

## REVIEW ARTICLE

# Spectral-Element and Adjoint Methods in Seismology

Jeroen Tromp<sup>1,\*</sup>, Dimitri Komatitsch<sup>2</sup> and Qinya Liu<sup>3</sup>

<sup>1</sup> *Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA.*

<sup>2</sup> *Department of Geophysical Modeling and Imaging, University of Pau, 64013 Pau Cedex, France.*

<sup>3</sup> *Scripps Institution of Oceanography, University of California San Diego, La Jolla CA 92093-0225, USA.*

Received 14 September 2006; Accepted (in revised version) 6 June 2007

Available online 14 September 2007

---

**Abstract.** We provide an introduction to the use of the spectral-element method (SEM) in seismology. Following a brief review of the basic equations that govern seismic wave propagation, we discuss in some detail how these equations may be solved numerically based upon the SEM to address the *forward problem* in seismology. Examples of synthetic seismograms calculated based upon the SEM are compared to data recorded by the Global Seismographic Network. Finally, we discuss the challenge of using the remaining differences between the data and the synthetic seismograms to constrain better Earth models and source descriptions. This leads naturally to adjoint methods, which provide a practical approach to this formidable computational challenge and enables seismologists to tackle the *inverse problem*.

**AMS subject classifications:** 74S05, 74S30, 86A15, 86A22

**Key words:** Spectral-element method, adjoint methods, seismology, inverse problems, numerical simulations.

---

## Contents

1	Introduction	2
2	Basic theory of seismology	2
3	Spectral-element method	7
4	Adjoint methods	18
5	Discussion and conclusions	28

---

\*Corresponding author. *Email addresses:* [jtromp@gps.caltech.edu](mailto:jtromp@gps.caltech.edu) (J. Tromp), [dimitri.komatitsch@univ-pau.fr](mailto:dimitri.komatitsch@univ-pau.fr) (D. Komatitsch), [liuqy@ucsd.edu](mailto:liuqy@ucsd.edu) (Q. Liu)

## 1 Introduction

The spectral-element method (SEM) has been used for more than two decades in computational fluid dynamics [48], but it has only recently gained popularity in seismology. Initially the method was applied to 2D seismic wave propagation problems [15,50], but currently the SEM is widely used for 3D regional [21, 28, 29, 33, 37, 55] and global [10–12,30,31,35,36] seismic wave propagation. Recent reviews of the SEM in seismology may be found in [38] and [13].

Like a classical finite-element method, the SEM is based upon an integral or weak implementation of the equation of motion. It combines the accuracy of the global pseudospectral method with the flexibility of the finite-element method. The wavefield is typically represented in terms of high-degree Lagrange interpolants, and integrals are computed based upon Gauss-Lobatto-Legendre quadrature, which leads to a simple explicit time scheme that lends itself very well to calculations on parallel computers.

The current challenge lies in harnessing these numerical capabilities to enhance the quality of tomographic images of the Earth's interior, in conjunction with improving models of the rupture process during an earthquake. [62] demonstrated that this problem may be solved iteratively by numerically calculating the derivative of a waveform misfit function. The construction of this derivative involves the interaction between the wavefield for the current model and a wavefield obtained by using the time-reversed waveform differences between the data and the current synthetics as simultaneous sources. Only two numerical simulations are required to calculate the gradient of the misfit function: one for the current model and a second for the time-reversed differences between the data and the current synthetics. [60] generalized the calculation of the derivative of a misfit function by introducing the concept of an 'adjoint' calculation. The acoustic theory developed by [62] was extended to the anelastic wave equation by [63,64]. Applications of the theory may be found in [1, 2, 17, 23, 45, 46, 49, 61].

The purpose of this article is to review the use of the SEM in seismology, and to illustrate the powerful combination of the SEM for the *forward problem*, i.e., given a 3D Earth model and a (finite) source model accurately simulate the associated ground motions, with adjoint methods for the *inverse problem*, i.e., using the remaining differences between the data and the simulations to improve source and Earth models.

## 2 Basic theory of seismology

### 2.1 Earth models

Seismologists have determined the average, spherically symmetric structure of the Earth with a high degree of accuracy. A typical one-dimensional (1D) model is the Preliminary Reference Earth Model (PREM) [20], shown in Fig. 1. Such an isotropic, elastic Earth model may be characterized in terms of three parameters: the distribution of mass density  $\rho$ , the compressional wave speed  $\alpha$ , and the shear wave speed  $\beta$ . Rather than the

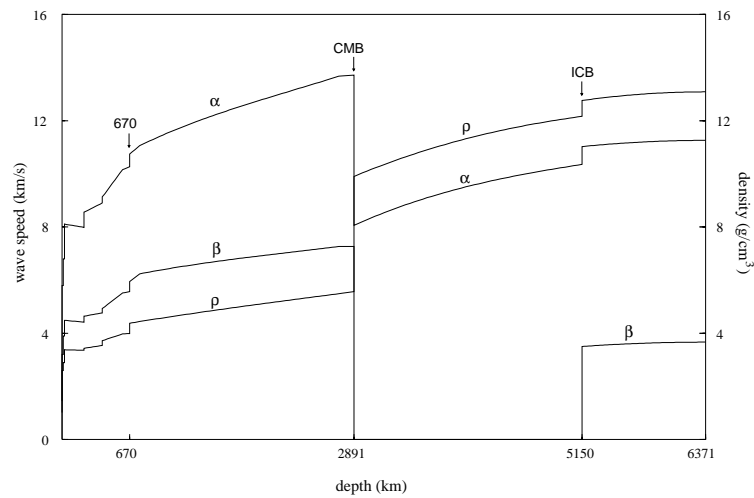


Figure 1: Compressional-wave speed  $\alpha$ , shear-wave speed  $\beta$ , and density  $\rho$  in the spherically symmetric, isotropic Preliminary Reference Earth Model (PREM) [20]. The locations of the inner-core boundary (ICB) and the core-mantle boundary (CMB) are marked. The model is capped by a 3 km thick uniform oceanic layer. There are also a number of solid-solid boundaries, including a Moho discontinuity between the crust and mantle at a depth of 24.4 km, and upper-mantle discontinuities at depths of 220, 400 and 670 km.

compressional and shear wave speeds, one may also use the bulk modulus, or incompressibility,  $\kappa$ , and the shear modulus, or rigidity,  $\mu$ . These two sets of parameters are related by  $\alpha = [(\kappa + \frac{4}{3}\mu) / \rho]^{1/2}$  and  $\beta = (\mu / \rho)^{1/2}$ . The PREM parameters exhibit a number of discontinuities as a function of depth, for example at the ocean floor, the Moho (the boundary between the crust and the mantle), across the upper mantle phase transitions, at the core-mantle boundary (CMB), and at the inner core boundary (ICB).

The lateral variations superimposed on typical 1D Earth models are of order a few percent for compressional wave speed and as much as  $\pm 8\%$  for shear wave speed in the shallow mantle. For example, Fig. 2 shows map views of 3D shear-wave speed model S20RTS [52] at various depths throughout the mantle. The goal of seismic tomography is to map lateral variations in the Earth’s mantle and crust, and to relate such three-dimensional (3D) heterogeneity to variations in temperature and composition. Thus, these maps help constrain the composition and dynamics of the Earth’s interior.

With modern numerical methods and computers, seismologists are now able to simulate seismic wave propagation in 3D Earth models, such as the one shown in Fig. 2, at unprecedented resolution and accuracy [30, 31, 35, 67].

## 2.2 Constitutive relationships

Let us denote the displacement field induced by an earthquake by  $\mathbf{s}$ . The (symmetric) strain tensor associated with the displacement  $\mathbf{s}$  is given by

$$\boldsymbol{\epsilon} = \frac{1}{2}[\nabla\mathbf{s} + (\nabla\mathbf{s})^T], \tag{2.1}$$

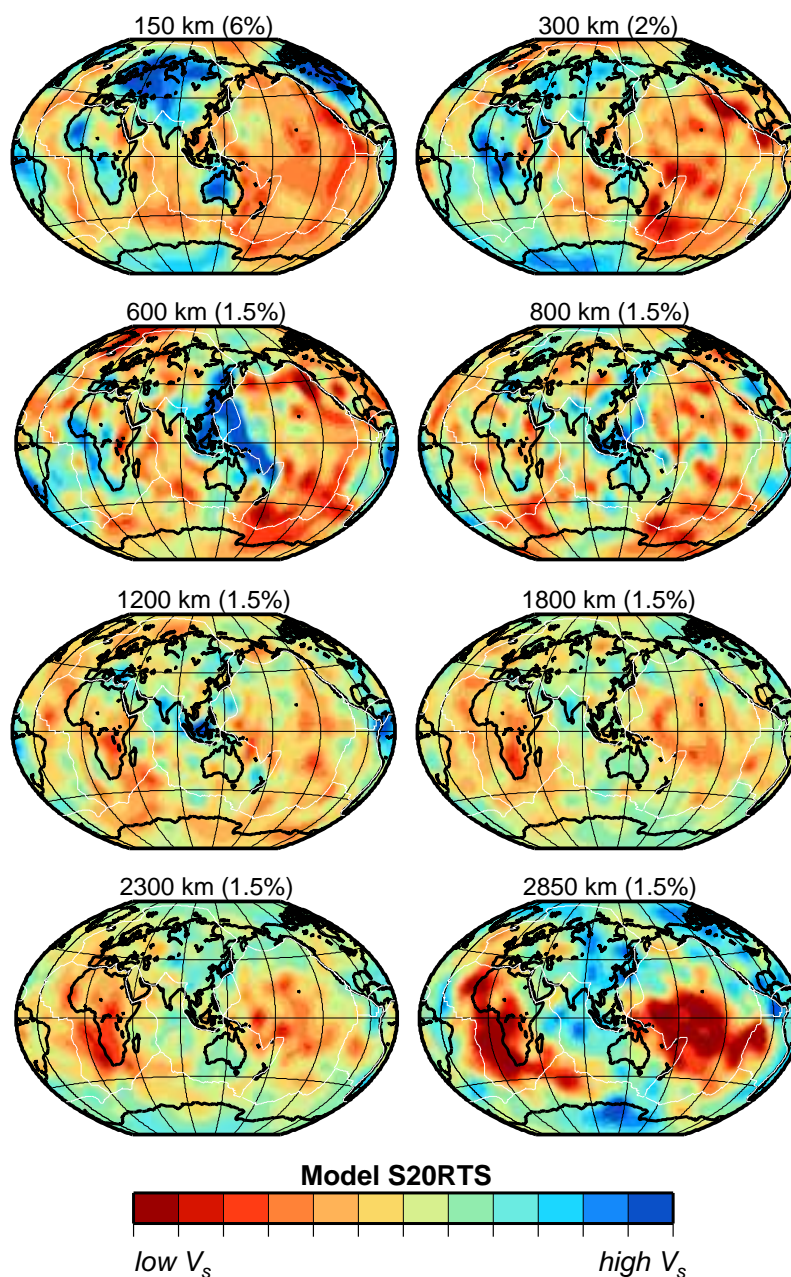


Figure 2: Map views of 3D mantle model S20RTS [52] at depths of 150 km, 300 km, 600 km, 1200 km, 1800 km, 2300 km, and 2850 km. The range in shear-wave speed perturbations is indicated in parentheses above each map. White lines represent plate boundaries and black lines continents. The shear-wave speed  $\beta$  ( $V_s$ ) in regions colored blue (red) is higher (lower) than in PREM (Fig. 1). Shear-wave speed variations in the upper 200 km of the mantle range from  $-8\%$  to  $8\%$ . Note that in the shallow mantle (i.e., at 200 km depth), high shear-wave speeds are associated with old (cold) cratons, whereas low shear-wave speeds correspond to mid-oceanic ridges, where hot material wells up. In the lowermost mantle (e.g., at 2850 km depth) there exists a distinct ring of fast shear-wave speeds around the Pacific, whereas underneath Africa and the Pacific one can clearly distinguish low shear-wave speed anomalies associated with large-scale upwellings (superplumes). *Courtesy of Jeroen Ritsema.*

where a superscript  $T$  denotes the transpose. The most general linear constitutive relationship between the stress tensor  $\mathbf{T}$  and the strain tensor (2.1) involves the fourth-order elastic tensor  $\mathbf{c}$ :

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

or in index notation  $T_{ij} = c_{ijkl} \epsilon_{kl}$ . Because both the stress and the strain tensor are symmetric, and due to certain thermodynamic considerations [3, 18], the elements of the elastic tensor exhibit the following symmetries:  $c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}$ . These symmetries reduce the number of independent elastic parameters from 81 to 21.

In an isotropic, elastic Earth model the elastic tensor may be expressed in terms of just two elastic parameters: the bulk modulus  $\kappa$  and the shear modulus  $\mu$ . In index notation we have

$$c_{ijkl} = \left(\kappa - \frac{2}{3}\mu\right) \delta_{jk} \delta_{lm} + \mu(\delta_{jl} \delta_{km} + \delta_{jm} \delta_{kl}), \quad (2.3)$$

and Hooke's law (2.2) reduces to

$$\mathbf{T} = \left(\kappa - \frac{2}{3}\mu\right) \text{tr}(\boldsymbol{\epsilon}) \mathbf{I} + 2\mu \boldsymbol{\epsilon}. \quad (2.4)$$

Here  $\mathbf{I}$  denotes the  $3 \times 3$  identity tensor. For mathematical and notational convenience, we will sometimes use the anisotropic constitutive relationship (2.2), rather than the more practical isotropic version (2.4). A simple isotropic Earth model supports basic shear and compressional waves, but wave propagation in an anisotropic medium is much richer, involving widely observed phenomena such as shear-wave splitting.

In the fluid regions of the Earth model, e.g., in the oceans and in the outer core, the shear modulus vanishes, and Hooke's law (2.4) reduces to

$$\mathbf{T} = \kappa \text{tr}(\boldsymbol{\epsilon}) \mathbf{I}, \quad (2.5)$$

which implies that the pressure associated with the fluid motions is given by  $-\kappa \text{tr}(\boldsymbol{\epsilon})$ .

The Earth is not a perfectly elastic body, and effects due to attenuation should be incorporated. In an anelastic medium, the stress  $\mathbf{T}$  at time  $t$  is determined by the entire strain history  $\boldsymbol{\epsilon}(t')$ . In its most general anisotropic form Hooke's law becomes [3, 18]:

$$\mathbf{T}(t) = \int_{-\infty}^t \partial_t \mathbf{c}(t-t') : \boldsymbol{\epsilon}(t') dt'. \quad (2.6)$$

In this article we will avoid the mathematical and numerical complications associated with attenuation, but suffice it to say that these effects can be readily accommodated in most numerical simulations.

### 2.3 Equation of motion

Having introduced various possibilities for the constitutive relationship relating stress and strain, let us introduce the seismic equation of motion. Consider an Earth model

with volume  $\Omega$  and outer free surface  $\partial\Omega$ . The displacement wavefield  $\mathbf{s}(\mathbf{x}, t)$ , where  $\mathbf{x}$  denotes material points in the Earth model and  $t$  time, is determined by the seismic wave equation [3, 18]

$$\rho \partial_t^2 \mathbf{s} - \nabla \cdot \mathbf{T} = \mathbf{f}, \quad (2.7)$$

where  $\rho$  denotes the distribution of mass density, and the stress  $\mathbf{T}$  is determined in terms of the strain through one of the constitutive relationships discussed in the previous section.

On the Earth's free surface  $\partial\Omega$  the traction must vanish:

$$\hat{\mathbf{n}} \cdot \mathbf{T} = \mathbf{0} \quad \text{on } \partial\Omega, \quad (2.8)$$

where  $\hat{\mathbf{n}}$  denotes the unit outward normal on the surface. On solid-solid boundaries, such as the Moho or upper mantle discontinuities, both the traction  $\hat{\mathbf{n}} \cdot \mathbf{T}$  and the displacement  $\mathbf{s}$  need to be continuous, whereas on fluid-solid boundaries, such as the ocean floor, the CMB and the ICB, both the traction  $\hat{\mathbf{n}} \cdot \mathbf{T}$  and the normal component of displacement  $\hat{\mathbf{n}} \cdot \mathbf{s}$  need to be continuous.

When our modeling domain is not the entire Earth, seismic energy needs to be absorbed on the fictitious boundaries of the domain. To accomplish this one usually uses a paraxial equation to damp the wavefield on the edges [14, 51]. In recent years, a significantly more efficient absorbing condition called the Perfectly Matched Layer (PML) has been introduced [6], which is now being used in regional numerical simulations [5, 16, 22, 32]. In this article we will largely ignore the mathematical and numerical complications associated with absorbing boundary conditions.

In addition to the boundary condition (2.8), the seismic wave equation (2.7) must be solved subject to the initial conditions

$$\mathbf{s}(\mathbf{x}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}(\mathbf{x}, 0) = \mathbf{0}. \quad (2.9)$$

Finally, the force  $\mathbf{f}$  in (2.7) represents the earthquake. In the case of a simple point source it may be written in terms of the *moment tensor*  $\mathbf{M}$  as [3, 18]

$$\mathbf{f} = -\mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) S(t), \quad (2.10)$$

where the location of the point source is denoted by  $\mathbf{x}_s$ ,  $\delta(\mathbf{x} - \mathbf{x}_s)$  denotes the Dirac delta distribution located at  $\mathbf{x}_s$ , and  $S(t)$  denotes the source-time function. Mathematically, an earthquake may be represented in terms of a *double-couple* source, a representation that leads to the introduction of the moment tensor and the point source representation (2.10).

At periods longer than about 150 s, self-gravitation and rotation start to play a role in global seismic wave propagation. The equation of motion for a rotating, self-gravitating Earth model is significantly more complicated than (2.7) [18]. Nevertheless, numerical simulations frequently do take these complications into account. In this article we will not incorporate these long-period complications for the sake of simplicity.

For 1D models, i.e., models that vary only as a function of depth, semi-analytical techniques have been developed to calculate the wavefield generated by a point source



[3, 18, 27]. Such calculations are still widely used, and provide an excellent benchmark for more ambitious 3D numerical simulations.

## 2.4 Weak formulation

For computational purposes, one frequently works with an integral or *weak* formulation of the problem. It is obtained by taking the dot product of the momentum equation (2.7) — the *strong* form of the equation of motion — with an arbitrary test vector  $\mathbf{w}$ , and integrating by parts over the volume  $\Omega$  of the Earth:

$$\int_{\Omega} \rho \mathbf{w} \cdot \partial_t^2 \mathbf{s} d^3 \mathbf{x} = - \int_{\Omega} \nabla \mathbf{w} : \mathbf{T} d^3 \mathbf{x} + \mathbf{M} : \nabla \mathbf{w}(\mathbf{x}_s) S(t). \quad (2.11)$$

Eq. (2.11) is equivalent to the strong formulation (2.7) because it holds for any test vector  $\mathbf{w}$ . The term on the left hand side gives rise to the *mass matrix* in finite-element parlance, and the first term on the right is related to the *stiffness matrix*. The second term on the right is related to the source term (2.10), which has been integrated explicitly by using the properties of the Dirac delta distribution. Note that only first-order spatial derivatives of the displacement field and the test vector are involved in the weak form (2.11), but that the temporal derivatives are of second order.

It is important to appreciate that the traction-free surface condition (2.8) is imposed naturally and automatically during the integration by parts, because the contour integral over the free surface simply vanishes. In other words, the free-surface condition is a natural condition of the problem. In the context of regional simulations an additional term appears in (2.11), which represents the absorption of energy on the artificial boundaries of the domain [29].

In an Earth model with fluid and solid regions, one uses a domain decomposition approach in which one solves separate wave equations in the fluid and solid regions, which are coupled at the fluid-solid boundaries by imposing continuity of traction and the normal component of displacement [30]. These complications are beyond the scope of this article.

## 3 Spectral-element method

In this section we briefly outline how the spectral-element technique may be used to numerically solve the equations that govern the propagation of seismic waves. This summary is largely based upon the introductory article by [29].

### 3.1 Meshing

We begin by subdividing the model volume  $\Omega$  into a number of non-overlapping elements  $\Omega_e$ ,  $e = 1, \dots, n_e$ , such that  $\Omega = \cup_e^{n_e} \Omega_e$ . In a finite-element method (FEM) [25, 72] a variety of elements, e.g., tetrahedra, hexahedra, pyramids or prisms, may be used, but a

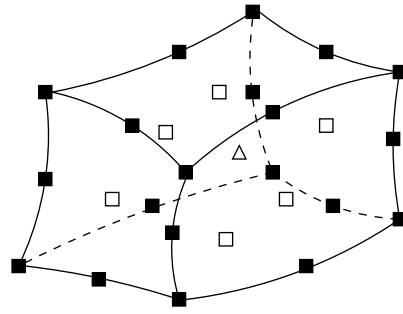


Figure 3: The geometry of a hexahedral finite element (deformed cube) may be defined in terms of its eight corners, the eight corners plus the 12 edge centers (the 20 black squares), or the eight corners plus the 12 edge centers plus the 6 face centers (the 6 open squares) plus the center (the open triangle). In a classical FEM, these 8, 20, or 27 anchors are used to define the shape of the element as well as for the interpolation of functions. In a SEM, the anchors are only used to define the geometry of the elements, but not for the interpolation and integration of functions. Instead, functions are represented in terms of high-degree Lagrange polynomials on Gauss-Lobatto-Legendre interpolation points, as illustrated in Fig. 4. *Courtesy of [29].*

3D SEM is generally restricted to hexahedral elements (deformed cubes). The SEM has been extended to include triangles in two dimensions [34, 44, 57, 65], but this leads to complications that are beyond the scope of this article.

Points  $\mathbf{x} = (x, y, z)$  within each hexahedral element  $\Omega_e$  may be uniquely related to points  $\xi = (\xi, \eta, \zeta)$ ,  $-1 \leq \xi, \eta, \zeta \leq 1$ , in a reference cube based upon the invertible mapping

$$\mathbf{x}(\xi) = \sum_{a=1}^M \mathbf{x}_a N_a(\xi). \quad (3.1)$$

The  $a = 1, \dots, M$  anchors  $\mathbf{x}_a = \mathbf{x}(\xi_a, \eta_a, \zeta_a)$  and shape functions  $N_a(\xi)$  define the geometry of an element  $\Omega_e$ . For example, the geometry of hexahedral elements may be controlled by  $M = 8, 20$ , or 27 anchors, as illustrated in Fig. 3.

Hexahedral shape functions  $N_a(\xi)$  are typically products of degree 1 or 2 Lagrange polynomials. In general, the  $n+1$  Lagrange polynomials of degree  $n$  are defined in terms of  $n+1$  control points  $-1 \leq \xi_\alpha \leq 1$ ,  $\alpha = 0, \dots, n$ , by

$$h_\alpha(\xi) = \frac{(\xi - \xi_0) \cdots (\xi - \xi_{\alpha-1})(\xi - \xi_{\alpha+1}) \cdots (\xi - \xi_n)}{(\xi_\alpha - \xi_0) \cdots (\xi_\alpha - \xi_{\alpha-1})(\xi_\alpha - \xi_{\alpha+1}) \cdots (\xi_\alpha - \xi_n)}. \quad (3.2)$$

Note that as a result of this definition, the Lagrange polynomials return either zero or one at any given control point:

$$h_\alpha(\xi_\beta) = \delta_{\alpha\beta}, \quad (3.3)$$

where  $\delta$  denotes the Kronecker delta. Fig. 4 illustrates these characteristics for degree 4 Lagrange polynomials. In the context of shape functions, the two Lagrange polynomials of degree 1 with two control points  $\xi = -1$  and  $\xi = 1$  are

$$h_0(\xi) = \frac{1}{2}(1 - \xi), \quad h_1(\xi) = \frac{1}{2}(1 + \xi),$$



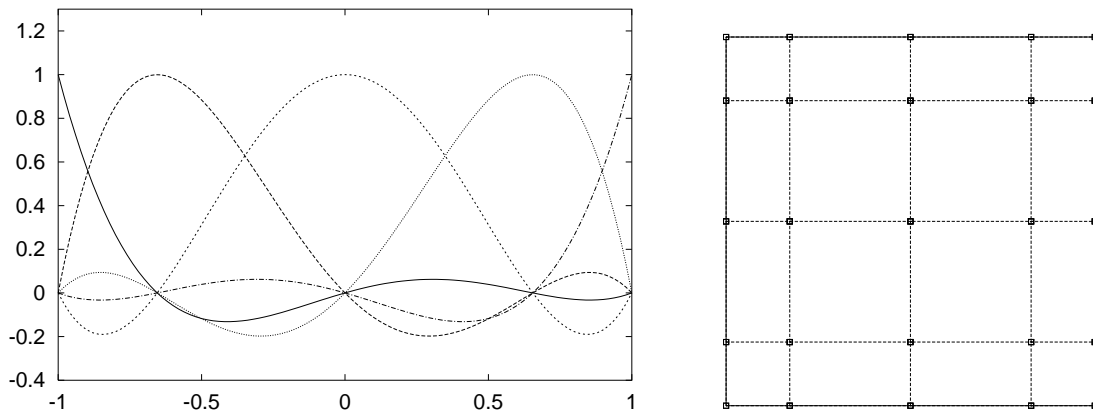


Figure 4: Left: The five Lagrange interpolants of degree  $N=4$  on the reference segment  $[-1,1]$ . The  $N+1=5$  Gauss-Lobatto-Legendre (GLL) points, determined by eqn. (3.7), can be distinguished along the horizontal axis. Note that the first and last GLL points are exactly -1 and 1. All Lagrange polynomials are, by definition, equal to 1 or 0 at each of the GLL points, in accordance with eqn. (3.3). In a FEM, shape functions  $N_a, j=1, \dots, M$ , are typically triple products of degree 1 or 2 Lagrange polynomials, and functions are interpolated on a hexahedral element, such as the one shown in Fig. 3, in terms of these low-degree polynomials. In a SEM, the geometry of elements is captured by the same low-degree shape functions, but functions are represented in terms of triple products of high-degree Lagrange polynomials (typically of degree 4–10), one for each direction in the reference cube. Right: When Lagrange polynomials of degree  $n$  are used to discretize functions on an element, each 3D spectral element contains a grid of  $(n+1)^3$  GLL points, and each 2D face of an element contains a grid of  $(n+1)^2$  GLL points, as illustrated here for the degree 4 polynomials shown on the left. *Courtesy of [38].*

and the three Lagrange polynomials of degree 2 with three control points  $\zeta = -1, \zeta = 0$ , and  $\zeta = 1$  are

$$h_0(\zeta) = \frac{1}{2}\zeta(\zeta - 1), \quad h_1(\zeta) = 1 - \zeta^2, \quad h_2(\zeta) = \frac{1}{2}\zeta(\zeta + 1).$$

The weak form (2.11) involves volume integrals over elements  $\Omega_e$ . Using the mapping (3.1), an element of volume  $d^3\mathbf{x} = dx dy dz$  within a given element  $\Omega_e$  is related to an element of volume  $d^2\zeta = d\zeta d\eta d\zeta$  in the reference cube by

$$d^3\mathbf{x} = dx dy dz = J d\zeta d\eta d\zeta = J d^3\zeta, \tag{3.4}$$

where the Jacobian  $J$  of the mapping (3.1) is given by

$$J = \left| \frac{\partial(x, y, z)}{\partial(\zeta, \eta, \zeta)} \right|. \tag{3.5}$$

To calculate the Jacobian  $J$ , we need the partial derivative matrix  $\partial\mathbf{x}/\partial\zeta$ , which is obtained by differentiating (3.1):

$$\frac{\partial\mathbf{x}}{\partial\zeta} = \sum_{a=1}^M \mathbf{x}_a \frac{\partial N_a}{\partial\zeta}. \tag{3.6}$$

From (3.6) we conclude that partial derivatives of the shape functions are determined analytically in terms of Lagrange polynomials of degree 1 or 2 and their derivatives. The

elements  $\Omega_e$  should be constructed in such a way that the Jacobian  $J$  never vanishes, which poses strong constraints on the mesh generation process. This ensures that the mapping from the reference cube to the element,  $\mathbf{x}(\boldsymbol{\xi})$ , is unique and invertible, i.e.,  $\boldsymbol{\xi}(\mathbf{x})$  is well-defined.

### 3.2 Representation and integration of functions on elements

To solve the weak form of the seismic wave equation (2.11), integrations over the volume  $\Omega$  are subdivided in terms of smaller integrals over the hexahedral elements  $\Omega_e$ . This section is concerned with the representation and integration of functions on such elements.

We have seen in the previous section that the shape of the elements can be defined in terms of low-degree Lagrange polynomials. In a traditional FEM, low-degree polynomials are also used as basis functions for the representation of functions on an element. In a SEM, a higher-degree Lagrange interpolant is used to express functions on the elements, and the control points  $\xi_\alpha$ ,  $\alpha = 0, \dots, n$ , needed in the definition of the Lagrange polynomials of degree  $n$  (3.2) are chosen to be the  $n+1$  Gauss-Lobatto-Legendre (GLL) points, which are the roots of the equation [8]

$$(1 - \xi^2)P'_n(\xi) = 0, \quad (3.7)$$

where  $P'_n$  denotes the derivative of the Legendre polynomial of degree  $n$ . Eq. (3.7) implies that  $+1$  and  $-1$  are always GLL points, regardless of the degree  $n$ . Therefore, some GLL points always lie exactly on the boundaries of the elements.

A SEM typically uses Lagrange polynomials (3.2) of degree 4–10 for the interpolation of functions. As an example, the 5 Lagrange polynomials of degree 4 are shown in Figs. 4 (Left), and 4 (Right) illustrates the distribution of the associated GLL points on the face of a hexahedral element. As we will see, the combination of Lagrange interpolants with a particular integration rule leads to an exactly diagonal mass matrix, which in turn leads to a simple explicit time scheme that lends itself very well to numerical simulations on parallel computers.

#### 3.2.1 Polynomial representation on elements

In the weak form of the wave equation (2.11), we expand functions  $f$ , e.g., a component of the displacement field  $\mathbf{s}$  or the test vector  $\mathbf{w}$ , in terms of degree- $n$  Lagrange polynomials (3.2) with GLL control points determined by (3.7):

$$f(\mathbf{x}(\xi, \eta, \zeta)) = \sum_{\alpha=0}^n \sum_{\beta=0}^n \sum_{\gamma=0}^n f^{\alpha\beta\gamma} h_\alpha(\xi) h_\beta(\eta) h_\gamma(\zeta), \quad (3.8)$$

where  $f^{\alpha\beta\gamma} = f(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma))$  denotes the value of the function  $f$  at the GLL point  $\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma)$ . In a SEM, polynomials of degree 4 or 5 provide the best trade off between accuracy and time-integration stability [56].

The weak form (2.11) involves gradients of the displacement field  $\mathbf{s}$  and the test vector  $\mathbf{w}$ . Using the polynomial representation (3.8), such gradients may be expressed as

$$\begin{aligned} \nabla f(\mathbf{x}(\xi, \eta, \zeta)) = & \sum_{i=1}^3 \hat{\mathbf{x}}_i \sum_{\alpha=0}^n \sum_{\beta=0}^n \sum_{\gamma=0}^n f^{\alpha\beta\gamma} [h'_\alpha(\xi)h_\beta(\eta)h_\gamma(\zeta)\partial_i\xi \\ & + h_\alpha(\xi)h'_\beta(\eta)h_\gamma(\zeta)\partial_i\eta + h_\alpha(\xi)h_\beta(\eta)h'_\gamma(\zeta)\partial_i\zeta], \end{aligned} \quad (3.9)$$

where a prime denotes the derivative of a Lagrange polynomial, and where we have used the index notation  $\partial_i = \partial_{x_i}$ ,  $i = 1, 2, 3$ , and  $x_1 = x$ ,  $x_2 = y$ , and  $x_3 = z$ . The matrix  $\partial\xi/\partial\mathbf{x}$  is obtained by inverting the matrix  $\partial\mathbf{x}/\partial\xi$ . Because in a SEM one uses higher-degree polynomials for interpolation, the derivatives calculated based upon (3.9) tend to be more accurate than those used in low-order finite-element or finite-difference methods.

### 3.2.2 Integration over elements

The next step is to evaluate the integrals in the weak form (2.11) at the elemental level. In a SEM, a Gauss-Lobatto-Legendre integration rule is used for this purpose, because, as we will see, this leads to a diagonal mass matrix when used in conjunction with GLL interpolation points. Based upon this approach, integrations over elements  $\Omega_e$  may be approximated as

$$\begin{aligned} \int_{\Omega_e} f(\mathbf{x}) d^3\mathbf{x} = & \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 f(\mathbf{x}(\xi, \eta, \zeta)) J(\xi, \eta, \zeta) d\xi d\eta d\zeta \\ = & \sum_{\alpha=0}^n \sum_{\beta=0}^n \sum_{\gamma=0}^n \omega_\alpha \omega_\beta \omega_\gamma f^{\alpha\beta\gamma} J^{\alpha\beta\gamma}, \end{aligned} \quad (3.10)$$

where  $J^{\alpha\beta\gamma} = J(\xi_\alpha, \eta_\beta, \zeta_\gamma)$  denotes the value of the Jacobian  $J$  of the mapping at a GLL point, and  $\omega_\alpha$ ,  $\alpha = 0, \dots, n$ , denote the  $n+1$  quadrature weights associated with the GLL integration. To facilitate the integration of functions and their partial derivatives over the elements, the values of the inverse Jacobian matrix  $\partial\xi/\partial\mathbf{x}$  need to be stored at the  $(n+1)^3$  GLL integration points.

### 3.3 Assembly

We have seen that during the SEM meshing process the model is subdivided in terms of non-overlapping hexahedral elements. Functions on the elements are sampled at the GLL points of integration. As illustrated in Fig. 5, GLL points on the sides, edges, or corners of an element are shared amongst neighboring elements. Because neighboring elements share GLL points, the need arises to distinguish between points that define an element, the *local mesh*, and the collective points in the model, the *global mesh*. Therefore, one needs to determine a mapping between GLL points in the local mesh and grid points in the global mesh. Fortunately, efficient finite-element routines are available for this

purpose. The contributions from all the elements that share a common global grid point need to be summed during each time step. In a traditional FEM this is referred to as the *assembly* of the system. On parallel computers assembly is an expensive part of the calculation, because information from each element needs to be shared with neighboring elements, an operation that involves communication between distinct CPUs.

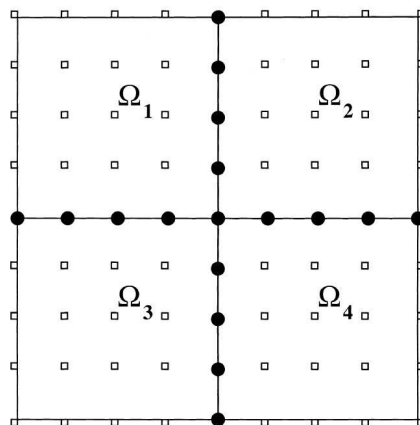


Figure 5: Illustration of the local and global meshes for a four-element 2D spectral-element discretization with polynomial degree  $N=4$ . Each spectral element contains  $(N+1)^2=25$  GLL points that constitute the local mesh for each element. These points are non-evenly spaced, but have been drawn evenly spaced here for simplicity. In the global mesh, points lying on edges or corners (as well as on faces in 3D) are shared between elements. The contributions to the global system of degrees of freedom, computed separately on each element, have to be summed at these common points represented by black dots. Exactly two elements share points on an edge in 2D, while corners can be shared by any number of elements depending on the topology of the mesh, which may be non-structured. *Courtesy of [29].*

### 3.4 Meshing the globe

Fig. 6 shows an example of a conforming, unstructured hexahedral SEM mesh for the entire globe [30]. The mesh is based upon an analytical mapping between the cube and the sphere called the *cubed sphere* [10, 53, 54]. Each of the six chunks that constitute the cubed sphere is meshed in such a way that it matches perfectly with its neighbors. The mesh is doubled in size once below the Moho, a second time below the 670 km discontinuity (see the close-up in Fig. 6), and a third time just above the ICB. Each of the six chunks has  $240 \times 240$  elements at the free surface and, as a result of the three doublings,  $30 \times 30$  elements at the ICB. The coordinate singularity at the Earth's center is avoided by placing a small cube at the center of the inner core [10]. The mesh in Fig. 6 honors all discontinuities in PREM [20] (see Fig. 1), and contains a total of approximately 2.6 million spectral elements. In each spectral element we use a polynomial degree  $N=4$  for the expansion of the Lagrange interpolants at the GLL points, i.e., each element contains  $(N+1)^3=125$  local points, and the global mesh contains 180 million points.

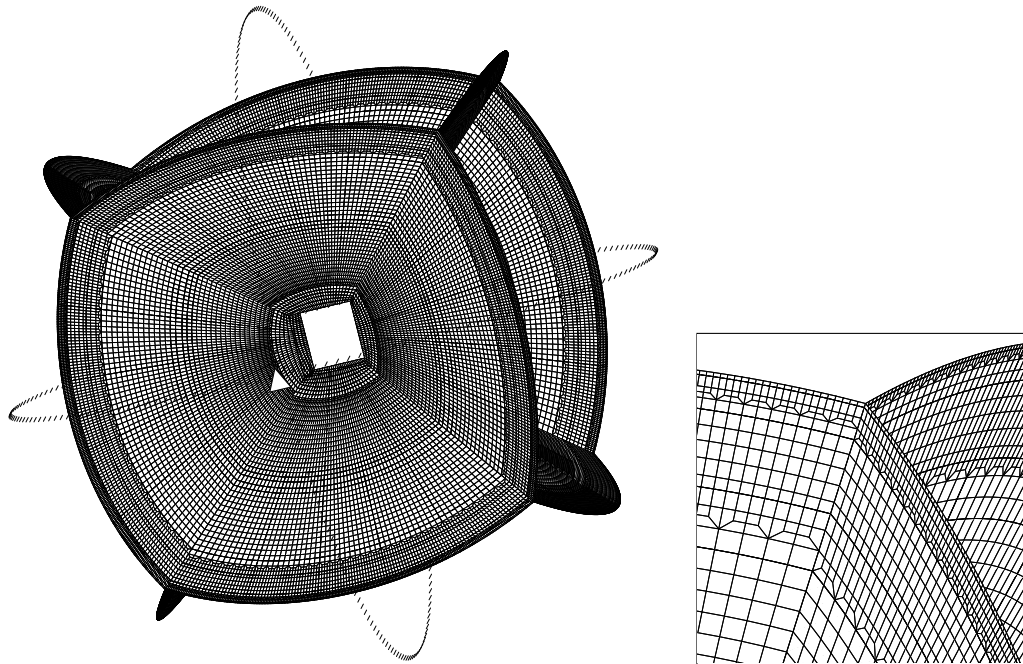


Figure 6: Left: Example of a spectral-element mesh in the mantle. Right: Close-up of the mesh doublings in the upper mantle. The spectral-element mesh is conforming, i.e., every side of every element matches up perfectly with the side of a neighboring element, but non-structured, i.e., the number of elements that share a given point can vary and take any value. The mesh honors first-order discontinuities in PREM at depths of 24.4 km, 220 km, 400 km, and 670 km, the CMB, and the ICB; it also honors second-order discontinuities in PREM at 600 km, 771 km, and at the top of  $D''$ . The mesh is doubled in size once below the Moho, a second time below the 670 km discontinuity, and a third time just above the ICB. Each of the six chunks that constitute the cubed sphere has  $240 \times 240$  elements at the free surface and  $30 \times 30$  elements at the ICB. The central cube in the inner core has been removed for clarity of viewing. The triangle indicates the location of the source, situated on the equator and the Greenwich meridian. Rings of receivers with a  $2^\circ$  spacing along the equator and the Greenwich meridian are shown by the dashes. *Courtesy of [30].*

Any typical 3D crustal and/or mantle model, e.g., S20RTS shown in Fig. 2, may now be superimposed on the mesh. Surface topography & bathymetry and the Earth's ellipticity may be accommodated by stretching or squishing the mesh. The largest global simulation to date was performed on the Earth Simulator at the Japan Marine Science & Technology Center (JAMSTEC) in Yokohama, Japan. It involved 4056 processors, 13.7 billion global grid points, and required 7 terabytes of memory [68].

### 3.5 Mesh partitioning and load-balancing

The global mesh in Fig. 6 is too large to fit in memory on a single computer. We therefore implement the SEM on parallel computers by partitioning the mesh into slices of elements, such that each processor in the parallel machine is only responsible for the elements in one particular slice. Each of the six chunks is divided in  $N \times N$  slices, for a total of



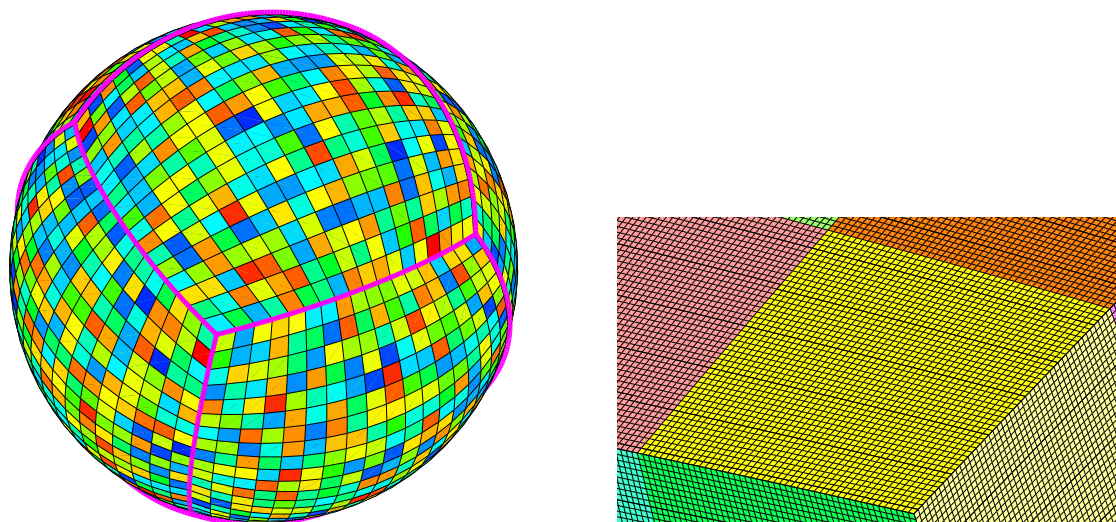


Figure 7: The spectral-element method uses a mesh of hexahedral finite elements on which the wavefield is interpolated by high-degree Lagrange polynomials on Gauss-Lobatto-Legendre (GLL) integration points. In order to perform the calculations on a parallel computer with distributed memory, the mesh is split into slices based upon a regular domain-decomposition topology. Each slice is handled by a separate processor. Adjacent slices located on different processors exchange information about common faces and edges based upon a message-passing methodology. The figure on the left shows a global view of the mesh at the surface, illustrating that each of the six sides of the cubed sphere mesh is divided into  $18 \times 18$  slices, shown here with different colors, for a total of 1944 slices. The elements within each slice reside in memory on a single processor of the parallel machine. The figure on the right shows a close-up of the mesh of  $48 \times 48$  spectral elements at the surface of each slice. Within each surface spectral element we use  $5 \times 5 = 25$  GLL grid points, which translates into an average grid spacing of 2.9 km (i.e.,  $0.026^\circ$ ) on the surface of the Earth. *Courtesy of [36].*

$6N^2$  processors. The *de facto* standard for distributed programming is to use a message-passing programming methodology based upon the ‘Message Passing Interface’ (MPI) protocol [24, 47]. Fig. 7 illustrates how a SEM mesh may be split into  $6 \times 18^2 = 1944$  slices for a parallel calculation on 1944 processors. The key to this distribution process is to end up with a mesh partitioning that is *load-balanced*, such that every MPI slice contains approximately the same number of spectral-elements, and every processor involved in the calculation performs roughly the same number of operations per time step.

At the edges of a slice results need to be communicated to its neighbors. Therefore, on a parallel computer assembly involves communication between adjacent mesh slices. The real benefits of the SEM become abundantly clear on large parallel machine: the diagonal mass matrix leads to fully explicit time-marching schemes, which implies that the compute nodes spend most of their time performing computations, and communication of results between nodes represents only a small fraction of the simulation time. For this reason the application has scaled very well with the number of processors, which ranges from a few tens of CPUs on a typical PC cluster to several thousands of CPUs on machines like the Earth Simulator and BlueGene.



### 3.6 Application of the SEM in seismology

At this point we have collected all the necessary ingredients to solve the weak form of the seismic wave equation (2.11) based upon the SEM. To accomplish this, we begin by calculating the SEM *mass matrix*, which is obtained from the integral on the left hand side of (2.11). First, we expand the displacement field  $\mathbf{s}$  and the test vector  $\mathbf{w}$  in terms of Lagrange polynomials:

$$\mathbf{s}(\mathbf{x}(\xi, \eta, \zeta), t) = \sum_{i=1}^3 \hat{\mathbf{x}}_i \sum_{\sigma=0}^n \sum_{\tau=0}^n \sum_{\nu=0}^n s_i^{\sigma\tau\nu}(t) h_\sigma(\xi) h_\tau(\eta) h_\nu(\zeta), \tag{3.11}$$

$$\mathbf{w}(\mathbf{x}(\xi, \eta, \zeta)) = \sum_{i=1}^3 \hat{\mathbf{x}}_i \sum_{\sigma=0}^n \sum_{\tau=0}^n \sum_{\nu=0}^n w_i^{\alpha\beta\gamma} h_\alpha(\xi) h_\beta(\eta) h_\gamma(\zeta). \tag{3.12}$$

Eqs. (3.11) and (3.12) imply that the SEM is a Galerkin method, because the displacement and test vectors are expanded in terms of the same basis function. Next, we substitute the interpolations (3.11) and (3.12) in the integral on the left hand side of (2.11), using GLL quadrature (3.10), to obtain at the elemental level

$$\begin{aligned} \int_{\Omega_e} \rho \mathbf{w} \cdot \partial_t^2 \mathbf{s} d^3 \mathbf{x} &= \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \rho(\mathbf{x}(\xi)) \mathbf{w}(\mathbf{x}(\xi)) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}(\xi), t) J(\xi) d^3 \xi \\ &= \sum_{\alpha=0}^n \sum_{\beta=0}^n \sum_{\gamma=0}^n \omega_\alpha \omega_\beta \omega_\gamma J^{\alpha\beta\gamma} \rho^{\alpha\beta\gamma} \sum_{i=1}^3 w_i^{\alpha\beta\gamma} \dot{s}_i^{\alpha\beta\gamma}, \end{aligned} \tag{3.13}$$

where a dot denotes differentiation with respect to time, and  $\rho^{\alpha\beta\gamma} = \rho(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma))$ . Note that the density may vary across the element. By independently setting factors of  $w_1^{\alpha\beta\gamma}$ ,  $w_2^{\alpha\beta\gamma}$ , and  $w_3^{\alpha\beta\gamma}$  equal to zero, since the weak formulation (2.11) must hold for any test vector  $\mathbf{w}$ , we obtain independent equations for each component of acceleration  $\dot{s}_i^{\alpha\beta\gamma}(t)$  at grid point  $(\xi_\alpha, \eta_\beta, \zeta_\gamma)$ . The value of acceleration at each point of a given element,  $\dot{s}_i^{\alpha\beta\gamma}(t)$ , is simply multiplied by the factor  $\omega_\alpha \omega_\beta \omega_\gamma \rho^{\alpha\beta\gamma} J^{\alpha\beta\gamma}$ , which means that the elemental mass matrix is exactly diagonal. This is one of the principal ideas behind the SEM, and the main motivation for the choice of Lagrange interpolation at the GLL points used in conjunction with GLL numerical quadrature.

The main differences between finite-element and spectral-element methods are the polynomial degree of the basis functions, the choice of integration rule, and the nature of the time-marching scheme. In a FEM one tends to use low-degree basis functions and Gauss quadrature. In a SEM one uses higher-degree basis functions and GLL quadrature to obtain better resolution as well as a diagonal mass matrix. For the SEM this leads to simple explicit time schemes, as opposed to the numerically more expensive implicit time schemes used in FEMs.

It is important to note that even under ideal circumstances the GLL quadrature rule is exact only for integrands which are polynomials of degree  $2n - 1$ . Since integration in the

SEM involves the product of two polynomials of degree  $n$  —the displacement and the test function— the integration of the resulting polynomial of degree  $2n$  is never exact. As in a FEM, for deformed elements there are additional errors related to curvature [42], and the same is true for elements with heterogeneous material properties. Thus, in a SEM a diagonal mass matrix is obtained by a process of subintegration. In this respect the SEM is related to FEMs in which mass lumping is used to avoid the costly resolution of the non-diagonal system resulting from the use of Gauss quadrature [15].

To determine the SEM *stiffness matrix*, we need to evaluate the first integral on the right hand side of (2.11). The first step is to calculate the nine elements of the displacement gradient  $\nabla \mathbf{s}$  on the element  $\Omega_e$ . We have, using index notation,

$$\begin{aligned} \partial_i s_j(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma), t) = & \left[ \sum_{\sigma=0}^n s_j^{\sigma\beta\gamma}(t) h'_\sigma(\xi_\alpha) \right] \partial_i \xi(\xi_\alpha, \eta_\beta, \zeta_\gamma) + \left[ \sum_{\sigma=0}^n s_j^{\alpha\sigma\gamma}(t) h'_\sigma(\eta_\beta) \right] \partial_i \eta(\xi_\alpha, \eta_\beta, \zeta_\gamma) \\ & + \left[ \sum_{\sigma=0}^n s_j^{\alpha\beta\sigma}(t) h'_\sigma(\zeta_\gamma) \right] \partial_i \zeta(\xi_\alpha, \eta_\beta, \zeta_\gamma). \end{aligned} \quad (3.14)$$

This calculation requires knowledge of the nine elements of the inverse Jacobian matrix  $\partial \xi / \partial \mathbf{x}$ . Next, one calculates the six elements of the symmetric stress tensor  $\mathbf{T}$  on the element:

$$\mathbf{T}(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma), t) = \mathbf{c}(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma)) : \nabla \mathbf{s}(\mathbf{x}(\xi_\alpha, \eta_\beta, \zeta_\gamma), t). \quad (3.15)$$

This requires knowledge of the previously calculated displacement gradient (3.14) and of the elastic tensor  $\mathbf{c}$  at the GLL integration points. The formulation is not limited to isotropic media or to anisotropic media with a high degree of symmetry, as is frequently the case for other numerical methods. Furthermore, the Earth model may be fully heterogeneous, i.e.,  $\rho$  and  $\mathbf{c}$  need not be constant inside an element. The integrand  $\nabla \mathbf{w} : \mathbf{T}$  in the stiffness term in (2.11) may be written in the form

$$\nabla \mathbf{w} : \mathbf{T} = \sum_{i,j=1}^3 T_{ij} \partial_j w_i = \sum_{i,k=1}^3 \left( \sum_{j=1}^3 T_{ij} \partial_j \xi_k \right) \frac{\partial w_i}{\partial \xi_k} = \sum_{i,k=1}^3 F_{ik} \frac{\partial w_i}{\partial \xi_k}, \quad (3.16)$$

where

$$F_{ik} = \sum_{j=1}^3 T_{ij} \partial_j \xi_k. \quad (3.17)$$

The next step is to calculate the nine matrix elements  $F_{ik}$  at the GLL integration points:  $F_{ik}^{\sigma\tau\nu} = F_{ik}(\mathbf{x}(\xi_\sigma, \eta_\tau, \zeta_\nu))$ ; this requires knowledge of the stress tensor  $\mathbf{T}$  computed in (3.15) and of the inverse Jacobian matrix  $\partial \xi / \partial \mathbf{x}$ . The stiffness term in (2.11) may now be rewritten at the elemental level as

$$\int_{\Omega_e} \nabla \mathbf{w} : \mathbf{T} d^3 \mathbf{x} = \sum_{i,k=1}^3 \int_{\Omega_e} F_{ik} \frac{\partial w_i}{\partial \xi_k} d^3 \mathbf{x} = \sum_{i,k=1}^3 \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 F_{ik} \frac{\partial w_i}{\partial \xi_k} J_e d\xi d\eta d\zeta. \quad (3.18)$$

Upon substituting the test vector (3.12) in (3.18) and using the GLL integration rule, we find that the elemental SEM stiffness matrix is determined by

$$\int_{\Omega_e} \nabla \mathbf{w} : \mathbf{T} d^3 \mathbf{x} = \sum_{\alpha, \beta, \gamma=0}^n w_i^{\alpha\beta\gamma} \left[ \omega_\beta \omega_\gamma \sum_{\alpha'=0}^n \omega_{\alpha'} J_e^{\alpha'\beta\gamma} F_{i1}^{\alpha'\beta\gamma} h'_\alpha(\xi_{\alpha'}) + \omega_\alpha \omega_\gamma \sum_{\beta'=0}^n \omega_{\beta'} J_e^{\alpha\beta'\gamma} F_{i2}^{\alpha\beta'\gamma} h'_\beta(\eta_{\beta'}) + \omega_\alpha \omega_\beta \sum_{\gamma'=0}^n \omega_{\gamma'} J_e^{\alpha\beta\gamma'} F_{i3}^{\alpha\beta\gamma'} h'_\gamma(\zeta_{\gamma'}) \right]. \quad (3.19)$$

Finally, to complete the SEM implementation of the weak form we need to evaluate the second term on the right hand side of (2.11). This source term may be expressed as

$$\mathbf{M} : \nabla \mathbf{w} = \sum_{i,j=1}^3 M_{ij} \partial_j w_i = \sum_{i,k=1}^3 \left( \sum_{j=1}^3 M_{ij} \partial_j \xi_k \right) \frac{\partial w_i}{\partial \xi_k} = \sum_{i,k=1}^3 G_{ik} \frac{\partial w_i}{\partial \xi_k}, \quad (3.20)$$

where

$$G_{ik} = \sum_{j=1}^3 M_{ij} \partial_j \xi_k. \quad (3.21)$$

Upon defining  $G_{ik}^{\sigma\tau\nu} = G_{ik}(\mathbf{x}(\xi_\sigma, \eta_\tau, \zeta_\nu))$  and using the test vector (3.12), we obtain

$$\begin{aligned} \mathbf{M} : \nabla \mathbf{w}(\mathbf{x}_s) &= \sum_{\alpha, \beta, \gamma=0}^n w_i^{\alpha\beta\gamma} \left\{ \sum_{\sigma, \tau, \nu=0}^n h_\sigma(\xi_{\alpha_s}) h_\tau(\eta_{\beta_s}) h_\nu(\zeta_{\gamma_s}) \right. \\ &\quad \times \left[ G_{i1}^{\sigma\tau\nu} h'_\alpha(\xi_{\alpha_s}) h_\beta(\eta_{\beta_s}) h_\gamma(\zeta_{\gamma_s}) + G_{i2}^{\sigma\tau\nu} h_\alpha(\xi_{\alpha_s}) h'_\beta(\eta_{\beta_s}) h_\gamma(\zeta_{\beta_s}) \right. \\ &\quad \left. \left. + G_{i3}^{\sigma\tau\nu} h_\alpha(\xi_{\alpha_s}) h_\beta(\eta_{\beta_s}) h'_\gamma(\zeta_{\gamma_s}) \right] \right\}, \quad (3.22) \end{aligned}$$

where  $\mathbf{x}(\xi_{\alpha_s}, \eta_{\beta_s}, \zeta_{\gamma_s}) = \mathbf{x}_s$  denotes the point source location.

At this stage we have all the necessary ingredients for time-marching the weak form (2.11). The collective displacement vectors at all the grid points in the global mesh are referred to as the *global degrees of freedom* of the system and will be denoted by the vector  $\mathbf{U}$ , and the corresponding global test vector is denoted by  $\mathbf{W}$ . At the global level, Eqs. (3.13), (3.19), and (3.22) lead to expressions of the form  $\mathbf{W}^T \mathbf{M} \ddot{\mathbf{U}}$ ,  $\mathbf{W}^T \mathbf{K} \mathbf{U}$ , and  $\mathbf{W}^T \mathbf{F}$ , where  $\mathbf{M}$  denotes the global mass matrix,  $\mathbf{K}$  the global stiffness matrix, and  $\mathbf{F}$  the global source vector, which combine into a discrete equation that holds for all test vectors  $\mathbf{W}$ . The ordinary differential equation that governs the time dependence of the global system may thus be written in the symbolic form

$$\mathbf{M} \ddot{\mathbf{U}} = -\mathbf{K} \mathbf{U} + \mathbf{F}. \quad (3.23)$$

To take full advantage of the fact that the global mass matrix is diagonal, time discretization of the second-order ordinary differential equation (3.23) is achieved based upon

a classical explicit second-order finite-difference scheme, which is a particular case of the more general Newmark scheme [25]. This scheme is conditionally stable, and the Courant stability condition is governed by the maximum value of the ratio between the compressional wave speed and the grid spacing. The main numerical cost associated with the SEM is related to small local matrix-vector products, not the time scheme.

### 3.7 Examples of SEM simulations in seismology

The SEM has been extensively benchmarked against discrete wavenumber methods for layercake models [29] and normal-mode synthetics for spherically symmetric Earth model PREM (Fig. 1) [9, 30, 31]. Fig. 8 shows an example of such a benchmark for the June 9, 1994, magnitude 8.2 Bolivia earthquake recorded at station PAS in Pasadena, CA. The match between the SEM and mode synthetics on all three components is excellent. These kinds of benchmarks for spherically symmetric Earth models are very challenging, because they involve solid-fluid domain decomposition and coupling, attenuation, anisotropy, self-gravitation, and the effect of the ocean layer. Thus far, only the SEM has been capable of accurately incorporating all of these effects.

The SEM can now be used with confidence to simulate global seismic wave propagation in fully 3D Earth models. As an example, data recorded by the Global Seismographic Network and SEM synthetic seismograms for the November 3, 2002, Denali fault, Alaska, earthquake are shown in Fig. 9. Note that the SEM synthetic seismograms for mantle model S20RTS [52] (Fig. 2), accurate at periods of 5 s and longer, capture the dispersion of the Rayleigh surface waves reasonably well, but that the synthetics for spherically symmetric model PREM (Fig. 1) do not, as can be expected.

Our next goal is to use the remaining differences between the data and the SEM synthetics, e.g., Fig. 9 (Right), to improve models of the Earth's mantle and kinematic representations of the earthquake. In other words, we want to address the inverse problem. This brings us to the next topic of this paper, which is *adjoint methods*. Our introduction to adjoint methods in seismology is largely based upon the article by [39].

## 4 Adjoint methods

The objective will be to minimize some measure of the remaining differences between the data and SEM synthetics. There are numerous ways in which to characterize such differences, e.g., cross-correlation traveltime and amplitude anomalies, multi-taper phase and amplitude measurements, or straight waveform differences. Here we choose to minimize least-squares waveform differences. Therefore, we seek to minimize the waveform misfit function

$$\chi = \frac{1}{2} \sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]^2 dt, \quad (4.1)$$

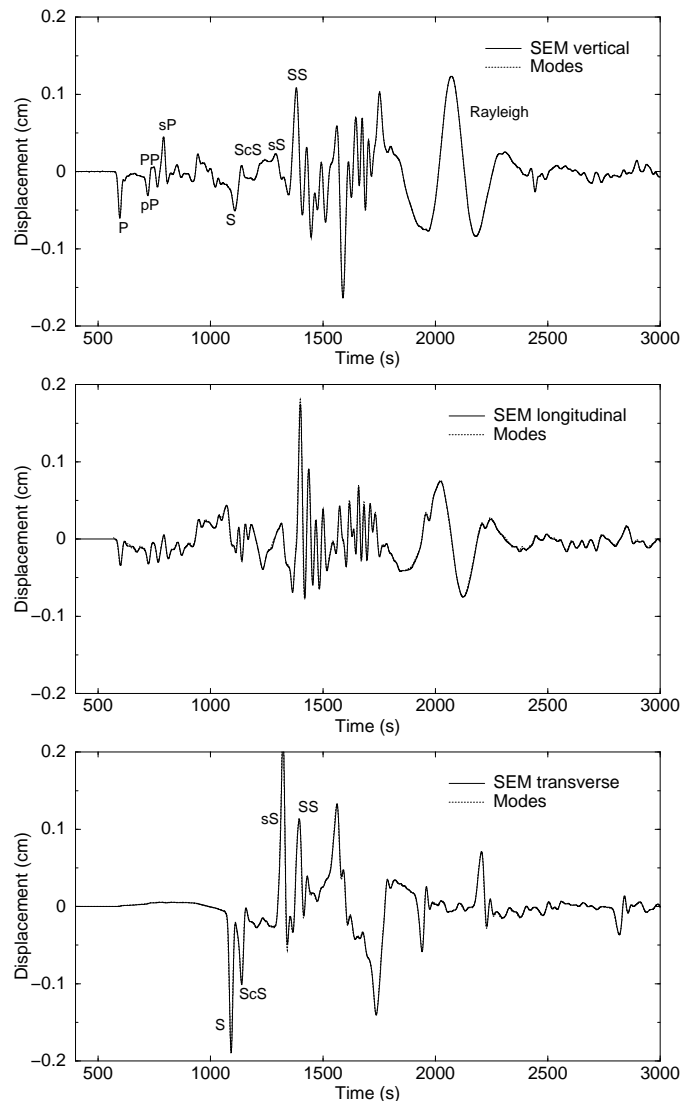


Figure 8: SEM (solid line) and normal-mode mode (dotted line) synthetic seismograms in PREM (shown in Fig. 1) for the great magnitude 8.2 Bolivia earthquake of June 9, 1994, recorded at SCSN station PAS in Pasadena, California. The depth of the event is 647 km. Top: vertical component. Middle: longitudinal component. Bottom: transverse component. The synthetics are accurate at periods of 18 s and longer. *Courtesy of [30].*

where the interval  $[0, T]$  denotes the time series of interest, and  $\mathbf{d}(\mathbf{x}_r, t)$  denotes the observed 3-component displacement vector, e.g., the black seismogram in Fig. 9 (Right), and  $\mathbf{s}(\mathbf{x}_r, t)$  denotes the synthetic displacement at receiver location  $\mathbf{x}_r$  as a function of time  $t$ , e.g., the red seismogram in Fig. 9 (Right). In practice, both the data  $\mathbf{d}$  and the synthetics  $\mathbf{s}$  will be windowed, filtered, and possibly weighted on the time interval  $[0, T]$ . In what follows we will implicitly assume that such filtering operations have been per-

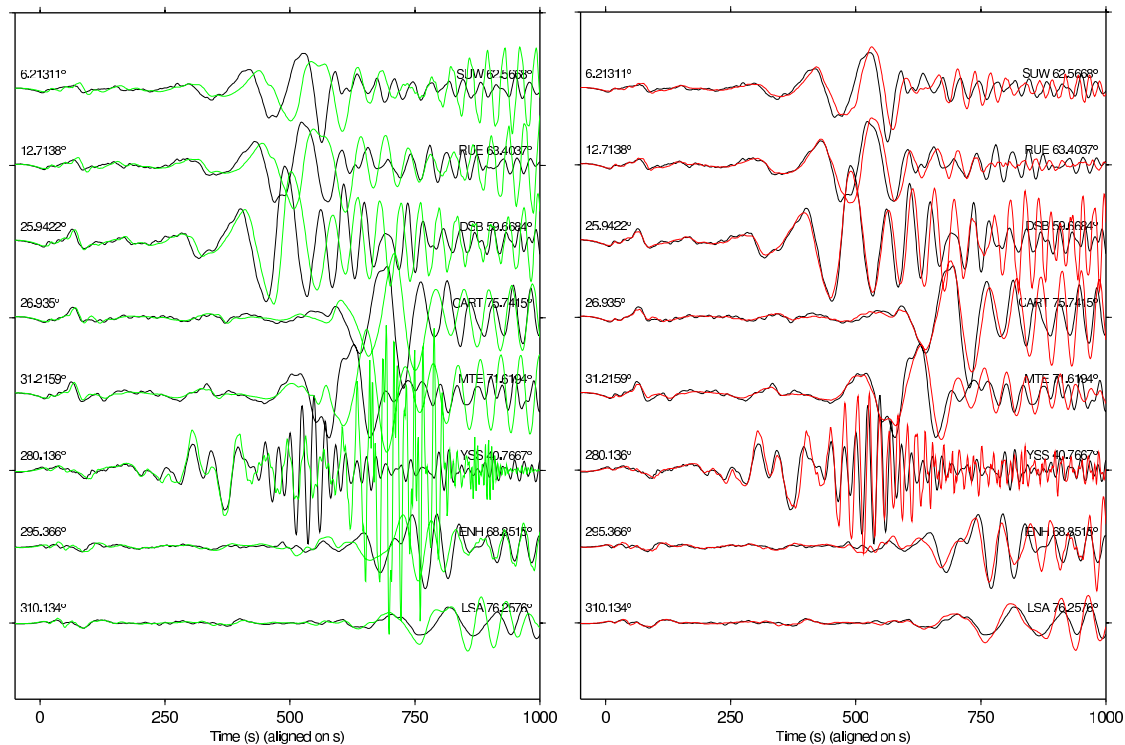


Figure 9: Left: Comparison of transverse component data (black line) and spectral-element synthetic seismograms (green line) for spherically symmetric model PREM (shown in Fig. 1) for the November 3, 2002, Denali fault, Alaska, earthquake. The synthetic seismograms and the data are low-pass filtered at 5 s. The source azimuth measured clockwise from due North is indicated on the left of each trace, and the station name and epicentral distance are on the right. Records are aligned on the S wave. Right: Comparison of the same transverse component data (black line) and spectral-element synthetic seismograms (red line) for a 3D Earth model that includes mantle model S20RTS (shown in Fig. 2) and crustal model crust2.0 [4]. *Courtesy of [68].*

formed, i.e., the symbols  $\mathbf{d}$  and  $\mathbf{s}$  will denote processed data and synthetics, respectively.

Following [39], we minimize the misfit function (4.1) subject to the constraint that the synthetic displacement field  $\mathbf{s}$  satisfies the seismic wave equation (2.7). Mathematically, this implies the *PDE-constrained* minimization of the action

$$\chi = \frac{1}{2} \sum_r \int_0^T [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)]^2 dt - \int_0^T \int_{\Omega} \lambda \cdot (\rho \partial_t^2 \mathbf{s} - \nabla \cdot \mathbf{T} - \mathbf{f}) d^3 \mathbf{x} dt, \quad (4.2)$$

where the vector Lagrange multiplier  $\lambda(\mathbf{x}, t)$  remains to be determined. Upon taking the variation of the action (4.2), using Hooke's law (2.2), integrating terms involving spatial and temporal derivatives of both  $\mathbf{s}$  and the variation  $\delta \mathbf{s}$  by parts, and perturbing the free surface boundary condition (2.8) and the initial conditions (2.9), we obtain after some



algebra

$$\begin{aligned} \delta\chi = & \int_0^T \int_{\Omega} \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta\mathbf{s}(\mathbf{x}, t) d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} (\delta\rho\lambda \cdot \partial_t^2 \mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda \cdot \delta\mathbf{f}) d^3\mathbf{x} dt - \int_0^T \int_{\Omega} [\rho\partial_t^2 \lambda - \nabla \cdot (\mathbf{c} : \nabla\lambda)] \cdot \delta\mathbf{s} d^3\mathbf{x} dt \\ & - \int_{\Omega} [\rho(\lambda \cdot \partial_t \delta\mathbf{s} - \partial_t \lambda \cdot \delta\mathbf{s})]_T d^3\mathbf{x} - \int_0^T \int_{\partial\Omega} \hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) \cdot \delta\mathbf{s} d^2\mathbf{x} dt, \end{aligned} \quad (4.3)$$

where the notation  $[f]_T$  means  $f(T)$ , for any enclosed function  $f$ .

In the absence of perturbations in the model parameters  $\delta\rho$ ,  $\delta\mathbf{c}$ , and  $\delta\mathbf{f}$ , the variation in the action (4.3) is stationary with respect to perturbations in displacement  $\delta\mathbf{s}$  provided the Lagrange multiplier  $\lambda$  satisfies the equation

$$\rho\partial_t^2 \lambda - \nabla \cdot (\mathbf{c} : \nabla\lambda) = \sum_r [\mathbf{s}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}_r), \quad (4.4)$$

subject to the free surface boundary condition

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla\lambda) = \mathbf{0} \quad \text{on } \partial\Omega, \quad (4.5)$$

and the end conditions

$$\lambda(\mathbf{x}, T) = \mathbf{0}, \quad \partial_t \lambda(\mathbf{x}, T) = \mathbf{0}. \quad (4.6)$$

More generally, provided the Lagrange multiplier  $\lambda$  is determined by Eqs. (4.4)-(4.6), the variation in the action (4.3) reduces to

$$\delta\chi = - \int_0^T \int_{\Omega} (\delta\rho\lambda \cdot \partial_t^2 \mathbf{s} + \nabla\lambda : \delta\mathbf{c} : \nabla\mathbf{s} - \lambda \cdot \delta\mathbf{f}) d^3\mathbf{x} dt. \quad (4.7)$$

This equation relates changes in the misfit function  $\delta\chi$  to changes in the model parameters  $\delta\rho$ ,  $\delta\mathbf{c}$ , and  $\delta\mathbf{f}$  in terms of the original wavefield  $\mathbf{s}$  determined by (2.7)-(2.9) and the Lagrange multiplier wavefield  $\lambda$  determined by (4.4)-(4.6).

To understand the nature of the Lagrange multiplier wavefield, we define the *adjoint wavefield*  $\mathbf{s}^\dagger$  in terms of the Lagrange multiplier wavefield  $\lambda$  by

$$\mathbf{s}^\dagger(\mathbf{x}, t) \equiv \lambda(\mathbf{x}, T - t). \quad (4.8)$$

Thus the adjoint wavefield  $\mathbf{s}^\dagger$  is equal to the time-reversed Lagrange multiplier wavefield  $\lambda$ . The adjoint wavefield  $\mathbf{s}^\dagger$  is determined by the set of equations

$$\rho\partial_t^2 \mathbf{s}^\dagger - \nabla \cdot \mathbf{T}^\dagger = \sum_r [\mathbf{s}(\mathbf{x}_r, T - t) - \mathbf{d}(\mathbf{x}_r, T - t)] \delta(\mathbf{x} - \mathbf{x}_r), \quad (4.9)$$

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

$$\mathbf{T}^\dagger = \mathbf{c} : \nabla\mathbf{s}^\dagger. \quad (4.10)$$

The adjoint wave equation (4.9) is subject to the free surface boundary condition

$$\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = \mathbf{0} \quad \text{on } \partial\Omega, \quad (4.11)$$

and the initial conditions

$$\mathbf{s}^\dagger(\mathbf{x}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}^\dagger(\mathbf{x}, 0) = \mathbf{0}. \quad (4.12)$$

Upon comparing (4.9)-(4.12) with (2.7)-(2.9), we see that the adjoint wavefield  $\mathbf{s}^\dagger$  is determined by exactly the same wave equation, boundary conditions, and initial conditions as the regular wavefield, with the exception of the source term: the regular wavefield is determined by the earthquake source  $\mathbf{f}$ , given by (2.10), whereas the adjoint wavefield is generated by using the time-reversed differences between the synthetics  $\mathbf{s}$  and the data  $\mathbf{d}$  at the receivers as simultaneous sources.

In terms of the adjoint wavefield  $\mathbf{s}^\dagger$ , the gradient of the misfit function (4.7) may be rewritten in the form

$$\delta\chi = \int_{\Omega} (\delta \ln \rho K_{\rho} + \delta \mathbf{c} :: \mathbf{K}_{\mathbf{c}}) d^3 \mathbf{x} + \int_0^T \int_{\Omega} \mathbf{s}^\dagger \cdot \delta \mathbf{f} d^3 \mathbf{x} dt, \quad (4.13)$$

where we have defined the kernels

$$K_{\rho}(\mathbf{x}) = - \int_0^T \rho(\mathbf{x}) \mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) dt, \quad (4.14)$$

$$\mathbf{K}_{\mathbf{c}}(\mathbf{x}) = - \int_0^T \nabla \mathbf{s}^\dagger(\mathbf{x}, T-t) \nabla \mathbf{s}(\mathbf{x}, t) dt. \quad (4.15)$$

Realizing that  $\delta \mathbf{c}$  and  $\mathbf{K}_{\mathbf{c}}$  are both fourth-order tensors, we introduced the notation  $\delta \mathbf{c} :: \mathbf{K}_{\mathbf{c}} = \delta c_{ijkl} K_{c_{ijkl}}$  in (4.13).

The perturbation to the point source (2.10) may be written in the form

$$\delta \mathbf{f} = -\delta \mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) S(t) - \mathbf{M} \cdot \nabla [\delta \mathbf{x}_s \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s)] S(t) - \mathbf{M} \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) \delta S(t), \quad (4.16)$$

where  $\delta \mathbf{M}$  denotes the perturbed moment tensor,  $\delta \mathbf{x}_s$  the perturbed point source location, and  $\delta S(t)$  the perturbed source-time function. Upon substituting (4.16) into the gradient of the misfit function (4.13), using the properties of the Dirac delta distribution, we obtain

$$\begin{aligned} \delta\chi = & \int_{\Omega} (\delta \ln \rho K_{\rho} + \delta \mathbf{c} :: \mathbf{K}_{\mathbf{c}}) d^3 \mathbf{x} + \int_0^T \delta \mathbf{M} : \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) S(t) dt \\ & + \int_0^T \mathbf{M} : (\delta \mathbf{x}_s \cdot \nabla_s) \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) S(t) dt + \int_0^T \mathbf{M} : \boldsymbol{\epsilon}^\dagger(\mathbf{x}_s, T-t) \delta S(t) dt, \end{aligned} \quad (4.17)$$

where

$$\boldsymbol{\epsilon}^\dagger = \frac{1}{2} [\nabla \mathbf{s}^\dagger + (\nabla \mathbf{s}^\dagger)^T], \quad (4.18)$$

denotes the adjoint strain tensor and a superscript  $T$  denotes the transpose.

In an isotropic Earth, using (2.3), we obtain

$$\delta \mathbf{c} :: \mathbf{K}_c = \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa, \tag{4.19}$$

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

$$K_\mu(\mathbf{x}) = - \int_0^T 2\mu(\mathbf{x}) \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) dt, \tag{4.20}$$

$$K_\kappa(\mathbf{x}) = - \int_0^T \kappa(\mathbf{x}) [\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)] [\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \tag{4.21}$$

where

$$\mathbf{D} = \frac{1}{2} [\nabla \mathbf{s} + (\nabla \mathbf{s})^T] - \frac{1}{3} (\nabla \cdot \mathbf{s}) \mathbf{I}, \tag{4.22}$$

$$\mathbf{D}^\dagger = \frac{1}{2} [\nabla \mathbf{s}^\dagger + (\nabla \mathbf{s}^\dagger)^T] - \frac{1}{3} (\nabla \cdot \mathbf{s}^\dagger) \mathbf{I}, \tag{4.23}$$

denote the traceless strain deviator and its adjoint, respectively.

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

$$\delta \ln \rho K_\rho + \delta \ln \mu K_\mu + \delta \ln \kappa K_\kappa = \delta \ln \rho K_{\rho'} + \delta \ln \beta K_\beta + \delta \ln \alpha K_\alpha, \tag{4.24}$$

where

$$K_{\rho'} = K_\rho + K_\kappa + K_\mu, \tag{4.25}$$

$$K_\beta = 2 \left( K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right), \tag{4.26}$$

$$K_\alpha = 2 \left( \frac{\kappa + \frac{4}{3} \mu}{\kappa} \right) K_\kappa. \tag{4.27}$$

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

#### 4.1 Numerical implementation of the adjoint method

In order to perform the time integration involved in the calculation of the kernels (4.25), (4.26), and (4.27), we require simultaneous access to the forward wavefield  $\mathbf{s}$  at time  $t$  and the adjoint wavefield  $\mathbf{s}^\dagger$  at time  $T-t$ . This rules out the possibility of carrying both the forward and the adjoint simulation simultaneously in one spectral-element simulation, because both wavefields would only be available at a given time  $t$ . A brute-force solution

is to run the forward simulation, save the entire forward wavefield as a function of space and time, and then launch the adjoint simulation while performing the time integration by accessing time slice  $t$  of the adjoint wavefield while reading back the corresponding time slice  $T-t$  of the forward wavefield stored on the hard drive. For large 3D simulations this approach poses a serious storage problem.

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

$$\rho \partial_t^2 \mathbf{s} = \nabla \cdot (\mathbf{c} : \nabla \mathbf{s}) + \mathbf{f} \quad \text{in } V, \quad (4.28)$$

$$\mathbf{s}(\mathbf{x}, T) \quad \text{and} \quad \partial_t \mathbf{s}(\mathbf{x}, T) \quad \text{given}, \quad (4.29)$$

$$\hat{\mathbf{n}} \cdot (\mathbf{c} : \nabla \mathbf{s}) = \mathbf{0} \quad \text{on } \partial\Omega. \quad (4.30)$$

This initial and boundary value problem can be solved to reconstruct  $\mathbf{s}(\mathbf{x}, t)$  for  $T \geq t \geq 0$  the same way the forward wave equation is solved. Technically, the only difference between solving the backward wave equation versus solving the forward wave equation is a change in the sign of the time step parameter  $\Delta t$ . If we carry both the backward and the adjoint simulation simultaneously in memory during the spectral-element simulation, we have access to the forward wavefield at time  $t$  and the adjoint wavefield at time  $T-t$ , which is exactly what we need to perform the time integrations involved in the construction of the kernels (4.25), (4.26), and (4.27). An advantage of this approach is that only the wavefield at the last time step of the forward simulation needs to be stored and read back for the reconstruction of  $\mathbf{s}(\mathbf{x}, t)$  and the construction of the kernels.

## 4.2 Sensitivity kernels

The key ingredients of the adjoint approach are the sensitivity kernels (4.25), (4.26), and (4.27). What do these kernels look like? Clearly they are the result of the integrated interaction between the regular wavefield  $\mathbf{s}$  and the adjoint wavefield  $\mathbf{s}^\dagger$ . The adjoint wavefield is simultaneously generated at all the receivers based upon some measure of the difference between the data and the synthetics, e.g., the time-reversed waveform difference  $[\mathbf{s}(T-t) - \mathbf{d}(T-t)]$ . In principle, we can send back the entire time-reversed waveform difference, e.g., the differences between the data and synthetics for the 2002 Denali earthquake shown in Fig. 9 (Right). However, we are free to choose any time window of interested, and so to develop some intuition for the nature of these kernels, let us consider the simplest arrival in the seismogram, which is the first arriving compressional wave called the P wave (for 'primary'). The arrival time of this wave is controlled by the compressional wave speed  $\alpha$ , and thus we expect this wave to be sensitive to compressional wave speed perturbations  $\delta \ln \alpha$ , but not to shear wave speed perturbations  $\delta \ln \beta$ , nor density perturbations  $\delta \ln \rho$ .

We simulate three-component seismograms for the June 9, 1994, Bolivia Earthquake. This magnitude 8.2 earthquake occurred at a depth of 647 km and is one of

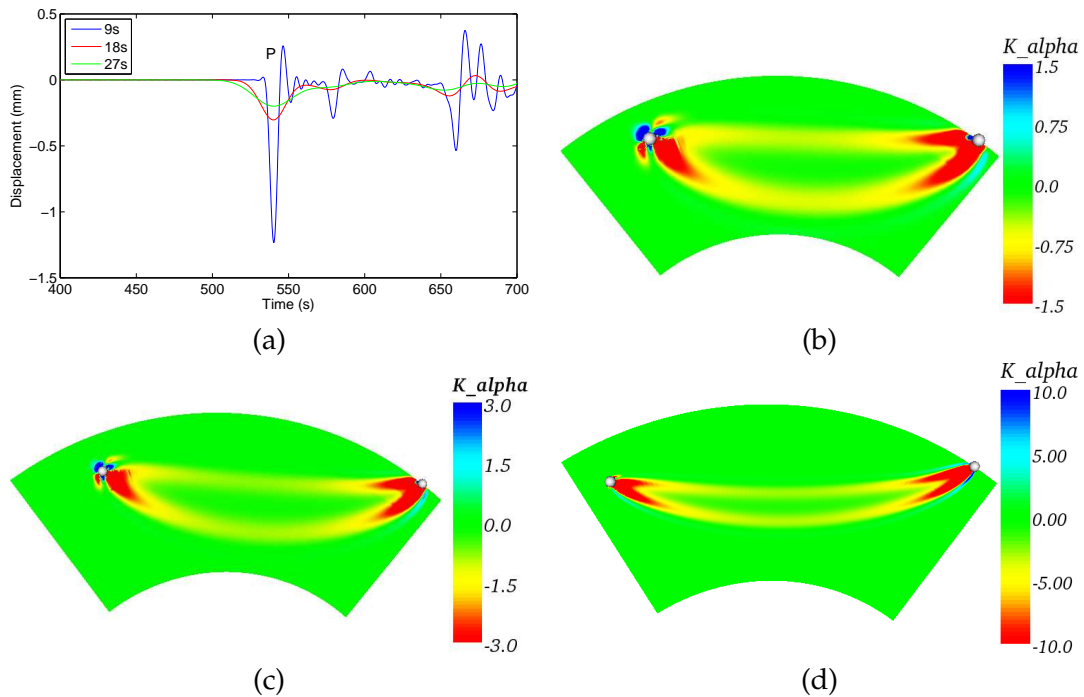


Figure 10: (a) Vertical component synthetic velocity seismograms recorded at an epicentral distance of 60 for simulations accurate down to periods of 27 s (green), 18 s (red) and 9 s (blue), respectively. (b) Source-receiver cross-section of the  $K_\alpha$  kernel, defined by (4.27), for a 27 s P wave recorded at a station at an epicentral distance of 60°. The source and receiver locations are denoted with two small white circles. The unit of the sensitivity kernels is  $10^7 \text{ s/km}^3$  throughout this paper. (c)  $K_\alpha$  kernel for an 18 s P wave recorded at a station at an epicentral distance of 60°. (d)  $K_\alpha$  kernel for a 9 s P wave recorded at a station at an epicentral distance of 60°. *Courtesy of [40].*

the largest deep events in modern recording history. As a source we use the event location and the centroid-moment tensor (CMT) solution from the Harvard CMT catalog ([www.seismology.harvard.edu](http://www.seismology.harvard.edu)). Fig. 10(a) shows vertical component synthetic seismograms recorded by a receiver at an epicentral distance of 60°. To investigate the finite-frequency characteristics of the kernels, the synthetic seismogram is low-pass filtered three times, with corners at 27 s, 18 s and 9 s, respectively. Figs. 10(b), (c) and (d) show source-receiver cross-sections of the corresponding kernels  $K_\alpha$  defined by (4.27) calculated based upon the adjoint method. Notice that as the frequency content of the adjoint signal is increased, the kernel becomes skinnier. In fact, in the high-frequency limit the kernel will collapse on to the P wave geometrical raypath. Note also that the kernels have a ‘hole’ along this raypath. The characteristic ‘banana’ shape of the kernel in the source-receiver plane and the ‘donut’ shape of the kernel in a cross-section perpendicular to the source-receiver plane prompted [43] to refer to these kernels as ‘banana-donut kernels’. As shown by [19], the size of the donut hole decreases with increasing resolution, in accordance with the scaling relation  $\text{width} \sim \sqrt{\lambda L}$ , where  $L$  denotes the length of the raypath and  $\lambda$  the wavelength.

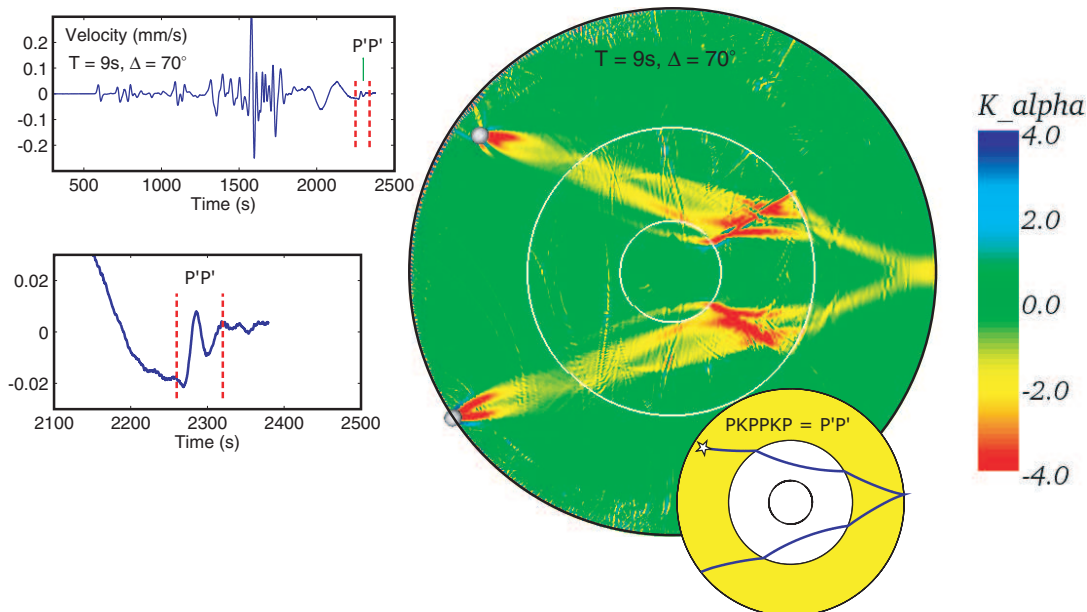


Figure 11: Top left: Vertical component velocity seismogram recorded at a distance of  $60^\circ$ . The shortest period in the simulations is 9 s. The tiny P'P' phase is labeled. Bottom left: Last 300 s of the seismogram shown above blown up by a factor of 12. The P'P' phase is labeled. Right:  $K_\alpha$  kernel, defined by (4.27), for the P'P' phase identified on the left. The inset shows the P'P' ray geometry in blue. The radii of the ICB and CMB are indicated by the concentric black circles. *Courtesy of [40].*

The adjoint approach can readily be used to look at more exotic arrivals. As an example, we generate an 9 s  $K_\alpha$  sensitivity kernel, defined by (4.27), for the P'P' arrival (also known as PKPPKP): a PKP phase reflected off the Earth's surface. This phase, which arrives around 2300 s, is nearly unidentifiable in the complete vertical component seismogram, as shown in Fig. 11 (Top left). However, when the last few hundred seconds of the seismogram is blown up by a factor of 12, we can identify the P'P' phase clearly at the tail end of the surface waves, as illustrated in Fig. 11 (Bottom left). The corresponding  $K_\alpha$  kernel is shown in Fig. 11 (Right), which clearly follows the PKPPKP raypath.

For simple 1D Earth models, banana-donut kernels may be calculated relatively inexpensively based upon ray-theory [19,26,71], or, slightly more expensively, based upon normal-mode methods [69,70]. However, for 3D Earth models one needs to use an adjoint approach, i.e., for tomography involving 3D reference Earth models one must resort to fully numerical techniques. A nice feature of the adjoint approach is that we need not be able to 'label' a particular pulse in the seismogram, i.e., we need not have any knowledge of the raypath associated with this specific pulse. By performing the forward and adjoint simulations we automatically obtain the 3D sensitivity associated with this particular pulse in the seismogram. Sometimes the kernels may be readily identified with a specific geometrical raypath, but frequently the sensitivity kernels are much richer, as in the case of the P'P' kernel shown in Fig. 11.



### 4.3 Toward adjoint tomography

How do we use these finite-frequency kernels to address the inverse problem? To answer this question it is important to recognize that in the adjoint approach we do not need to calculate individual banana-donut kernels for each measurement. If  $N_{\text{events}}$  denotes the number of earthquakes,  $N_{\text{stations}}$  the number of stations, and  $N_{\text{picks}}$  the number of measurements at that station, such an approach would require  $N_{\text{events}} \times N_{\text{stations}} \times N_{\text{picks}}$  simulations, i.e., one simulation for each banana-donut kernel corresponding to one particular pick. Typical tomographic inversions involve hundreds of earthquakes recorded by hundreds of three-component receivers, and in each time series one typically makes several picks. This implies that one would have to calculate tens of thousands of banana-donut kernels corresponding to all the individual measurements, something that is currently prohibitively expensive from a numerical stand point.

To circumvent this computational burden, the adjoint approach is to measure as many arrivals as possible in three component seismograms from all available stations for any given earthquake. Ideally, every component at every station will have a number of arrivals suitable for measurement, for example in terms of frequency-dependent phase and amplitude anomalies. During the adjoint simulation, each component of every receiver will transmit simultaneously its measurements in reverse time, and the interaction between the so generated adjoint wavefield and the forward wavefield results in a misfit kernel for that particular event. As discussed by [66] and illustrated by [61], this earthquake-specific ‘event kernel’ is essentially a sum of weighted banana-donut kernels, with weights determined by the corresponding measurement, and is obtained based upon just two 3D simulations. These two simulations take about three times the computation time of a regular forward simulation, because the adjoint simulation involves both the advance of the adjoint wavefield and the reconstruction of the forward wavefield. By summing these event kernels one obtains the ‘summed event kernel’, which highlights where the current 3D model is inadequate and enables one to iteratively obtain an improved Earth model, for example based upon a non-linear conjugate gradient approach. The number of 3D simulations at each conjugate gradient step scales linearly with the number of earthquakes  $N_{\text{events}}$ , but is independent of the number of receivers  $N_{\text{stations}}$  or the number of measurements  $N_{\text{picks}}$ .

Entrapment into local minima is common in the conjugate gradient method, as addressed in [1, 2]. Such local minima may be avoided by using multi-scale methods [7]. Alternatively, one can also try to avoid local minima by starting at longer periods, which constrain the long-wavelength heterogeneity, gradually moving to shorter periods, which constrain smaller-scale structures.

Finally, note that we can also use (4.17) to perform adjoint source inversions or joint source and structural inversions. Traditional CMT source inversions involve nine forward simulations to obtain the necessary Fréchet derivatives [41]. Therefore, if the number of conjugate gradient iterations based upon the derivative (4.17) is less than nine, adjoint source inversions can be more efficient than a brute force CMT inversion. The

adjoint approach has the additional advantage that it need make no assumptions about the location and timing of the event, nor, for finite sources, about its directivity.

## 5 Discussion and conclusions

The spectral-element method has enabled seismologists to simulate 3D seismic wave propagation at unprecedented resolution and accuracy. Furthermore, advances in high-performance computing have made such simulations relatively fast and inexpensive. Due to this confluence of events, we have reached a point where seismologists are beginning to use 3D simulations to address inverse problems, and adjoint methods, already widely used in the atmospheric and ocean sciences, provide a powerful means of making this practical.

We have demonstrated that tomographic inversions involve 3D kernels that may be calculated based upon interactions between the wavefield for the current model and an adjoint wavefield obtained by using time-reversed signals as simultaneous sources at the receivers. For every earthquake, these kernels may be obtained based upon just two simulations, such that one iteration in the non-linear inverse problem involves a total number of simulations that equals twice the number of events. There are five main advantages of the adjoint approach in seismology. First, the kernels are calculated on-the-fly by carrying the adjoint wavefield and the regular wavefield in memory at the same time. This doubles the memory requirements for the simulation but avoids the storage of Green's functions for all events and stations as a function of space and time. Second, the kernels can be calculated for fully 3D reference models, something that is critical in highly heterogeneous settings, e.g., in regional seismology or exploration geophysics. Third, the approach scales linearly with the number of earthquakes but is independent of the number of receivers and the number of arrivals that are used in the inversion. Fourth, any time window where the data and the synthetics have significant amplitudes and match reasonably well is suitable for a measurement. One does not need to be able to label the arrival, e.g., identify it as P or P'P', because the adjoint simulation will reveal how this particular measurements 'sees' the Earth model, and the resulting 3D sensitivity kernel will reflect this view. Finally, the cost of the simulation is independent of the number of model parameters, i.e., one can consider fully anisotropic Earth models with 21 elastic parameters for practically the same computational cost as an isotropic simulation involving just two elastic parameters. [58] and [59] used this approach to calculate finite-frequency sensitivity kernels for surface and body waves, respectively.

Recently, [61] successfully applied the adjoint approach in conjunction with a conjugate gradient method in a 2D surface wave study, and [2] used a similar technique for a 3D regional wave propagation problem. We anticipate much more widespread use of these techniques in the future.

The software used to perform the spectral-element simulations shown in this article is freely available via [www.geodynamics.org](http://www.geodynamics.org).

## Acknowledgments

Broadband data were obtained from the IRIS Data Management Center. This material is based in part upon work supported by the National Science Foundation under grant EAR-0309576. This is contribution No. 9157 of the Division of Geological & Planetary Sciences (GPS), California Institute of Technology. The numerical simulations for this research were performed on Caltech's GPS Division Dell cluster.

## References

- [1] V. Akçelik, G. Biros and O. Ghattas, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, Proceedings of the ACM/IEEE Supercomputing SC'2002 conference, 2002, published on CD-ROM and at [www.sc-conference.org/sc2002](http://www.sc-conference.org/sc2002).
- [2] V. Akçelik, J. Bielak, G. Biros, I. Epanomeritakis, A. Fernández, O. Ghattas, E. J. Kim, J. López, D. O'Hallaron, T. Tu and J. Urbanic, High resolution forward and inverse earthquake modeling on terascale computers, Proceedings of the ACM/IEEE Supercomputing SC'2003 conference, 2003, published on CD-ROM and at [www.sc-conference.org/sc2003](http://www.sc-conference.org/sc2003).
- [3] K. Aki and P. G. Richards, Quantitative Seismology, Theory and Methods, W. H. Freeman, San Francisco, 1980.
- [4] C. Bassin, G. Laske and G. Masters, The current limits of resolution for surface wave tomography in North America, EOS Trans. AGU, 81 (2000), F897.
- [5] U. Basu and A. K. Chopra, Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: Theory and finite-element implementation, Comput. Method. Appl. Mech. Engrg., 192 (2003), 1337-1375.
- [6] J. P. Bérenger, A perfectly matched layer for the absorption of electromagnetic waves, J. Comput. Phys., 114 (1994), 185-200.
- [7] C. Bunks, F. Saleck, S. Zaleski and G. Chavent, Multiscale seismic waveform inversion, Geophysics, 60 (1995), 1457-1473.
- [8] C. Canuto, M. Y. Hussaini, A. Quarteroni and T. A. Zang, Spectral Methods in Fluid Dynamics, Springer-Verlag, New York, 1988.
- [9] Y. Capdeville, E. Chaljub, J. P. Vilotte and J. P. Montagner, Coupling the spectral element method with a modal solution for elastic wave propagation in global Earth models, Geophys. J. Int., 152 (2003), 34-67.
- [10] E. Chaljub, Modélisation numérique de la propagation d'ondes sismiques en géométrie sphérique: Application à la sismologie globale (Numerical modeling of the propagation of seismic waves in spherical geometry: Applications to global seismology), Ph.D. thesis, Université Paris VII Denis Diderot, Paris, France, 2000.
- [11] E. Chaljub and B. Valette, Spectral-element modeling of three-dimensional wave propagation in a self-gravitating Earth with an arbitrarily stratified outer core, Geophys. J. Int., 158 (2004), 131-141.
- [12] E. Chaljub, Y. Capdeville and J. P. Vilotte, Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, J. Comput. Phys., 187(2) (2003), 457-491.
- [13] E. Chaljub, D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette and G. Festa, Spectral element analysis in seismology, in: R.-S. Wu and V. Maupin (Eds.), IASPEI Monograph on

- Advances in Wave Propagation in Heterogeneous Media, Elsevier, Vol. 48, 2007, pp. 365-419.
- [14] R. Clayton and B. Engquist, Absorbing boundary conditions for acoustic and elastic wave equations, *Bull. Seismol. Soc. Am.*, 67 (1977), 1529-1540.
  - [15] G. Cohen, P. Joly and N. Tordjman, Construction and analysis of higher-order finite elements with mass lumping for the wave equation, in: R. Kleinman (Ed.), *Proceedings of the Second International Conference on Mathematical and Numerical Aspects of Wave Propagation*, SIAM, Philadelphia, 1993, pp. 152-160.
  - [16] F. Collino and C. Tsogka, Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media, *Geophysics*, 66(1) (2001), 294-307.
  - [17] E. Crase, A. Pica, M. Noble, J. McDonald and A. Tarantola, Robust elastic non-linear waveform inversion: Application to real data, *Geophysics*, 55 (1990), 527-538.
  - [18] F. A. Dahlen and J. Tromp, *Theoretical Global Seismology*, Princeton University Press, Princeton, 1988.
  - [19] F. A. Dahlen, G. Nolet and S.-H. Hung, Fréchet kernels for finite-frequency traveltime–I. Theory, *Geophys. J. Int.*, 141 (2000), 157-174.
  - [20] A. M. Dziewonski and D. L. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, 25 (1981), 297-356.
  - [21] E. Faccioli, F. Maggio, R. Paolucci and A. Quarteroni, 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method, *J. Seismol.*, 1 (1997), 237-251.
  - [22] G. Festa and J. P. Vilotte, The Newmark scheme as velocity-stress time-staggering: and efficient PML implementation for spectral element simulations of elastodynamics, *Geophys. J. Int.*, 161 (2005), 789-812.
  - [23] O. Gauthier, J. Virieux and A. Tarantola, Two-dimensional non-linear inversion of seismic waveforms: Numerical results, *Geophysics*, 51 (1986), 1387-1403.
  - [24] W. Gropp, E. Lusk and A. Skjellum, *Using MPI, Portable Parallel Programming with the Message-Passing Interface*, MIT Press, Cambridge, 1994.
  - [25] T. J. R. Hughes, *The Finite Element Method, Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall International, Englewood Cliffs, NJ, 1987.
  - [26] S.-H. Hung, F. A. Dahlen and G. Nolet, Fréchet kernels for finite-frequency traveltime–II. Examples, *Geophys. J. Int.*, 141 (2000), 175-203.
  - [27] B. L. N. Kennett, *Seismic Wave Propagation in Stratified Media*, Cambridge University Press, Cambridge, 1983.
  - [28] D. Komatitsch, *Méthodes spectrales et éléments spectraux pour l'équation de l'élastodynamique 2D et 3D en milieu hétérogène (Spectral and spectral-element methods for the 2D and 3D elastodynamics equations in heterogeneous media)*, Ph.D. thesis, Institut de Physique du Globe, Paris, France, 1997.
  - [29] D. Komatitsch and J. Tromp, Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.*, 139 (1999), 806-822.
  - [30] D. Komatitsch and J. Tromp, Spectral-element simulations of global seismic wave propagation–I. Validation, *Geophys. J. Int.*, 149 (2002), 390-412.
  - [31] D. Komatitsch and J. Tromp, Spectral-element simulations of global seismic wave propagation–II. 3-D models, oceans, rotation, and self-gravitation, *Geophys. J. Int.*, 150 (2002), 303-318.
  - [32] D. Komatitsch and J. Tromp, A Perfectly Matched Layer absorbing boundary condition for the second-order seismic wave equation, *Geophys. J. Int.*, 154 (2003), 146-153.
  - [33] D. Komatitsch and J. P. Vilotte, The spectral-element method: An efficient tool to simulate

- the seismic response of 2D and 3D geological structures, *Bull. Seismol. Soc. Am.*, 88(2) (1998), 368-392.
- [34] D. Komatitsch, R. Martin, J. Tromp, M. A. Taylor and B. A. Wingate, Wave propagation in 2-D elastic media using a spectral element method with triangles and quadrangles, *J. Comput. Acoust.*, 9(2) (2001), 703-718.
- [35] D. Komatitsch, J. Ritsema and J. Tromp, The spectral-element method, *Beowulf computing, and global seismology*, *Science*, 298 (2002), 1737-1742.
- [36] D. Komatitsch, S. Tsuboi, J. Chen and J. Tromp, A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the Earth Simulator, *Proceedings of the ACM/IEEE Supercomputing SC'2003 conference*, 2003.
- [37] D. Komatitsch, Q. Liu, J. Tromp, P. Süß, C. Stidham and J. H. Shaw, Simulations of ground motion in the Los Angeles Basin based upon the spectral-element method, *Bull. Seismol. Soc. Am.*, 94 (2004), 187-206.
- [38] D. Komatitsch, S. Tsuboi and J. Tromp, The spectral-element method in seismology, in: G. Nolet and A. Levander (Eds.), *The Seismic Earth*, AGU, Washington DC, 2005, pp. 205-227.
- [39] Q. Liu and J. Tromp, Finite-frequency kernels based upon adjoint methods, *Bull. Seismol. Soc. Am.*, 96 (2007), 2383-2397.
- [40] Q. Liu and J. Tromp, Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, (2007), submitted.
- [41] Q. Liu, J. Polet, D. Komatitsch and J. Tromp, Spectral-element moment tensor inversions for earthquakes in southern California, *Bull. Seismol. Soc. Am.*, 94 (2004), 1748-1761.
- [42] Y. Maday and E. M. Rønquist, Optimal error analysis of spectral methods with emphasis on non-constant coefficients and deformed geometries, *Comput. Method. Appl. Mech. Engrg.*, 80 (1990), 91-115.
- [43] H. Marquering, F. Dahlen and G. Nolet, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, 137 (1999), 805-815.
- [44] E. D. Mercerat, J. P. Vilotte and F. J. Sanchez-Sesma, Triangular spectral element simulation of two-dimensional elastic wave propagation using unstructured triangular grids, *Geophys. J. Int.*, 166 (2006), 679-698.
- [45] P. Mora, Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, 52 (1987), 1211-1228.
- [46] P. Mora, Elastic wave-field inversion of reflection and transmission data, *Geophysics*, 53 (1988), 750-759.
- [47] P. S. Pacheco, *Parallel Programming with MPI*, Morgan Kaufmann Press, San Francisco, 1997.
- [48] A. T. Patera, A spectral element method for fluid dynamics: Laminar flow in a channel expansion, *J. Comput. Phys.*, 54 (1984), 468-488.
- [49] R. G. Pratt, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics*, 64 (1999), 888-901.
- [50] E. Priolo, J. M. Carcione and G. Seriani, Numerical simulation of interface waves by high-order spectral modeling techniques, *J. Acoust. Soc. Am.*, 95(2) (1994), 681-693.
- [51] A. Quarteroni, A. Tagliani and E. Zampieri, Generalized Galerkin approximations of elastic waves with absorbing boundary conditions, *Comput. Method. Appl. Mech. Engrg.*, 163 (1998), 323-341.
- [52] J. Ritsema, H. J. Van Heijst and J. H. Woodhouse, Complex shear velocity structure imaged beneath Africa and Iceland, *Science*, 286 (1999), 1925-1928.



- [53] C. Ronchi, R. Ianoco and P. S. Paolucci, The “Cubed Sphere”: A new method for the solution of partial differential equations in spherical geometry, *J. Comput. Phys.*, 124 (1996), 93-114.
- [54] R. Sadourny, Conservative finite-difference approximations of the primitive equations on quasi-uniform spherical grids, *Monthly Weather Review*, 100 (1972), 136-144.
- [55] G. Seriani, 3-D large-scale wave propagation modeling by a spectral element method on a Cray T3E multiprocessor, *Comput. Method. Appl. Mech. Engrg.*, 164 (1998), 235-247.
- [56] G. Seriani and E. Priolo, A spectral element method for acoustic wave simulation in heterogeneous media, *Finite Elem. Anal. Des.*, 16 (1994), 337-348.
- [57] S. J. Sherwin and G. E. Karniadakis, A triangular spectral element method: applications to the incompressible Navier-Stokes equations, *Comput. Method. Appl. Mech. Engrg.*, 123 (1995), 189-229.
- [58] A. Sieminski, Q. Liu, J. Trampert and J. Tromp, Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, 168 (2007), 1153-1174.
- [59] A. Sieminski, Q. Liu, J. Trampert and J. Tromp, Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, 171 (2007), 368-389.
- [60] O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. Roy. Meteor. Soc.*, 113 (1987), 1311-1328.
- [61] C. H. Tape, Q. Liu and J. Tromp, Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves, *Geophys. J. Int.*, 168 (2007), 1105-1129.
- [62] A. Tarantola, Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49 (1984), 1259-1266.
- [63] A. Tarantola, Inversion of travel times and seismic waveforms, in: G. Nolet (Ed.), *Seismic Tomography*, Reidel Publishing, Dordrecht, 1987, pp. 135-157.
- [64] A. Tarantola, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure Appl. Geophys.*, 128 (1988), 365-399.
- [65] M. A. Taylor and B. A. Wingate, A generalized diagonal mass matrix spectral element method for non-quadrilateral elements, *Appl. Numer. Math.*, 33 (2000), 259-265.
- [66] J. Tromp, C. Tape and Q. Liu, Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels, *Geophys. J. Int.*, 160 (2005), 195-216.
- [67] S. Tsuboi, D. Komatitsch, J. Chen and J. Tromp, Spectral-element simulations of the November 3, 2002, Denali, Alaska earthquake on the Earth Simulator, *Phys. Earth Planet. Inter.*, 139 (2003), 305-312.
- [68] S. Tsuboi, D. Komatitsch and J. Tromp, Broadband modeling of global seismic wave propagation on the earth simulator using the spectral-element method, *J. Seismol. Soc. Jpn.*, 57 (2005), 321-329.
- [69] L. Zhao and T. H. Jordan, Structure sensitivities of finite-frequency seismic waves: A full-wave approach, *Geophys. J. Int.*, 165 (2006), 981-990.
- [70] L. Zhao, T. H. Jordan, K. B. Olsen and P. Chen, Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. Seismol. Soc. Am.*, 95 (2005), 2066-2080.
- [71] Y. Zhou, F. Dahlen and G. Nolet, 3-D sensitivity kernels for surface-wave observables, *Geophys. J. Int.*, 158 (2004), 142-168.
- [72] O. C. Zienkiewicz, *The Finite Element Method in Engineering Science*, 3rd ed., McGraw-Hill, New York, 1977.