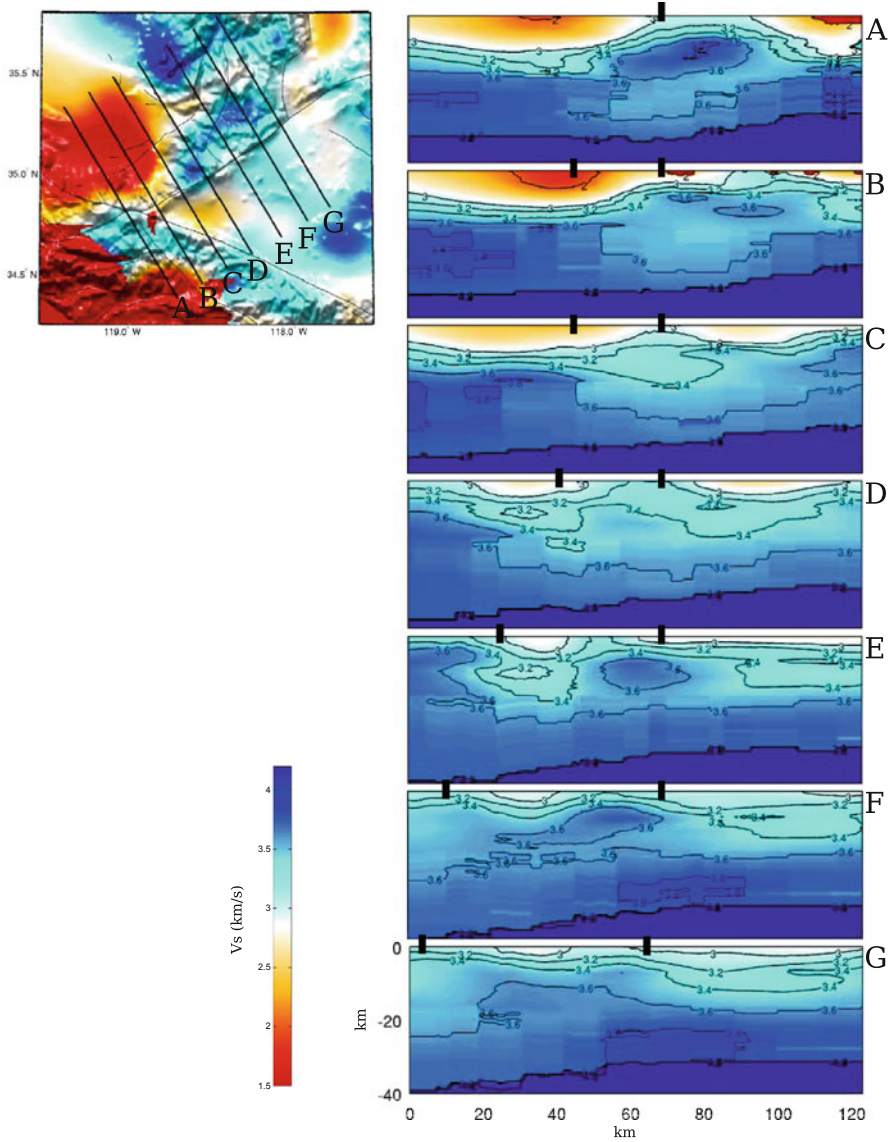


Fig. 6.35 Cross-sections of S-velocity in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

Fig. 6.38, I show a comparison of the 2-km S-velocity map-views of the three CVMs with the geologic map in and around SSJB. The three-dimensional geometry of the basin and velocity structures underneath the basin in CVM-S4.26 can be seen on the cross-section views in Figs. 6.39–6.41. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.42–6.47.

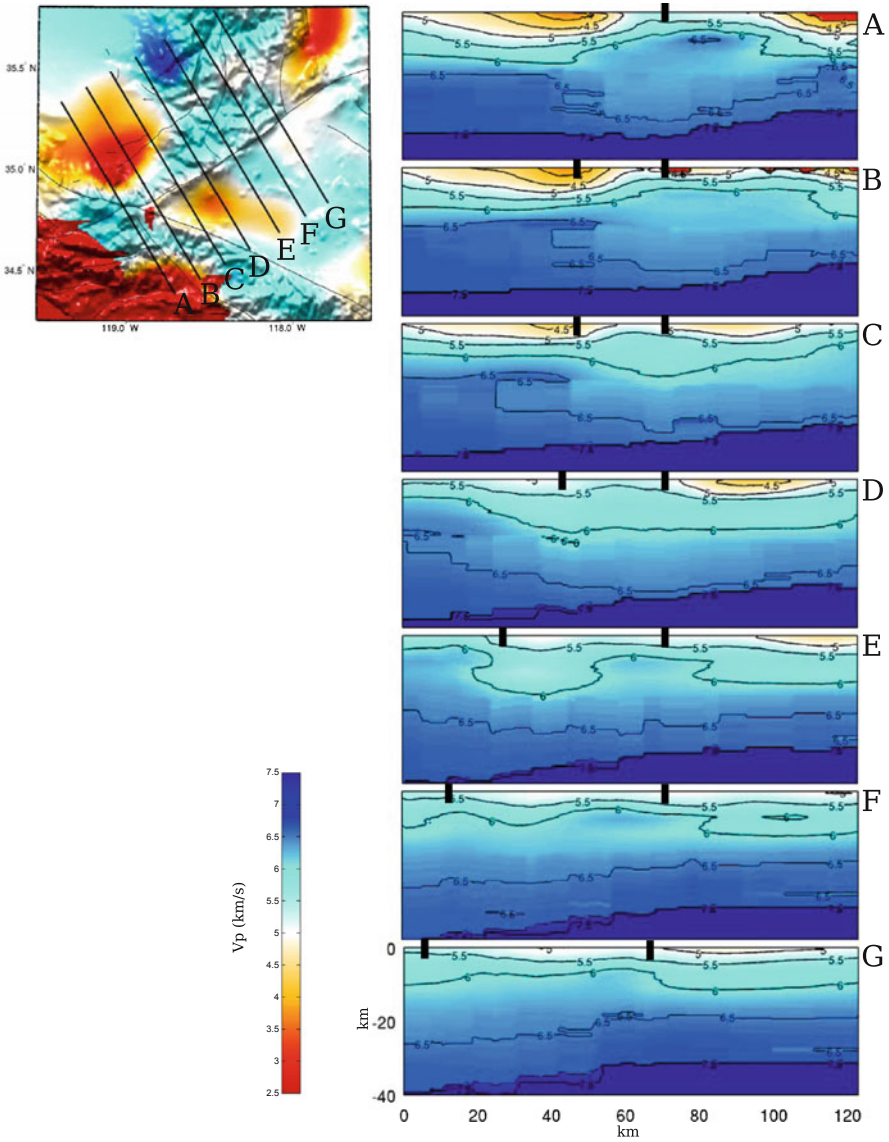


Fig. 6.36 Cross-sections of P-velocity in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

6.2.5.1 Basin Geometry

The SSJB is mostly absent in our starting model CVM-S4 (Figs. 6.38, 6.42–6.44). It is well constructed in our model CVM-S4.26 through our iterative full-3D tomography process (Figs. 6.38, 6.39–6.41). At shallow depths, the low-velocity region

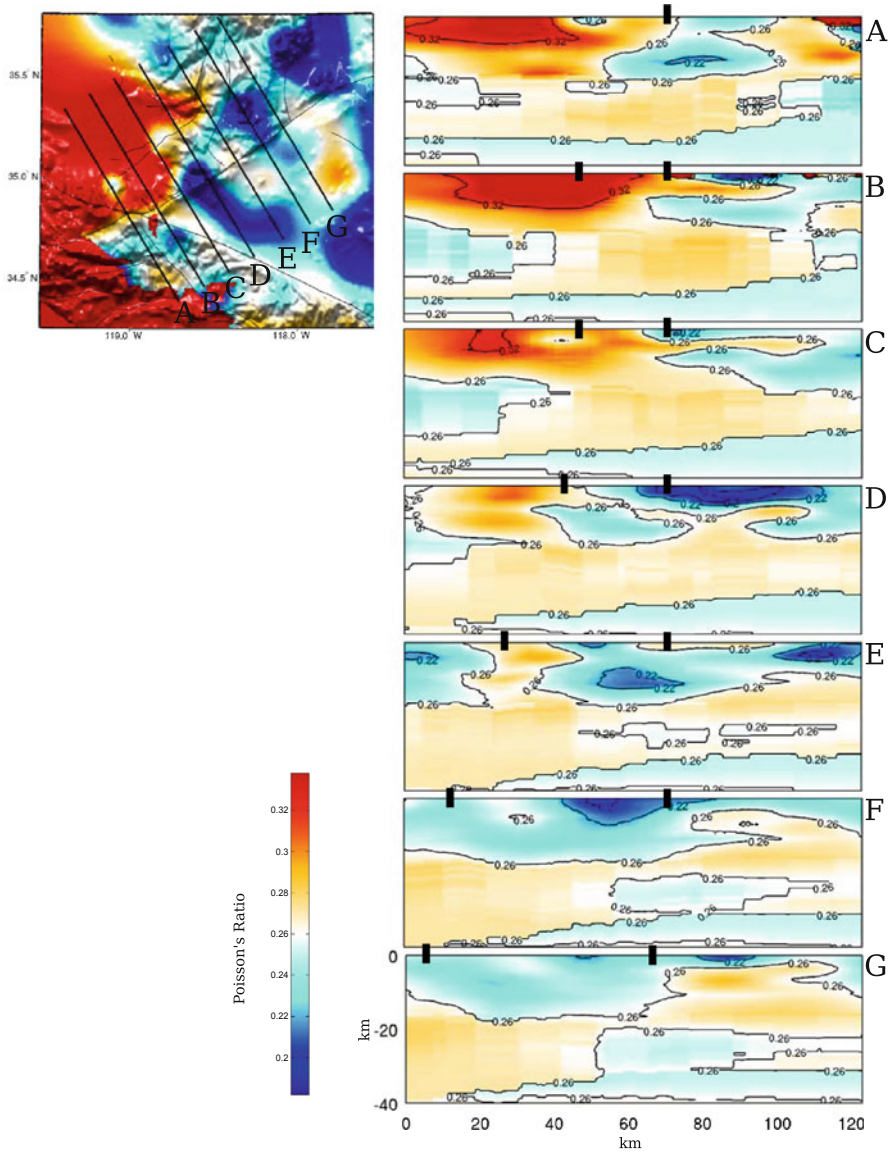


Fig. 6.37 Cross-sections of Poisson's ratio in CVM-H11.9 along the 7 profiles (A-G). The format is identical to that in Fig. 6.29.

associated with the SSJB shows excellent correlation with the distribution of sediments (Fig. 6.38). The bottom of the basin is relatively flat and marked by a sharp velocity gradient at around 8–10 km depth. The average P-velocity, S-velocity and Poisson's ratio inside the basin in CVM-S4.26 are about 4.2 km/s, 2.2 km/s and 0.31, respectively. On the west side, the SSJB extends slightly across the surface

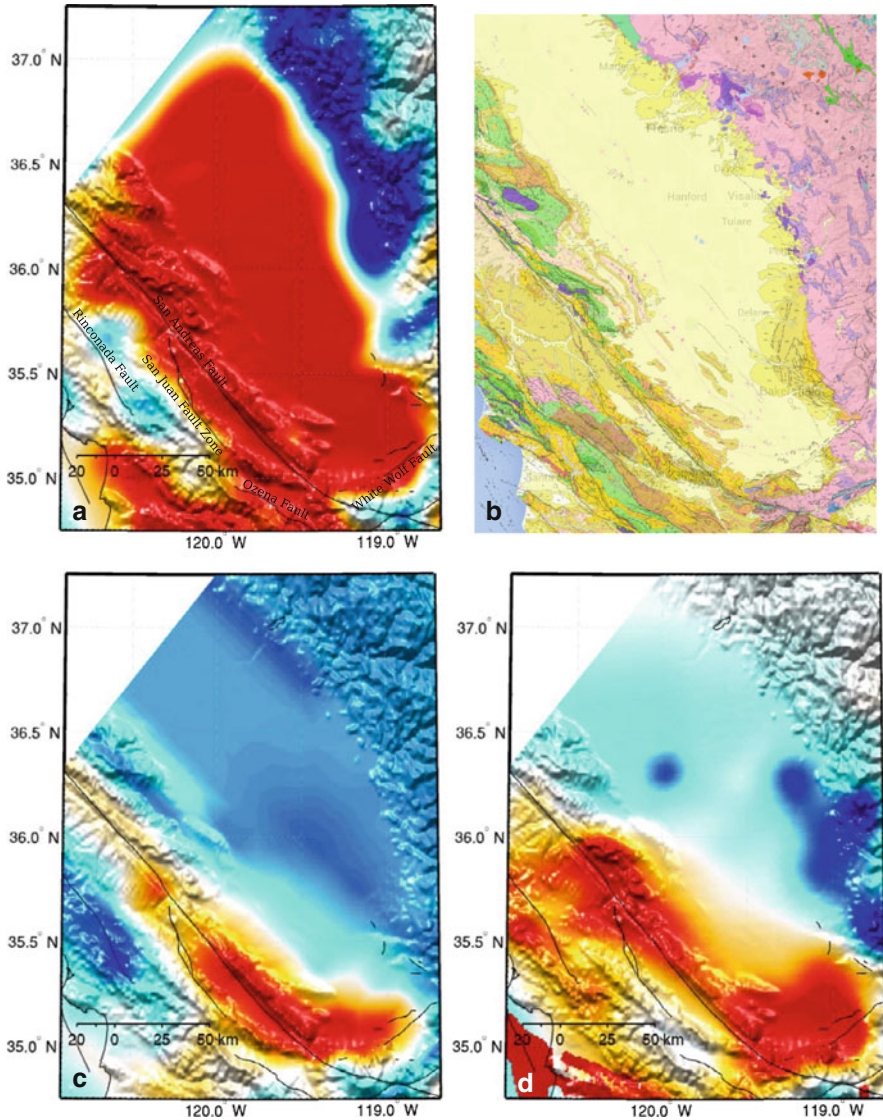


Fig. 6.38 S-velocity at 2-km depth in CVM-S4.16 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in the southern San Joaquin Basin. The color scale is identical to that in the 2-km map-view in Fig. 6.1.

trace of the San Andreas Fault into the Salinian block. The western boundary of the basin dips to the east. On the east side, the SSJB is bounded by the western Sierra Nevada Foothills on the surface. Except at very shallow depths (< 2 km), the eastern boundary of the basin dips to the west. To the south, the SSJB extends slightly across the surface trace of the White Wolf Fault and stops at the surface trace of the Garlock Fault. The southern boundary dips to the north.

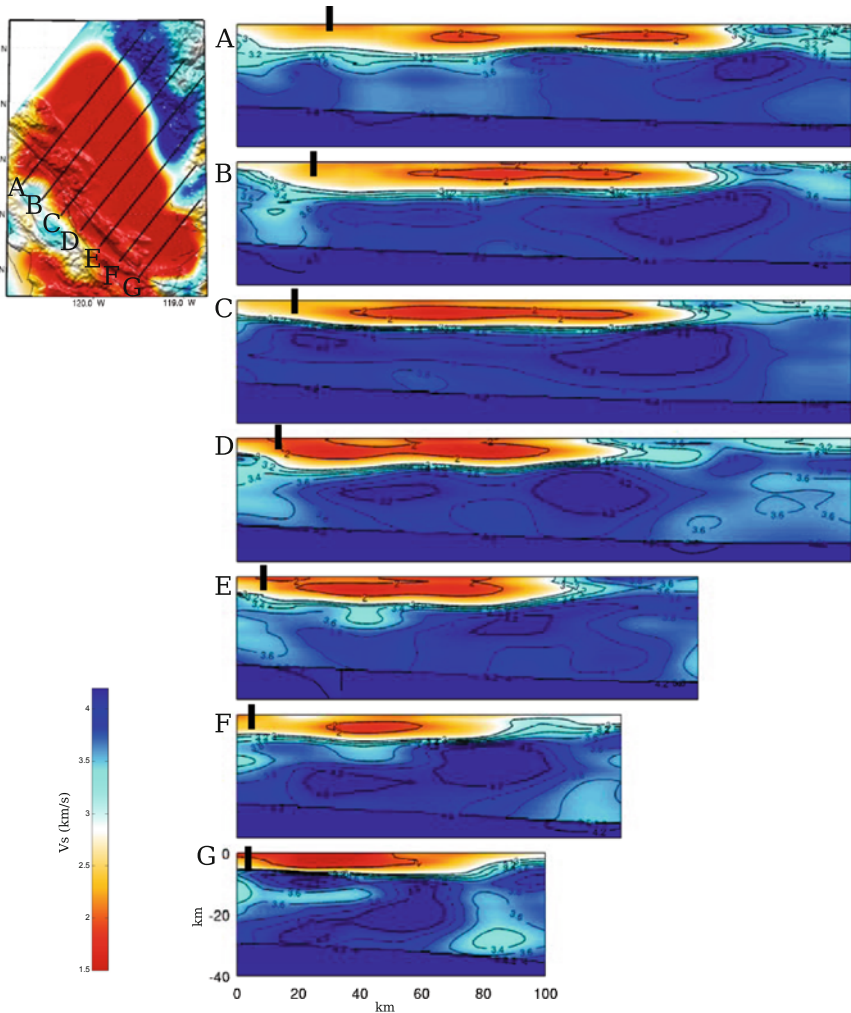


Fig. 6.39 Cross-sections of the S-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of the surface traces of San Andreas Fault.

6.2.5.2 Great Valley Ophiolite

The high-velocity anomalies underneath the SSJB have different origins. The thicker high-velocity body lying under the eastern side of the basin was interpreted as the mafic western SNB generated in Early Cretaceous in Sect. 6.2.4.1. In addition to the evidences mentioned in Sect. 6.2.4.1, the Lower Cretaceous petrofacies in the

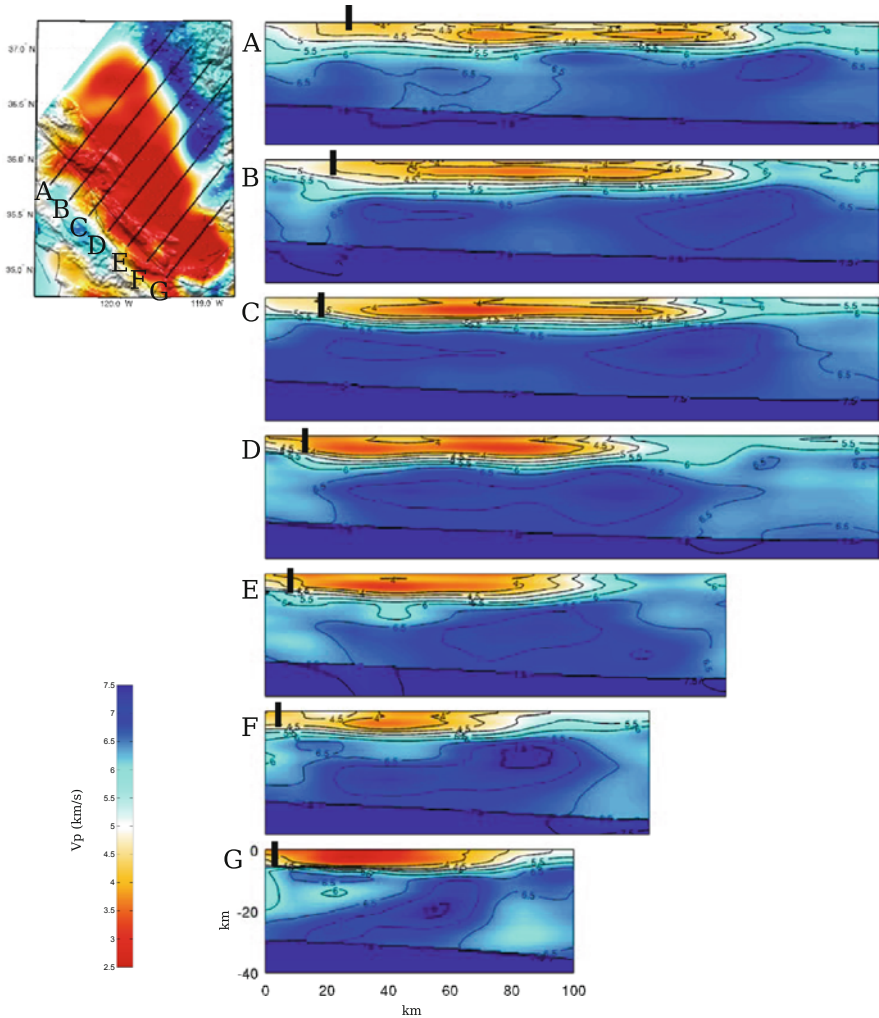


Fig. 6.40 Cross-sections of the P-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

Great Valley Sequence show abundant detritus derived from this mafic western SNB (Dickinson and Rich 1972). A rough estimate of the western extent of this mafic SNB underneath the SSJB was drawn on Figs. 6.20–6.22.

The mafic western SNB is connected with a thinner, slab-like high-velocity layer lying under the western part of the SSJB (profile B-D in Fig. 6.39). The maximum P-velocity and S-velocity inside this thin high-velocity layer reach 7.2 km/s and 4.2 km/s, respectively, at around 20-km depth. The velocities beneath this thin high-velocity layer reduce slightly and then increase again at the Moho. In general the Poisson’s ratio is slightly higher in this thin high-velocity layer than in the mafic western SNB (Fig. 6.41). This thin high-velocity layer might be associated

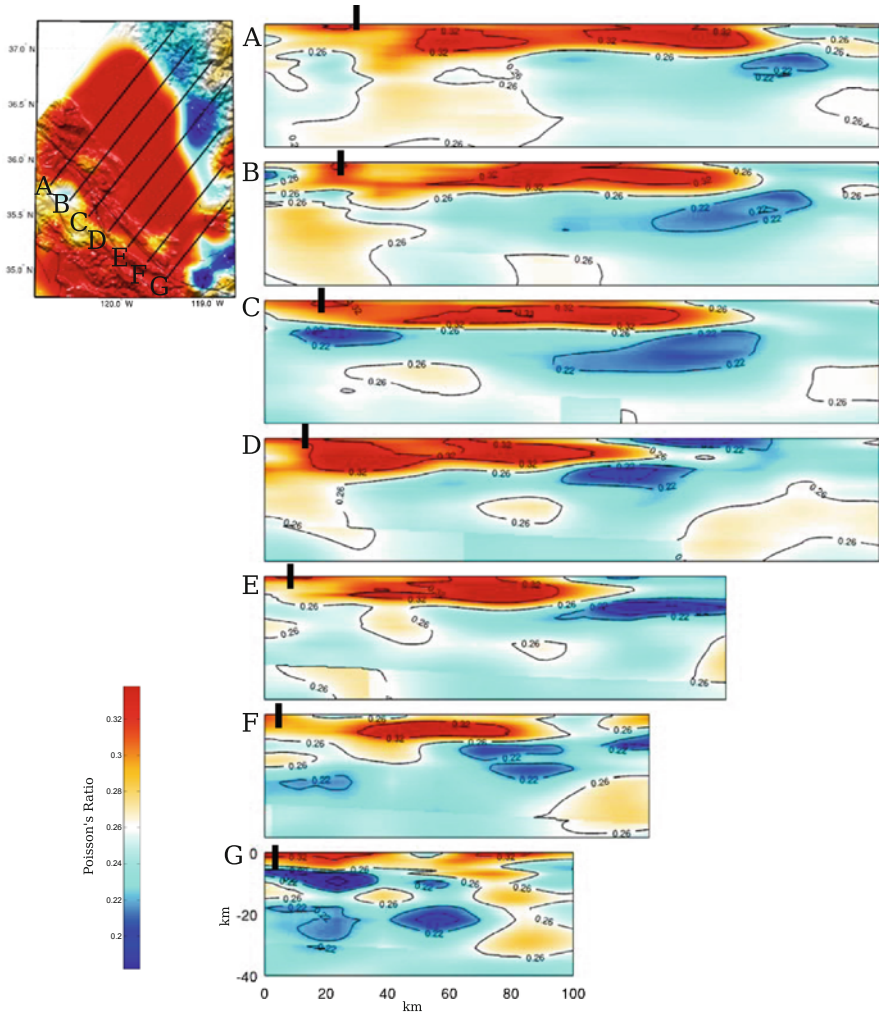


Fig. 6.41 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

with the Great Valley Ophiolite (GVO) (Bailey et al. 1970; Page 1981; Godfrey and Klempner 1998). It has been suggested that the GVO was formed by back-arc spreading behind an east-facing arc and the arc and the back arc “docked” on the continental margin during the Nevadan orogeny in southern Great Valley (e.g., Godfrey and Klempner 1998). This thin high-velocity layer in CVM-S4.26 might be the obducted oceanic crust.

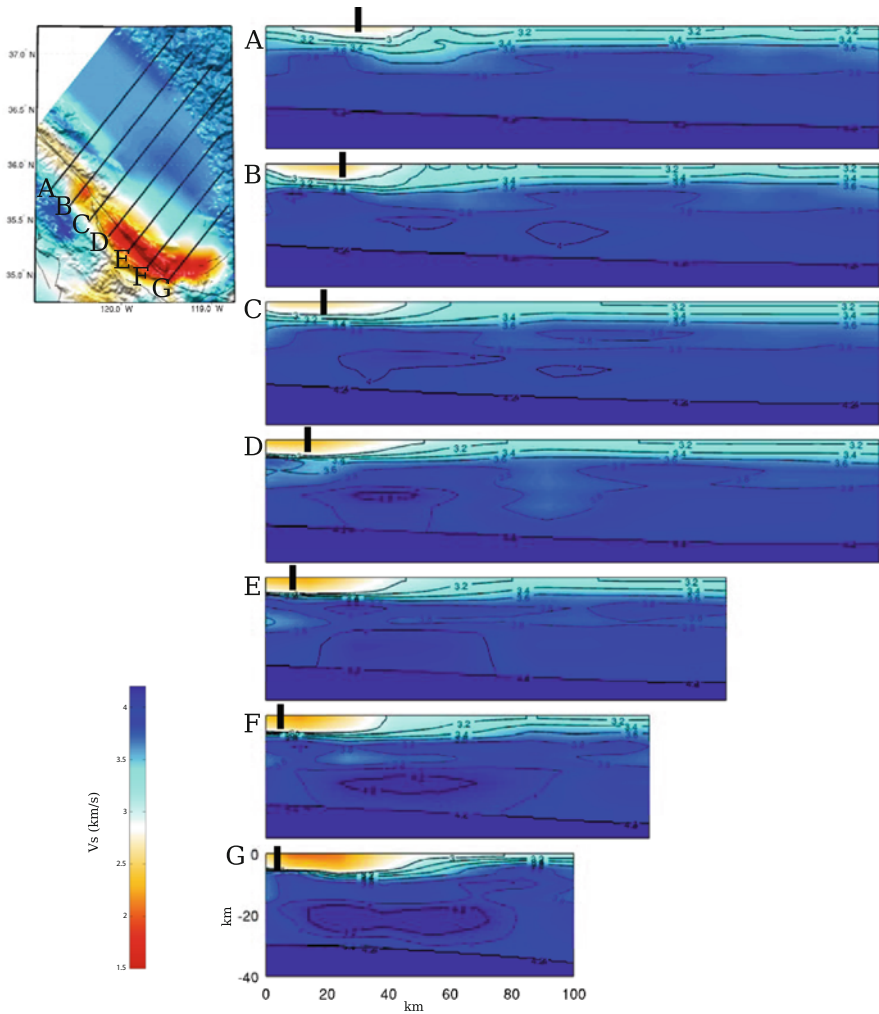


Fig. 6.42 Cross-sections of the S-velocity in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

6.2.6 Southern Coast Ranges

Our tomography box covers the southern portion of the Coast Ranges, which belongs to the Monterey terrane (i.e., the region bounded by the Pacific Ocean on the west, the San Andreas Fault (SAF) on the east, Cape Mendocino on the north and the Transverse Ranges on the south) as defined in (Ducea et al. 2009). The Monterey terrane has its origin in southern California and was translated to its current location by the San Andreas Fault system in Neogene (Page 1981; Saleeby 2003). It is mainly composed of the Salinian block and the Nacimiento block (e.g., Fig. 1 in Ducea et al.

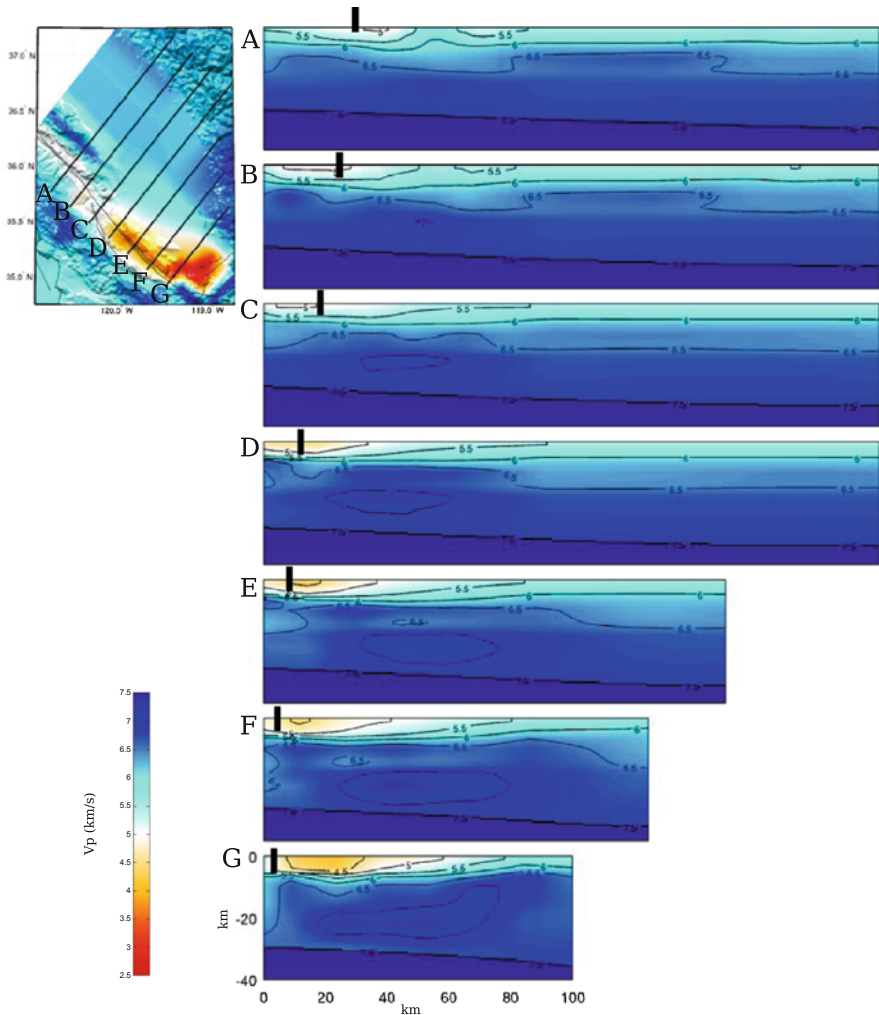


Fig. 6.43 Cross-sections of the P-velocity in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

2009). The Salinian block was part of the Sierra-Mojave-Peninsular batholith belt and might have been located on the west of the Mojave block in pre-SAF time (Ross 1976, 1984; Hall 1991; Silver and Mattinson 1986; Saleeby 2003). The Nacimiento basement mainly consists of Franciscan Complex rocks (Gilbert 1973; Hall 1991), which are metamorphosed trench sediments. There are outcrops of the Coast Range Ophiolite structurally above the Franciscan units in the Nacimiento block (Vedder et al. 1983; Shervais et al. 2004).

In Fig. 6.48, I show a comparison of the 2-km S-velocity map-views of the three CVMs with the geologic map in and around the southern Coast Ranges.

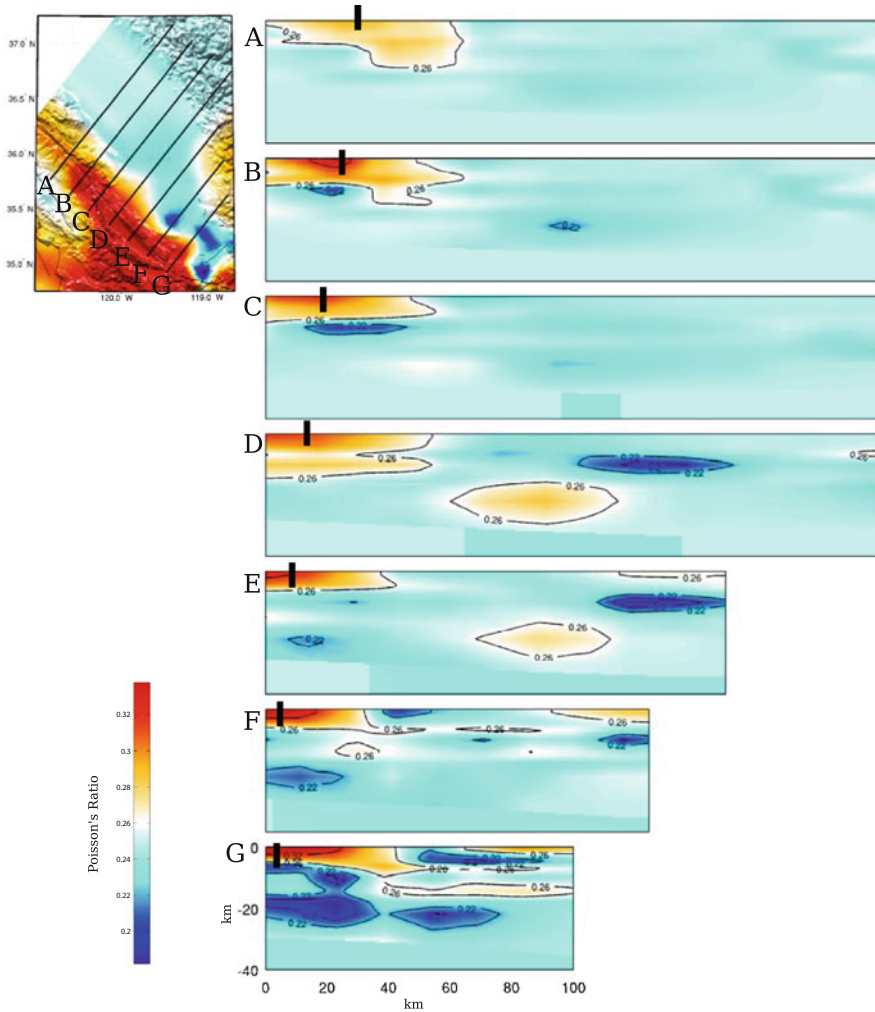


Fig. 6.44 Cross-sections of the Poisson's ratio in CVM-S4 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

Low-velocity anomalies in CVM-S4.26 are well correlated with distributions of sediments. The high-velocity anomaly between the Nacimiento-Rinconada Fault and the San Juan Fault Zone is correlated with outcrops of the Salinian basement (Fig. 2 in Ducea et al. 2009). The intermediate- to high-velocities on the west of the Nacimiento Fault north of the Pismo Beach might be correlated with outcrops of the Franciscan Complex of the Nacimiento block (Fig. 2 in Ducea et al. 2009). Cross-section views of the S-velocity, P-velocity and Poisson's ratio in CVM-S4.26 along 8 profiles crossing the southern Coast Ranges are shown in Figs. 6.49–6.51. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.52–6.57.

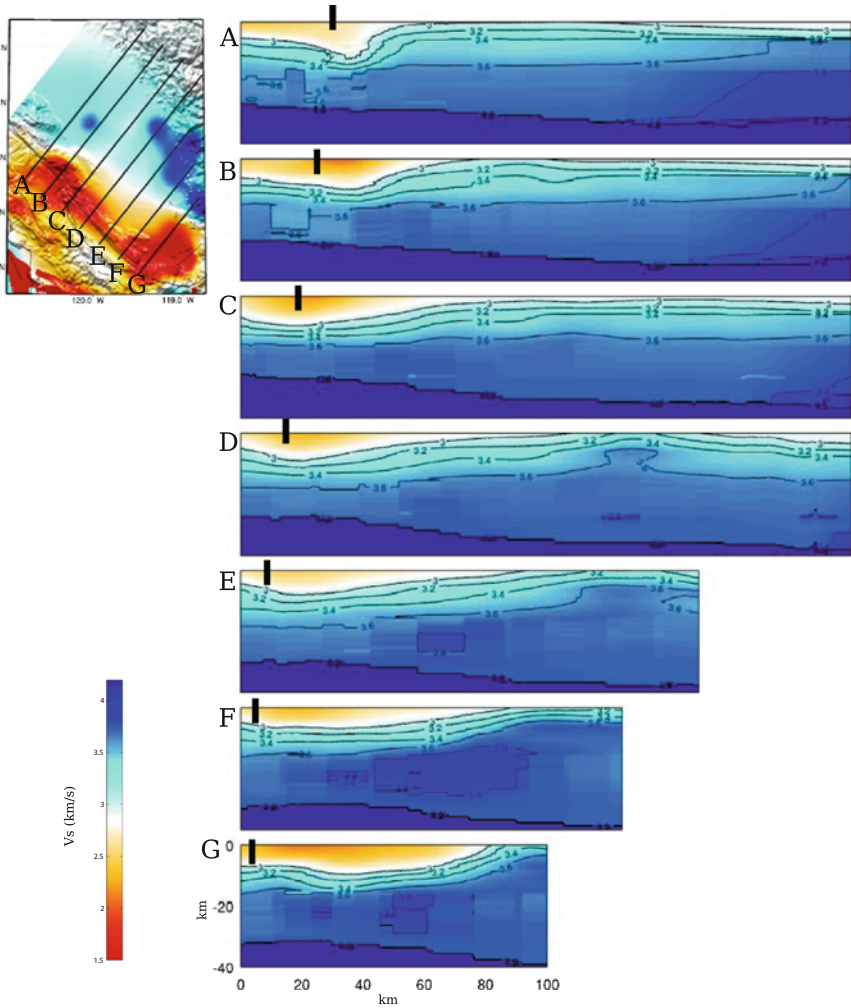


Fig. 6.45 Cross-sections of the S-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

6.2.6.1 Salinian Block

The dominant low-velocity feature inside the Salinian block is the Cuyama Basin lying at the southern tip of the Salinian block. It is mostly absent in our starting model CVM-S4 (Fig. 6.48c), but well imaged in CVM-S4.26 (Fig. 6.48a). The Cuyama Basin includes thick Upper Cretaceous through Quaternary sedimentary sequence in the Carrizo Plain, the Caliente Range and the thinner sediments in the Cuyama Valley (Davis et al. 1988). To the west of the Ozena Fault, the sediments become thinner (Figs. 6.49–6.50). The sedimentary section

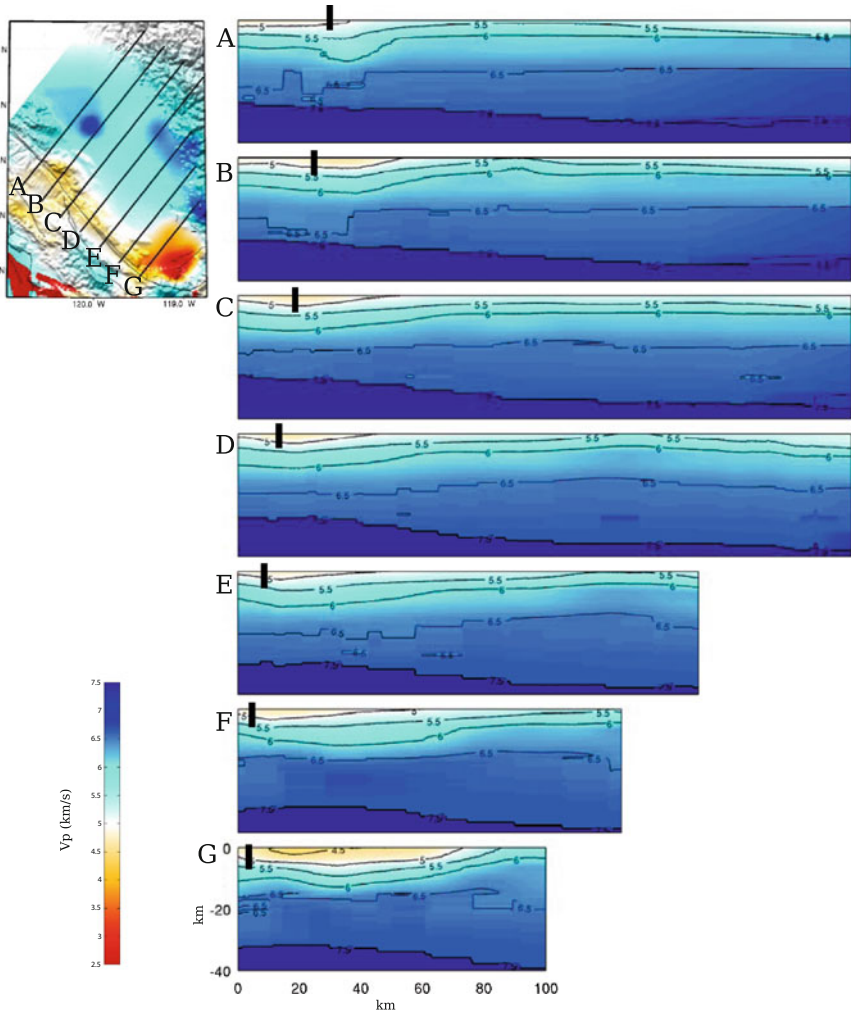


Fig. 6.46 Cross-sections of the P-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

in the Cuyama Basin consists of marine clastic rocks of late Cretaceous, Paleocene and Eocene and thick Middle and Upper Cenozoic sedimentary rocks deposited in marine and subaerial environments (Davis et al. 1988). The bottom of the basin is marked by sharp velocity gradients in CVM-S4.26 and the maximum thickness of the Cuyama Basin is about 5 km (Figs. 6.49–6.50), which is in broad agreement with the cross-sections in (Davis et al. 1988). The average P- and S-velocity inside the basin are about 4.4 km/s and 2.4 km/s, respectively, and the average Poisson's ratio is about 0.3.

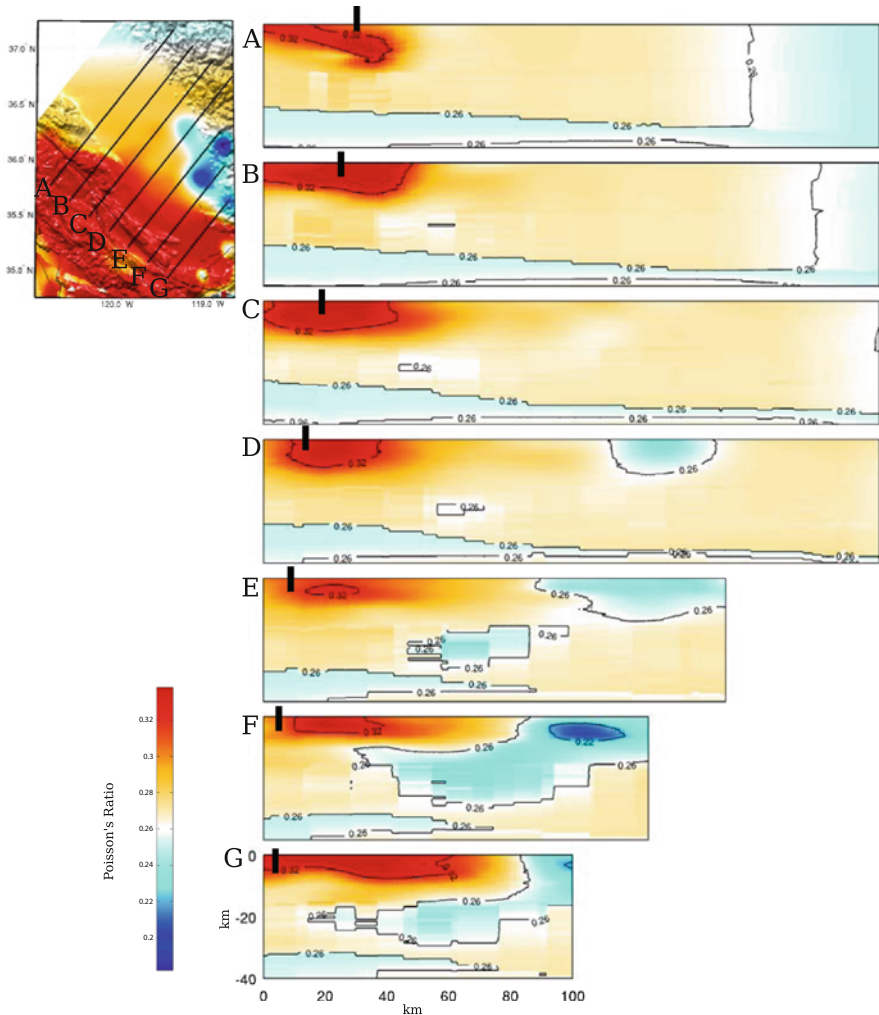


Fig. 6.47 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 7 profiles (A-G) crossing the southern San Joaquin Basin. The format is identical to that in Fig. 6.39.

To the north of the Cuyama Basin, between the Nacimiento-Rinconada Fault and the San Juan Fault, is a region where the Salinian basement crops out (Ducea et al. 2009). This region is shown as a high-velocity anomaly in CVM-S4.26 (Fig. 6.48, profile B-D in Figs. 6.49–6.50). The Salinian basement is mostly composed of mid-Cretaceous granite plutons mixed with older high-temperature metasedimentary rocks. The average composition of the Salinian arc in the upper crust is granodiorite (Ross 1975). In CVM-S4.26, the upper-crustal high-velocity block lying underneath the Salinian basement outcrops has maximum P- and S-velocity of about 6.2 km/s and 3.7 km/s, respectively, which is roughly consistent with the granodiorite composition (Christensen 1996).

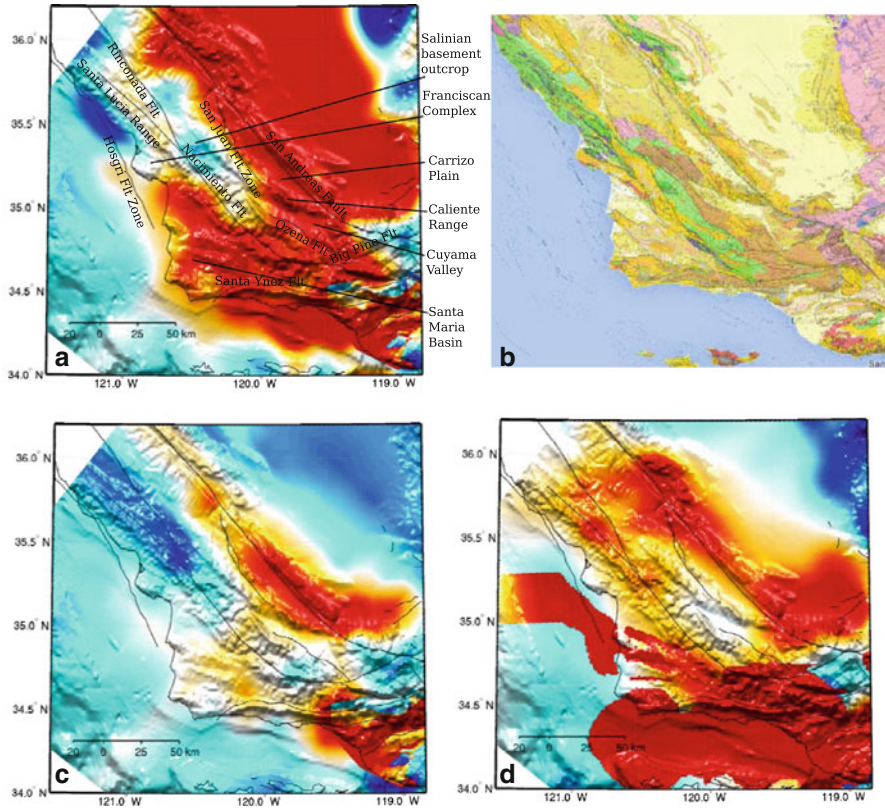


Fig. 6.48 Map-views of the S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) in and around the southern Coast Ranges. The color scale is identical to that in the 2-km map-view in Fig. 6.1.

6.2.6.2 Nacimiento Block

To the west of the Nacimiento Fault and north of the Santa Ynez Fault is the Nacimiento block (Fig. 6.48), which is also known as the Sur-Obispo composite terrane (Vedder et al. 1983; Hall 1991; Ducea et al. 2009). The basement mainly consists of Franciscan Complex rocks with fragments of Late Jurassic Coast Range Ophiolite (Hopson et al. 1981; Vedder et al. 1983; Shervais et al. 2004). The sedimentary cover is mostly Uppermost Cretaceous to Cenozoic (Hall 1991).

In CVM-S4.26, the dominant feature is the low-velocity region associated with the Santa Maria Basin (Fig. 6.48, profile C-G in Figs. 6.49–6.51) in the Pliocene to Quaternary fold and thrust belt (Namson and Davis 1990). It is mainly composed of Lower Miocene through Quaternary volcanic, volcanoclastic, deep-marine, shallow-water and nonmarine sedimentary rocks (Namson and Davis 1990). The basin extends to the southern Santa Lucia Range in the north and to the western

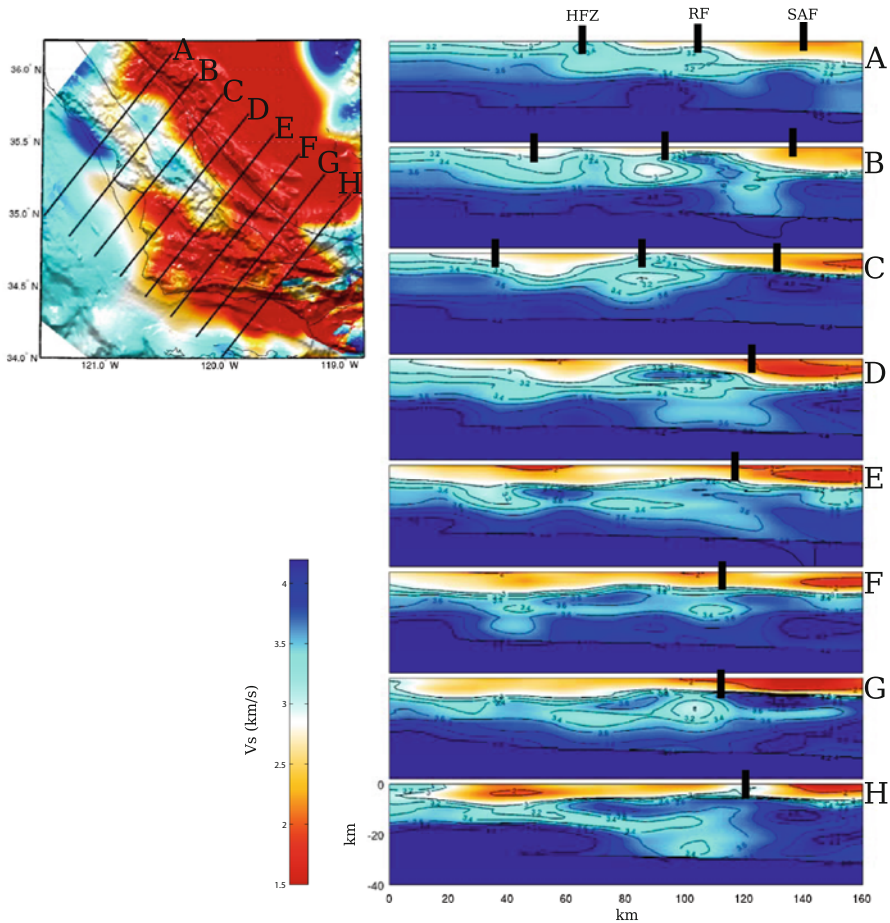


Fig. 6.49 Cross-sections of the S-velocity in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. The *black dots* on the cross-sections show the locations of the surface traces of San Andreas Fault (SAF), the Rinconada Fault (RF) and the Hosgri Fault Zone (HFZ). Locations of the faults are marked on Fig. 6.48.

Santa Ynez Mountains in the south and it is bounded by the San Rafael Mountains to the east and extends westward to the offshore region (Fig. 6.48). The bottom of the basin is marked by sharp velocity gradients in CVM-S4.26 and has a highly variable depth with maximum depth exceeding 6 km (profile C-G in Figs. 6.49–6.51). The average P-velocity and S-velocity inside the basin are about 4.8 km/s and 2.5 km/s, respectively, and the average Poisson’s ratio is about 0.32. The lowest velocities and highest Poisson’s ratio are reached at around the center of the basin, which overlies a high-velocity body with maximum P- and S-velocity exceeding

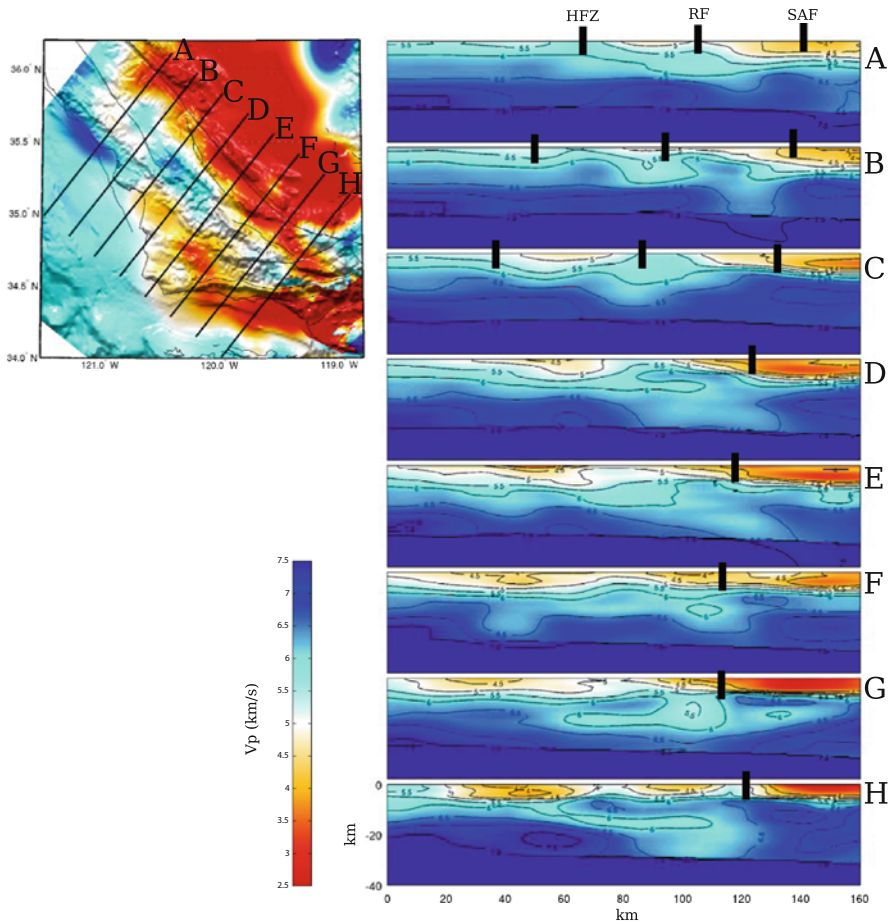


Fig. 6.50 Cross-sections of the P-velocity in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

6.1 km/s and 3.8 km/s, respectively (between profile E and F in Figs. 6.49–6.50). The seismic velocities of this high-velocity body are in agreement with the laboratory measurements of the ophiolite at Point Sal (Nichols et al. 1980). Outcrops of the Coast Range Ophiolite are found in the Santa Lucia and San Rafael Mountains and west of the town of Santa Maria (Ducea et al. 2009). This high-velocity body might be a fragment of the Coast Range Ophiolite. The subsidence of the Santa Maria Basin may have been driven, perhaps in part, by the isostatic response.

At about mid-crustal depth (17 km), a low-velocity layer extends from the Nacimiento block eastward into the Salinian block (profile D-H in Figs. 6.49–6.51). The velocities and thickness of this low-velocity layer vary from profile to profile. On average, the thickness is about 5 km, the P-velocity is 5.8–6.0 km/s and the S-velocity is 3.0–3.2 km/s. Perhaps this low-velocity layer is related to the schist of

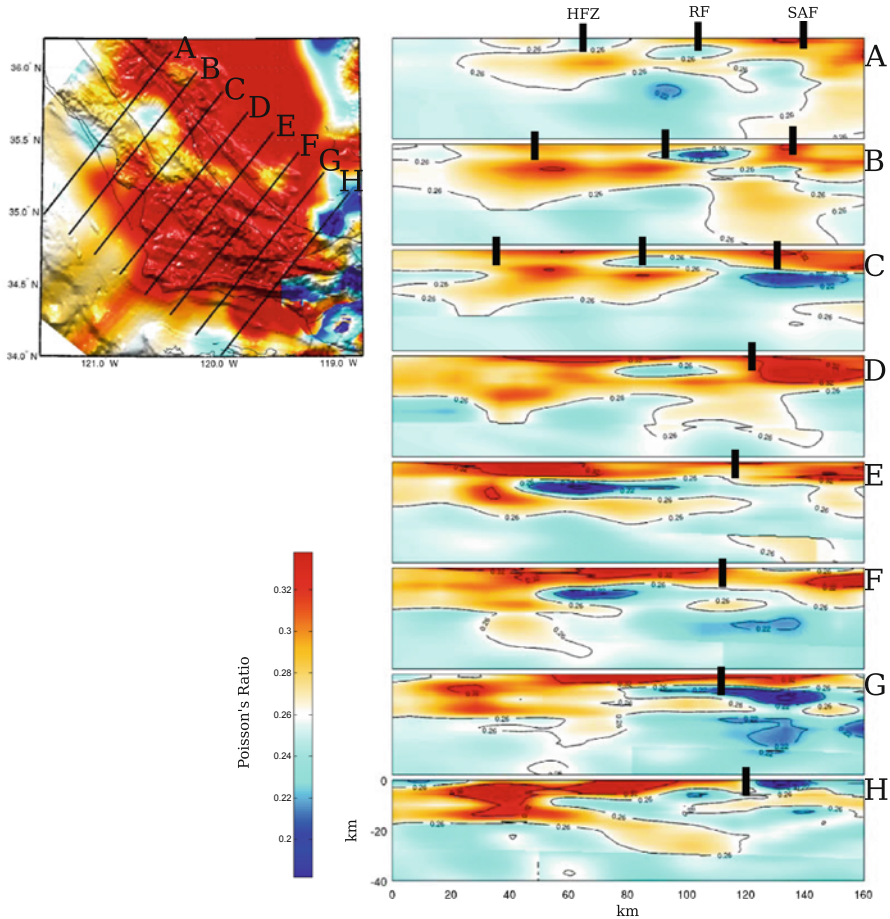


Fig. 6.51 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

Sierra de Salinas (Barth et al. 2003; Kidder and Ducea 2006; Ducea et al. 2009), which crops out southwest of the Salinas Valley lying slightly to the north of our tomography box. The Sierra de Salinas schist, as well as the Pelona, Orcopia and Rand schists in the Mojave and San Gabriel region, represent the accretion wedge and/or forearc sediments underplated beneath the continental arc during the Laramide shallow subduction (e.g., Jacobson et al. 2000; Saleeby 2003; Ducea et al. 2009). The Sur Fault, which is the northern extension of the Nacimiento Fault, has been interpreted as a major late Cretaceous thrust fault that brought the Salinian block outboard over the accretionary wedge (Hall 1991; Ducea et al. 2009). The Nacimiento Fault might have played a similar role in late Cretaceous as the Sur Fault to the north.

In the lower-crust, there are fragments of high-velocity bodies with maximum P- and S-velocity exceeding 7.0 km/s and 4.1 km/s, respectively (Figs. 6.49–6.50). These high-velocity bodies might be due to variations in the Moho depth (i.e., our

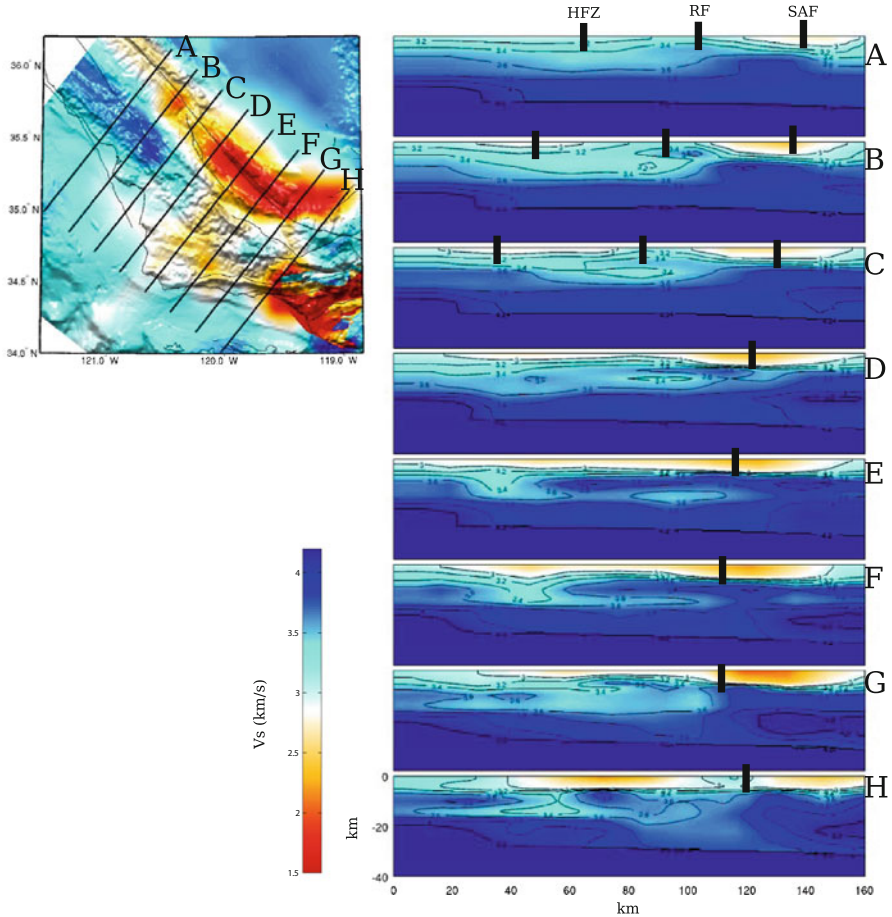


Fig. 6.52 Cross-sections of the S-velocity in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

inversion was trying to accommodate variations in the Moho depth by adding high-velocity bodies around the Moho in the starting model). Another possibility is that they represent trapped fragments of the oceanic crust associated with the Farallon plate and/or the younger Monterey plate (e.g., Miller et al. 1992).

6.2.7 Transverse Ranges

The Transverse Ranges province consists of young east-west trending mountain ranges and valleys that clearly traverse the northwest-southeast tectonic grain in most of the southern California. The western Transverse Ranges (Fig. 6.58) are bounded by the Santa Ynez Mountains (Santa Ynez Fault) on the north and the Channel Islands—Santa Monica Mountains (Santa Monica Fault) on the south.

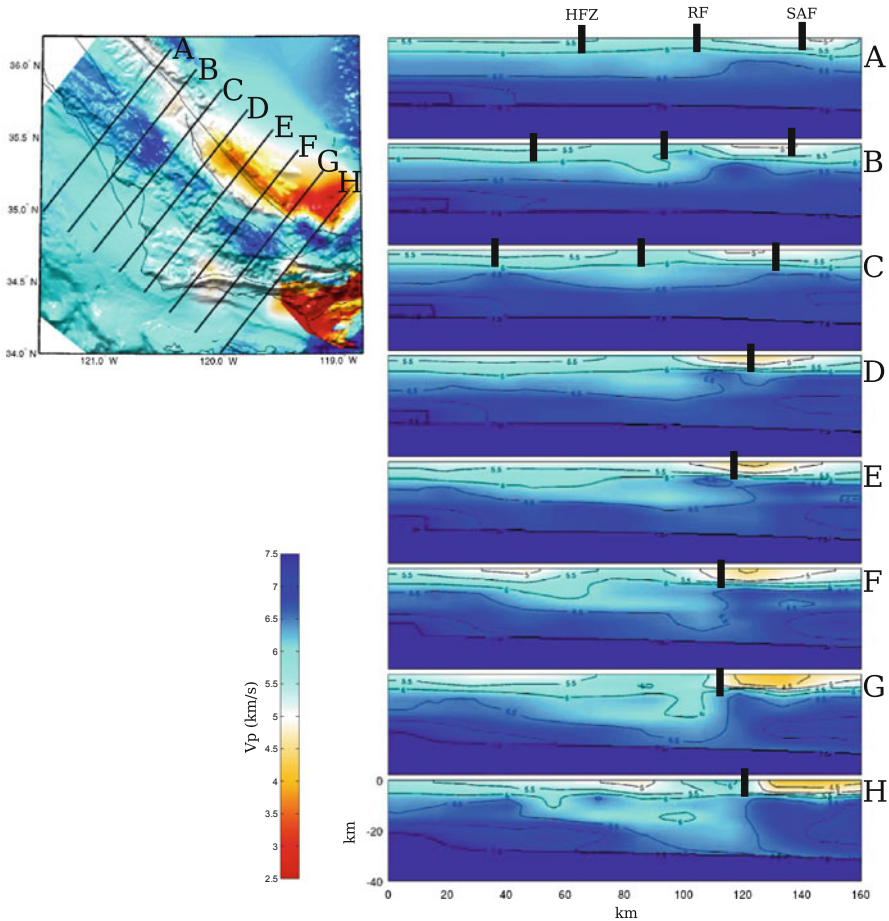


Fig. 6.53 Cross-sections of the P-velocity in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

Between the two mountain ranges lie the Ventura Basin and its off-shore extension, the Santa Barbara Channel. This region is bounded by the San Andreas Fault on the east and by the off-shore Santa Lucia Escarpment on the west.

The Los Angeles Region Seismic Experiment phase I and II (LARSE-I&II) active-source seismic survey lines crossed the western Transverse Ranges along two different profiles. Detailed comparisons between cross-sections of CVM-S4.26, CVM-S4 and CVM-H11.9 along the LARSE-I&II profiles with the 2D refraction tomography models from LARSE-I&II (Lutter et al. 1999, 2004) have been documented in (Lee et al. 2014a, b). In general, at the upper- to mid-crustal depths, CVM-S4.26 shows better agreement with the LARSE-I&II refraction tomography results than the other two CVMs (Fig. 8 in Lee et al. 2014b and Fig. 1 in Lee et al.

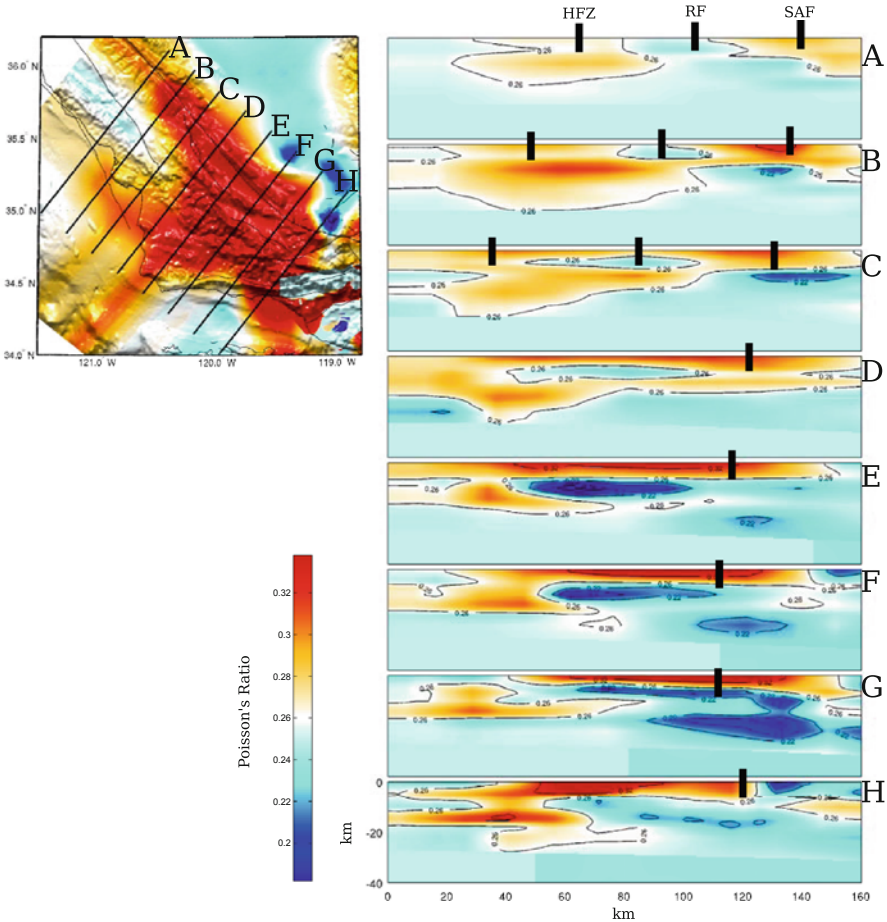


Fig. 6.54 Cross-sections of the Poisson's ratio in CVM-S4 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

2014a). At mid- to lower-crustal depths, CVM-S4.26 shows structural features that are well correlated with the deep reflectors mapped in e.g., (Fuis et al. 2003).

The Eastern Transverse Ranges block (Fig. 6.59) is bounded by the San Andreas Fault on the southwest and the Pinto Mountain Fault on the north. The southern and eastern boundaries do not exactly coincide with single fault traces. The eastern boundary follows approximately the low-to-high S-velocity transition in Fig. 6.59a, which is roughly aligned with the Sheep Hole Fault. The southern boundary is approximately along the Salton Creek Fault between the Orocopia Mountains on the north and the Chocolate Mountains on the south. Previous studies suggest that the southern boundary is a zone of complex distributed strain with possibly non-rigid deformation (e.g., Richard 1993; Dickinson 1996). In the Chocolate Mountains south of the Salton Creek Fault, there still exist east-west-trending sinistral faults

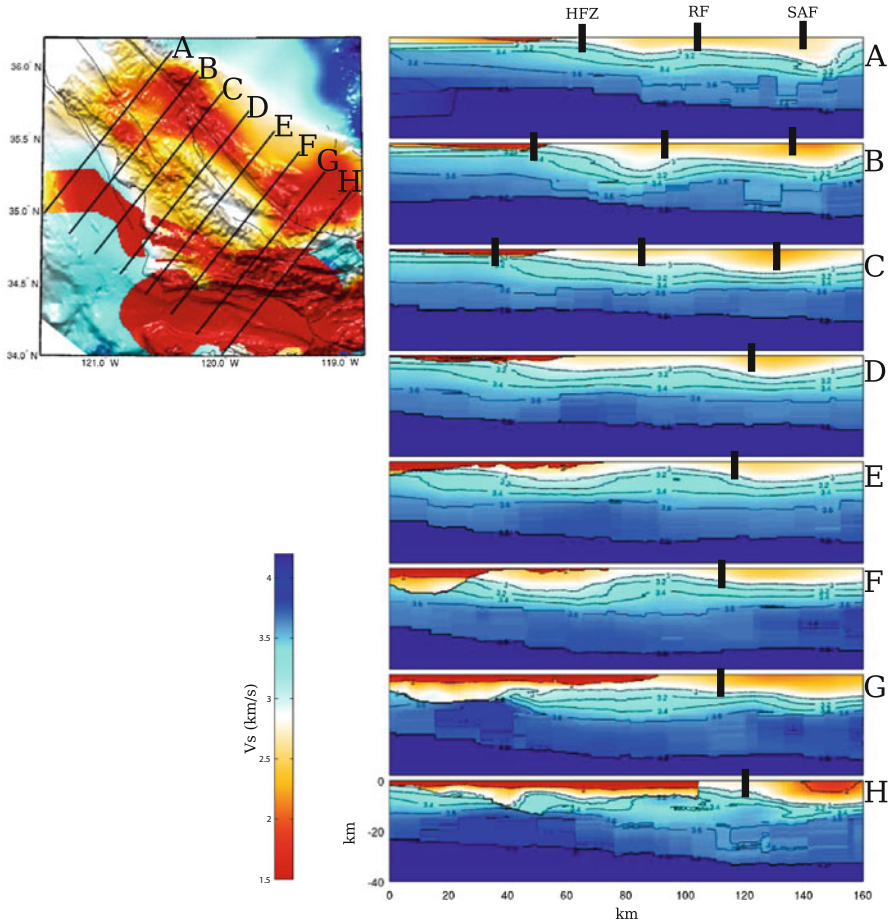


Fig. 6.55 Cross-sections of the S-velocity in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

that are typical in the eastern Transverse Ranges (e.g., Darin and Dorsey 2013). The southern boundary is also located around a low-to-high S-velocity transition on Fig. 6.59a.

The exact tectonic mechanisms for producing the Transverse Ranges are still a major research topic in current geological studies in this region. One possible mechanism involves clockwise rotations of crustal blocks associated with the development of the San Andreas Fault system followed by north-south convergence. By mid-Miocene, the development of the San Andreas Fault system had produced significant transrotational deformation in this region (e.g., Luyendyk et al. 1980; Bird and Rosenstock 1984; Hornafius et al. 1986; Humphreys and Hager 1990; Luyendyk 1991; Dickinson 1996). In the western Transverse Ranges, paleomagnetic data suggest cumulative clockwise rotation of about 70° – 90° (e.g., Kamerling and Luyendyk

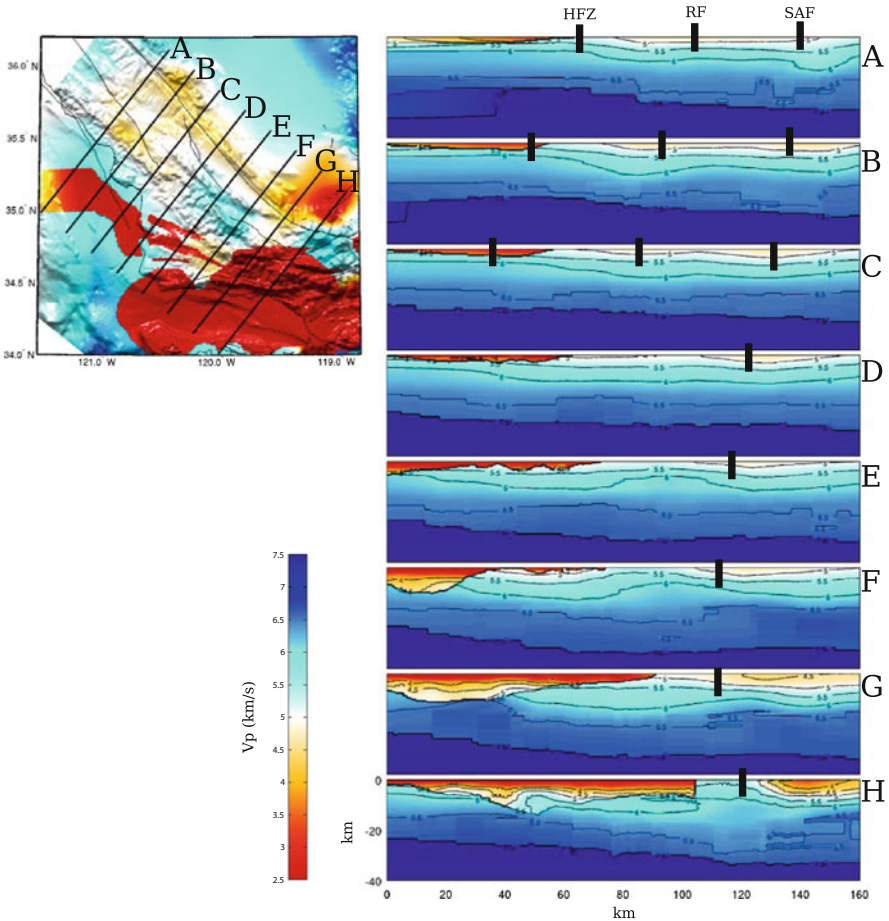


Fig. 6.56 Cross-sections of the P-velocity in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

1985; Terres and Luyendyk 1985; Hornafius et al. 1986; Luyendyk 1991), which is in agreement with earlier geological studies of disrupted lithotectonic belts (e.g., Jones et al. 1976; Crouch 1979) and sediments of the Eocene Poway Group (Yeats et al. 1974). For the eastern Transverse Ranges, paleomagnetic data record about 40° of clockwise rotation between 10 and 4.5 Ma (Carter et al. 1987; Powell 1993; Richard 1993; Dickinson 1996) which is in agreement with geophysical constraints in (Langenheim and Powell 2009). In Pliocene to Pleistocene, the southern segment of the Pacific-North America plate boundary shifted eastward into the Gulf of California and the plate boundary had to break a new path across southern California (e.g., Atwater and Stock 1998). This reconfiguration of the plate boundary produced substantial transpressional deformation that uplifted and deformed the Transverse Ranges to their current state (e.g., Luyendyk et al. 1980; Dickinson

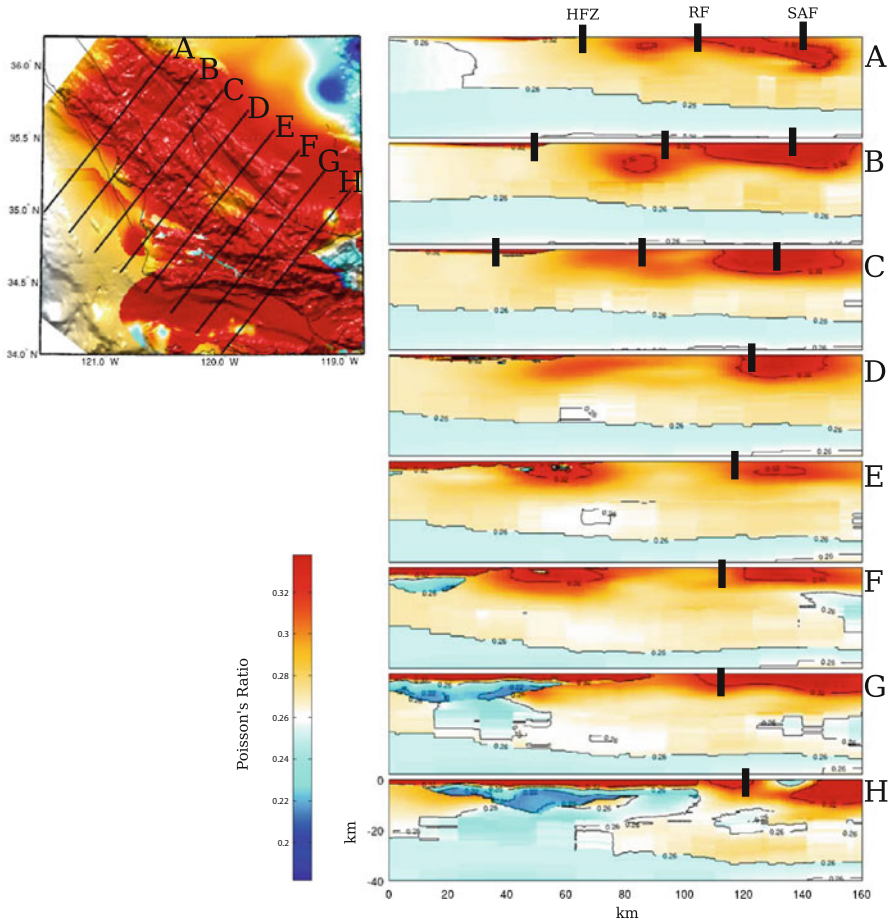


Fig. 6.57 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 8 profiles (A-H) crossing the southern Coast Ranges. The format is identical to that in Fig. 6.49.

1996). One consequence of the block rotations was the opening of Neogene basins at the join of the rotated and unrotated blocks (e.g., Luyendyk et al. 1980). This mechanism explains, maybe in part, the opening of some of the sedimentary basins along the north and south borders of the Transverse Ranges, such as the Santa Maria Basin (Sect. 6.2.6.2), the Cuyama Basin (Sect. 6.2.6.1) and the Los Angeles Basin (Sect. 6.2.8.1) (Blake Jr et al. 1978; Luyendyk et al. 1980). An alternative, core complex-like crustal extension mechanism was also suggested to explain the opening of some of these basins in e.g., (Crouch and Suppe 1993).

In the western Transverse Ranges, surface rock types include Mesozoic ophiolitic and Franciscan rocks, Mesozoic through Paleogene shelf sediments of the Great Valley type and Neogene sedimentary and volcanic rocks (e.g., Keller and Prothero

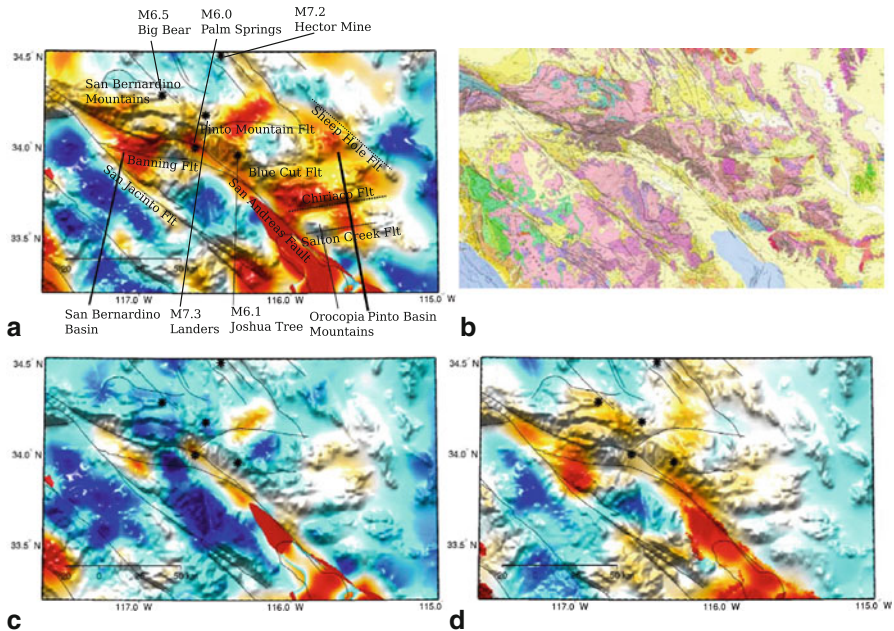


Fig. 6.59 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) of the eastern Transverse Ranges region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1. Black stars indicate the epicenters of several major earthquakes in this region.

1987). The S-velocity at 2-km depth in CVM-S4.26 shows good agreement with surface distribution of sediments and rock types (Fig. 6.58). The correlation among the 3 CVMs is also quite high at shallow depths. The seismic velocities in CVM-S4.26 are also well correlated with gravity and magnetic anomalies in this region (e.g., Fig. 3b in Lee et al. 2014b), which also show the east-west trend of the geomorphic features. In the eastern Transverse Ranges, surface rock types include Pliocene and Quaternary sediments, Proterozoic and Mesozoic plutonic and metamorphic rocks (Powell 1981, 1982; Langenheim and Powell 2009). A series of large and medium-sized earthquakes occurred in the eastern Transverse Ranges during the 1980's and early 1990's (Fig. 6.59), including the 1986 M6.0 Palm Springs earthquakes, the 1992 M6.1 Joshua Tree earthquake, the 1992 M7.3 Landers earthquake, the 1992 M6.5 Big Bear earthquake and the 1999 M7.2 Hector Mine earthquake. These large earthquakes and the numerous aftershocks following them may have disturbed the velocity structures in this region.

In addition to the cross-sections along the LARSE I&II profiles shown in (Lee et al. 2014a, b) in Figs. 6.60–6.62 I show cross-sections of CVM-S4.26 along 13 north-south profiles crossing the western Transverse Ranges and 10 north-south profiles crossing the eastern Transverse Ranges. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.63–6.68.

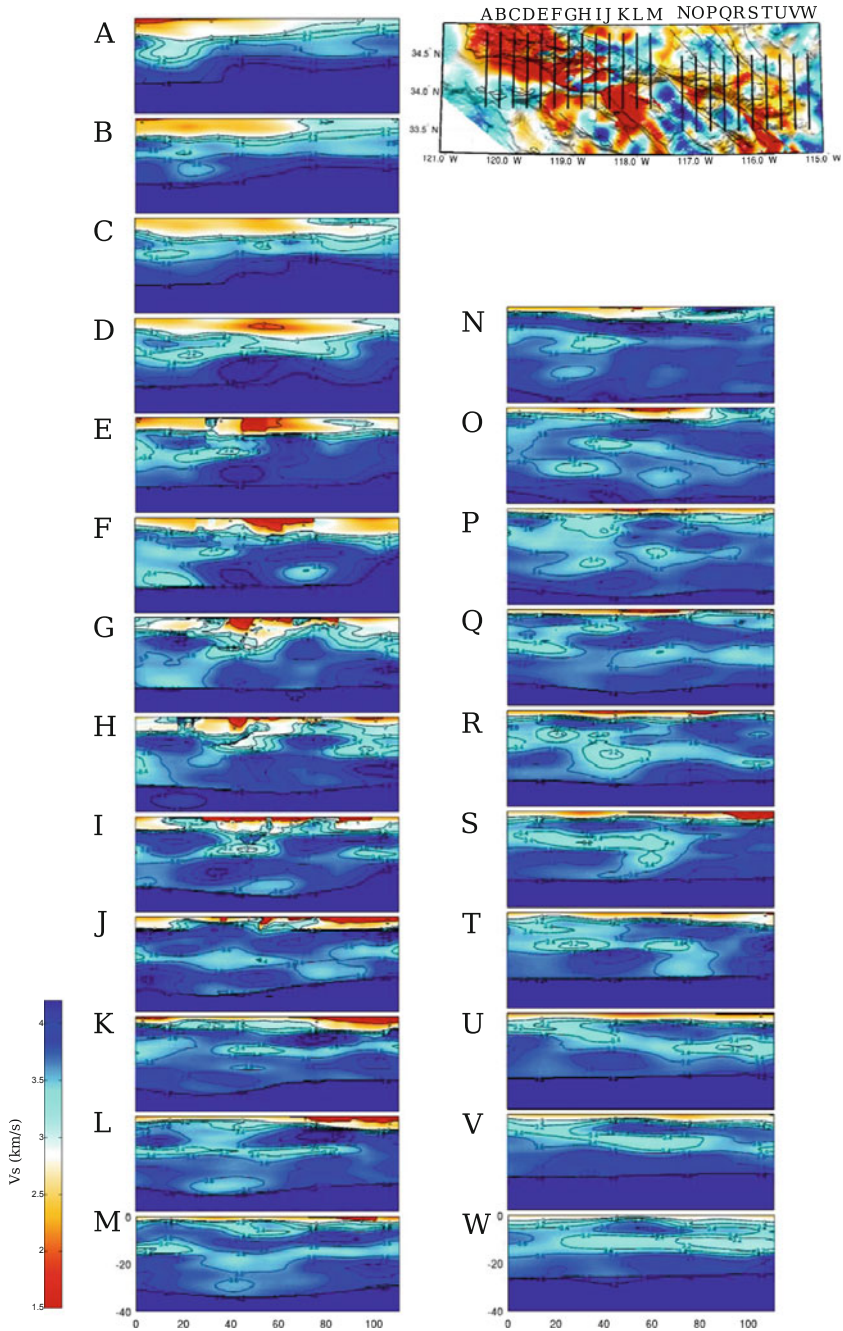


Fig. 6.60 Cross-sections of the S-velocity in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*.

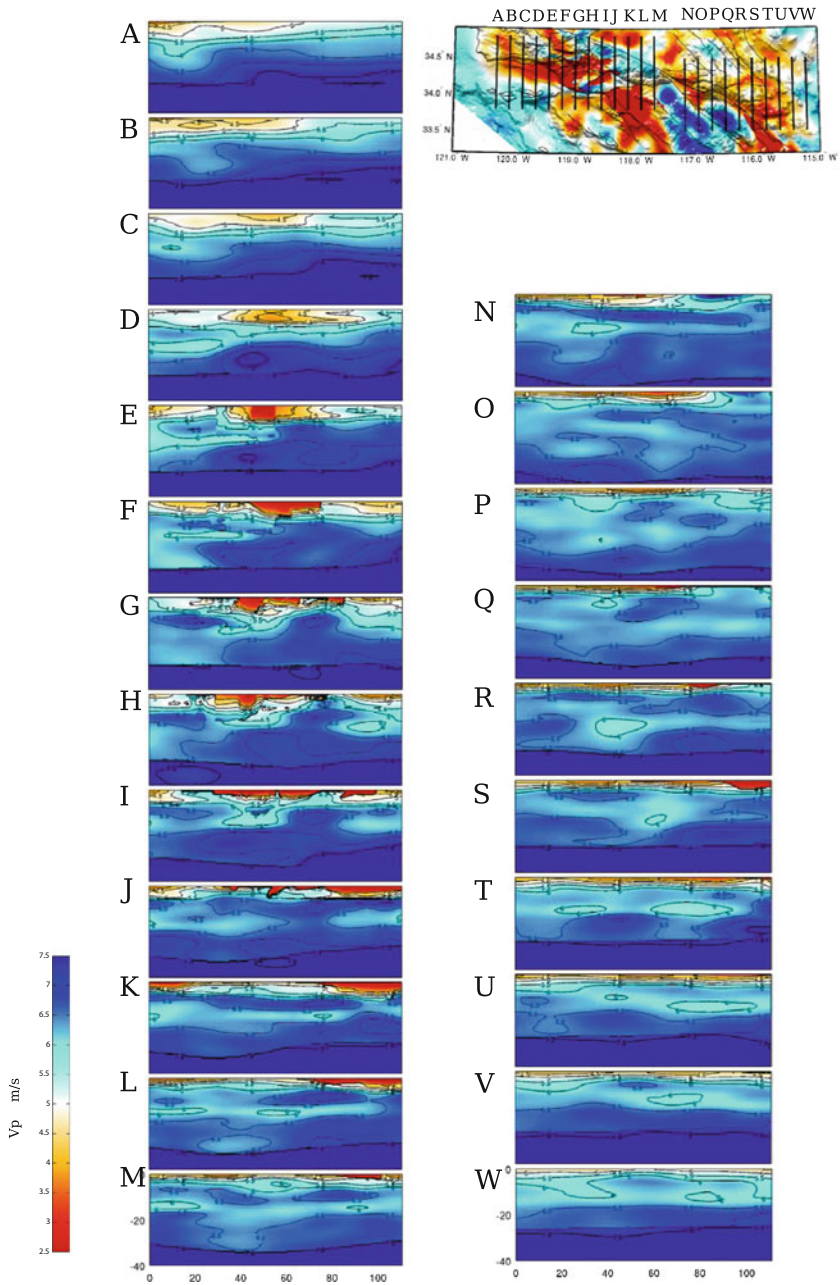


Fig. 6.61 Cross-sections of the P-velocity in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

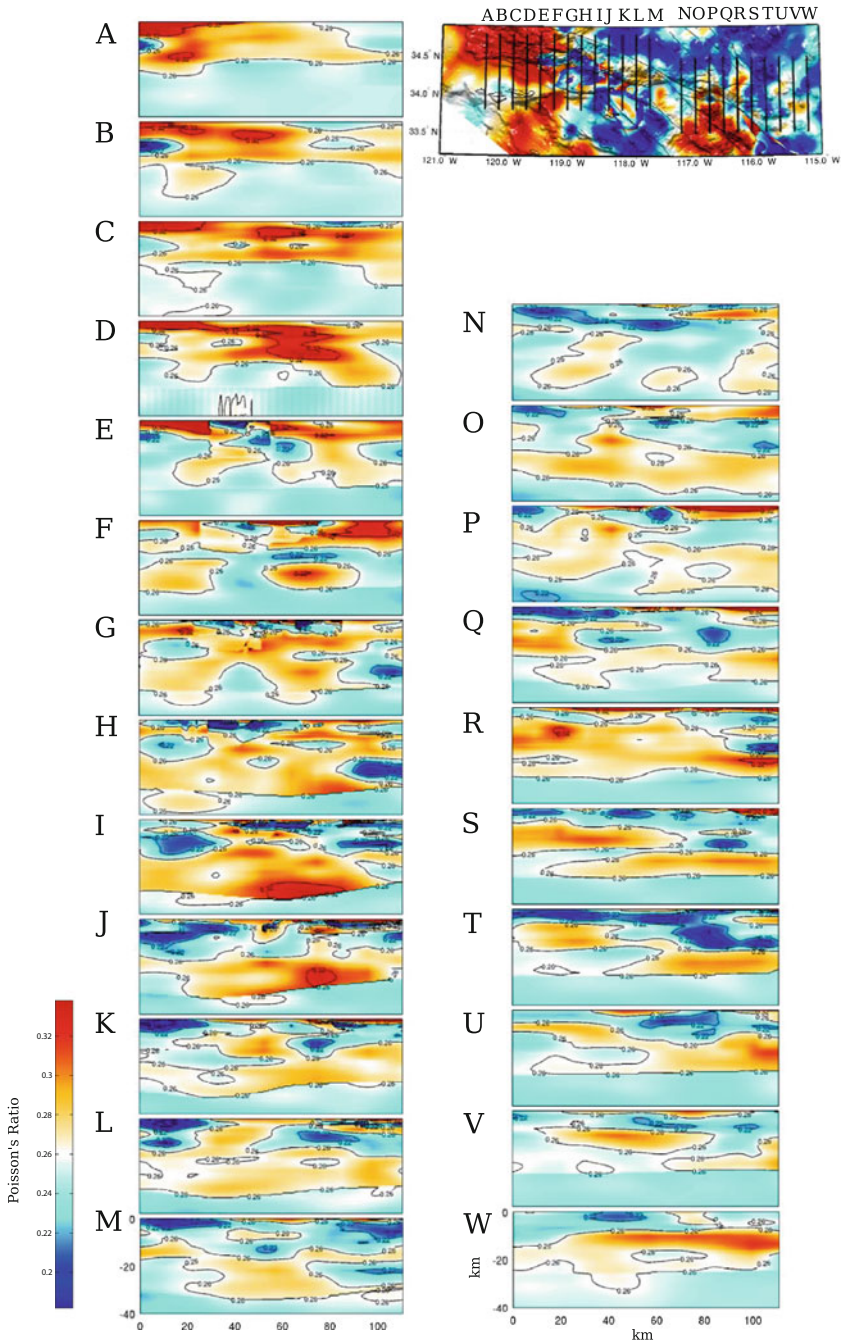


Fig. 6.62 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

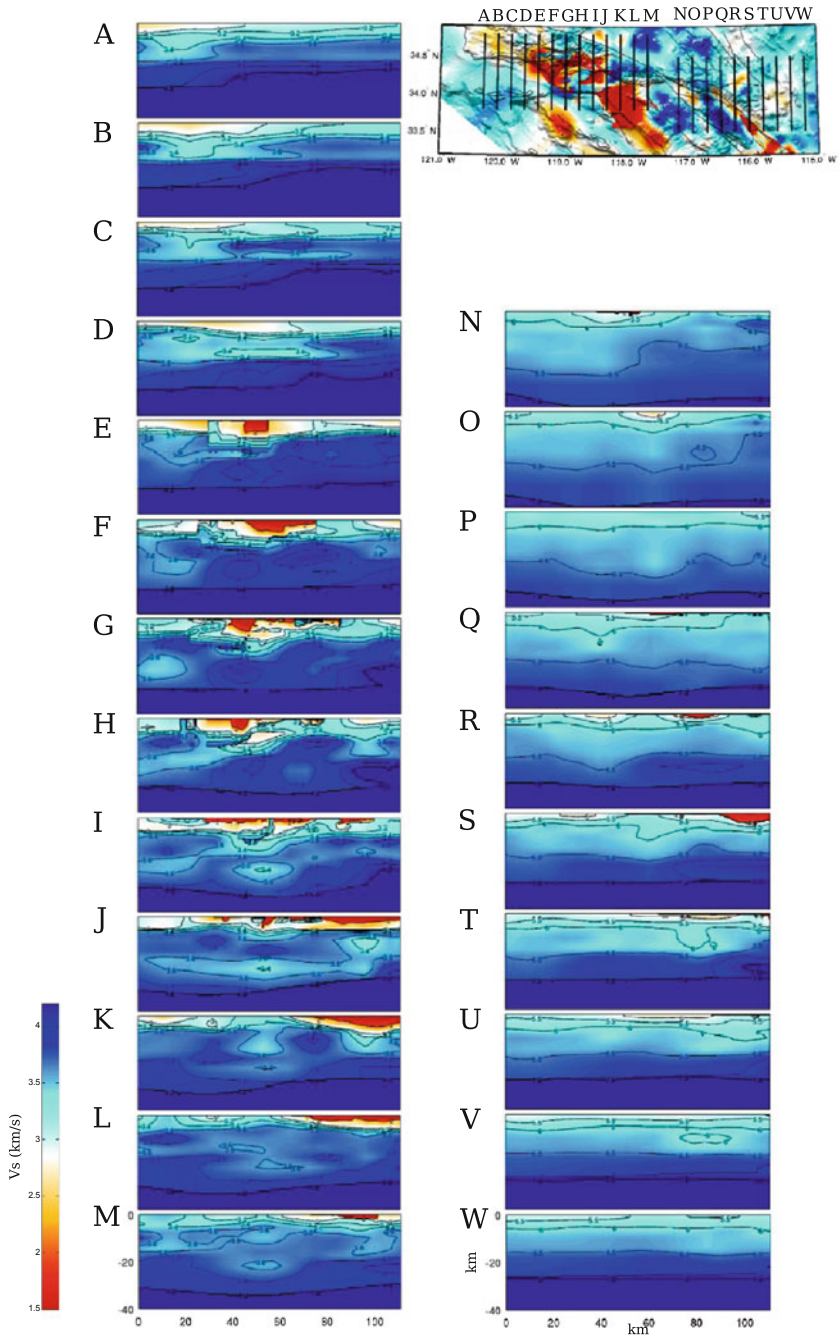


Fig. 6.63 Cross-sections of the S-velocity in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

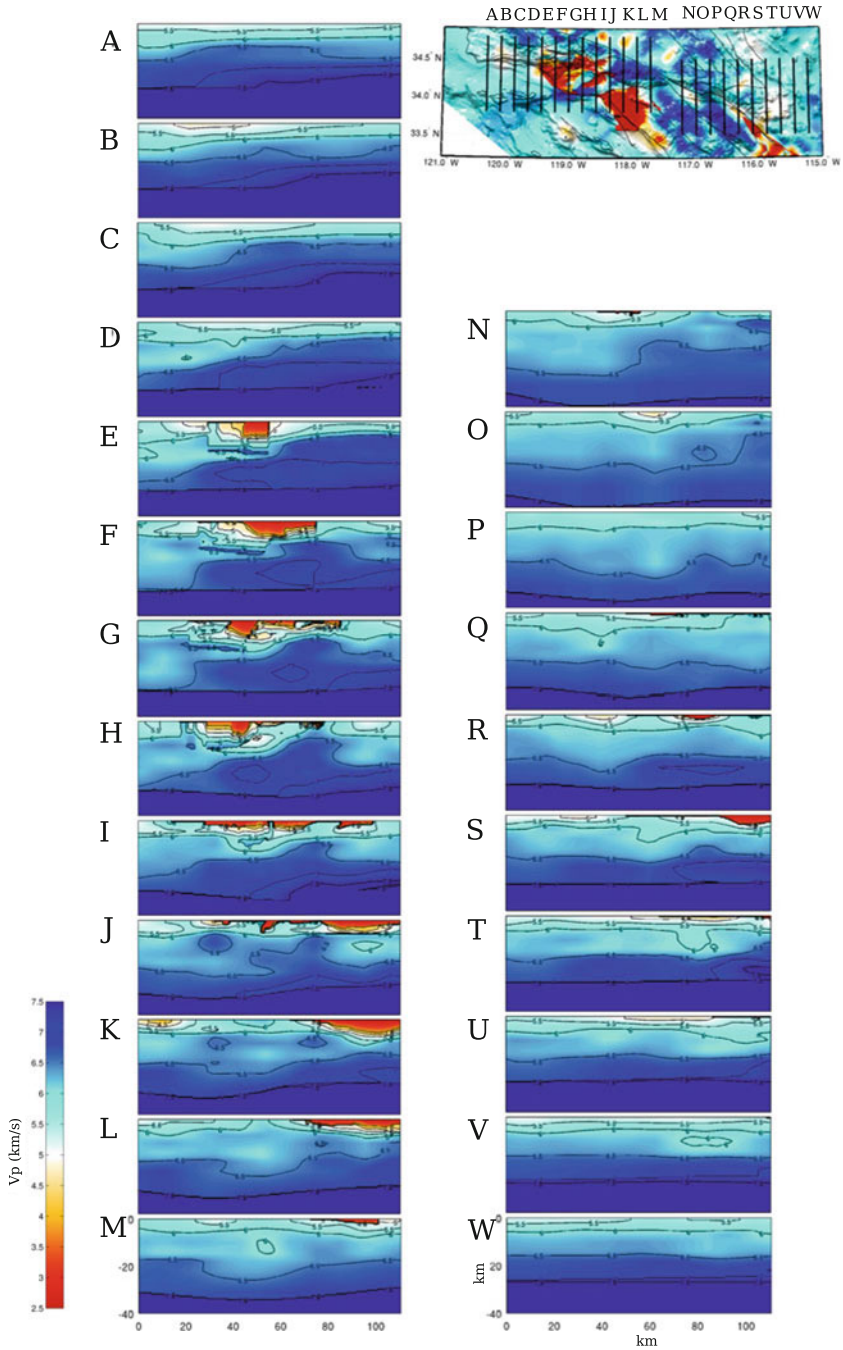


Fig. 6.64 Cross-sections of the P-velocity in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

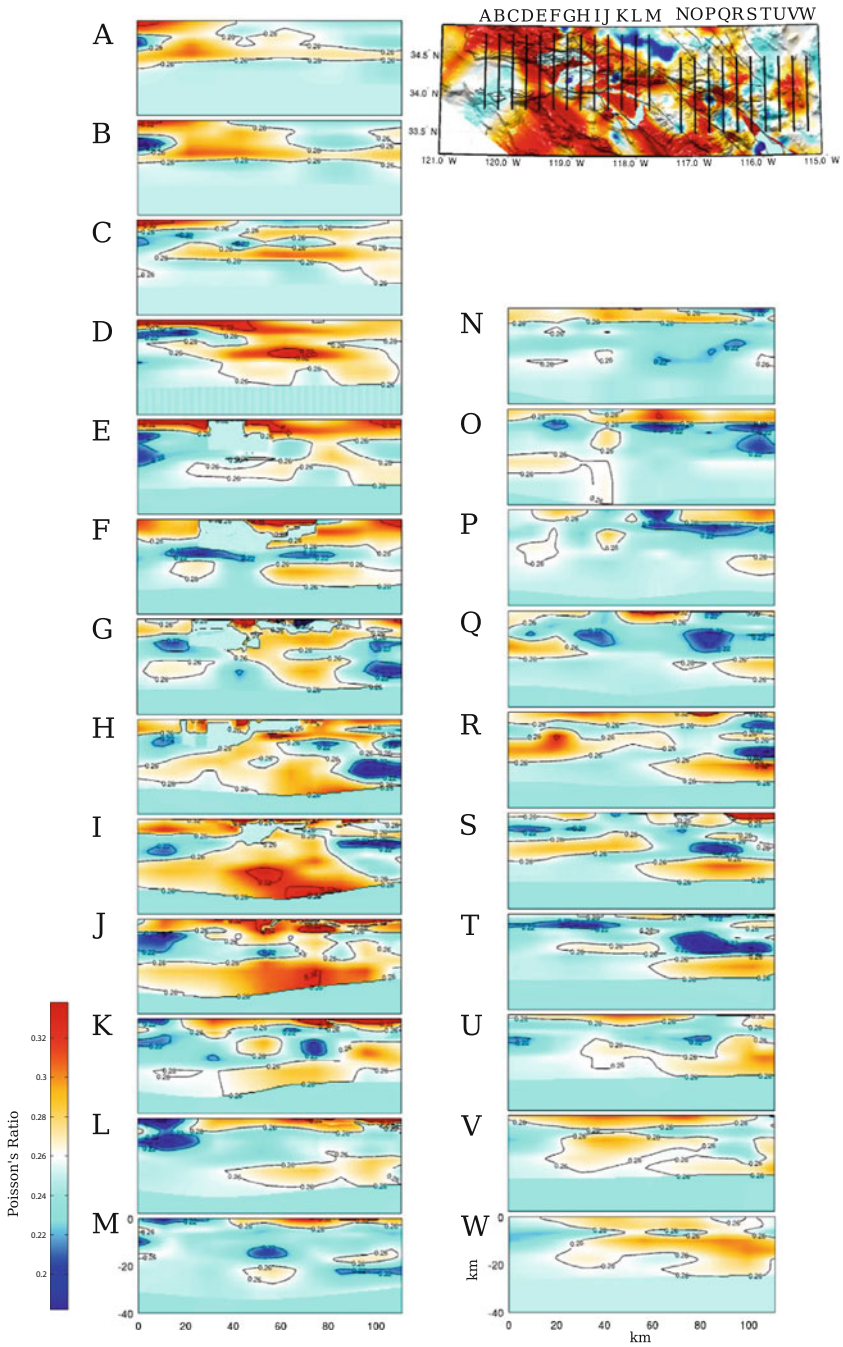


Fig. 6.65 Cross-sections of the Poisson's ratio in CVM-S4 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

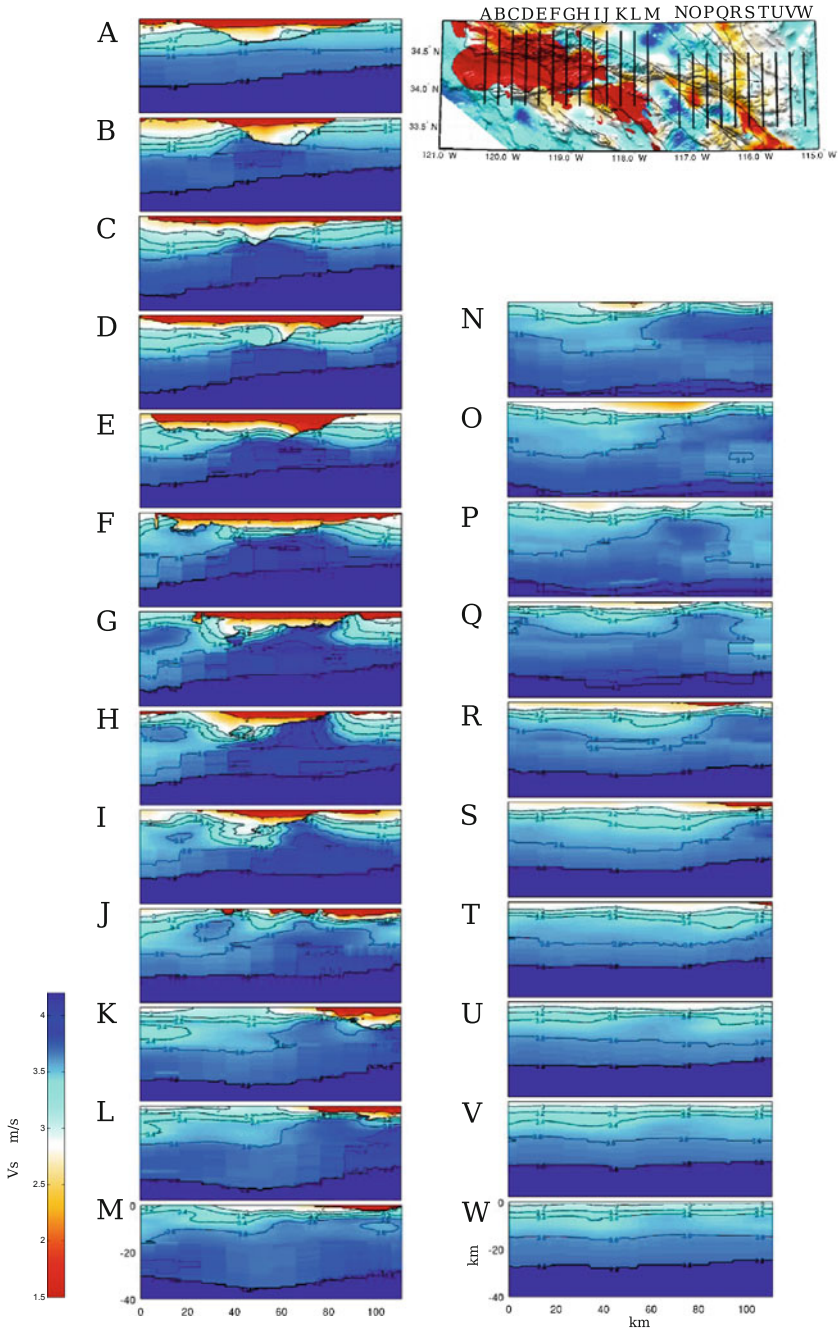


Fig. 6.66 Cross-sections of the S-velocity in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

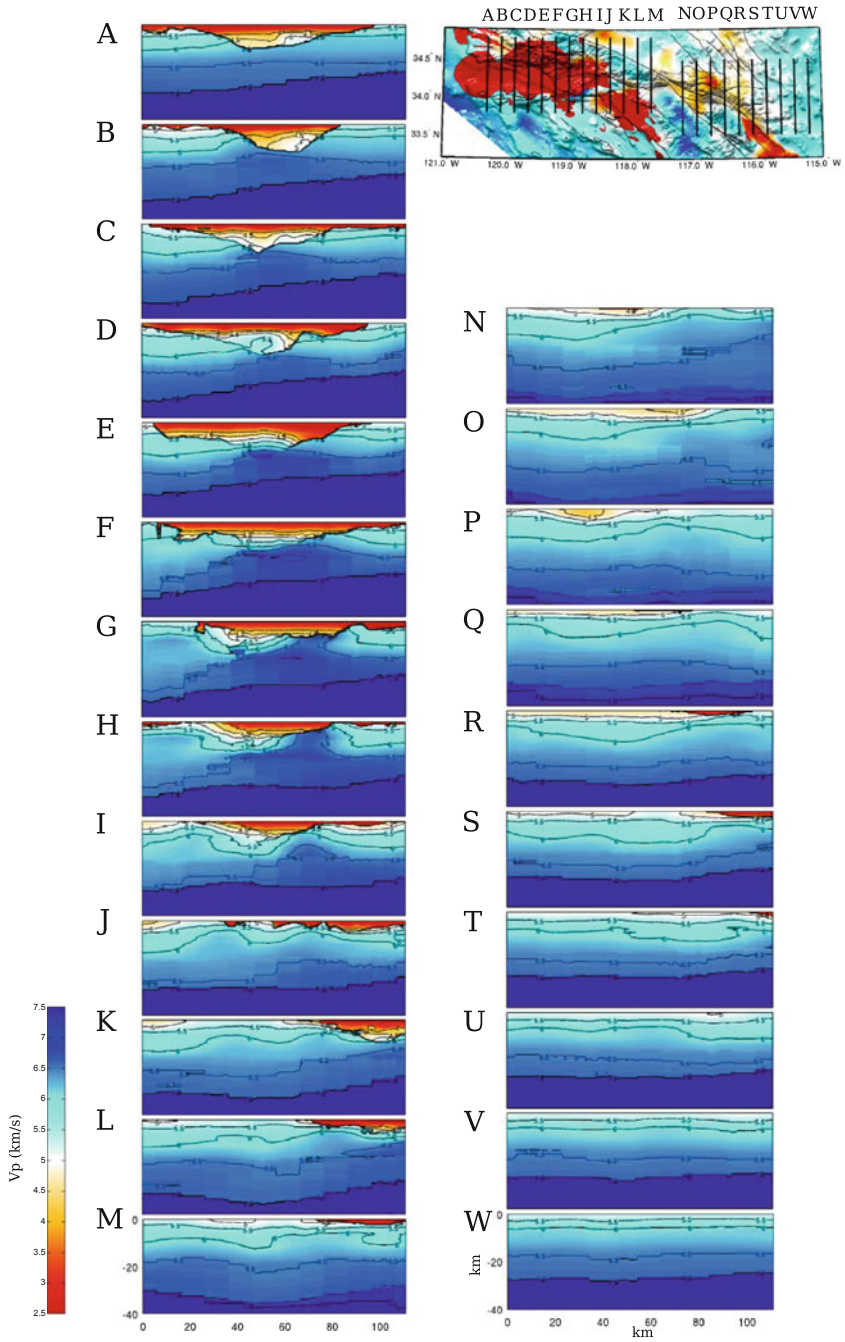


Fig. 6.67 Cross-sections of the P-velocity in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

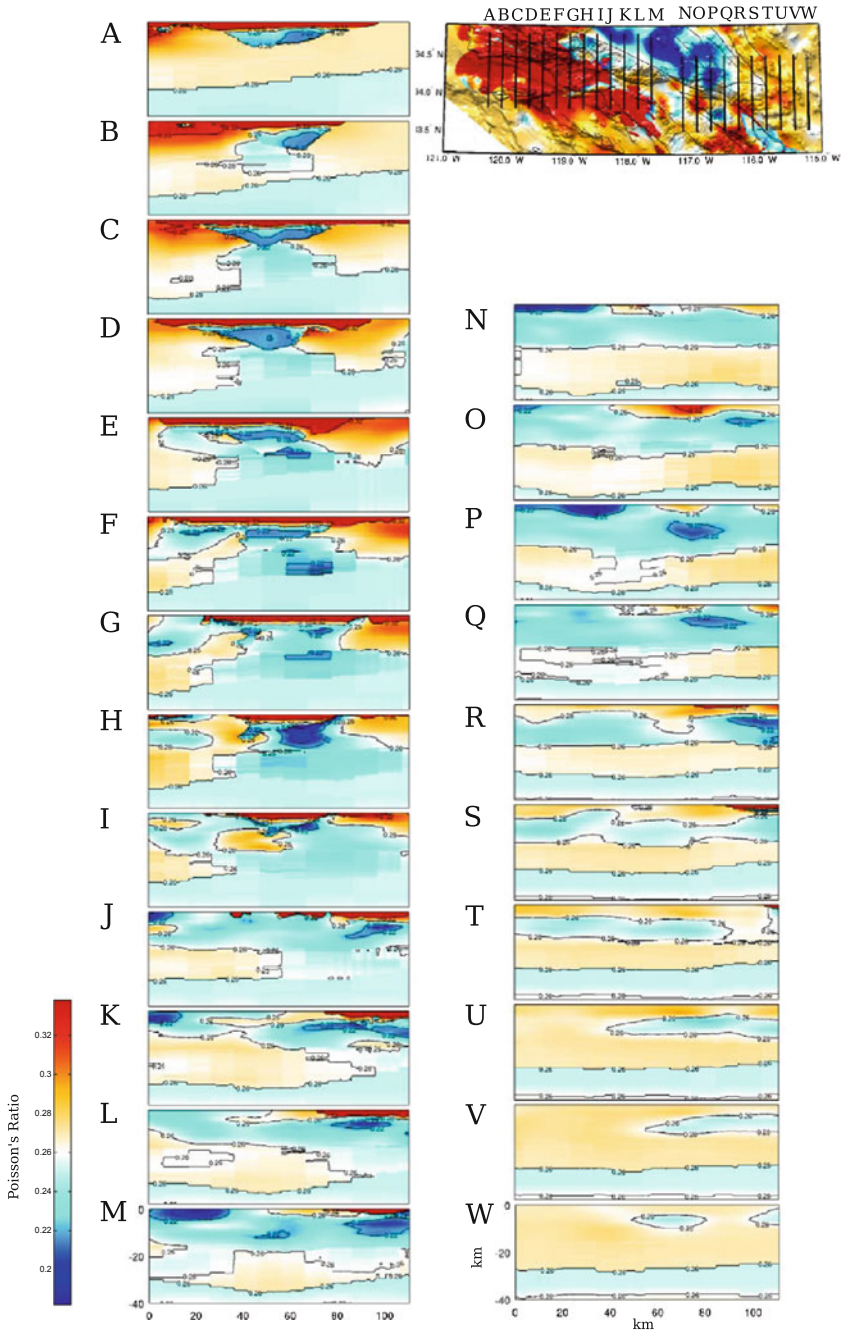


Fig. 6.68 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 23 profiles (A-W) crossing the western and eastern Transverse Ranges. The format is identical that in Fig. 6.60.

6.2.7.1 Ventura Basin—Santa Barbara Channel

The Ventura Basin lies in the main structural downwarp of a major fold and thrust belt that developed in late Pliocene (Keller 1995). The offshore extension of the structural downwarp is the Santa Barbara Channel. The sedimentary sequence mainly consists of Cretaceous to Pleistocene, mostly marine, sediments. In our starting model CVM-S4, the Ventura Basin is associated with a low-velocity anomaly (Figs. 6.63–6.64). Compared with our starting model, CVM-S4.26 has slightly higher velocities (i.e., positive perturbation) inside the Ventura Basin and much lower velocities in the Santa Barbara Channel (Figs. 6.60–6.61). This extension of the low-velocity anomaly from the Ventura Basin to the Santa Barbara Channel was also observed in the adjoint tomography in (Tape et al. 2010), in which an earlier version of CVM-H without the Santa Barbara Channel sedimentary basin was used as the starting model.

The relative perturbations between CVM-S4.26 and CVM-S4 show correlations with geological units surrounding the Ventura Basin (Fig. 6.58). To the east of the basin, positive perturbations are observed in the Topatopa Mountains, the Soledad Basin and the Sierra Pelona and negative perturbations are observed in the Santa Clarita Valley. To the southeast of the basin, negative perturbations are observed in the Santa Susana Mountains and the Simi Valley, which also contains thick sediments, and positive perturbations are associated with the Santa Monica Mountains and the San Fernando Valley. South of the Santa Monica Mountains, negative perturbations are associated with the offshore Santa Monica Basin.

6.2.7.2 San Bernardino Basin

The San Bernardino Basin is located next to the northwest edge of the eastern Transverse Ranges (Fig. 6.59). It was formed as a pull-apart structure between the San Andreas Fault and the San Jacinto Fault Zone in Pleistocene (Morton and Matti 1993). The sediments inside the basin are mainly Tertiary and Quaternary alluvial deposits (Dutcher and Garrett 1963). Compared with our starting model CVM-S4, the velocities inside the basin and to the northeast of the basin were reduced by 10–15 % while the velocities to the southwest of the basin were increased by 5–10 % in CVM-S4.26. At shallow depths, the velocities inside and around the basin in CVM-S4.26 are more consistent with the basin model of (Süss and Shaw 2003). The basin stops abruptly to the southwest along the San Jacinto Fault Zone. The sharp basin boundary is more obvious in CVM-S4.26 than in CVM-H11.9. The thickness of the sediments reduces more gradually toward northeast.

The basin geometry in CVM-S4.26 is more consistent with the one determined using water well and seismic refraction data in (Frankel 1993) and also with the model obtained using potential field and seismic data in (Anderson et al. 2004) than the basin geometries in CVM-S4 and CVM-H11.9. The basin models in (Frankel 1993; Anderson et al. 2004) and CVM-S3 (an earlier version of the CVM-S series velocity model for Southern California) were used in wave-propagation simulations

for the 1999 M7.1 Hector Mine earthquake in (Graves and Wald 2004). In that study, the synthetic seismograms computed using the basin models of (Frankel 1993; Anderson et al. 2004) provided substantially better fit to the observed seismograms at frequencies up to 0.5 Hz than those computed using CVM-S3, which did not have the more complex basin geometries present in the other two basin models.

6.2.8 Inner Continental Borderland

The Inner Continental Borderland (ICB) is the region bounded by the western Transverse Ranges (Santa Cruz Island Fault—Santa Monica Fault) on the north, by the California Outer Continental Borderland (East Santa Cruz Basin Fault) on the west and by the Peninsular Ranges on the east (Fig. 6.69). It includes the Los Angeles Basin, where the Los Angeles Metropolitan area lies. The Outer Borderland to the west and the western Transverse Ranges to the north are mainly composed of the Great-Valley-type forearc-basin sequence. The ICB has a quite different composition. Studies on the rock outcrops, core samples from oil wells and seismic reflection/refraction profiles have shown that the ICB basement is mainly composed of medium- to high-grade metamorphic rocks known as the Catalina Schist (e.g., Crouch and Suppe 1993, Bohannon and Geist 1998; ten Brink et al. 2000).

Several mechanisms have been suggested for the origin of the ICB. One possibility is that the ICB was emplaced by northward translation of an exotic terrain along strike-slip faults (e.g., Vedder 1987). The recent transpressional deformation in the region may have brought mid- and lower-crustal rocks to shallow depths (e.g., Mann and Gordon 1996).

The model proposed in (Crouch and Suppe 1993) relates the formation of the ICB with the clockwise rotation of the western Transverse Ranges. Because the western Transverse Ranges block was deeply embedded in the continent and the rotation was mostly around a northern pivot point, a gap opened on the east side of the rotating block. This gap was then filled from below by igneous rocks and the metamorphosed Franciscan accretionary wedge rocks underplated during the subduction of the Farallon Plate. This upwelling process was similar to the metamorphic core complex extensions found in the Basin and Range Province and elsewhere (e.g., Wernicke et al. 1992). The Catalina Schist is therefore the exhumed Franciscan rocks, originally buried at mid-crustal depths. The layers above the Franciscan rocks, which were mainly Great Valley forearc sequence and its underlying Coast Range Ophiolite basement, now comprise the Santa Monica Mountains and Simi Hills to the north and San Joaquin Hills to the east of the ICB. The faults surrounding the ICB, including the Newport-Inglewood Fault and the Whittier-Elsinore Fault, were originally high-angle normal faults terminating at a major landward-dipping detachment, along which the lower Franciscan layer was separated from the upper Great Valley forearc and its ophiolite basement layers.

In Fig. 6.69, I show comparisons of the S-velocity at 2-km depth in CVM-S4.26, CVM-S4 and CVM-H11.9 with a geologic map in the ICB region. The geologic map

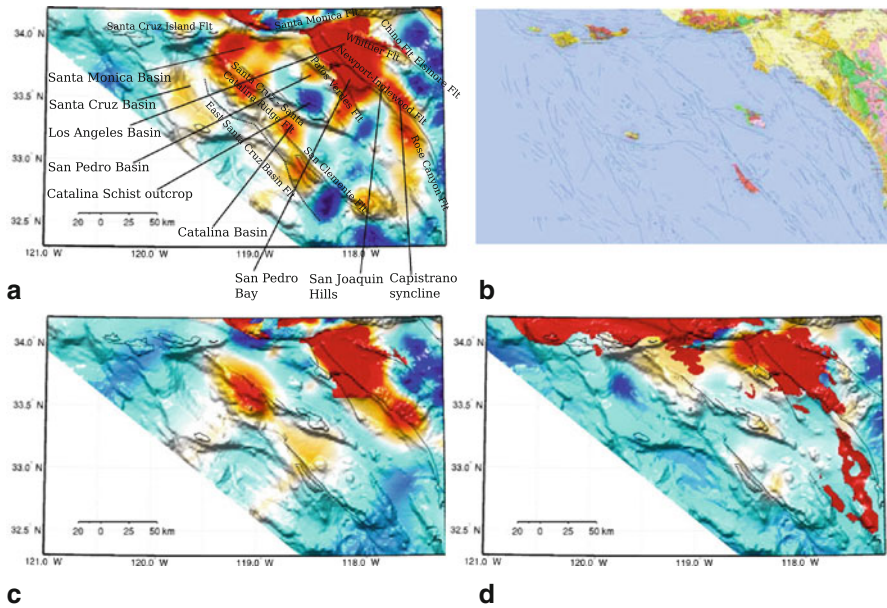


Fig. 6.69 S-velocity at 2-km depth in CVM-S4.26 (**a**), CVM-S4 (**c**) and CVM-H11.9 (**d**) compared with the geologic map (**b**) of the Inner Continental Borderland region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1.

does not cover the submerged areas. In the onshore areas, low-velocities in CVM-S4.26 show good correlation with the distribution of sediments on the geologic map. The Los Angeles Basin is shown as the major low-velocity anomaly on all three CVMs. In the offshore areas, CVM-S4.26 shows more pronounced low-velocity anomalies that are correlated with offshore basins than the other two CVMs. One of the high-velocity anomalies in CVM-S4.26 is located around the Santa Catalina Island and might be associated with the Catalina Schist outcrops that have been extensively studied in e.g., (Woodford 1924; Bailey et al. 1941; Platt 1975; Legg 1991). In Figs. 6.70–6.72, I show cross-sections along 11 profiles crossing the ICB region in CVM-S4.26. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.73–6.78.

6.2.8.1 Los Angeles Basin

The sediments inside the Los Angeles Basin are up to 10 km thick and consist of Upper Cretaceous marine clastic rocks, a thick section of Tertiary marine sedimentary and volcanic rocks and Quaternary marine and nonmarine clastic deposits (Wright 1991). In (Chen et al. 2007), a 3D crustal structure model for the Los Angeles Basin area, LAF3D, was obtained through full-3D tomography based on the scattering-integral method (F3DT-SI) using CVM-S3 as the starting model. The

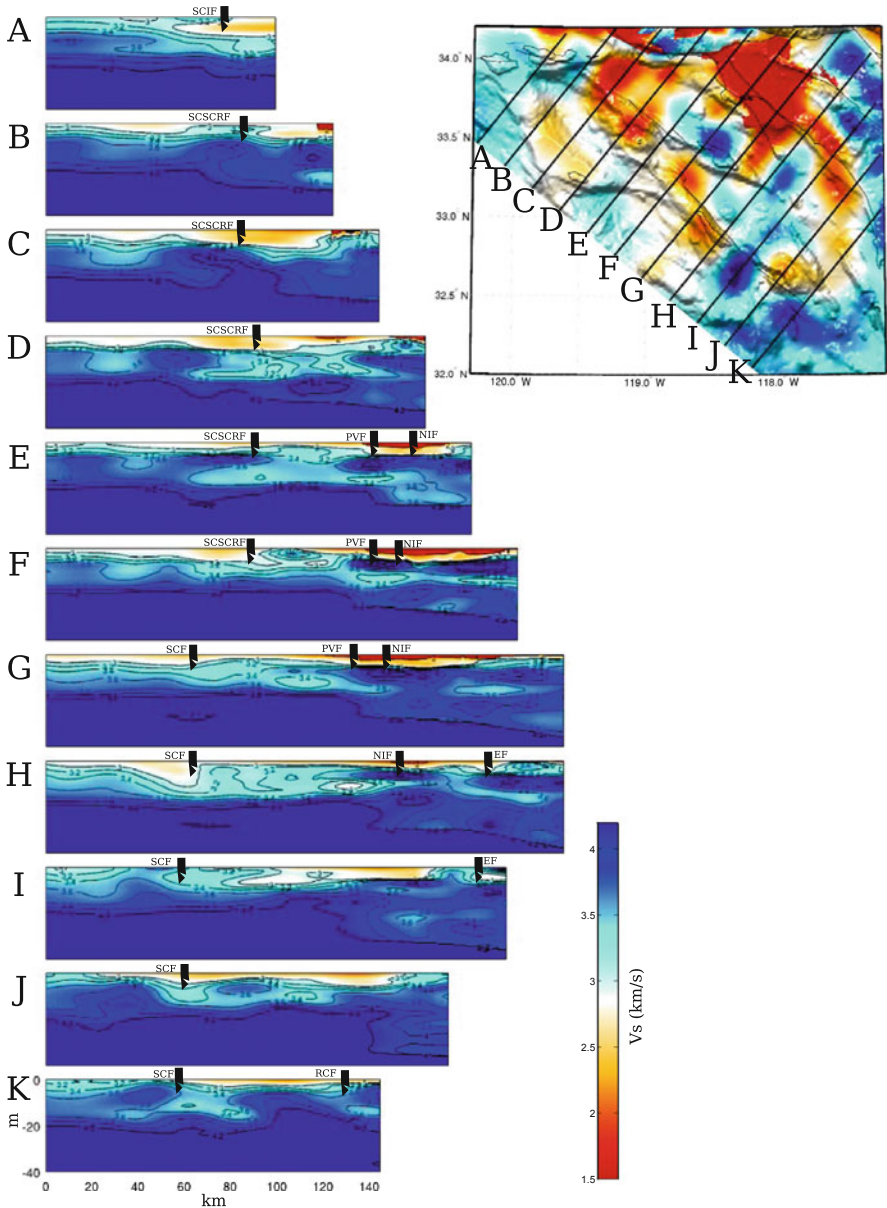


Fig. 6.70 Cross-sections of the S-velocity in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-right corner*. Major faults are shown as *black dots* on the cross-sections. Elsinore Fault (*EF*), Newport-Inglewood Fault (*NIF*), Palos Verdes Fault (*PVF*), Rose Canyon Fault (*RCF*), Santa Cruz Island Fault (*SCIF*), Santa Cruz—Santa Catalina Ridge Fault (*SCSCRF*).

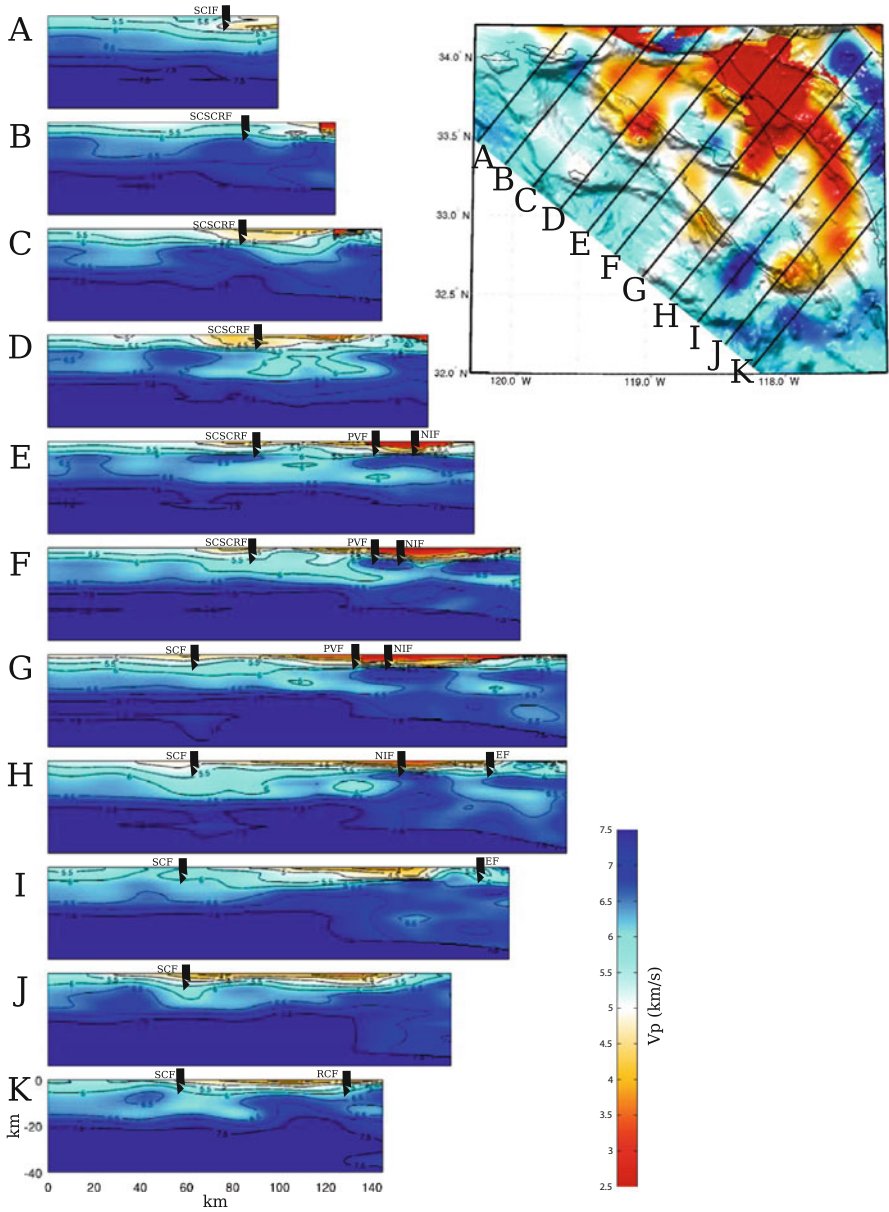


Fig. 6.71 Cross-sections of the P-velocity in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

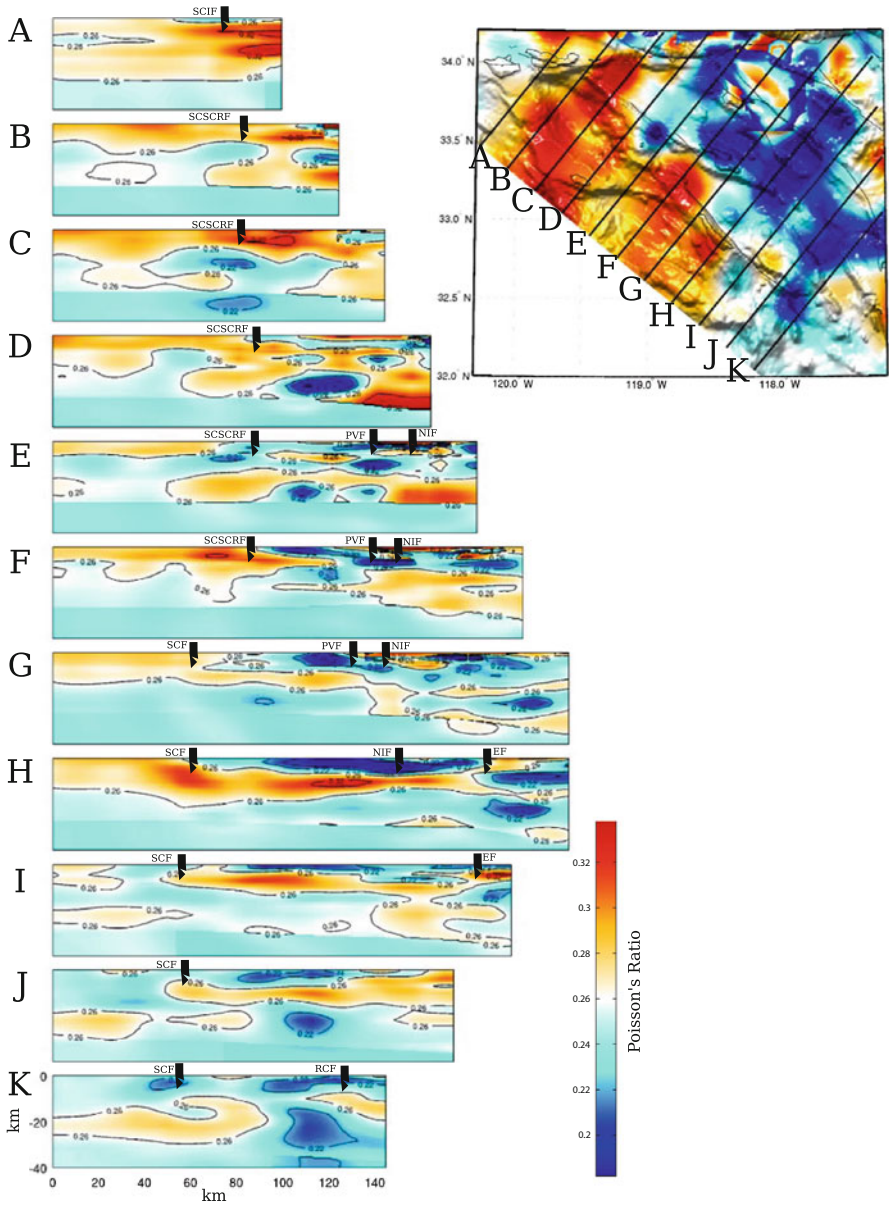


Fig. 6.72 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

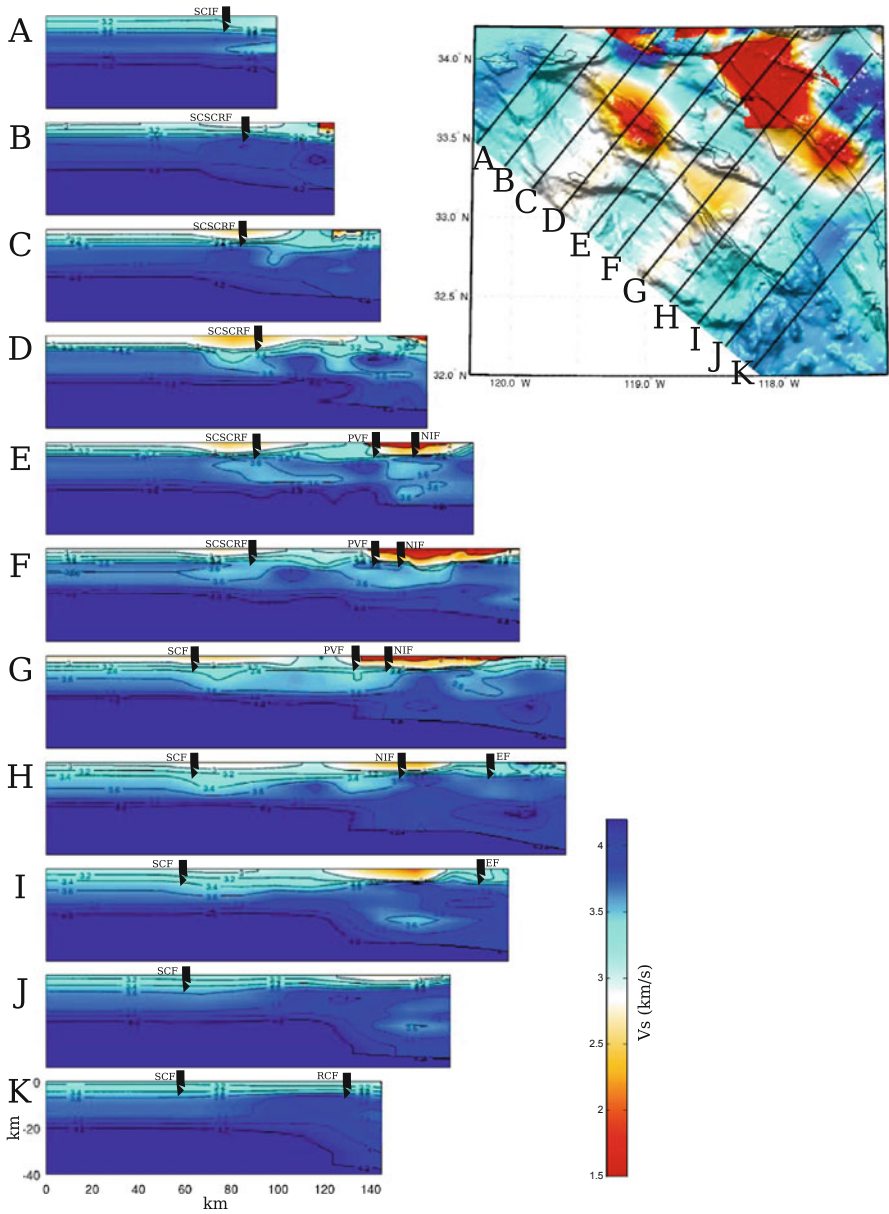


Fig. 6.73 Cross-sections of the S-velocity in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

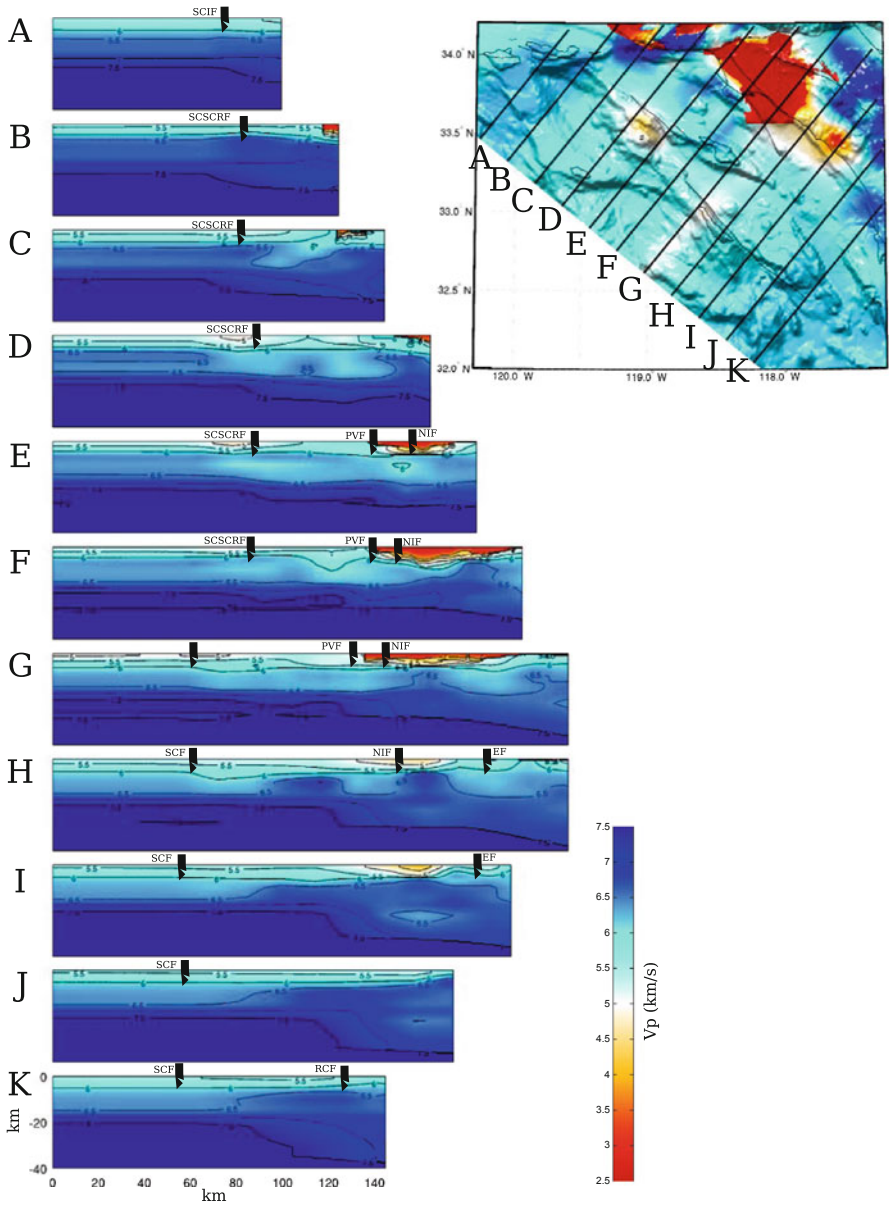


Fig. 6.74 Cross-sections of the P-velocity in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

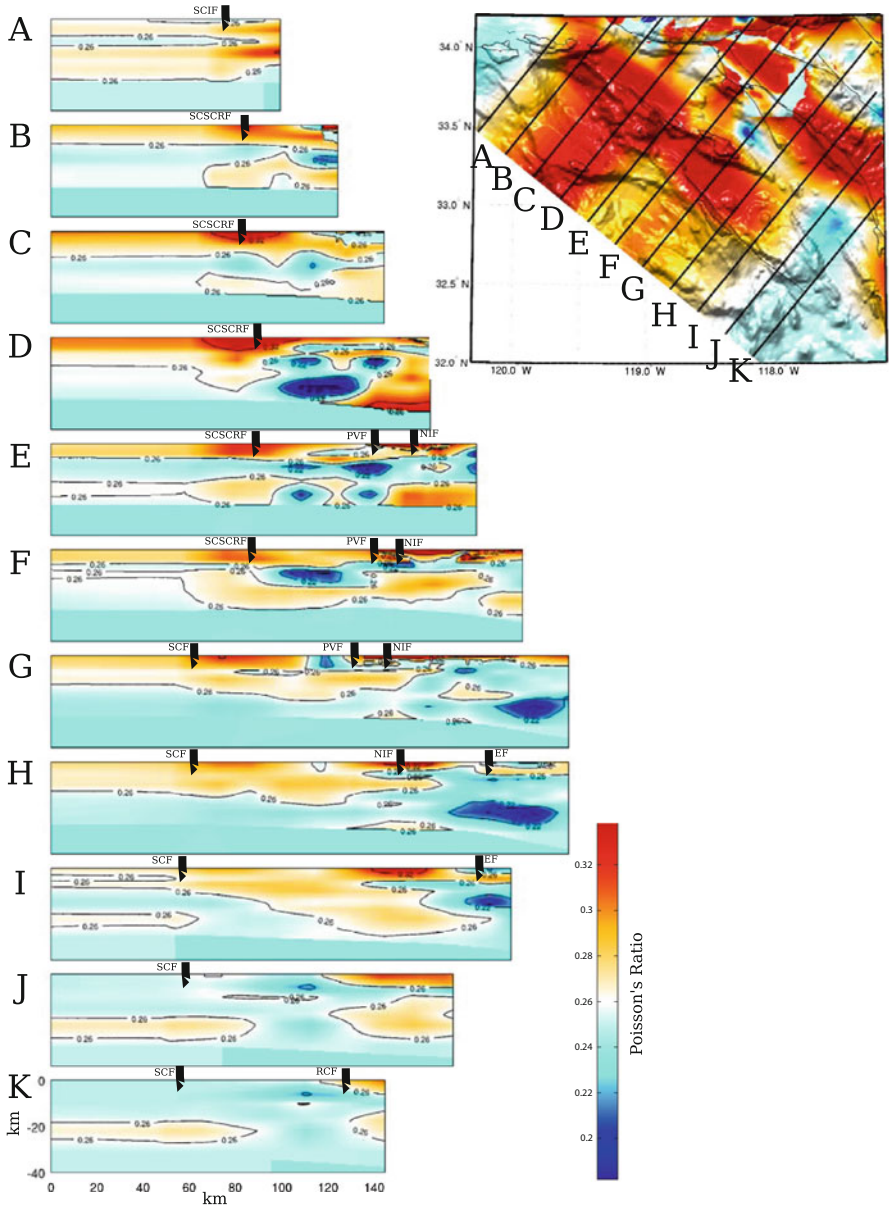


Fig. 6.75 Cross-sections of the Poisson's ratio in CVM-S4 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

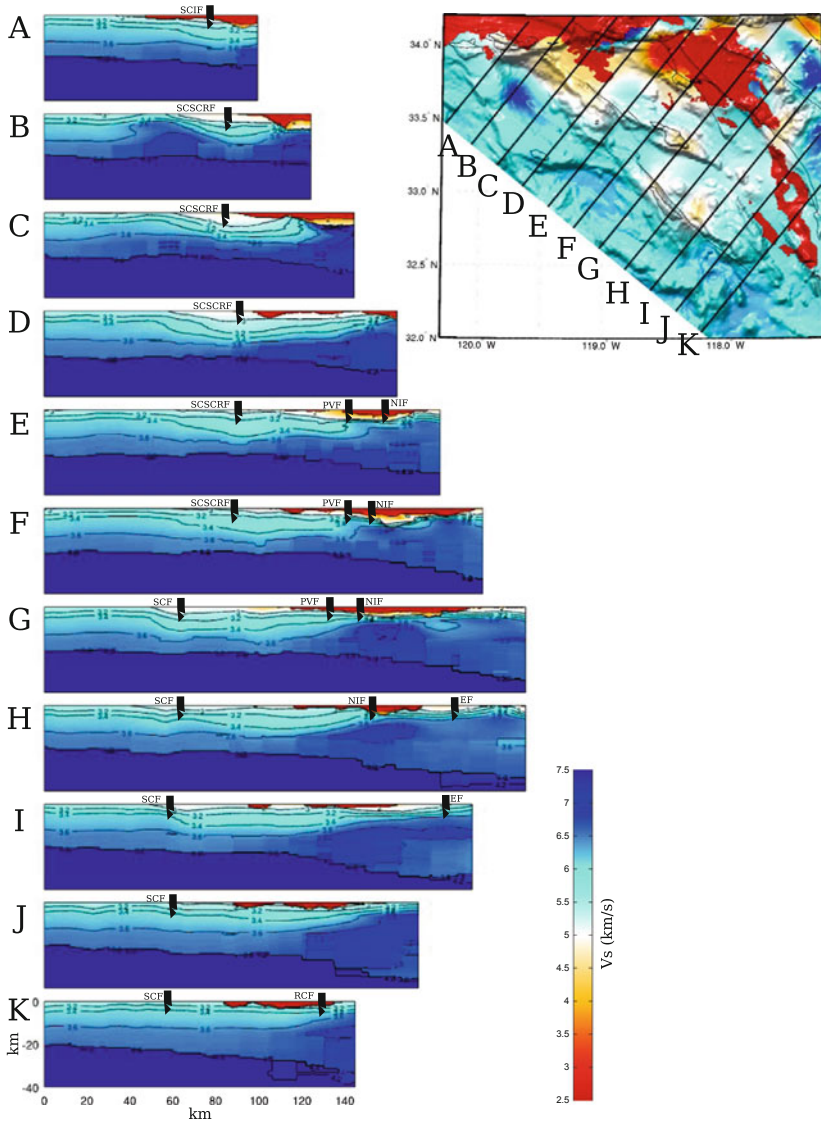


Fig. 6.76 Cross-sections of the S-velocity in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

seismograms used in that inversion had a maximum frequency of 1.2 Hz, which is significantly higher than the frequency band used for deriving CVM-S4.26, which was up to 0.2 Hz. LAF3D shows structural details that do not exist in CVM-S4.26. A general observation made in (Chen et al. 2007) was that LAF3D slightly increased the velocities inside the basin and decreased the velocities around the border of the

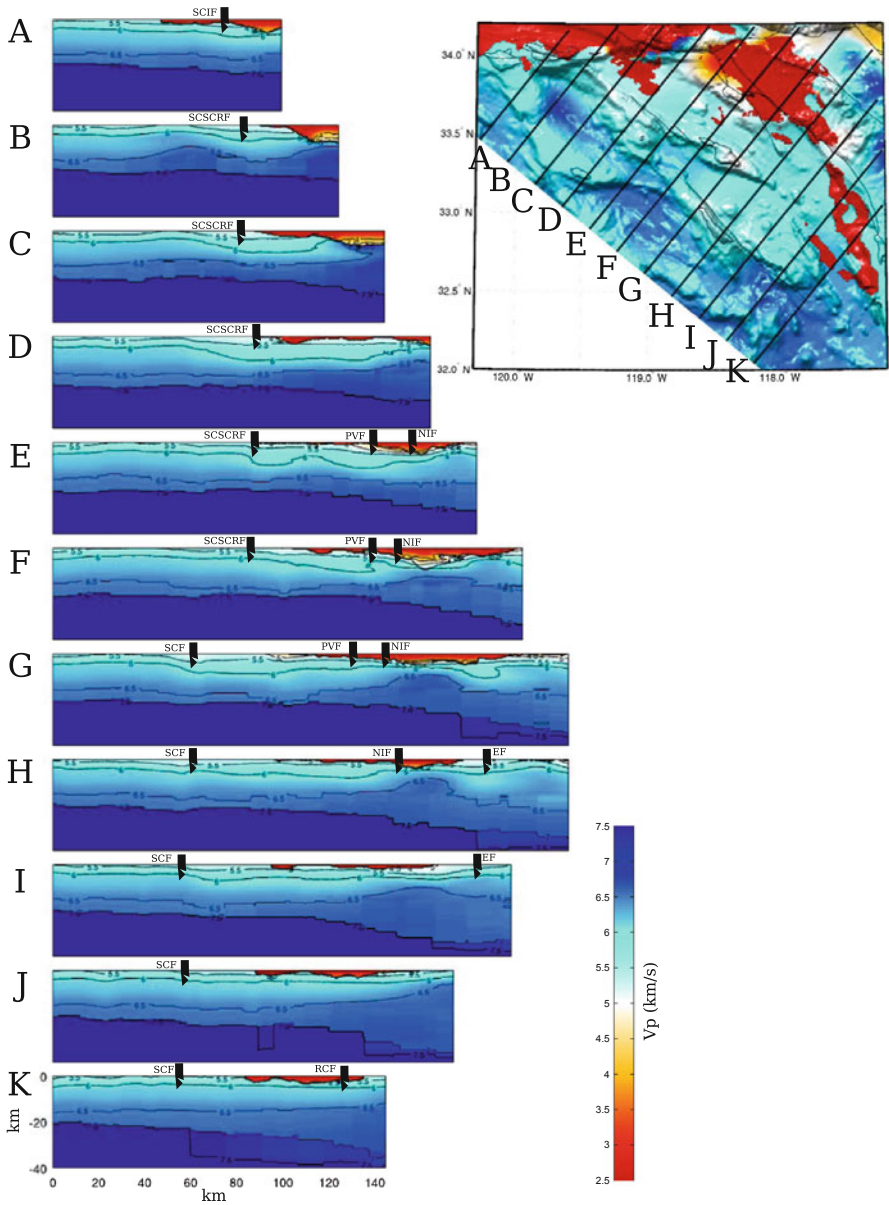


Fig. 6.77 Cross-sections of the P-velocity in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

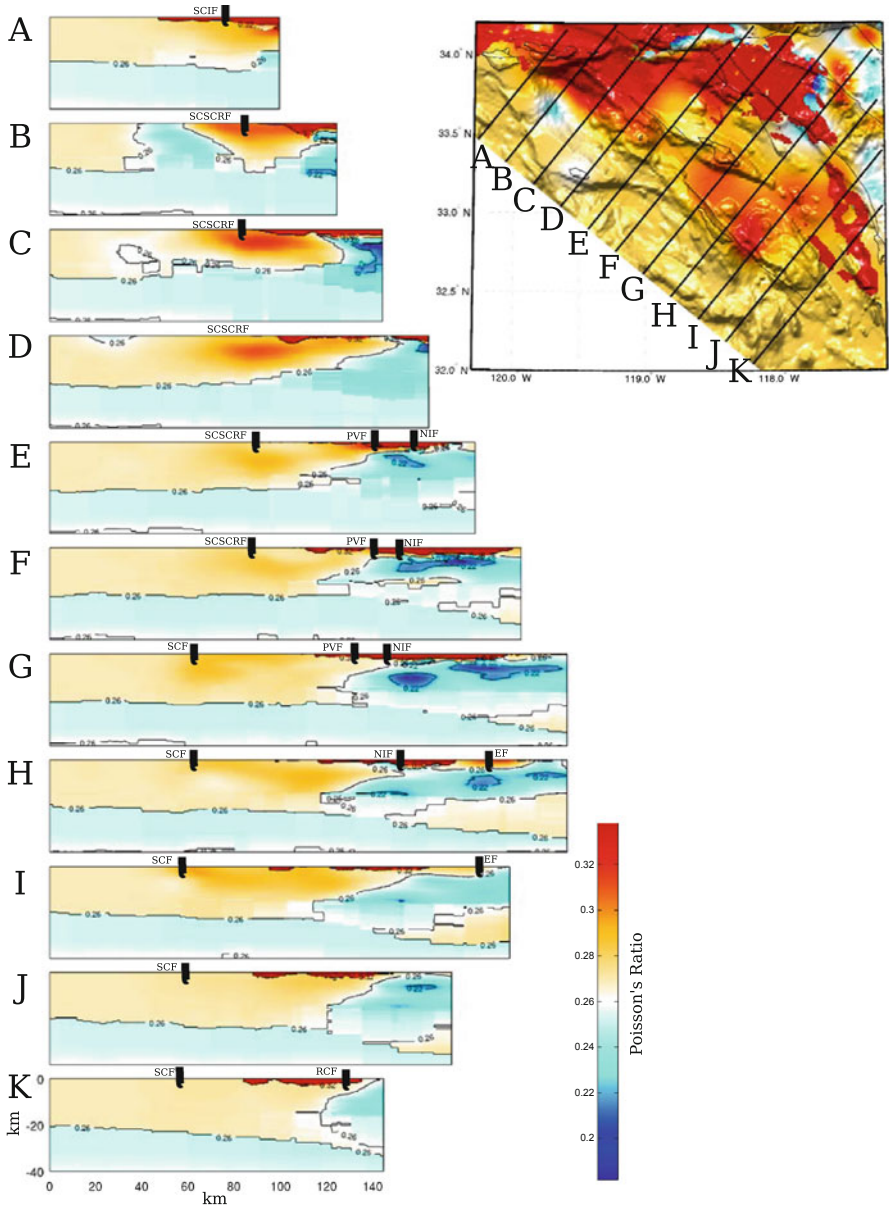


Fig. 6.78 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 11 profiles (A-K) crossing the Inner Continental Borderland. The format is identical to that in Fig. 6.70.

basin, which brought CVM-S3 into better agreement with the Harvard model for this basin (Süss and Shaw 2003). The perturbations between CVM-S4.26 and CVM-S4 are still consistent with this general observation made in (Chen et al. 2007). Positive perturbations are observed in the interior of the basin and negative perturbations are observed around the basin boundary. At shallow depths, in areas southwest of the Newport-Inglewood Fault extending towards the Rose Canyon Fault and in the San Pedro Bay extending offshore, the velocities have been reduced by more than 10 % in CVM-S4.26 compared with CVM-S4 (Fig. 6.69). In particular, the velocities around the San Joaquin Hills were increased slightly and those around the Capistrano syncline are reduced. These perturbations also brought the basin structure closer to that in (Süss and Shaw 2003).

Beneath the basin, CVM-S4.26 has a high-velocity layer with P- and S-velocity exceeding 7.0 km/s and 4.2 km/s, respectively (profile E-H in Figs. 6.70–6.71). This high-velocity layer is not visible in CVM-S4 or CVM-H11.9 (Figs. 6.73–6.74, 6.76–6.77), but can be seen on the active-source refraction tomography model along the LARSE-I profile (e.g., Fig. 8 in Lee et al. 2014b; Lutter et al. 1999). This high-velocity layer might be fragments of the Coast Range Ophiolite basement originally sandwiched between the Great Valley forearc basin on the top and the Franciscan accretionary wedge on the bottom. If this high-velocity layer also has high density, the formation of the Los Angeles Basin may have been driven, perhaps in part, by the isostatic response.

6.2.8.2 Offshore Region

The offshore region in ICB is believed to be underlain by mainly Catalina Schist, which is composed of mainly metamorphosed Franciscan accretionary wedge (e.g., Vedder 1987; Crouch and Suppe 1993; ten Brink et al. 2000). Beneath the sediments and above the Moho, the average P-velocity in CVM-S4.26 is between 5.5 km/s and 6.0 km/s (average S-velocity in CVM-S4.26 is between 3.2 km/s and 3.4 km/s), which is in broad agreement with the results in (ten Brink et al. 2000) and also consistent with laboratory measurements of the Pelona Schist around the Mojave region (Pellerin and Christensen 1998), which is believed to be similar to the Catalina Schist. This schist layer extends from the ICB eastward, passing underneath the Los Angeles Basin and goes into the Mojave Desert (Sect. 6.2.9.2; Fig. 10 in Lee et al. 2014b).

Several large offshore faults can be correlated with strong velocity gradients on the cross-sections for CVM-S4.26. For instance, the Santa Cruz—Santa Catalina Ridge Fault might be dipping landward and can be correlated with sharp bends on the velocity contours on profile B-F in Figs. 6.70–6.71. The San Clemente Fault might also be correlated with sharp bends in the velocity contours (profile G-K in Figs. 6.70–6.71). The East Santa Cruz Basin Fault might also be correlated with the velocity contours in all the profiles shown in Figs. 6.70–6.71.

6.2.9 Mojave Desert Region

The Mojave Desert region (Fig. 6.79) is bounded to the southwest by the San Andreas Fault, to the north by the Garlock fault and to the east by the southern extension of the Death Valley fault zone (e.g., Glazner et al. 2002). It contains the Eastern California Shear Zone (ECSZ), a set of northwest-trending dextral faults along the western edge of the Basin and Range Province. The ECSZ and the Walker Lane Belt north of the Garlock Fault accommodate up to 25 % of present-day relative motion between the Pacific and the North American plates (e.g., McClusky et al. 2001; Unruh et al. 2003).

In pre-Neogene time, the Salinian block (Sect. 6.2.6.1) and the San Gabriel terrane were immediately adjacent to the western Mojave block (e.g., Huffman 1972; Saleeby 2003). Collectively, the Mojave, Salinian, San Gabriel and the immediately adjacent southern Colorado River Desert basement complexes were termed the Mojave Desert—Salinian Batholith (MSB) in (Saleeby 2003) and the Southern California Batholith (SCB) in (Luffi et al. 2009). Different from the Sierra Nevada Batholith to the north and the Peninsular Ranges Batholith to the south, the batholithic root under most of SCB was tectonically removed by the shallow segment of the Farallon plate during the Laramide orogeny from latest Cretaceous to early Paleogene (e.g., Coney and Reynolds 1977; Dickinson and Snyder 1978; Bird 1988; Saleeby et al. 2003; Saleeby 2003). Thermobarometric constraints indicate that the depth of the tectonic truncation was at 0.8–1.0 GPa (about 30-km depth) conditions (e.g., Sharry 1981; Pickett and Saleeby 1993; Jacobson 1995). During mid- to late-Laramide time, regions affected by the flat slab experienced vigorous gravitational and extensional collapse facilitated by water-induced weakening and retrograde metamorphism in the overlying quartzofeldspathic batholithic crust (e.g., Wernicke et al. 1987; Malin et al. 1995; Saleeby 2003). The high water-flux is thought to come from the underplated schist protolith (Bebout and Barton 1989). The tectonic erosion of the batholithic root and the denudation of the overlying mid- to upper-crustal felsic batholith tapered the thick batholithic column to a much thinner section and uplifted the remaining batholithic column and the underplated schist to upper- and mid-crustal depths (e.g., Malin et al. 1996; Saleeby et al. 2003; Saleeby 2003). From about 70 Ma to about 20 Ma, the mid- to lower-crust and mantle lithosphere under western Mojave were reconstructed (Luffi et al. 2009). In Early Miocene between about 24 Ma and 18 Ma, the central Mojave experienced east-west extension and volcanism, which resulted in the central Mojave metamorphic core complex (e.g., Glazner et al. 1989; Walker et al. 1990; Martin et al. 1993; Fletcher et al. 1995; Glazner et al. 2002). From about 12 Ma to present, the Mojave region has experienced right-lateral shear along strike-slip faults, which may have accommodated up to 100 km of slip so far (e.g., McQuarrie and Wernicke 2005).

In Fig. 6.79, I show comparisons of the S-velocity at 2-km depth in CVM-S4.26, CVM-S4 and CVM-H11.9 with a geologic map of the Mojave Desert region. Low-velocity anomalies in CVM-S4.26 are well correlated with distribution of sediments and Tertiary to Quaternary volcanic rocks. Low-velocity anomalies east of the

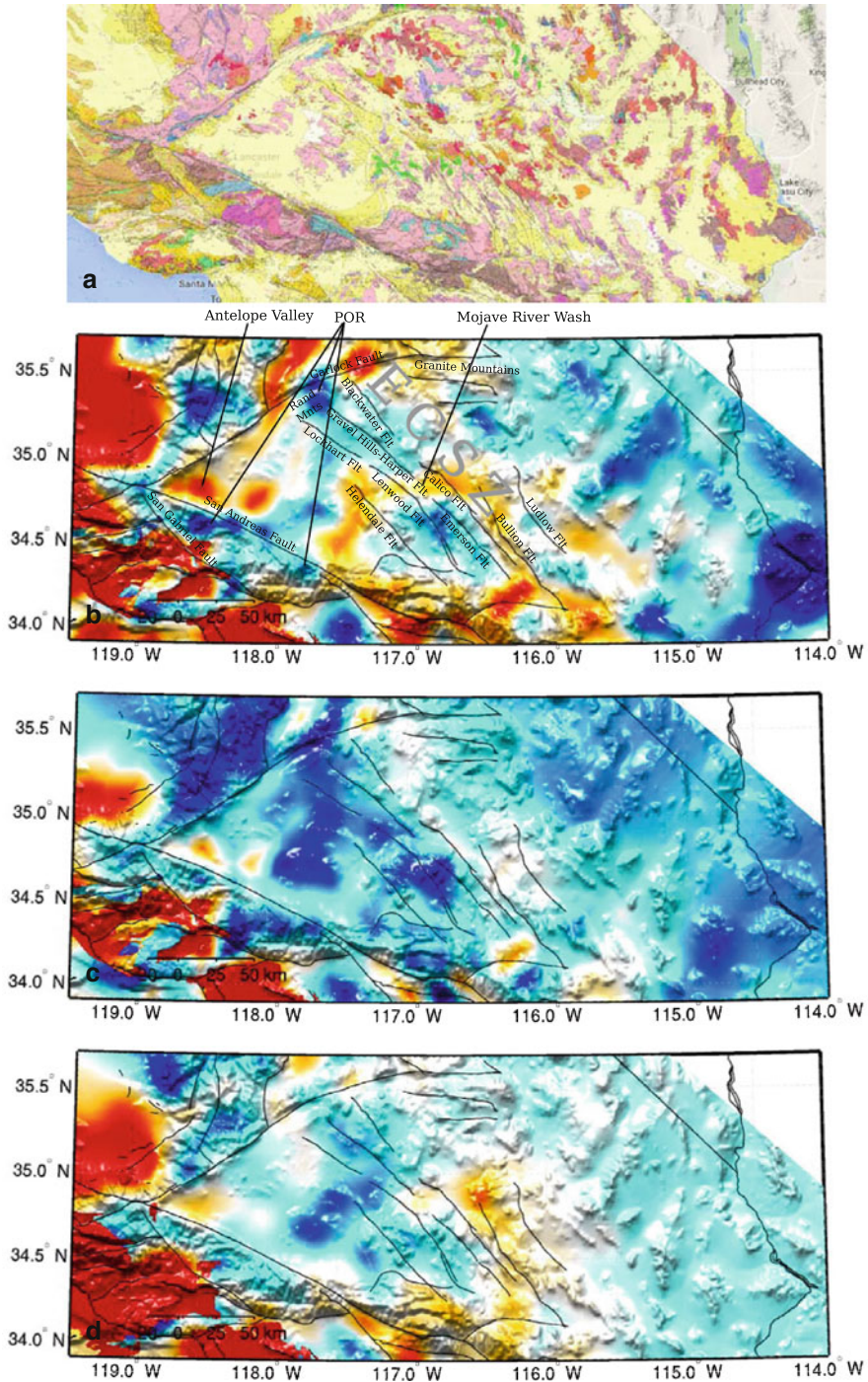


Fig. 6.79 S-velocity at 2-km depth in CVM-S4.26 (b), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (a) of the Mojave Desert region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1.

Camp-Rock—Homestead Valley—Emerson Fault Zone might be due to the higher heat flow caused by Quaternary volcanism (e.g., Bonner et al. 2003; Luffi et al. 2009). Outcrops of the Pelona-Orocopia-Rand (POR) schists were found around the northern and southwestern boundaries of the Mojave block. The location of the Orocopia Mountains was shown in Fig. 6.59. Other POR outcrop locations are indicated on Fig. 6.79. The high-velocity anomaly located at the Rand Mountains might be related to the Tehachapi anomaly (Sect. 6.2.4.2), if the left-lateral slip on the Garlock fault is restored. This possible correlation is also visible at larger depths (e.g., 12-km and 14-km map-views in Figs. 6.1–6.2). In Figs. 6.80–6.82, I show cross-sections of S-velocity, P-velocity and Poisson's ratio in CVM-S4.26 along 13 profiles crossing the Mojave Desert region. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.83–6.88.

6.2.9.1 Antelope Valley

The Antelope Valley is mainly filled with Quaternary and Tertiary volcanic and sedimentary rocks. Much of the 2-km thick sedimentary section may have been deposited during the Quaternary (e.g., Ponti 1985). The Quaternary basin sediments were mainly derived from the Transverse Ranges on the southwest and the Tehachapi Mountains to the northwest. Compared with CVM-S4, the velocities in CVM-S4.26 are reduced in the central part of the valley and the basin is extended more to the northeast along the Garlock Fault (Fig. 6.79). In addition to the cross-sections shown in Figs. 6.80–6.82, the cross-section along the LARSE-II profile, which also crosses the Antelope Valley, is shown in Fig. 8 in (Lee et al. 2014b). Along the LARSE-II profile, CVM-S4.26 shows that the thickness of the sediments reaches the maximum in the central valley and reduces on both the north and the south sides of the valley, which is in broad agreement with the 2D refraction tomography results of (Lutter et al. 2004) and also results in oil-test wells presented in (Fuis et al. 2003).

6.2.9.2 POR Schist

In CVM-S4.26, there is a pervasive low-velocity layer existing at 12–17 km depth. This low-velocity layer is observed under western Mojave, southernmost Sierra Nevada, the Salinian block and the San Gabriel terrane. It is most pronounced under western Mojave. The lateral extent of this low-velocity layer coincides with regions affected by the shallow Farallon slab during the Laramide orogeny (e.g., Saleeby 2003). The average P- and S-velocities in this layer are reduced by about 11 % compared with the average velocities in the layer above and the average Poisson's ratio increases about 8 % (Fig. 10 in Lee et al. 2014b). The average velocities and Poisson's ratio in this low-velocity layer are similar to those found in the Catalina Schist in the Inner Continental Borderland (Sect. 6.2.8.2). The depth of this mid-crustal

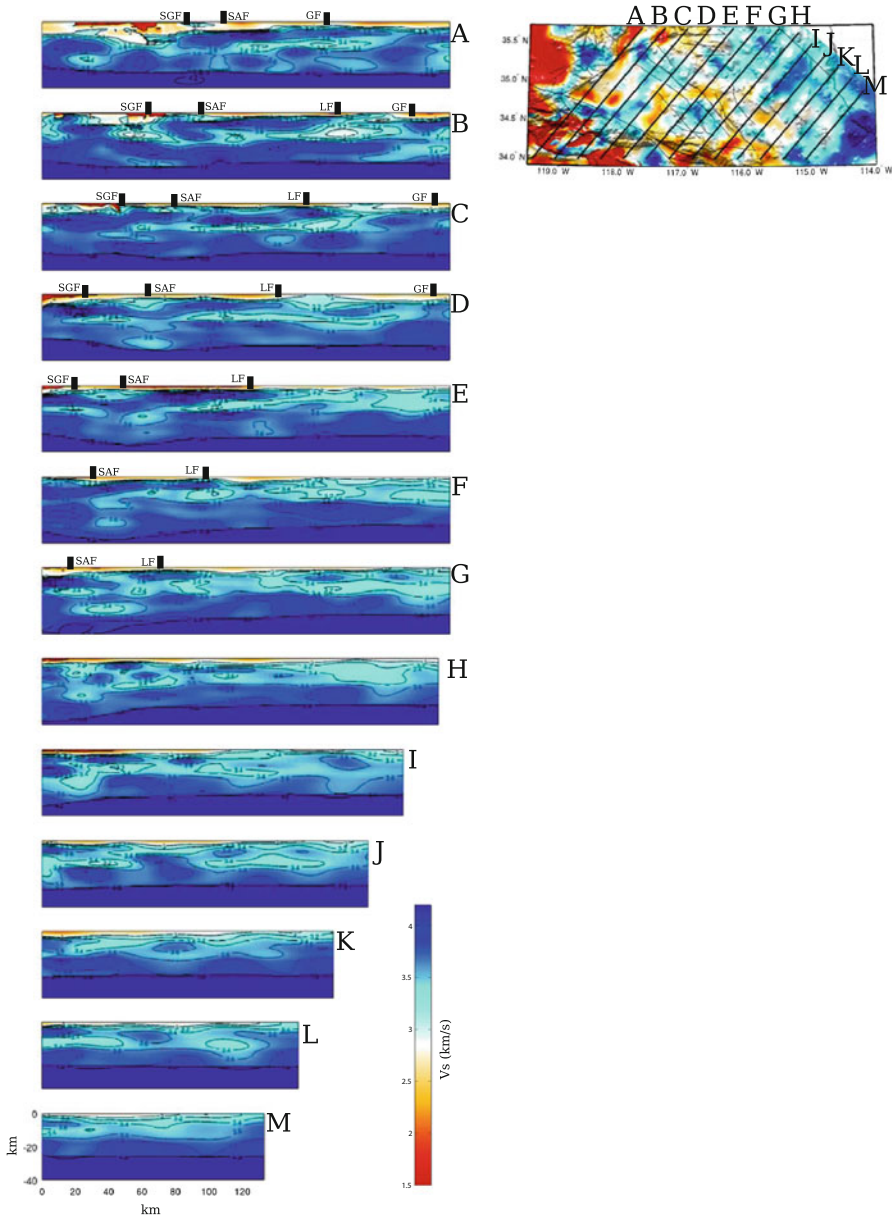


Fig. 6.80 Cross-sections of the S-velocity in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The map-view on the *upper-right corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-right corner*. Major faults are shown as black dots on the cross-sections. Garlock Fault (*GF*), Lockhart-Lenwood Fault (*LF*), San Andreas Fault (*SAF*), San Gabriel Fault (*SGF*).

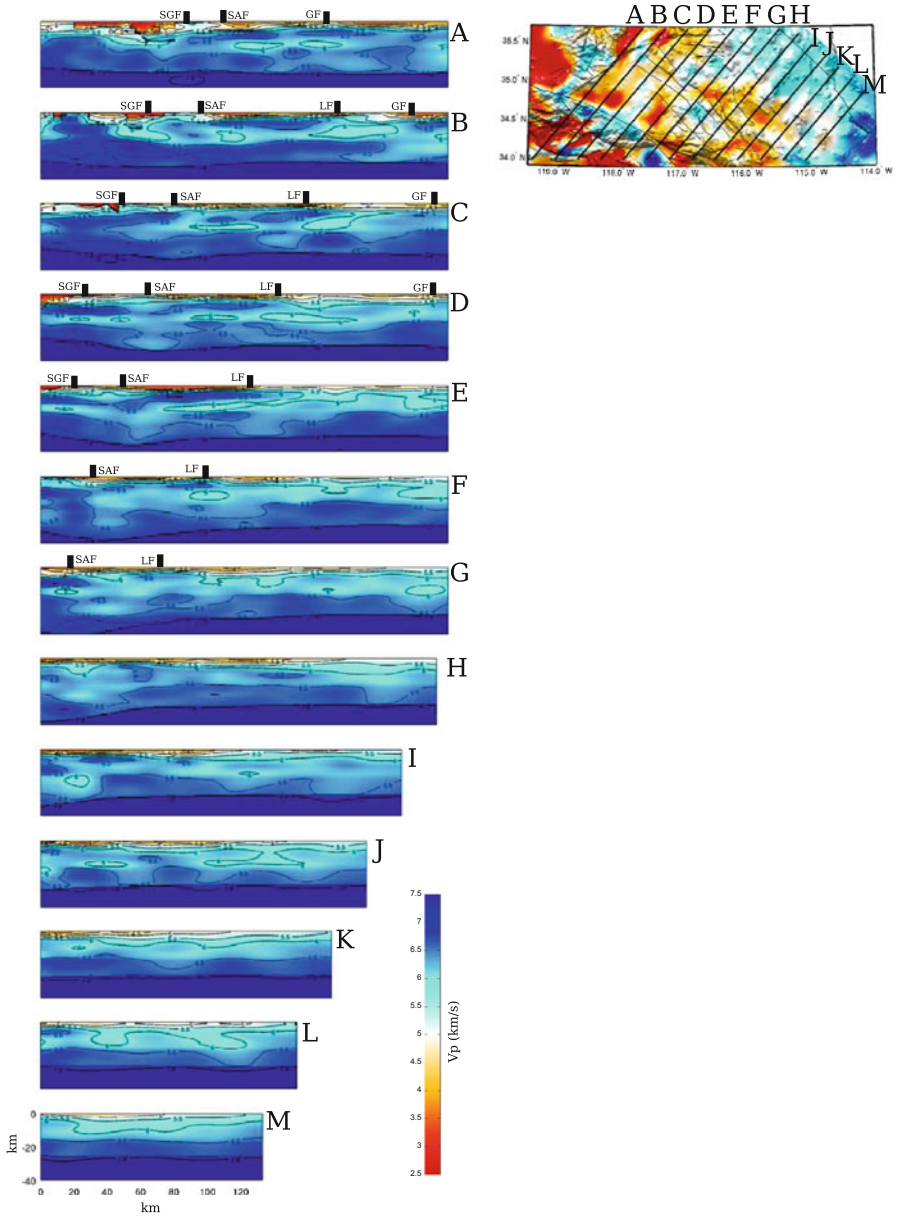


Fig. 6.81 Cross-sections of the P-velocity in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

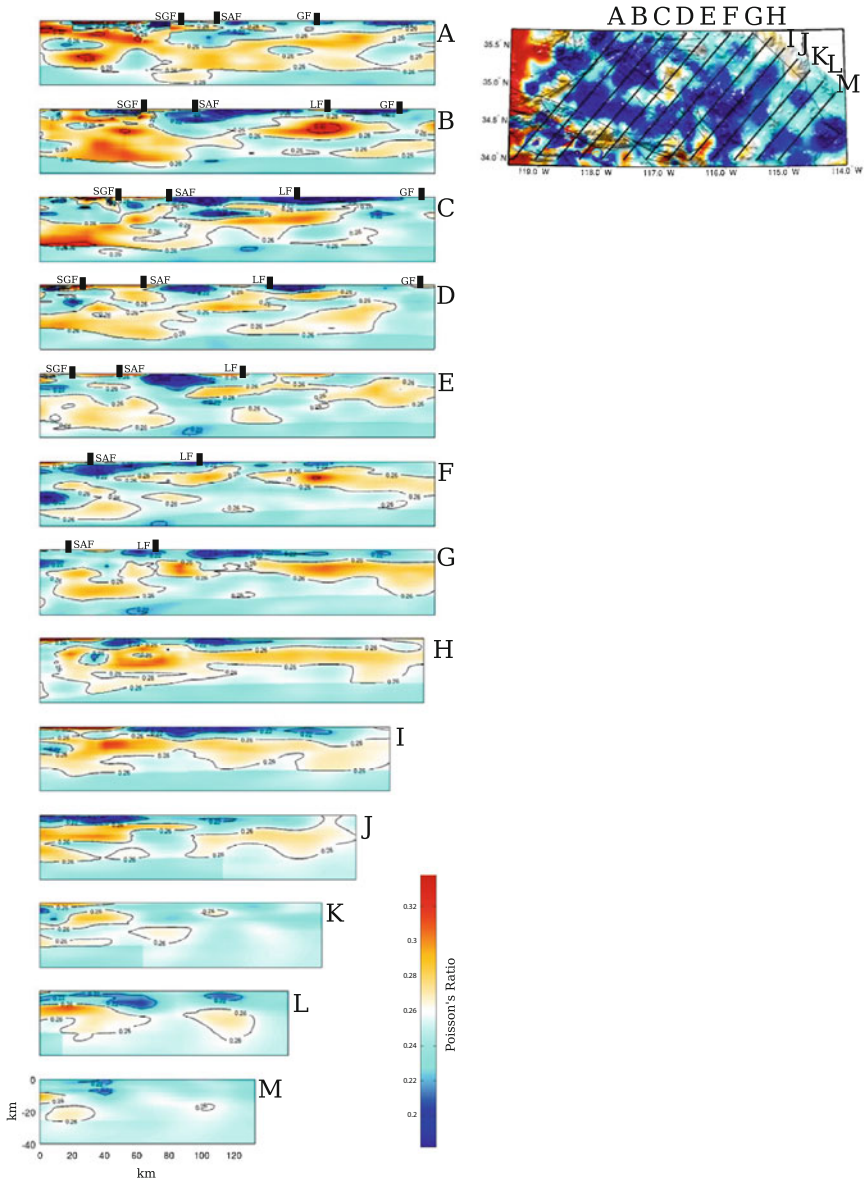


Fig. 6.82 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

low-velocity layer in the San Gabriel region coincides with strong mid-crustal reflectors mapped in e.g., (Ryberg and Fuis 1998). This mid-crustal low-velocity layer might be related to the underplated Franciscan accretionary wedge sediments metamorphosed into the POR schist.

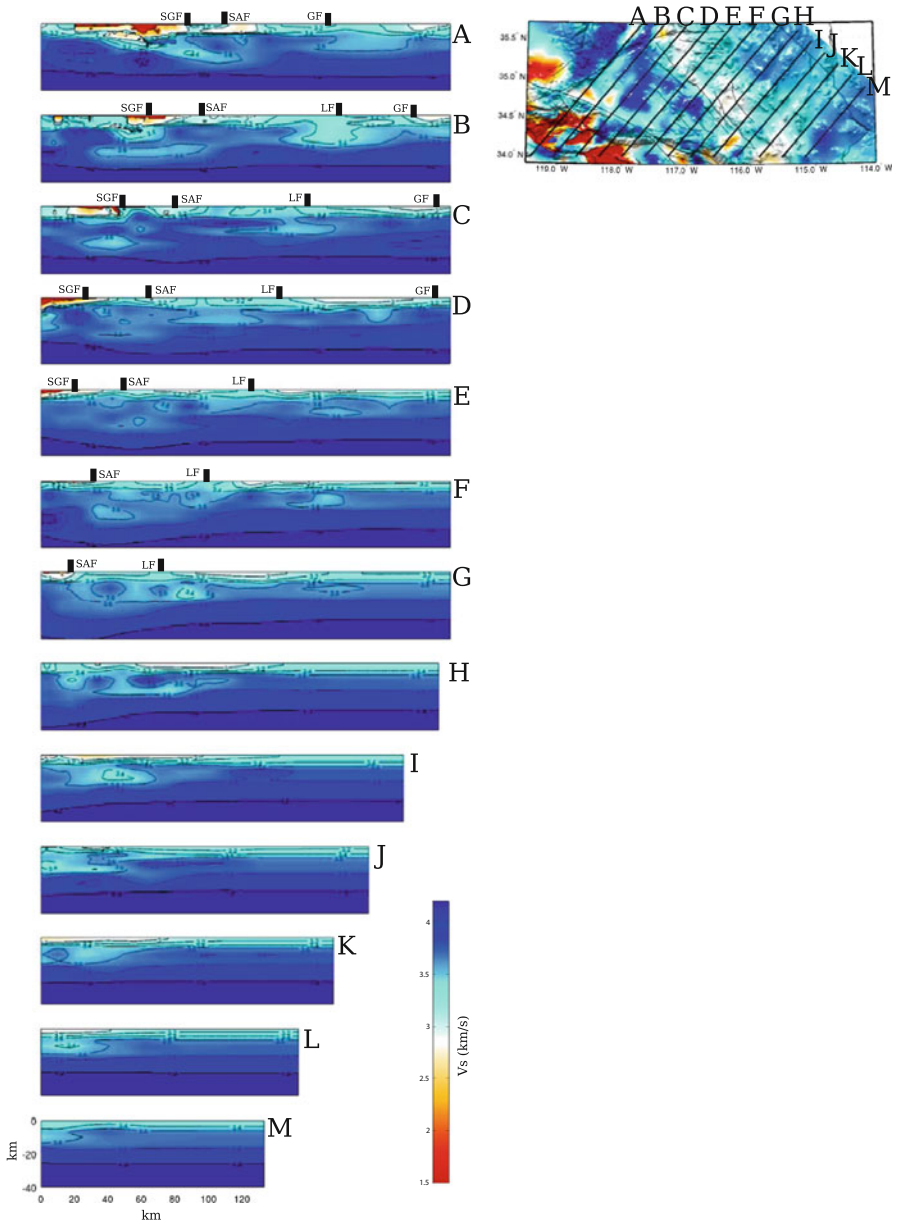


Fig. 6.83 Cross-sections of the S-velocity in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

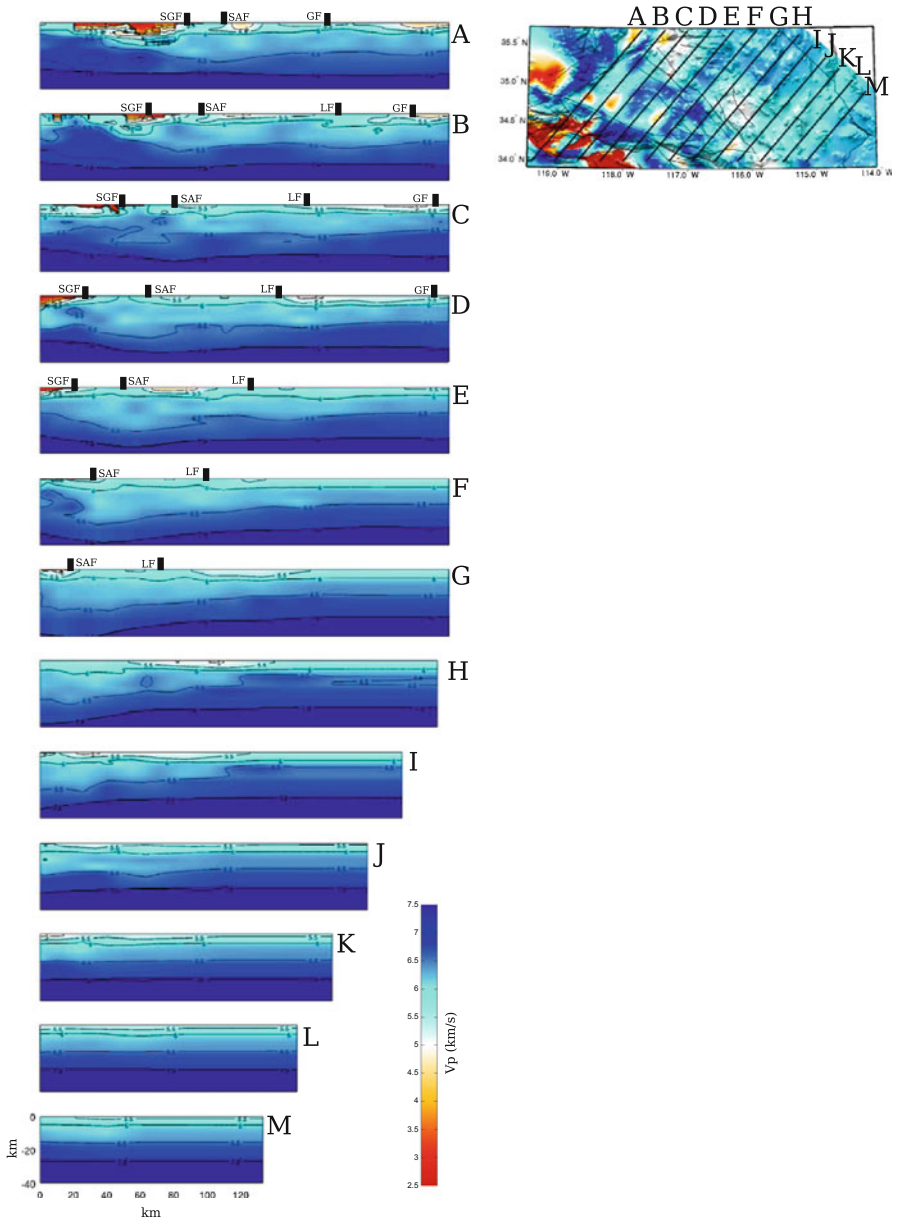


Fig. 6.84 Cross-sections of the P-velocity in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

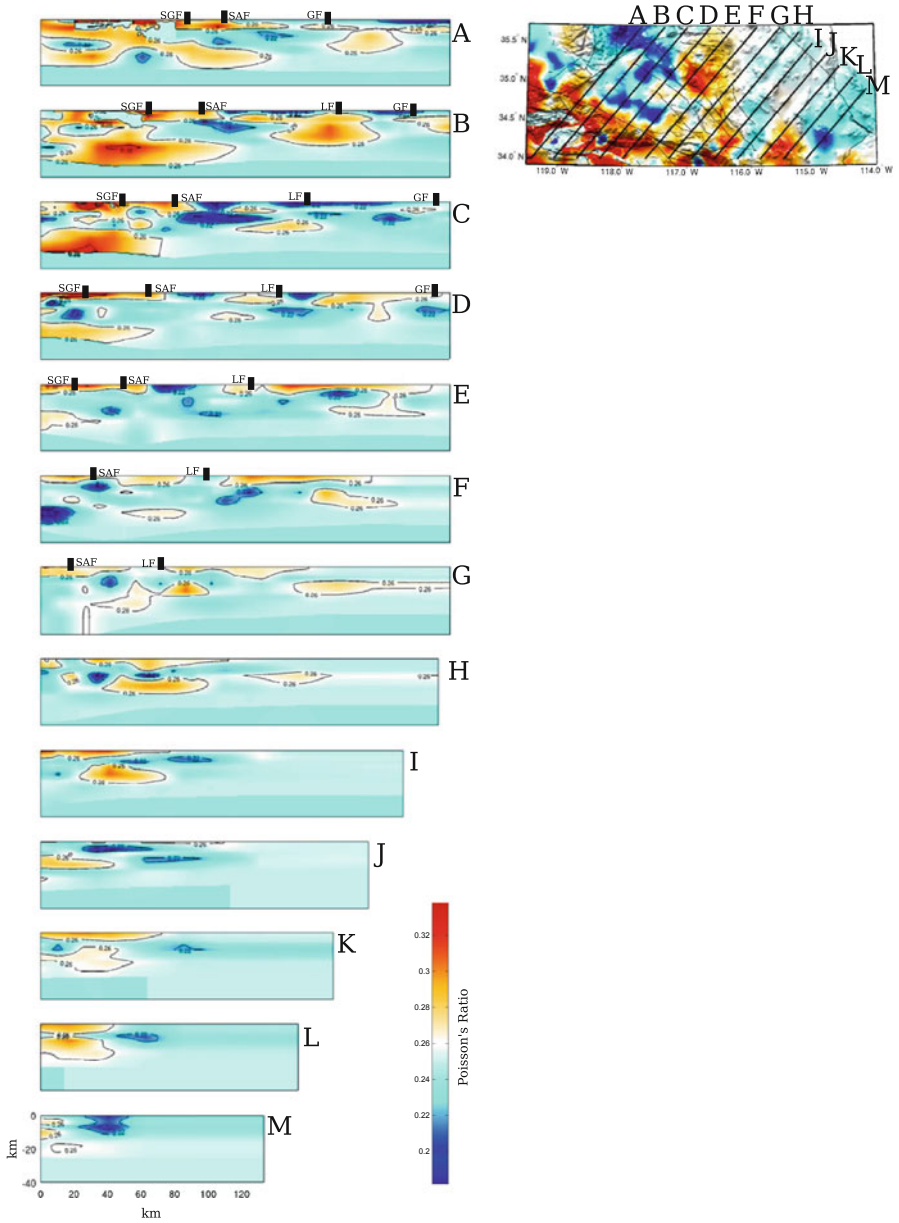


Fig. 6.85 Cross-sections of the Poisson's ratio in CVM-S4 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

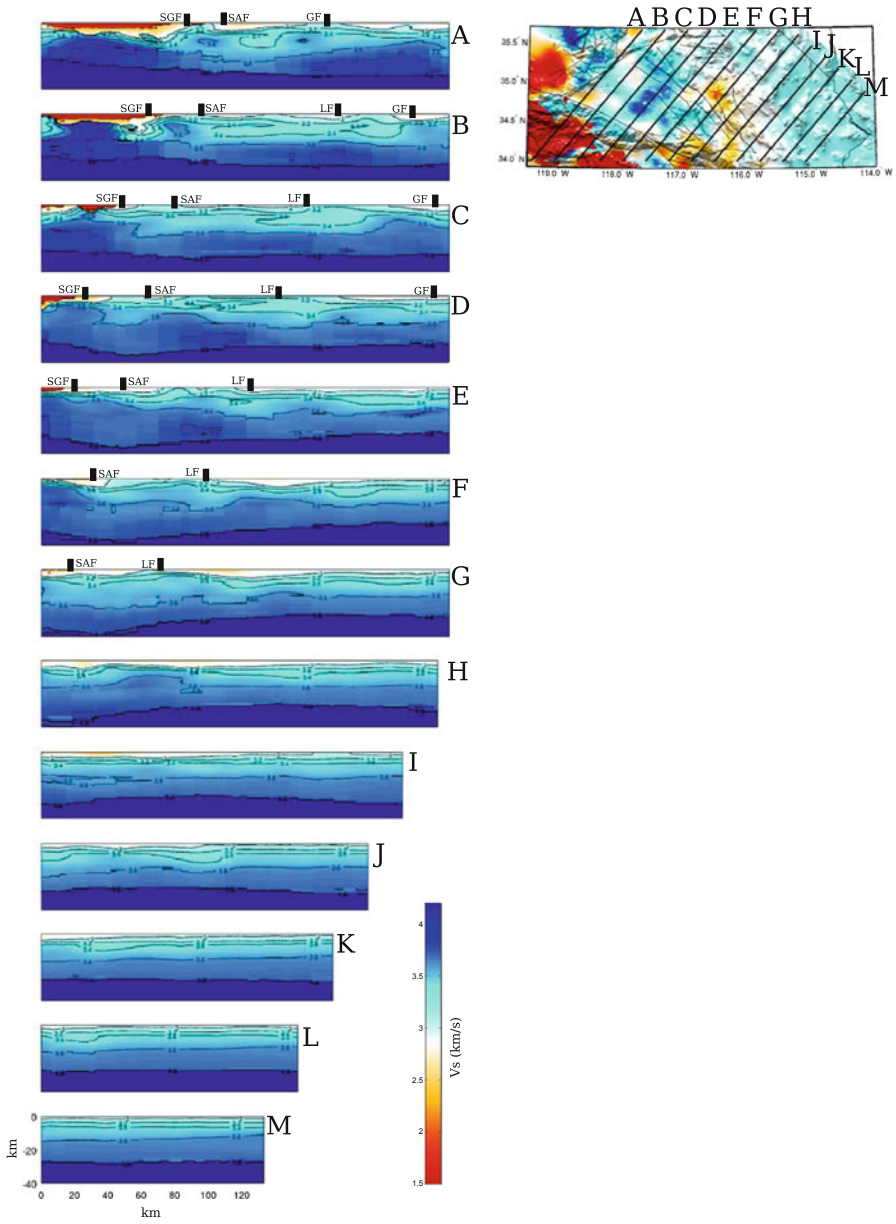


Fig. 6.86 Cross-sections of the S-velocity in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

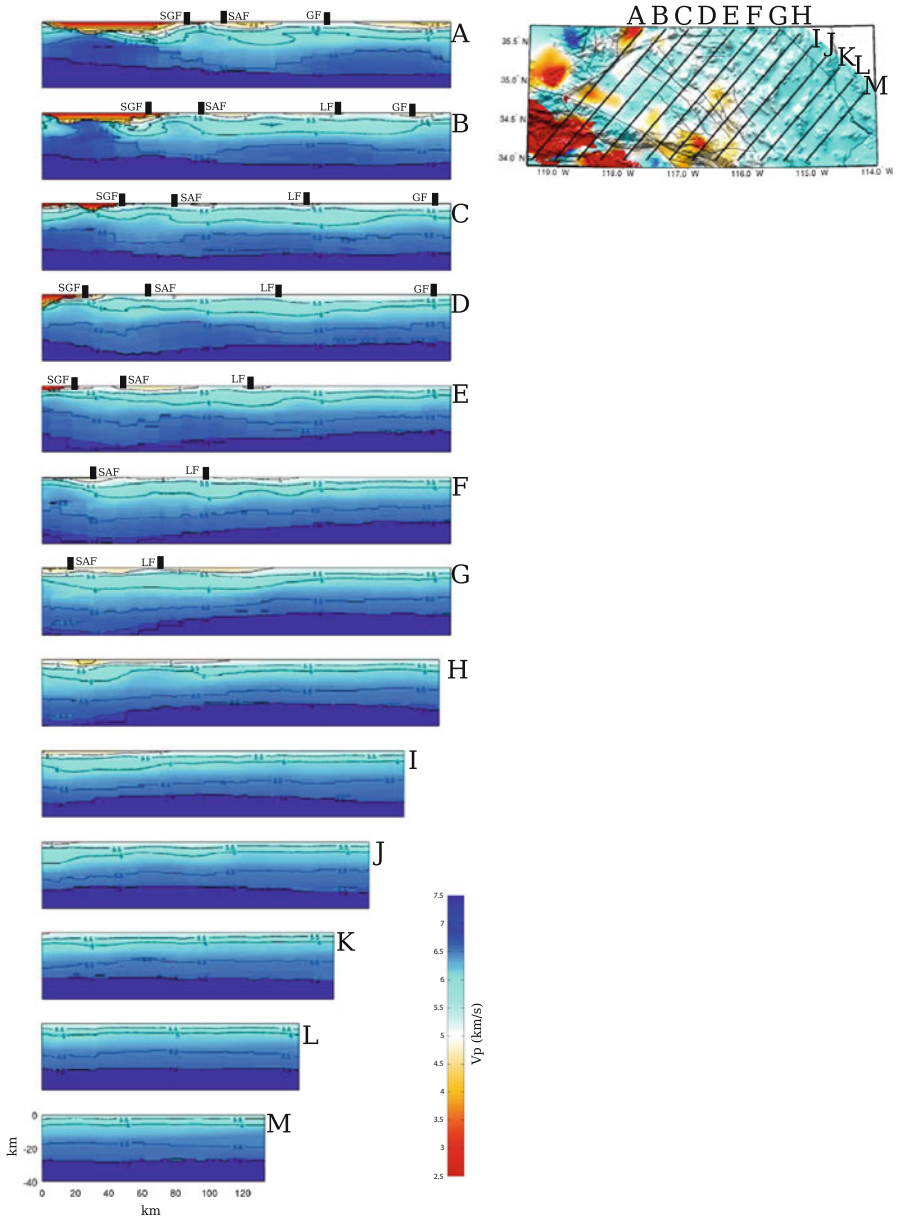


Fig. 6.87 Cross-sections of the P-velocity in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

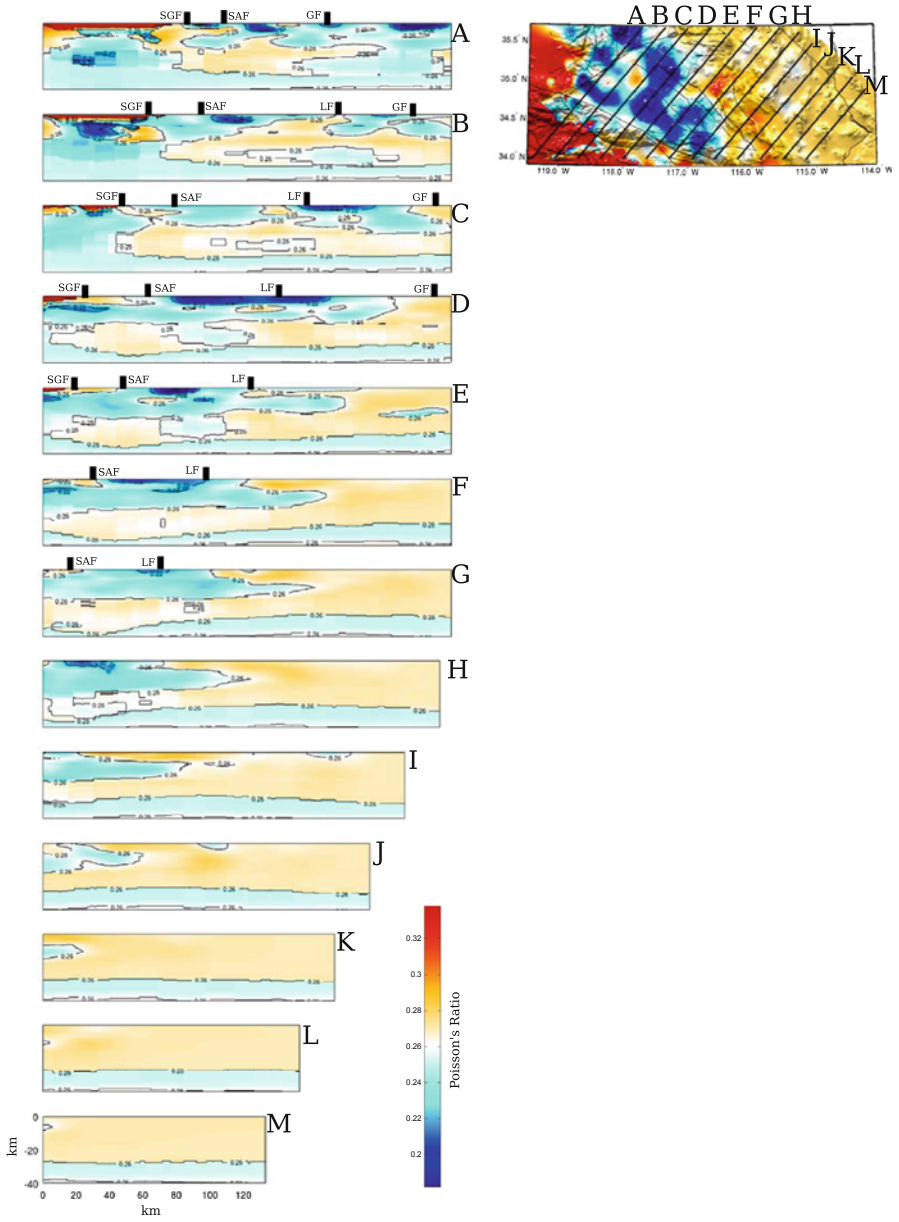


Fig. 6.88 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 13 profiles (A-M) crossing the Mojave Desert region. The format is identical to that in Fig. 6.80.

In CVM-S4.26, there exists a high-velocity layer immediately above the mid-crustal low-velocity layer within the 5–10 km depth range. The average P- and S-velocities in this high-velocity layer reach 6.7 km/s and 3.9 km/s, respectively, under southernmost Sierra Nevada and western Mojave. This high-velocity layer might have a significant amount of garnet granulite residues resulting from the partial melting of hydrous mafic to intermediate composition rocks at about 1.0 GPa (around 33-km depth) conditions (Saleeby et al. 2003). It was lying in the lower crust just above the seismic Moho and escaped the flat-slab tectonic truncation and was later uplifted to its current depth after the orogenic collapse. The low- to intermediate-velocity layer lying above the high-velocity layer may correspond to remaining felsic granitic rocks, which may have been highly attenuated during and after the orogenic collapse (Yan et al. 2005).

Below the mid-crustal low-velocity layer in western Mojave, the average P- and S-velocities in CVM-S4.26 are consistent with substantial portions of metabasalt. Laboratory measurements of P-velocity for the Pelona Schist range from 5.8 km/s for metaclastic rocks to 6.4 km/s for metabasaltic rocks (Pellerin and Christensen 1998) and there are indications that the proportion of the metabasalt increases with depth (e.g., Yan et al. 2005). Geochemical data from xenoliths in the Mojave region show evidences of MORB-like material lying under western Mojave (Luffi et al. 2009). It might be possible that substantial portions of oceanic crust from the Farallon plate were scrapped off and remain under western Mojave when the Farallon plate encountered the North American Craton during the Laramide shallow subduction (e.g., Luffi et al. 2009).

6.2.10 Peninsular Ranges—Salton Trough Region

The Peninsular Ranges Batholith (PRB) is part of the Mesozoic batholith belt emplaced along the western Cordilleran margin of North America. The entire PRB extends from south of the western Transverse Ranges to the southern end of the Baja California peninsula. Only the northern segment of the PRB is covered by our tomography box for CVM-S4.26 (Fig. 6.89). The PRB is highly heterogeneous. The plutons of the PRB have composition ranging from gabbro to granite with the predominant rock type of tonalite (e.g., Larsen 1948; Silver and Chappell 1988). Geochemical analysis shows that the PRB can be divided into compositionally distinct western and eastern sections along geochemical, mineralogic, isotopic and geophysical potential-field discontinuities (e.g., Gastil 1975; Baird and Miesch 1984; Todd and Shaw 1985; Langenheim et al. 2014). The western section is older and more mafic and has a wide range of compositions commonly including gabbros. The younger eastern section is more felsic and lacks gabbros. Several hypotheses have been proposed to explain the lateral variations within the PRB (e.g., Schmidt et al. 2002). One possibility is that the western and eastern sections were emplaced at two separate subduction zones and then accreted together in Cretaceous (e.g., Gastil et al. 1978). An alternative explanation is that the Mesozoic batholith

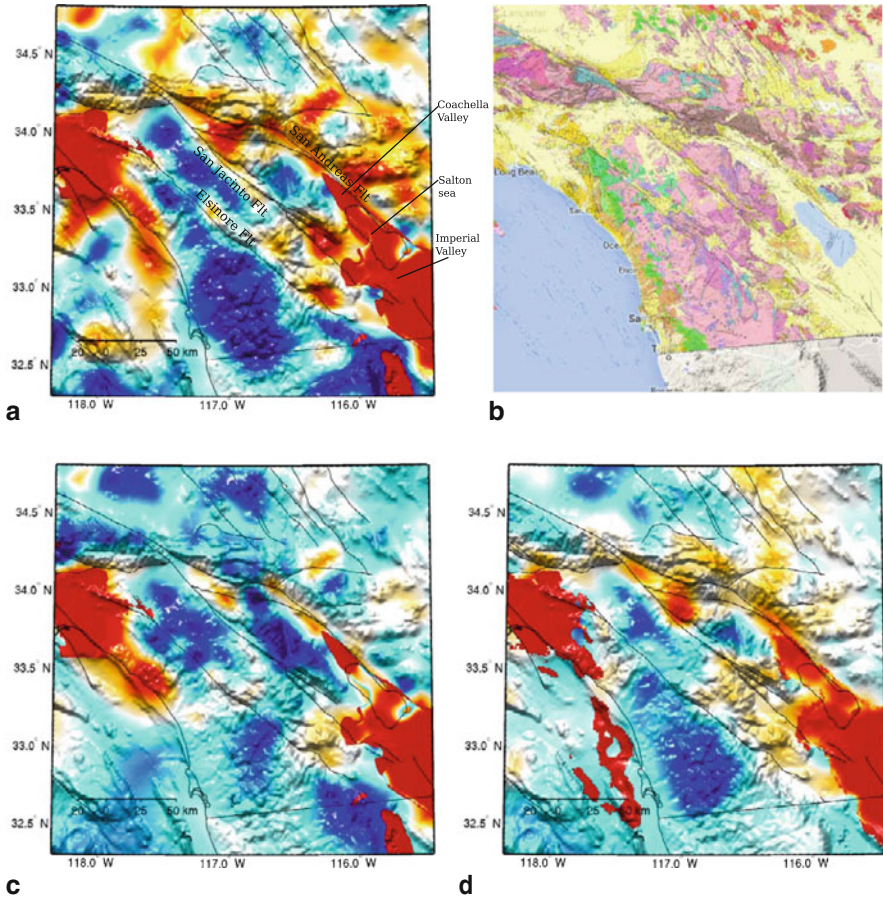


Fig. 6.89 S-velocity at 2-km depth in CVM-S4.26 (a), CVM-S4 (c) and CVM-H11.9 (d) compared with the geologic map (b) of the Peninsular Ranges–Salton Trough region. The color-scale is identical to that in the 2-km map-view in Fig. 6.1.

was emplaced across a pre-existing ocean–continental lithosphere transition (e.g., Walawender et al. 1991).

In Fig. 6.89, I show S-velocity at 2-km depth in the three CVMs compared with a geologic map for the Peninsular Ranges and Salton Trough region. The correlation between CVM-S4.26 and CVM-H11.9 is quite high for this region. CVM-S4.26 shows stronger lateral heterogeneities and CVM-H11.9 is relatively smoother. Cross-sections of S-velocity, P-velocity and Poisson’s ratio along 7 profiles crossing the PRB and the Salton Trough in CVM-S4.26 are shown in Figs. 6.90–6.92. The same cross-sections for CVM-S4 and CVM-H11.9 are shown in Figs. 6.93–6.98. CVM-S4.26 shows more heterogeneities in PRB. In the eastern section, the Moho becomes shallower towards east, which is consistent with results in (Ichinose

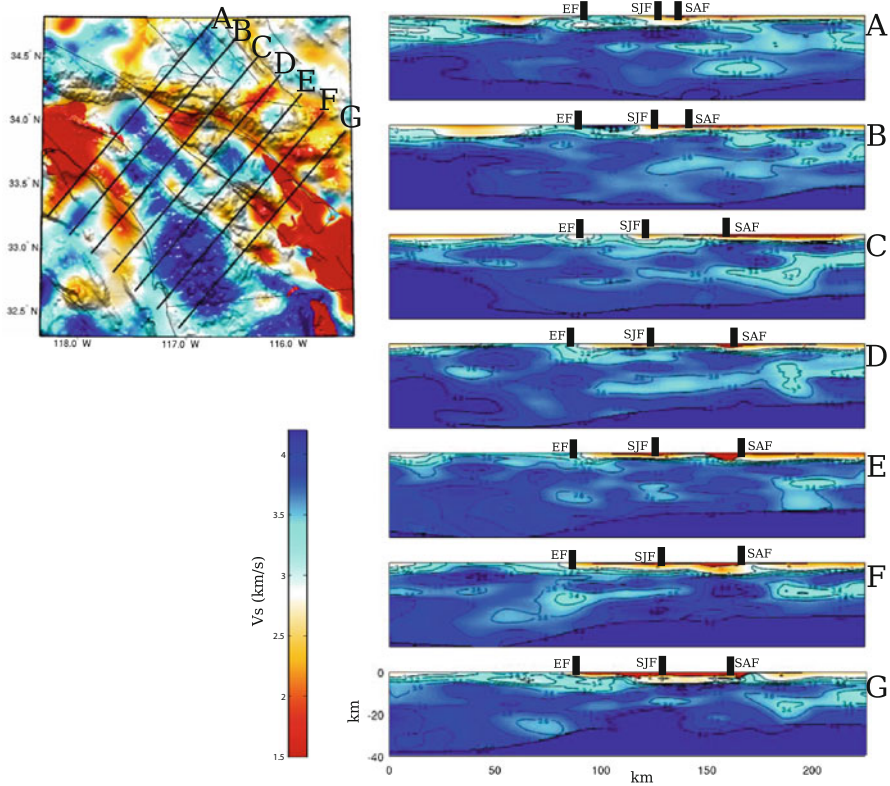


Fig. 6.90 Cross-sections of the S-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The map-view on the *upper-left corner* shows the S-velocity at 2-km depth in the same color-scale as the 2-km map-view in Fig. 6.1. The color-scale for the cross-sections is shown on the *lower-left corner*. Major faults are shown as black dots on the cross-sections. Elsinore Fault (EF); San Jacinto Fault (SJF); San Andreas Fault (SAF).

et al. 1996). The laterally averaged velocity-depth profiles for the western and eastern sections were shown in Fig. 4 in (Lee et al. 2014b) and the western section has higher average velocities, consistent with a more mafic average composition.

The actively rifting Salton Trough was linked with the Gulf of California during late Miocene and early Pliocene and was later isolated from the gulf by sediments deposited by the Colorado River (Metzger 1968). The Salton Trough includes, from north to south, the Coachella Valley, the Salton sea, the Imperial Valley and the Mexicali Valley in Mexico (Fig. 6.89). Cross-sections of the three CVMs along the active-source Salton Seismic-Imaging Project (SSIP) lines 1, 3, 4, 5 and 6 were shown in Fig. 9 in (Lee et al. 2014b). The basin structures in CVM-S4.26 are in broad agreement with the P-velocity model obtained from 2D refraction tomography along those lines. Compared with our starting model CVM-S4, the velocities in

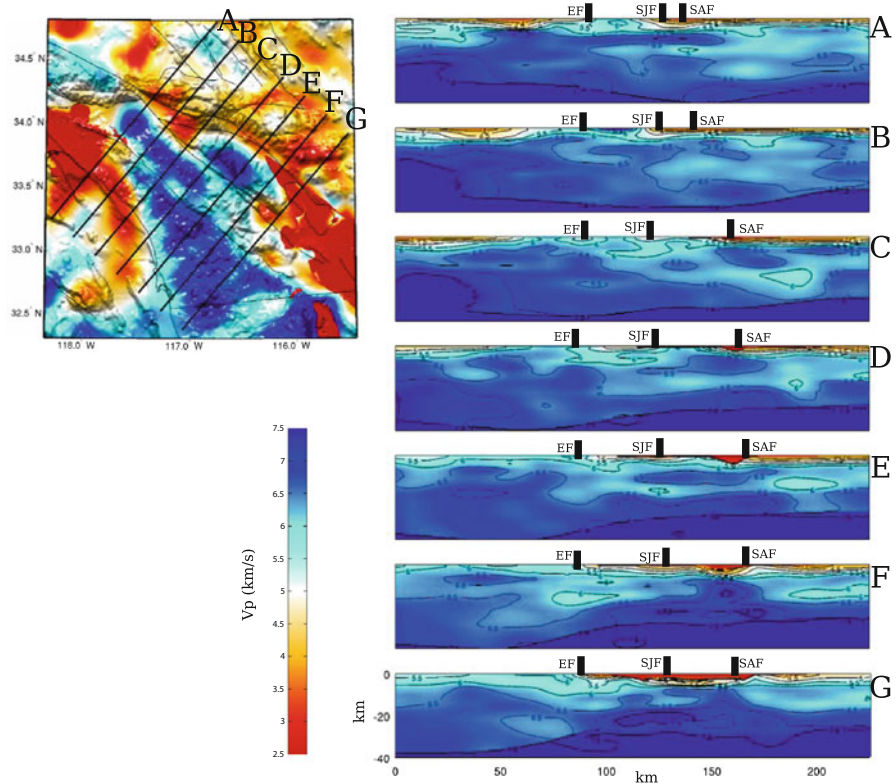


Fig. 6.91 Cross-sections of the P-velocity in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

the Salton sea and the Imperial Valley area were reduced by up to 20% above 5-km depth and increased by 5–10% below 5-km depth in CVM-S4.26. The Salton Trough region has relatively high thermal gradient, which can cause metamorphism in deep sediments. The relatively high velocities below about 5-km depth may come from metamorphosed sediments (Fuis and Mooney 1990). The velocities above 5-km depth in CVM-S4.26 are more consistent with the basin model of (Lovely et al. 2006), in which an extensive set of sonic logs from petroleum wells in the Imperial Valley and results from a seismic refraction study (Fuis et al. 1984) were used to constrain the basin structure.

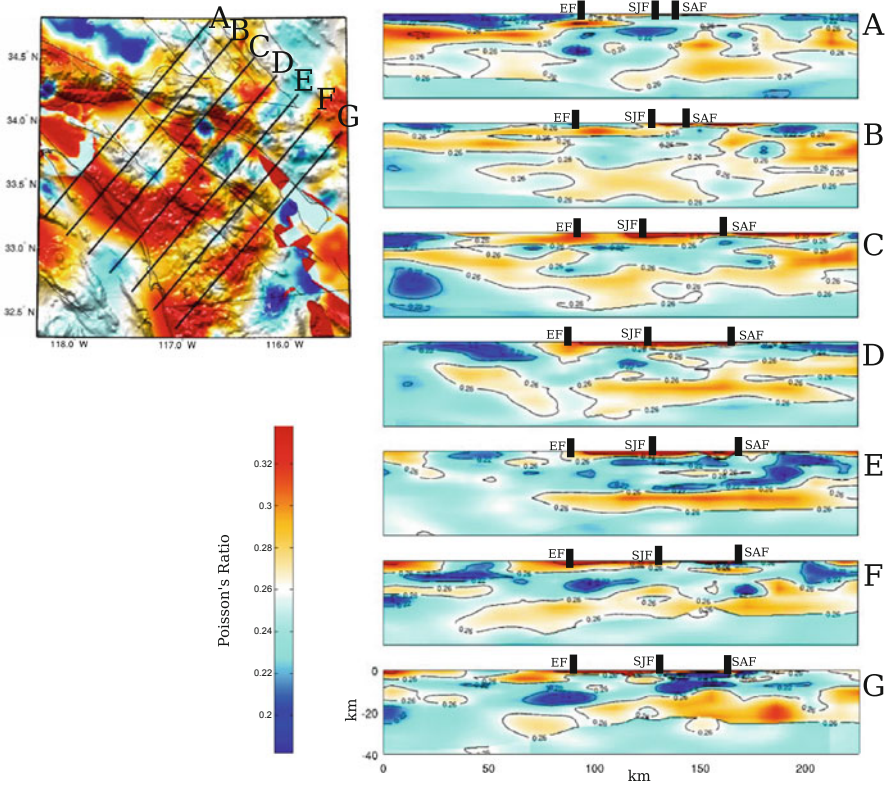


Fig. 6.92 Cross-sections of the Poisson's ratio in CVM-S4.26 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

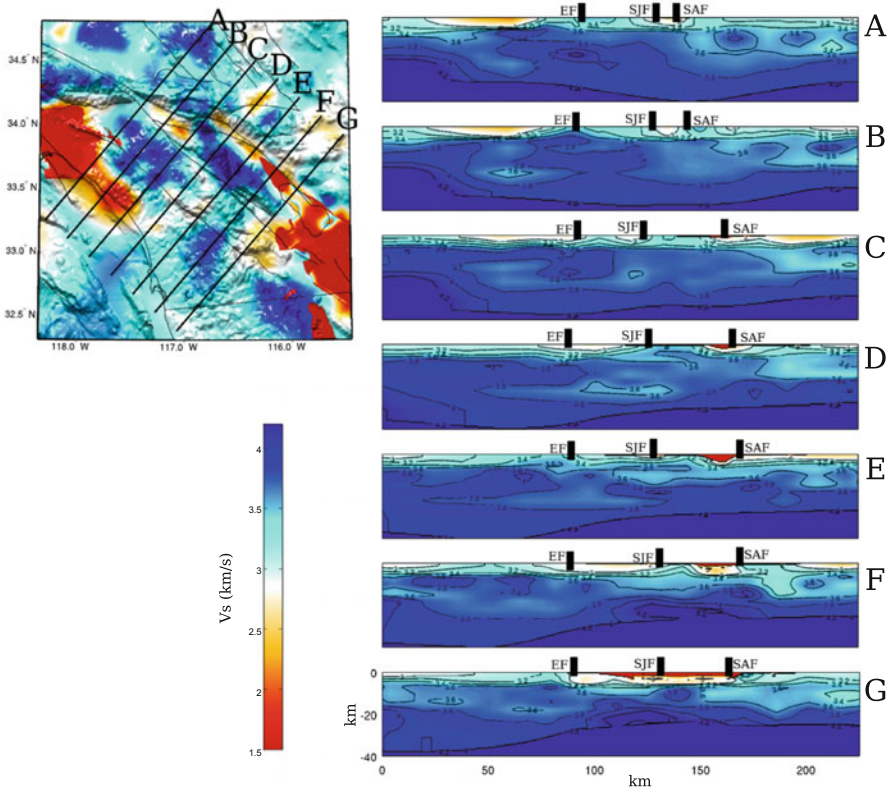


Fig. 6.93 Cross-sections of the S-velocity in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

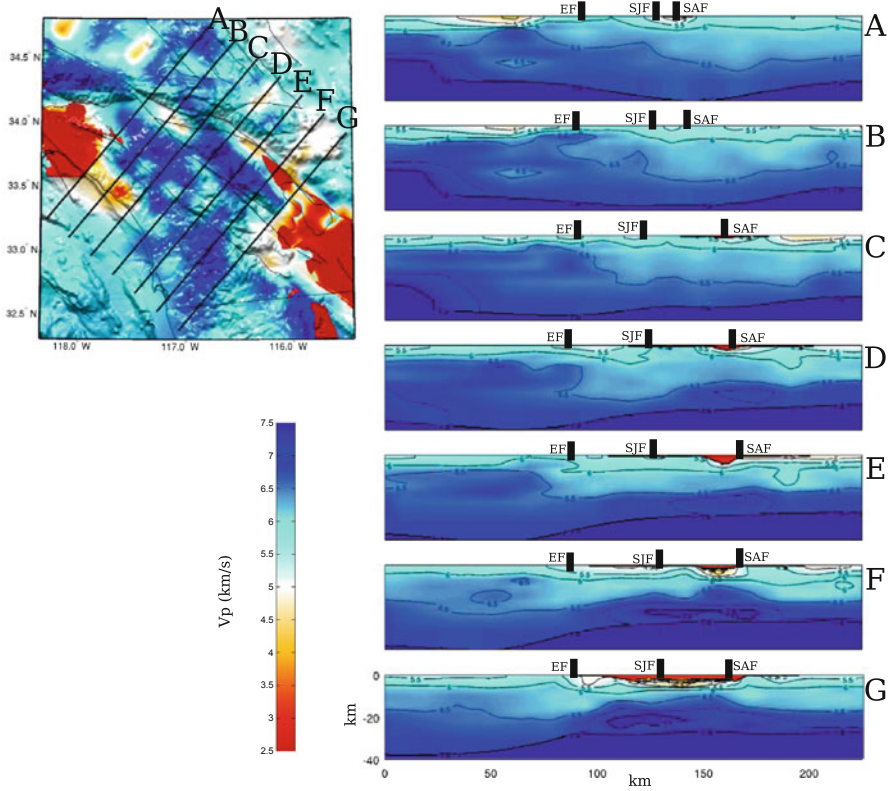


Fig. 6.94 Cross-sections of the P-velocity in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

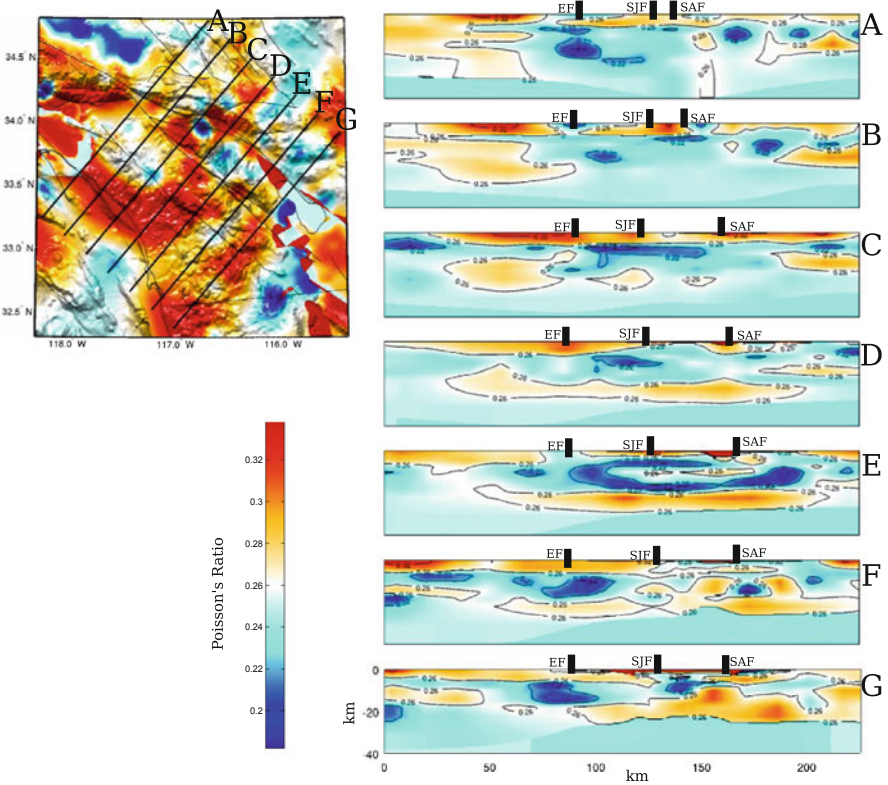


Fig. 6.95 Cross-sections of the Poisson's ratio in CVM-S4 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

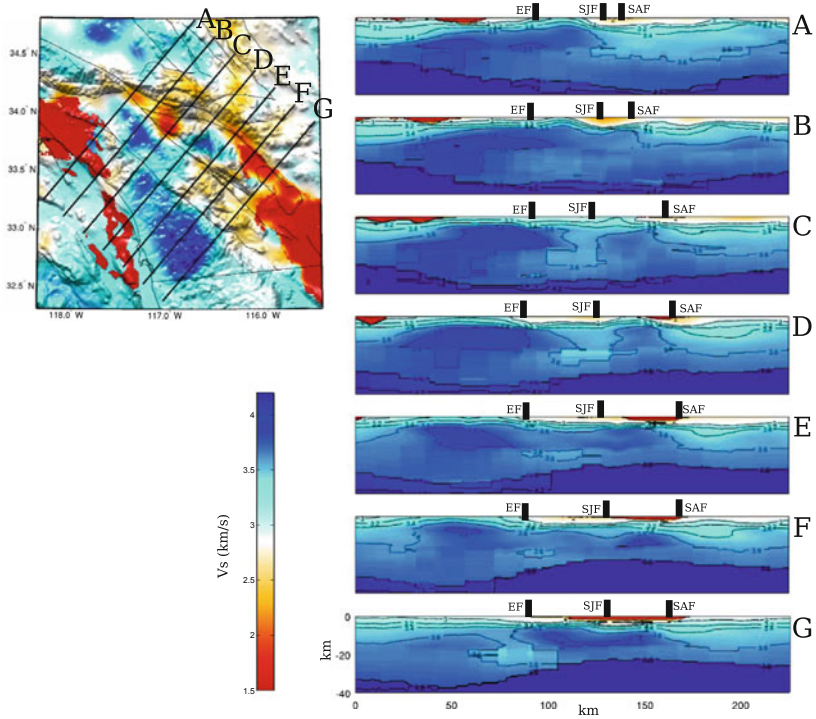


Fig. 6.96 Cross-sections of the S-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

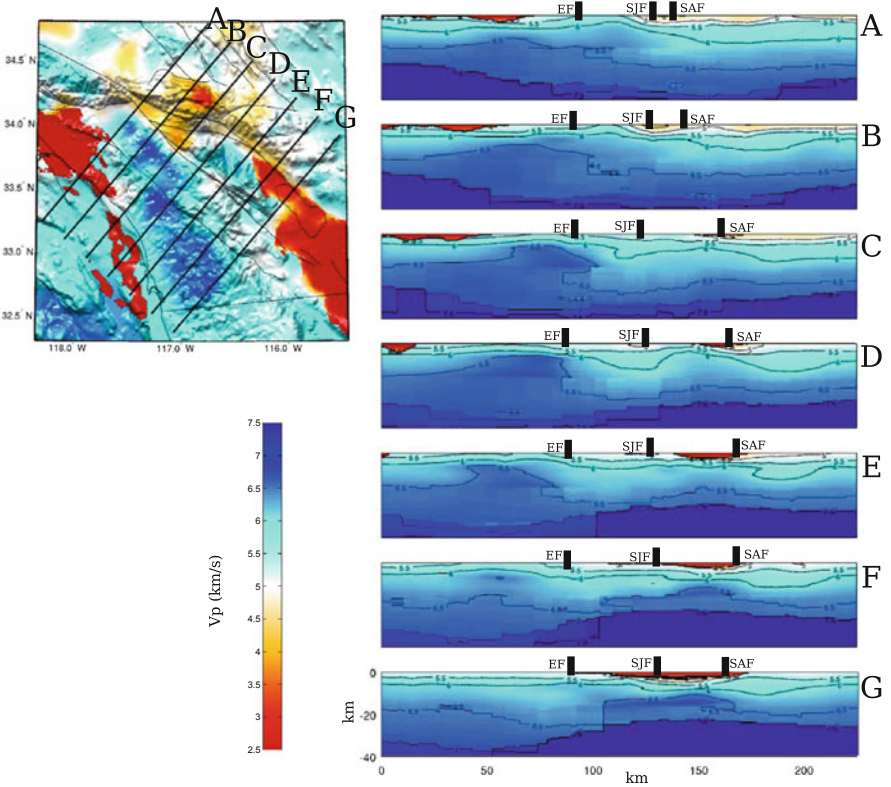


Fig. 6.97 Cross-sections of the P-velocity in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

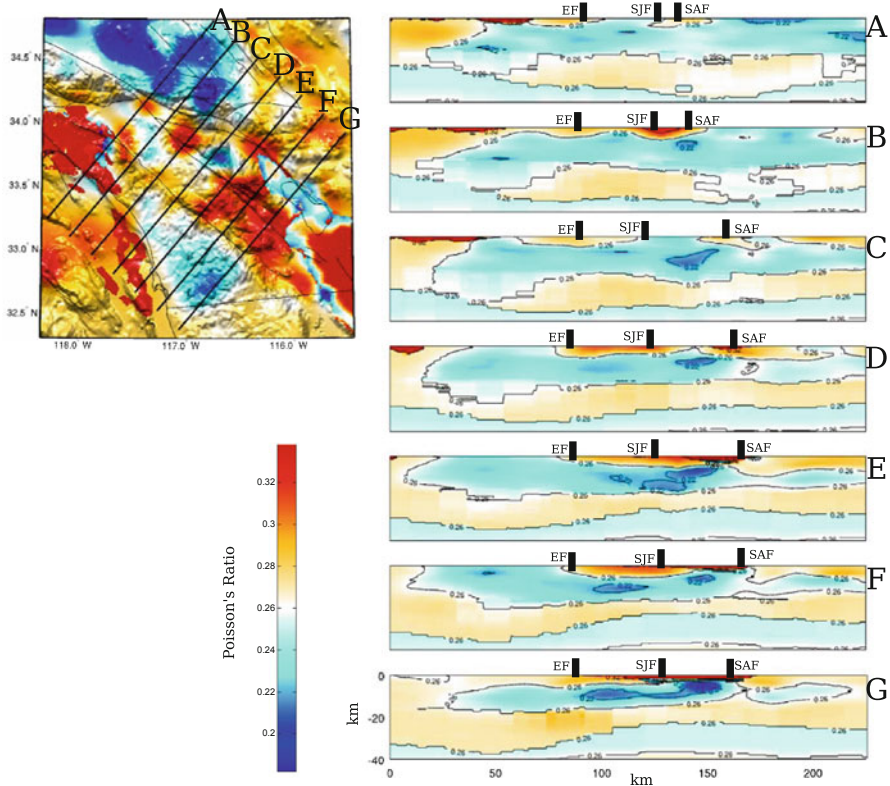


Fig. 6.98 Cross-sections of the Poisson's ratio in CVM-H11.9 along the 7 profiles (A-G) crossing the Peninsular Ranges and Salton Trough region. The format is identical to that in Fig. 6.90.

References

- Anderson, M., Matti, J., & Jachens, R. (2004). Structural model of the San Bernardino basin, California, from analysis of gravity, aeromagnetic, and seismicity data. *Journal of Geophysical Research*, 109(B4).
- Atwater, T., & Stock, J. (1998). Pacific-North America plate tectonics of the Neogene southwestern United States: An update. *International Geology Review*, 40(5), 375–402.
- Bailey, R. (1976). On the mechanisms of postsubsidence central doming and volcanism in resurgent cauldrons. *Geological Society of America Abstract with Programs*, 8, 567.
- Bailey, E. H., et al. (1941). *Mineralogy, petrology and geology of Santa Catalina Island, California*. California: Stanford University.
- Bailey, E. H., Blake, M., Jr., & Jones, D. L. (1970). On-land Mesozoic oceanic crust in California coast ranges. *US Geological Survey Professional Paper*, 700, C70–C81.
- Baird, A., & Miesch, A. (1984). Batholithic rocks of southern California: A model for the petrochemical nature of their source materials. Technical report, Geological Survey, Denver, CO (USA).

- Barth, A. P., Wooden, J. L., Grove, M., Jacobson, C. E., & Pedrick, J. N. (2003). U-Pb zircon geochronology of rocks in the Salinas Valley region of California: A reevaluation of the crustal structure and origin of the Salinian block. *Geology*, 31(6), 517–520.
- Battaglia, M., Roberts, C., & Segall, P. (1999). Magma intrusion beneath Long Valley caldera confirmed by temporal changes in gravity. *Science*, 285(5436), 2119–2122.
- Battaglia, M., Segall, P., Murray, J., Cervelli, P., & Langbein, J. (2003a). The mechanics of unrest at Long Valley caldera, California: 1. Modeling the geometry of the source using GPS, leveling and two-color EDM data. *Journal of Volcanology and Geothermal Research*, 127(3), 195–217.
- Battaglia, M., Segall, P., & Roberts, C. (2003b). The mechanics of unrest at Long Valley caldera, California. 2. Constraining the nature of the source using geodetic and micro-gravity data. *Journal of Volcanology and Geothermal Research*, 127(3), 219–245.
- Bebout, G. E., & Barton, M. D. (1989). Fluid flow and metasomatism in a subduction zone hydrothermal system: Catalina Schist terrane, California. *Geology*, 17(11), 976–980.
- Bird, P. (1988). Formation of the Rocky Mountains, western United States: A continuum computer model. *Science*, 239(4847), 1501–1507.
- Bird, P., & Rosenstock, R. W. (1984). Kinematics of present crust and mantle flow in southern California. *Geological Society of America Bulletin*, 95(8), 946–957.
- Blake, M. Jr., Campbell, R., Dibblee, T. Jr., Howell, D., Nilsen, T., Normark, W., Vedder, J., & Silver, E. (1978). Neogene basin formation in relation to plate-tectonic evolution of San Andreas fault system, California. *American Association of Petroleum Geologists Bulletin*, 62(3), 344–372.
- Black, R. A., Walker, J. D., & Baker, G. S. (2002). Three-dimensional gravity modeling and crustal-density variations, Panamint Range to the eastern Sierra Nevada, southeastern California. *Geological Society of America Memoirs*, 195, 229–241.
- Bohannon, R. G., & Geist, E. (1998). Upper crustal structure and Neogene tectonic development of the California continental borderland. *Geological Society of America Bulletin*, 110(6), 779–800.
- Bonner, J. L., Blackwell, D. D., & Herrin, E. T. (2003). Thermal constraints on earthquake depths in California. *Bulletin of the Seismological Society of America*, 93(6), 2333–2354.
- Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin Of The Seismological Society Of America*, 95(6), 2081–2092.
- Burchfiel, B., & Stewart, J. (1966). "Pull-apart" origin of the central segment of Death Valley, California. *Geological Society of America Bulletin*, 77(4), 439–442.
- Burchfiel, B., Lipman, P., & Zoback, M. (1992). *The Cordilleran orogen: Conterminous U.S. Geology of North America*. Boulder: Geological Society of America.
- Carter, J. N., Luyendyk, B. P., & Terres, R. R. (1987). Neogene clockwise tectonic rotation of the eastern Transverse Ranges, California, suggested by paleomagnetic vectors. *Geological Society of America Bulletin*, 98(2), 199–206.
- Cheadle, M., Czuchra, B., Byrne, T., Ando, C. J., Oliver, J., Brown, L., Kaufman, S., Malin, P., & Phinney, R. (1986). The deep crustal structure of the Mojave Desert, California, from COCORP seismic reflection data. *Tectonics*, 5(2), 293–320.
- Chen, J., & Tilton, G. (1991). Applications of lead and strontium isotopic relationships to the petrogenesis of granitoid rocks, central Sierra Nevada batholith, California. *Geological Society of America Bulletin*, 103(4), 439–447.
- Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, 97(4), 1094–1120.
- Christensen, N. I. (1996). Poisson's ratio and crustal seismology. *Journal of Geophysical Research*, 101(B2), 3139–3156.
- Christensen, N. I., & Mooney, W. D. (1995). Seismic velocity structure and composition of the continental crust: A global view. *Journal of Geophysical Research*, 100(B6), 9761–9788.
- Clemens-Knott, D. (1992). Geologic and isotopic investigations of the Early Cretaceous Sierra Nevada batholith, Tulare Co, CA and the Ivrea zone, NW Italian Alps: Examples of interaction between mantle derived magmas and continental crust. *California Institute of Technology*.

- Clemens-Knott, D., Saleeby, J., & Taylor Jr, H. (1991). O, Sr and Nd constraints on the evolution of the Early Cretaceous Sierra Nevada batholith. *Geological Society of America Abstract with Programs*, 23, 386.
- Clemens-Knott, D., & Saleeby, J. B. (1999). Impinging ring dike complexes in the Sierra Nevada batholith, California: Roots of the Early Cretaceous volcanic arc. *Geological Society of America Bulletin*, 111(4), 484–496.
- Clemens-Knott, D., Wolf, M. B., & Saleeby, J. B. (2000). Middle Mesozoic plutonism and deformation in the western Sierra Nevada foothills, California. *Geological Society of America Field Guide*, 2, 205–221.
- Coleman, D. S., Glazner, A. F., & Frost, T. P. (1992). Evidence from the Lamarck Granodiorite for rapid Late Cretaceous crust formation in California. *Science*, 258(5090), 1924–1926.
- Coney, P. J., & Reynolds, S. J. (1977). Cordilleran benioff zones. *Nature*, 270, 403–406.
- Crouch, J. K. (1979). Neogene tectonic evolution of the California Continental Borderland and western Transverse Ranges. *Geological Society of America Bulletin*, 90(4), 338–345.
- Crouch, J. K., & Suppe, J. (1993). Late Cenozoic tectonic evolution of the Los Angeles basin and inner California borderland: A model for core complex-like crustal extension. *Geological Society of America Bulletin*, 105(11), 1415–1434.
- Darin, M. H., & Dorsey, R. J. (2013). Reconciling disparate estimates of total offset on the southern San Andreas fault. *Geology*, 41(9), 975–978.
- Davis, T., Bazeley, W., Gordon, S., Lagoe, M., McIntosh, K., & Namson, J. (1988). Structure of Cuyama Valley, Caliente Range, and Carrizo Plain, and its significance to structural style of southern Coast Ranges and western Transverse Ranges. In Bazeley W. J. M. (Ed.), *Tertiary Tectonics and Sedimentation in the Cuyama Basin, San Luis Obispo, Santa Barbara, and Ventura Counties, California*, 141–158. Los Angeles, CA: Pacific Section, S.E.P.M.
- Day, S. M., & Bradley, C. R. (2001). Memory-efficient simulation of anelastic wave propagation. *Bulletin of the Seismological Society of America*, 91(3), 520–531.
- DePaolo, D. J. (1981). A neodymium and strontium isotopic study of the Mesozoic calc-alkaline granitic batholiths of the Sierra Nevada and Peninsular Ranges, California. *Journal of Geophysical Research*, 86(B11), 10470–10488.
- Dickinson, W. R. (1981). Plate tectonics and the continental margin of California. In W.G. Ernst (Ed.), *The geotectonic development of California: Englewood Cliffs* (pp. 1–28). New Jersey: Prentice-Hall.
- Dickinson, W. R. (1996). *Kinematics of transrotational tectonism in the California Transverse Ranges and its contribution to cumulative slip along the San Andreas transform fault system* (Vol. 305). Boulder, Colorado: Geological Society of America.
- Dickinson, W. R., & Rich, E. I. (1972). Petrologic intervals and petrofacies in the great valley sequence, sacramento valley, california. *Geological Society of America Bulletin*, 83(10), 3007–3024.
- Dickinson, W. R., & Snyder, W. S. (1978). Plate tectonics of the Laramide orogeny. *Geological Society of America Memoirs*, 151, 355–366.
- Ducea, M. N., & Saleeby, J. B. (1996). Buoyancy sources for a large, unrooted mountain range, the Sierra Nevada, California: Evidence from xenolith thermobarometry. *Journal of Geophysical Research*, 101(B4), 8229–8244.
- Ducea, M., & Saleeby, J. (1998). A case for delamination of the deep batholithic crust beneath the Sierra Nevada, California. *International Geology Review*, 40(1), 78–93.
- Ducea, M. N., Kidder, S., Chesley, J. T., & Saleeby, J. B. (2009). Tectonic underplating of trench sediments beneath magmatic arcs: The central California example. *International Geology Review*, 51(1), 1–26.
- Dutcher, L. C., & Garrett, A. (1963). *Geologic and hydrologic features of the San Bernardino area, California*. U.S. Geological Survey Water-Supply Paper 1419, 114p
- Ernst W. G. (1981). *The Geotectonic development of California*. Rubey Colloquium. New Jersey: Prentice Hall.

- Faulds, J. E., & Henry, C. D. (2008). Tectonic influences on the spatial and temporal evolution of the Walker Lane: An incipient transform fault along the evolving Pacific–North American plate boundary. *Ores and Orogenesis: Circum-Pacific Tectonics, Geologic Evolution, and Ore Deposits: Arizona Geological Society Digest*, 22, 437–470.
- Fichtner, A., Kennett, B., Igel, H., & Bunge, H. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, 179(3), 1703–1725.
- Fletcher, J. M., Bartley, J. M., Martin, M. W., Glazner, A. F., & Walker, J. D. (1995). Large-magnitude continental extension: An example from the central Mojave metamorphic core complex. *Geological Society of America Bulletin*, 107(12), 1468–1483.
- Fliedner, M. M., Ruppert, S., Group, S. S. N. C. D. W., et al. (1996). Three-dimensional crustal structure of the southern Sierra Nevada from seismic fan profiles and gravity modeling. *Geology*, 24(4), 367–370.
- Fliedner, M. M., Klemperer, S. L., & Christensen, N. I. (2000). Three-dimensional seismic model of the Sierra Nevada arc, California, and its implications for crustal and upper mantle composition. *Journal of Geophysical Research*, 105(B5), 10899–10921.
- Frankel, A. (1993). Three-dimensional simulations of ground motions in the San Bernardino Valley, California, for hypothetical earthquakes on the San Andreas fault. *Bulletin of the Seismological Society of America*, 83(4), 1020–1041.
- Fuis, G. S., & Mooney, W. D. (1990). Lithospheric structure and tectonics from seismic refraction and other data. *The San Andreas Fault System, California*, 1515, 206–236.
- Fuis, G., Mooney, W., Healy, J., McMechan, G., & Lutter, W. (1984). A seismic refraction survey of the imperial valley region, California. *Journal of Geophysical Research*, 89(B2), 1165–1189.
- Fuis, G. S., Clayton, R. W., Davis, P. M., Ryberg, T., Lutter, W. J., Okaya, D. A., Hauksson, E., Prodehl, C., Murphy, J. M., Benthien, M. L., et al. (2003). Fault systems of the 1971 San Fernando and 1994 Northridge earthquakes, southern California: Relocated aftershocks and seismic images from LARSE II. *Geology*, 31(2), 171–174.
- Gastil, R. G. (1975). Plutonic zones in the Peninsular Ranges of southern California and northern Baja California. *Geology*, 3(7), 361–363.
- Gastil, G., Morgan, G. J., & Krummenacher, D. (1978). Mesozoic history of peninsular California and related areas east of the Gulf of California. In Howell D. G. & McDougall K. A. (Eds.), *Mesozoic paleogeography of the western United States: Los Angeles, Pacific Section, Society of Economic Mineralogists and Paleontologists, Pacific Coast Paleogeography Symposium 2*, 107–115.
- Gazel, E., Plank, T., Forsyth, D. W., Bendersky, C., Lee, C.-T. A., & Hauri, E. H. (2012). Lithosphere versus asthenosphere mantle sources at the Big Pine Volcanic Field, California. *Geochemistry Geophysics Geosystems*, 13(6).
- Gilbert, W. G. (1973). Franciscan rocks near sur fault zone, northern Santa Lucia Range, California. *Geological Society of America Bulletin*, 84(10), 3317–3328.
- Glazner, A. F., Bartley, J. M., & Walker, J. D. (1989). Magnitude and significance of Miocene crustal extension in the central Mojave Desert, California. *Geology*, 17(1), 50–53.
- Glazner, A. F., Walker, J. D., Bartley, J. M., & Fletcher, J. M. (2002). Cenozoic evolution of the Mojave block of southern California. Geologic evolution of the Mojave Desert and southwestern basin and range. *Geological Society of America Memoir*, 195, 19–41.
- Godfrey, N. J., & Klemperer, S. L. (1998). Ophiolitic basement to a forearc basin and implications for continental growth: The Coast Range/Great Valley ophiolite, California. *Tectonics*, 17(4), 558–570.
- Graves, R. W., & Wald, D. J. (2004). Observed and simulated ground motions in the San Bernardino basin region for the Hector Mine, California, earthquake. *Bulletin of the Seismological Society of America*, 94(1), 131–146.
- Hall, C. A. (1991). Geology of the Point Sur–Lopez Point region, Coast Ranges, California: A part of the Southern California allochthon. *Geological Society of America Special Papers*, 266, 1–41.

- Hamilton, W. B. (1988). Detachment faulting in the Death Valley region, California and Nevada. *Geologic and hydrologic investigations of a potential nuclear waste disposal site at Yucca Mountain, southern Nevada: US Geological Survey Bulletin, 1790*, 51–85.
- Hammond, W. C., & Thatcher, W. (2007). Crustal deformation across the Sierra Nevada, northern Walker Lane, Basin and Range transition, western United States measured with GPS 2000–2004. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 112(B5), pp 1–26.
- Hammond, W. C., Kreemer, C., & Blewitt, G. (2009). Geodetic constraints on contemporary deformation in the northern Walker Lane: 3. Central Nevada seismic belt postseismic relaxation. *Geological Society of America Special Papers*, 447, 33–54.
- Hauksson, E. (2000). Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *Journal of Geophysical Research*, 105(B6), 13875–13903.
- Hauksson, E., & Unruh, J. (2007). Regional tectonics of the Coso geothermal area along the intracontinental plate boundary in central eastern California: Three-dimensional Vp and Vp/Vs models, spatial-temporal seismicity patterns, and seismogenic deformation. *Journal of Geophysical Research*, 112(B6).
- Hauksson, E., Small, P., Hafner, K., Busby, R., Clayton, R., Goltz, J., Heaton, T., Hutton, K., Kanamori, H., Polet, J., et al. (2001). Southern California Seismic Network: Caltech/USGS Element of TriNet 1997–2001. *Seismological Research Letters*, 72(6), 690–704.
- Hildreth, W. (2004). Volcanological perspectives on Long Valley, Mammoth Mountain, and Mono Craters: several contiguous but discrete systems. *Journal of Volcanology and Geothermal Research*, 136(3), 169–198.
- Hill, D. P. (2006). Unrest in Long Valley Caldera, California, 1978–2004. *Geological Society, London, Special Publications*, 269(1), 1–24.
- Hill, D. P., & Prejean, S. (2005). Magmatic unrest beneath Mammoth Mountain, California. *Journal of Volcanology and Geothermal Research*, 146(4), 257–283.
- Hopson, C. A., Mattinson, J. M., & Pessagno, E. A., Jr. (1981). Coast range ophiolite, western California. *The geotectonic development of California, Rubey, 1*, 418–510.
- Hornafius, J. S., Luyendyk, B. P., Terres, R., & Kamerling, M. (1986). Timing and extent of Neogene tectonic rotation in the western Transverse Ranges, California. *Geological Society of America Bulletin*, 97(12), 1476–1487.
- House, M. A., Wernicke, B. P., & Farley, K. A. (2001). Paleo-geomorphology of the Sierra Nevada, California, from (U-Th)/He ages in apatite. *American Journal of Science*, 301(2), 77–102.
- Howard, P., McAlee, J., & Gillespie, J. (1997). The role of fractures in surface and ground water flow: A field study of two canyons in the southeastern Sierra Nevada. *Geological Society of America*, 29, 20.
- Huang, H., Wang, L., Lee, E.-J., & Chen, P. (2012). An MPI-CUDA implementation and optimization for parallel sparse equations and least squares (LSQR). *Procedia Computer Science*, 9, 76–85.
- Huang, H., Dennis, J. M., Wang, L., & Chen, P. (2013). A scalable parallel LSQR algorithm for solving large-scale linear system for tomographic problems: A case study in seismic tomography. *Procedia Computer Science*, 18, 581–590.
- Huffman, O. F. (1972). Lateral displacement of upper Miocene rocks and the Neogene history of offset along the San Andreas fault in central California. *Geological Society of America Bulletin*, 83(10), 2913–2946.
- Humphreys, E. D., & Hager, B. H. (1990). A kinematic model for the late Cenozoic development of southern California crust and upper mantle. *Journal of Geophysical Research*, 95(B12), 19747–19762.
- Ichinose, G., Day, S., Magistrale, H., Prush, T., Vernon, F., & Edelman, A. (1996). Crustal thickness variations beneath the Peninsular ranges, southern California. *Geophysical Research Letters*, 23(22), 3095–3098.
- Ingersoll, R. V. (1982). Initiation and evolution of the Great Valley forearc basin of northern and central California, USA. *Geological Society, London, Special Publications*, 10(1), 459–467.

- Jacobson, C. (1995). Qualitative thermobarometry of inverted metamorphism in the Pelona and Rand Schists, southern California, using calciferous amphibole in mafic schist. *Journal of Metamorphic Geology*, 13(1), 79–92.
- Jacobson, C. E., Barth, A. P., & Grove, M. (2000). Late Cretaceous protolith age and provenance of the Pelona and Orocopia Schists, southern California: Implications for evolution of the Cordilleran margin. *Geology*, 28(3), 219–222.
- Jayko, A. S., & Bursik, M. (2012). Active transtensional intracontinental basins: Walker Lane in the western Great Basin. In C. Busby & A. A. Perez (Eds.), *Tectonics of sedimentary basins: Recent advances* (pp. 226–248). Chichester, West Sussex: Wiley.
- Jones, D. L., Blake, M., Jr., & Rangin, C. (1976). The four Jurassic belts of northern California and their significance to the geology of the southern California borderland. In Howell D. G. (Ed.), *Aspects of the geologic history of the California continental borderland: American Association of Petroleum Geologists Miscellaneous Publication 24*, 343–362.
- Kamerling, M. J., & Luyendyk, B. P. (1985). Paleomagnetism and Neogene tectonics of the northern Channel Islands, California. *Journal of Geophysical Research*, 90(B14), 12485–12502.
- Kanamori, H., Hauksson, E., & Heaton, T. (1991). TERRAScope and CUBE project at Caltech. *EOS Transactions American Geophysical Union*, 72(50), 564–564.
- Kane, M. F., & Pakiser, L. C. (1961). Geophysical study of subsurface structure in southern Owens Valley, California. *Geophysics*, 26(1), 12–26.
- Käuffel, P., Fichtner, A., & Igel, H. (2013). Probabilistic full waveform inversion based on tectonic regionalization—development and application to the Australian upper mantle. *Geophysical Journal International*, 193(1), 131.
- Keller, M. A. (1995). Ventura basin province (013). *National Assessment of United States Oil and Gas Resources—Results, Methodology, and Supporting Data* In D. Gautier, G. L. Dolton, K. I. Takahashi, & K. L. Varnes (Eds.), US Geological Survey Digital Data Series, 30.
- Keller, B., & Prothero, W. (1987). Western transverse ranges crustal structure. *Journal of Geophysical Research*, 92(B8), 7890–7906.
- Kidder, S., & Ducea, M. N. (2006). High temperatures and inverted metamorphism in the schist of Sierra de Salinas, California. *Earth and Planetary Science Letters*, 241(3), 422–437.
- Kistler, R. W. (1990). Two different lithosphere types in the Sierra Nevada, California. *Geological Society of America Memoirs*, 174, 271–282.
- Kistler, R. W., & Peterman, Z. E. (1973). Variations in Sr, Rb, K, Na, and initial Sr87/Sr86 in Mesozoic granitic rocks and intruded wall rocks in central California. *Geological Society of America Bulletin*, 84(11), 3489–3512.
- Lobotka, T. C., & Albee, A. L. (1990). Uplift and exposure of the Panamint metamorphic complex, California. *Geological Society of America Memoirs*, 176, 345–362.
- Langbein, J. O. (2003). Deformation of the Long Valley Caldera, California: Inferences from measurements from 1988 to 2001. *Journal of Volcanology and Geothermal Research*, 127(3), 247–267.
- Langenheim, V., & Powell, R. (2009). Basin geometry and cumulative offsets in the Eastern transverse ranges, southern California: Implications for transrotational deformation along the San Andreas fault system. *Geosphere*, 5(1), 1–22.
- Langenheim, V., Jachens, R., & Aiken, C. (2014). Geophysical framework of the Peninsular Ranges batholith—Implications for tectonic evolution and neotectonics. *Geological Society of America Memoirs*, 211, 1–20.
- Larsen, E. S. (1948). Batholith and associated rocks of Corona, Elsinore, and San Luis Rey quadrangles southern California. *Geological Society of America Memoirs*, 29, 1–185.
- Lawrence, J. F., & Prieto, G. A. (2011). Attenuation tomography of the western United States from ambient seismic noise. *Journal of Geophysical Research*, 116(B6).
- Lawrence, J. F., & Shearer, P. M. (2008). Imaging mantle transition zone thickness with SdS-SS finite-frequency sensitivity kernels. *Geophysical Journal International*, 174(1), 143–158.
- Lee, E.-J., & Chen, P. (2013). Automating seismic waveform analysis for full 3-D waveform inversions. *Geophysical Journal International*, 194(1), 572–589.

- Lee, E.-J., Chen, P., Jordan, T. H., & Wang, L. (2011). Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. *Geophysical Journal International*, 186(1), 311–330.
- Lee, E.-J., Huang, H., Dennis, J. M., Chen, P., & Wang, L. (2013). An optimized parallel LSQR algorithm for seismic tomography. *Computers & Geosciences*, 61, 184–197.
- Lee, E.-J., Chen, P., & Jordan, T. H. (2014a). Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes. *Seismological Research Letters*, 85(6), 311–330.
- Lee, E.-J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014b). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research*, 119(8), 6421–6451.
- Legg, M. R. (1991). Developments in understanding the tectonic evolution of the California Continental Borderland. In Osborne R. H. (Ed.), *From shoreline to abyss: Soc. Econ. Paleontol. Mineral., Spec. Publ. 46*, 291–312.
- Lovely, P., Shaw, J. H., Liu, Q., & Tromp, J. (2006). A structural VP model of the Salton Trough, California, and its implications for seismic hazard. *Bulletin of the Seismological Society of America*, 96(5), 1882–1896.
- Luffi, P., Saleeby, J. B., Lee, C.-T. A., & Ducea, M. N. (2009). Lithospheric mantle duplex beneath the central Mojave Desert revealed by xenoliths from Dish Hill, California. *Journal of Geophysical Research*, 114(B3).
- Lutter, W. J., Fuis, G. S., Thurber, C. H., & Murphy, J. (1999). Tomographic images of the upper crust from the Los Angeles basin to the Mojave Desert, California: Results from the Los Angeles Region Seismic Experiment. *Journal of Geophysical Research*, 104(B11), 25543–25565.
- Lutter, W., Fuis, G., Ryberg, T., Okaya, D., Clayton, R., Davis, P., Prodehl, C., Murphy, J., Langenheim, V., Benthien, M. L., et al. (2004). Upper crustal structure from the Santa Monica mountains to the Sierra Nevada, southern California: tomographic results from the Los Angeles regional seismic experiment, Phase II (LARSE II). *Bulletin of the Seismological Society of America*, 94(2), 619–632.
- Luyendyk, B. P. (1991). A model for Neogene crustal rotations, transtension, and transpression in southern California. *Geological Society of America Bulletin*, 103(11), 1528–1536.
- Luyendyk, B. P., Kamerling, M. J., & Terres, R. (1980). Geometric model for Neogene crustal rotations in southern California. *Geological Society of America Bulletin*, 91(4), 211–217.
- Ma, S., Prieto, G. A., & Beroza, G. C. (2008). Testing community velocity models for southern California using the ambient seismic field. *Bulletin of the Seismological Society of America*, 98(6), 2694–2714.
- Mack, S., Saleeby, J. B., & Farrell, J. E. (1979). Origin and emplacement of the Academy pluton, Fresno County, California. *Geological Society of America Bulletin*, 90(4 Part II), 633–694.
- Maggi, A., Tape, C., Chen, M., Chao, D., & Tromp, J. (2009). An automated time-window selection algorithm for seismic tomography. *Geophysical Journal International*, 178, 257–281.
- Magistrale, H., McLaughlin, K., & Day, S. (1996). A geology-based 3D velocity model of the Los Angeles basin sediments. *Bulletin of the Seismological Society of America*, 86(4), 1161–1166.
- Magistrale, H., Day, S., Clayton, R. W., & Graves, R. (2000). The SCEC Southern California reference three-dimensional seismic velocity model version 2. *Bulletin of the Seismological Society of America*, 90(6B), S65–S76.
- Malin, P., Goodman, E., Henyey, T., Li, Y., Okaya, D., & Saleeby, J. (1995). Significance of seismic reflections beneath a tilted exposure of deep continental crust, Tehachapi mountains, California. *Journal of Geophysical Research*, 100(B2), 2069–2087.
- Malin, P., Lou, M., & Rial, J. (1996). F-r waves: A second fault-guided mode with implications for fault property studies. *Geophysical Research Letters*, 23(24), 3547–3550.
- Mann, P., & Gordon, M. B. (1996). Tectonic Uplift and Exhumation of Blueschist Belts Along Transpressional Strike-Slip Fault Zones. In G. E. Bebout, D. W. Scholl, S. H. Kirby, & J. P. Platt (Eds.), *Subduction top to bottom. Amer. Geophys. Un. Monogr.*, 96, 143–154.

- Marcinkovich, C., & Olsen, K. (2003). On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme. *Journal of Geophysical Research*, 108(B5).
- Martin, M. W., Glazner, A. F., Walker, J. D., & Schermer, E. (1993). Evidence for right-lateral transfer faulting accommodating an echelon Miocene extension, Mojave Desert, California. *Geology*, 21(4), 355–358.
- McClusky, S., Bjornstad, S., Hager, B. H., King, R., Meade, B., Miller, M., Monastero, F., & Souter, B. (2001). Present day kinematics of the eastern California shear zone from a geodetically constrained block model. *Geophysical Research Letters*, 28(17), 3369–3372.
- McQuarrie, N., & Wernicke, B. P. (2005). An animated tectonic reconstruction of southwestern North America since 36 Ma. *Geosphere*, 1(3), 147–172.
- Metzger, D. (1968). The bouse formation (Pliocene) of the Parker-Blythe-Cibola area, Arizona and California. In *Geological Survey Research*, U.S. Geological Survey Professional Paper, pp. D126–D136.
- Miller, M. G., & Prave, A. (2002). Rolling hinge or fixed basin?: A test of continental extensional models in Death Valley, California, United States. *Geology*, 30(9), 847–850.
- Miller, D., Nilsen, T., & Bilodeau, W. (1992). Late cretaceous to early eocene geologic evolution of the us cordillera. *The Geology of North America*, 3, 205–260.
- Miller, M. M., Johnson, D. J., Dixon, T. H., & Dokka, R. K. (2001). Refined kinematics of the Eastern California shear zone from GPS observations, 1993–1998. *Journal of Geophysical Research*, 106(B2), 2245–2263.
- Monastero, F. C., Walker, J. D., Katzenstein, A. M., & Sabin, A. E. (2002). Neogene evolution of the Indian Wells Valley, east-central California. *Geologic Evolution of the Mojave Desert and Southwestern Basin and Range. Geological Society of America Memoir*, 195, 199–228.
- Moore, J. G. (1959). The quartz diorite boundary line in the western United States. *The Journal of Geology*, 67, 198–210.
- Morton, D. M., & Matti, J. C. (1993). Extension and contraction within an evolving divergent strike-slip fault complex: The San Andreas and San Jacinto fault zones at their convergence in southern California. *Geological Society of America Memoirs*, 178, 217–230.
- Namson, J., & Davis, T. L. (1990). Late cenozoic fold and thrust belt of the southern coast ranges and santa maria basin, california. *American Association of Petroleum Geologists Bulletin*, 74(4), 467–492.
- Nichols, J., Warren, N., Luyendyk, B. P., & Spudich, P. (1980). Seismic velocity structure of the ophiolite at Point Sal, southern California, determined from laboratory measurements. *Geophysical Journal International*, 63(1), 165–185.
- Norton, I. (2011). Two-stage formation of Death Valley. *Geosphere*, 7(1), 171–182.
- Ostdick, J. (1997). The hydrogeology of southwest Indian Wells Valley, Kern County, California: Evidence for extrabasinal, fracture-directed groundwater recharge from the adjacent Sierra Nevada Mountains. *Unpublished MS Thesis, California State University, Bakersfield*.
- Page, B. M. (1981). The southern coast ranges. *The geotectonic development of California*, 1, 329–417.
- Pellerin, C. L. M., & Christensen, N. I. (1998). Interpretation of crustal seismic velocities in the San Gabriel–Mojave region, southern California. *Tectonophysics*, 286(1), 253–271.
- Pickett, D. A., & Saleeby, J. B. (1993). Thermobarometric constraints on the depth of exposure and conditions of plutonism and metamorphism at deep levels of the Sierra Nevada Batholith, Tehachapi Mountains, California. *Journal of Geophysical Research*, 98(B1), 609–629.
- Pickett, D. A., & Saleeby, J. B. (1994). Nd, Sr, and Pb isotopic characteristics of Cretaceous intrusive rocks from deep levels of the Sierra Nevada batholith, Tehachapi Mountains, California. *Contributions to Mineralogy and Petrology*, 118(2), 198–215.
- Platt, J. (1975). Metamorphic and deformational processes in the Franciscan Complex, California: Some insights from the Catalina Schist terrane. *Geological Society of America Bulletin*, 86(10), 1337–1347.

- Ponti, D. J. (1985). The Quaternary alluvial sequence of the Antelope Valley, California. *Geological Society of America Special Papers*, 203, 79–96.
- Powell, R. E. (1981). *Geology of the crystalline basement complex, eastern Transverse Ranges, southern California: Constraints on regional tectonic interpretation*. PhD thesis, California Institute of Technology.
- Powell, R. E. (1982). *Crystalline basement terranes in the southern eastern Transverse Ranges, California*. Geologic excursions in the Transverse Ranges, Southern California. (pp. 109–136). Geological Society of America
- Powell, R. E. (1993). Balanced palinspastic reconstruction of pre-late Cenozoic paleogeology, southern California: Geologic and kinematic constraints on evolution of the San Andreas fault system. *Geological Society of America Memoirs*, 178, 1–106.
- Prejean, S., Ellsworth, W., Zoback, M., & Waldhauser, F. (2002). Fault structure and kinematics of the Long Valley Caldera region, California, revealed by high-accuracy earthquake hypocenters and focal mechanism stress inversions. *Journal of Geophysical Research*, 107(B12), ESE–9.
- Prejean, S., Stork, A., Ellsworth, W., Hill, D., & Julian, B. (2003). High precision earthquake locations reveal seismogenic structure beneath Mammoth Mountain, California. *Geophysical Research Letters*, 30(24), SDE4-1, 2247.
- Putirka, K., Jean, M., Cousens, B., Sharma, R., Torrez, G., & Carlson, C. (2012). Cenozoic volcanism in the Sierra Nevada and Walker Lane, California, and a new model for lithosphere degradation. *Geosphere*, 8(2), 265–291.
- Reasenber, P., Ellsworth, W., & Walter, A. (1980). Teleseismic evidence for a low-velocity body under the Coso Geothermal Area. *Journal of Geophysical Research*, 85(B5), 2471–2483.
- Richard, S. M. (1993). Palinspastic reconstruction of southeastern California and southwestern Arizona for the middle Miocene. *Tectonics*, 12(4), 830–854.
- Ritsema, J., & van Heijst, H.-J. (2002). Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes. *Geophysical Journal International*, 149(2), 482–489.
- Ross, D. C. (1975). *Modal and chemical data for granitic rocks of the gabilan range, central coast ranges, California*. US Geological Survey.
- Ross, D. (1976). Metagraywacke in SALINIAN block, central COAST RANGES, CALIFORNIA—possible correlative across SAN-ANDREAS FAULT. *Journal of Research of the US Geological Survey*, 4(6), 683–696.
- Ross, D. C. (1984). Possible correlations of basement rocks across the San Andreas, San Gregorio-Hosgri, and Rinconada-Reliz-King City faults, California. *U.S. Geological Survey Professional Paper*, 1317, 37.
- Ryberg, T., & Fuis, G. S. (1998). The san gabriel mountains bright reflective zone: Possible evidence of young mid-crustal thrust faulting in southern california. *Tectonophysics*, 286(1), 31–46.
- Saleeby, J. (1981). Ocean floor accretion and volcanoplutonic arc evolution of the Mesozoic Sierra Nevada. In W. G. Ernst (Ed.), *The geotectonic development of California* (pp. 132–181). New Jersey: Prentice-Hall
- Saleeby, J. B. (1990). Progress in tectonic and petrogenetic studies in an exposed cross-section of young (100 Ma) continental crust, southern Sierra Nevada, California. In M. H. Salisbury & D. M. Fountain (Eds.), *Exposed cross-sections of the continental crust* (pp. 137–158). New York: Springer.
- Saleeby, J. (2003). Segmentation of the Laramide slab—Evidence from the southern Sierra Nevada region. *Geological Society of America Bulletin*, 115(6), 655–668.
- Saleeby, J., & Sharp, W. (1980). Chronology of the structural and petrologic development of the southwest Sierra Nevada foothills, California. *Geological Society of America Bulletin*, 91(6 Part II), 1416–1535.
- Saleeby, J. B., Sams, D. B., & Kistler, R. W. (1987). U/Pb zircon, strontium, and oxygen isotopic and geochronological study of the southernmost Sierra Nevada batholith, California. *Journal of Geophysical Research*, 92(B10), 10443–10466.

- Saleeby, J., Ducea, M., & Clemens-Knott, D. (2003). Production and loss of high-density batholithic root, southern Sierra Nevada, California. *Tectonics*, 22(6).
- Schmidt, K., Wetmore, P., Johnson, S., & Paterson, S. (2002). Controls on orogenesis along an ocean-continent margin transition in the Jura-Cretaceous Peninsular Ranges batholith. *SPECIAL PAPERS-GEOLOGICAL SOCIETY OF AMERICA*, 49–72.
- Seccia, D., Chiarabba, C., De Gori, P., Bianchi, I., & Hill, D. (2011). Evidence for the contemporary magmatic system beneath Long Valley Caldera from local earthquake tomography and receiver function analysis. *Journal of Geophysical Research*, 116(B12).
- Serpa, L., & Pavlis, T. L. (1996). Three-dimensional model of the late Cenozoic history of the Death Valley region, southeastern California. *Tectonics*, 15(6), 1113–1128.
- Shapiro, N. M. (2004). Emergence of broadband rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(7), 4.
- Shapiro, N., Campillo, M., Stehly, L., & Ritzwoller, M. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, 307(5715), 1615–1618.
- Sharry, J. (1981). *The geology of the western Tehachapi Mountains, California*. PhD thesis, Massachusetts Institute of Technology.
- Shervais, J. W., Kimbrough, D. L., Renne, P., Hanan, B. B., Murchey, B., Snow, C. A., Zogman Schuman, M. M., & Beaman, J. (2004). Multi-stage origin of the coast range ophiolite, California: Implications for the life cycle of supra-subduction zone ophiolites. *International Geology Review*, 46(4), 289–315.
- Sigloch, K., & Nolet, G. (2006). Measuring finite-frequency body-wave amplitudes and travel-times. *Geophysical Journal International*, 167(1), 271–287.
- Silver, L., & Chappell, B. (1988). The Peninsular Ranges Batholith: An insight into the evolution of the Cordilleran batholiths of southwestern North America. *Transactions of the Royal Society of Edinburgh: Earth Sciences*, 79(2–3), 105–121.
- Silver, L., & Mattinson, J. (1986). Orphan Salinia has a home. *EOS Transactions American Geophysical Union*, 67(44), 1215.
- Sisson, T., Grove, T., & Coleman, D. (1996). Hornblende gabbro sill complex at Onion Valley, California, and a mixing origin for the Sierra Nevada batholith. *Contributions to Mineralogy and Petrology*, 126(1–2), 81–108.
- Snow, J. K., & Wernicke, B. P. (2000). Cenozoic tectonism, in the central Basin and Range: Magnitude, rate, and distribution of upper crustal strain. *American Journal of Science*, 300(9), 659.
- Stewart, J. H. (1983). Extensional tectonics in the Death Valley area, California: Transport of the Panamint Range structural block 80 km northwestward. *Geology*, 11(3), 153–157.
- Stewart, J. (1988). Tectonics of the Walker Lane belt, western Great Basin: Mesozoic and Cenozoic deformation in a zone of shear. *Metamorphism and crustal evolution of the western United States*, 7, 683–713.
- Surpless, B. E., Stockli, D. F., Dumitru, T. A., & Miller, E. L. (2002). Two-phase westward encroachment of Basin and Range extension into the northern Sierra Nevada. *Tectonics*, 21(1), 2–1.
- Süss, M. P., & Shaw, J. H. (2003). P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California. *Journal of Geophysical Research*, 108(B3).
- Tape, C., Liu, Q., & Tromp, J. (2007). Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves. *Geophysical Journal International*, 168(3), 1105–1129.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180(1), 433–462.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics. Philadelphia, 324p.

- ten Brink, U. S., Zhang, J., Brocher, T. M., Okaya, D. A., Klitgord, K. D., & Fuis, G. S. (2000). Geophysical evidence for the evolution of the California Inner Continental Borderland as a metamorphic core complex. *Journal of Geophysical Research*, 105(B3), 5835–5857.
- Terres, R. R., & Luyendyk, B. P. (1985). Neogene tectonic rotation of the San Gabriel region, California, suggested by paleomagnetic vectors. *Journal of Geophysical Research*, 90(B14), 12467–12484.
- Thyne, G. D., Gillespie, J. M., & Ostdick, J. R. (1999). Evidence for interbasin flow through bedrock in the southeastern Sierra Nevada. *Geological Society of America Bulletin*, 111(11), 1600–1616.
- Tizzani, P., Battaglia, M., Zeni, G., Atzori, S., Berardino, P., & Lanari, R. (2009). Uplift and magma intrusion at Long Valley caldera from InSAR and gravity measurements. *Geology*, 37(1), 63–66.
- Todd, V., & Shaw, S. (1985). S-type granitoids and an is line in the peninsular ranges batholith, southern California. *Geology*, 13(4), 231–233.
- Topping, D. J. (1993). Paleogeographic reconstruction of the Death Valley extended region: Evidence from Miocene large rock-avalanche deposits in the Amargosa Chaos Basin, California. *Geological Society of America Bulletin*, 105(9), 1190–1213.
- Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160(1), 195–216.
- Unruh, J., Humphrey, J., & Barron, A. (2003). Transtensional model for the Sierra Nevada frontal fault system, eastern California. *Geology*, 31(4), 327–330.
- van Leeuwen, T., & Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation traveltome tomography. *Geophysical Journal International*, 182(3), 1383–1394.
- Vedder, J. (1987). Regional geology and petroleum potential of the southern California borderland. In Scholl D. W., Grantz A and Vedder J. G. (Eds.), *Geology and Resource Potential of the Continental Margin of Western North America and Adjacent Ocean Basins ± Beaufort Sea to Baja California*. Circum-Pacific Council for Energy Mineral Resources, Earth Sci. Ser. 6, 403–447.
- Vedder, J., Howell, D., & McLean, H. (1983). Stratigraphy, sedimentation, and tectonic accretion of exotic terranes, southern Coast Ranges, California. *Studies in continental margin geology: American Association of Petroleum Geologists Memoir*, 34, 471–496.
- Von Huene, R., Bateman, P., & Ross, D. (1963). Indian Wells Valley, Owens Valley, Long Valley, and Mono Basin. In *Guidebook for seismological study tour, 1963 Meeting of IUGG: China Lake, California, US Naval Ordnance Test Station*, (pp. 57–110).
- Walawender, M. J., Girty, G. H., Lombardi, M., Kimbrough, D., Girty, M. S., & Anderson, C. (1991). A synthesis of recent work in the Peninsular Ranges batholith. *Geological Excursions in Southern California and Mexico*: San Diego, USA, San Diego State University, Department of Geological Sciences, 297–312.
- Walker, J. D., Bartley, J. M., & Glazner, A. F. (1990). Large-magnitude miocene extension in the central Mojave Desert: Implications for Paleozoic to Tertiary paleogeography and tectonics. *Journal of Geophysical Research*, 95(B1), 557–569.
- Weiland, C. M., Steck, L. K., Dawson, P. B., & Korneev, V. A. (1995). Nonlinear teleseismic tomography at Long Valley Caldera, using three-dimensional minimum travel time ray tracing. *Journal of Geophysical Research*, 100(B10), 20379–20390.
- Wernicke, B. P., England, P. C., Sonder, L. J., & Christiansen, R. L. (1987). Tectonomagmatic evolution of Cenozoic extension in the North American Cordillera. *Geological Society, London, Special Publications*, 28(1), 203–221.
- Wernicke, B. P., Walker, J. D., & Hodges, K. (1988). Field guide to the northern part of the Tucki Mountain fault system, Death Valley region, California. In *This extended land: Geological journeys in the Southern basin and range: Geological Society of America, Cordilleran section meeting, field trip guidebook* (pp. 58–63).
- Wernicke, B., et al. (1992). Cenozoic extensional tectonics of the US Cordillera. *The Geology of North America*, 3, 553–581.

- Wernicke, B., Clayton, R., Ducea, M., Jones, C. H., Parker, S., Ruppert, S., Saleeby, J., Snow, J. K., Squires, L., Flidner, M., et al. (1996). Origin of high mountains in the continents: The Southern Sierra Nevada. *Science*, 271(5246), 190–193.
- Wilson, C. K., Jones, C. H., & Gilbert, H. J. (2003). Single-chamber silicic magma system inferred from shear wave discontinuities of the crust and uppermost mantle, Coso geothermal area, California. *Journal of Geophysical Research*, 108(B5).
- Woodford, A. O. (1924). *The Catalina metamorphic facies of the Franciscan series*. California: University of California Press.
- Wright, T. L. (1991). Structural geology and tectonic evolution of the Los Angeles Basin, California. *Active Margin Basins*, 52, 35–134.
- Xu, Z., Chen, P., & Chen, Y. (2013). Sensitivity kernel for the weighted norm of the frequency-dependent phase correlation. *Pure and Applied Geophysics*, 170(3), 353–371.
- Yan, Z., Clayton, R. W., & Saleeby, J. (2005). Seismic refraction evidence for steep faults cutting highly attenuated continental basement in the central transverse ranges, California. *Geophysical Journal International*, 160(2), 651–666.
- Yeats, R. S., Cole, M. R., Merschat, W. R., & Parsley, R. M. (1974). Poway fan and submarine cone and rifting of the inner southern California borderland. *Geological Society of America Bulletin*, 85(2), 293–302.
- Zandt, G., Gilbert, H., Owens, T. J., Ducea, M., Saleeby, J., & Jones, C. H. (2004). Active foundering of a continental arc root beneath the southern Sierra Nevada in California. *Nature*, 431(7004), 41–46.
- Zhao, L., Jordan, T. H., Olsen, K. B., & Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models. *Bulletin of the Seismological Society of America*, 95(6), 2066–2080.

Index

A

ABC, 15
adjoint, 118
adjoint Green's function, 147
adjoint matrix, 101, 118
adjoint representation theorem, 162
adjoint source, 60
algebra, 113
amplitude reduction time, 216
amplitude-reduction time, 213, 270
anti-Hermitian, 118, 125
anti-self-adjoint, 118, 125
antilinear, 93
Arbitrary high-order DERivatives, 6
arithmetic average, 41
asymptotic ray series, 2
AWF, 67

B

basis, 95, 104
Beowulf cluster, 9
Born approximation, 185, 189
Born kernel, 180
Born series, 179
Born wavepaths, 183
boundary-element method, 7
bounded linear transform, 116
bounded operators, 115, 147

C

C-PML, 39
Cauchy sequences, 101
Cauchy-Schwarz inequality, 95
caustics, 4
CFL, 55
CFS-PML, 39
chain rule, 122
characteristic duration, 267
clock frequency, 8

coarse-grained, 33
complete, 104
complete vector space, 102
completeness relation, 100, 105, 106, 111, 114
complex vector space, 92
composition, 122
conjugate transpose, 101
conjunct, 124
continuous, 116
counting set, 105
CUDA, 11

D

data functional, 192
data sensitivity kernel, 193
Debye function, 25
denuisancing, 321
derivative of a distribution, 110
differentiable, 120
differential operator, 146
diffraction tomography, 183
Dirac delta function, 105, 108, 151
directivity velocity vector, 267
discontinuous-Galerkin, 6
distributed-memory, 9
distribution, 109, 147
divergence, 124
divergence theorem, 124, 138
dual, 98, 118
dual basis, 97
dual space, 97
dynamic ray tracing, 3

E

Eikonal equation, 3
engineering shear strain, 70
event kernel, 327
exponential integrator method, 20

F

final conditions, 139
 finite moment tensor, 267
 finite-difference, 5
 finite-dimensional, 95
 finite-element, 5
 finite-volume, 7
 FLOPS, 7
 formal adjoint, 124
 formally self-adjoint, 124
 FPGA, 10
 Fréchet derivative, 120
 Fréchet kernel, 164
 Fredholm equation, 175
 Fresnel zone, 233
 FWF, 67

G

Gâteaux derivative, 121
 Gauss-Lobatto-Legendre, 6
 Gaussian beam, 4
 Gaussian-Bayesian, 317
 generalized function, 109, 147
 generalized Green's identity, 124, 147, 155
 Generalized Maxwell Body, 22
 Generalized Zener Body, 22
 ghost layer, 52
 global grid index, 52
 GMT, 48
 GPU, 7
 gradient, 121
 Gram-Schmidt orthonormalization, 95
 Green's function, 146

H

harmonic average, 41
 Hermitian, 118, 124
 Hermitian conjugate, 101, 118
 Hermitian transpose, 101
 Hessian matrix, 123
 Hilbert space, 103
 homogeneous, 139
 HPC, 7

I

infinite-dimensional, 95
 inhomogeneous, 139
 inner product space, 93
 integral equation, 115
 integration kernel, 115, 146
 Intel Xeon Phi, 10

J

Jacobian matrix, 122, 313

K

kernel, 97
 Kronecker delta, 17, 95, 151

L

LAN, 9
 Leibniz rule, 122
 length, 94
 linear functional, 97
 linear operator, 114
 linear transformation, 97
 linearly independent, 94
 local grid index, 52
 LSQR, 318
 Lyapunov exponents, 4

M

Maslov method, 4
 maximum likelihood, 316
 memory variables, 20
 MPI, 9
 MPP, 9
 multiplication, 113

N

norm, 94
 normal vector, 95
 normalized vector, 95
 nuisance parameters, 321
 null space, 97

O

OpenCL, 11
 OpenMP, 9
 orthogonal, 95
 orthonormal basis, 95

P

Padé approximant, 21
 Parseval theorem, 113
 PBS, 66
 phase-delay time, 213, 216, 270
 PML, 15, 36
 positive definite, 93
 positive definite operator, 119
 positive operator, 119
 POSIX Threads, 9
 primary kernels, 192
 product rule, 122
 projection operator, 99
 propagator, 179
 pseudo-spectral, 5
 pullback, 100

Q

quality factor, 19

R

rate-of-relaxation function, 133

ray theory, 2

real vector space, 92

reciprocity relation, 149, 154

recording grid points, 64

representation theorem, 159

RGT, 60

roughing operator, 320

Rytov approximation, 226, 228

Rytov validity condition, 227

Rytov wavepaths, 233

S

scattered wavefield, 177

scattering-integral, 181

SCEC, 15

seismic tomography, 1

seismogram perturbation kernel, 194

self-adjoint, 118, 124, 126

sesquilinear, 93

sifting property, 108

SIMD, 7

skew-adjoint, 125

small-angle scattering, 229

SMP, 9

spaces, 91

spans, 95

spectral-element, 5

staggered-grid, 15

starting model, 1

stress imaging, 34

Sturm-Liouville operator, 127

sub-mesh, 46, 51, 66

subspace, 92

surface term, 124

symmetric, 124

T

Taylor expansion, 123

test functions, 109

transformation, 97

transport equation, 3

transpose, 100

transverse isotropy, 251

U

UTM, 45

V

vector space, 91

vectors, 91

Volterra equation, 175

W

wavefield perturbation kernel, 195

weight function, 105