A Review of Adjoint Methods for Computing Derivatives Used in Wave Field Inversion

William Menke

Lamont-Doherty Earth Observatory, Palisades NY

MENKE@LDEO.COLUMBIA.EDU

July 2016

Abstract. The wave field imaging techniques that have so revolutionized seismic tomography are predicated on our ability to efficiently compute the derivative of error with respect to a model parameter describing Earth structure. The error quantifies the quality of fit between the observed and predicted data, which can be either the wave field itself ("waveform inversion") or some quantity derived from it (e.g. finite frequency travel times). Computation of the derivatives is an essential part of the inversion process and usually the most computationally-intensive part of it. Adjoint Methods use a mathematical manipulation drawn from the theory of linear operators to reorganize the calculation, substantially improving its efficiency. We review Adjoint Methods and present simplified derivations of its essential formulas. The concept of the adjoint field is developed, using two complementary techniques: a direct method based on substitution and an implicit one based on Lagrange multipliers. We then show how the introduction of the adjoint field changes the scaling of the derivative calculation, from one proportional to the number of model parameters (which can be very large) to one proportional to the number of receivers (which is typically more modest in size). We derive derivative formula for four types of data: the wave field itself, finite frequency travel times, wave field power, and the cross-convolution measure. In each case, we first develop the general formula and then apply it to the special case of a weakly-heterogeneous medium with a constant background structure.

1. Introduction

Wave field inversion is the process of inferring Earth structure and/or source parameters from measurements of the seismic wave field. Structural parameters include material properties such as density, compressional velocity and shear velocity and the positions of interfaces. Source parameters include the time histories and spatial patterns of forces and the seismic moment density associated with faulting. The measurements (data) might be the displacement of the wave field, or any of several quantities derived from it, such as finite-frequency travel times Marquering et al. [1999] and cross-convolution measures [Menke and Levin, 2003].

Wave field inversion has been developed by many researchers over the last fifty years and has many different implementations. Most variants employ the principle that the best estimate of the parameters is the one that matches the data to its theoretical prediction. Wave field inversion becomes a nonlinear optimization problem when the misfit between theory and observation is quantified by a formally-defined error (such as the least squares error) and a wide range of well-understood techniques are available to solve it. Among these are iterative methods, which start with an initial estimate of the model, which only poorly fits the data, and successively perturb it to achieve a better fit. Two types of iterative methods are in common use: Newton's method [e.g. Deuflhard, 2004] and the Gradient-Descent method [e.g. Snyman, 2005]. The former method requires the derivative G_{ij} of the predicted data d_i with respect to a model parameter m_j, and the latter requires the derivative g_j of the error with respect to a model parameter:

$$G_{ij} = \frac{\partial d_i}{\partial m_j} \quad \text{and} \quad g_j = \frac{\partial E}{\partial m_j} \tag{1.1 a,b}$$

Most of the work of wave field inversion (and the cleverness needed to avoid that work) is expended during the computation of these derivatives. These derivatives are often referred to as *sensitivity kernels*, because they quantify how sensitive the predicted data and error are to small changes in the model.

Like most inversion methods, wave field inversion is predicated on the ability to solve the forward problem; that is, to simulate (predict) the seismic wave field in an arbitrary heterogeneous medium and with a realistic source. Wave field simulation only became practical when the efficiency of computation increased to the point where a complete calculation could be completed in a few hours. Wave field inversion requires many such simulations; the trick is to reduce the number to a manageable level.

A simplistic analysis based on the finite difference approximation indicates that the time needed to compute a full set of partial derivatives might scale with the number of model parameters M in the Earth model. For example, $M + 1$ simulations are need to compute all M elements of g_j:

$$g_j = \frac{\partial E}{\partial m_j} \approx \frac{E(m_j + \Delta m) - E(m_j)}{\Delta m} \quad \text{with} \quad j = 1, \cdots, M \tag{1.2}$$

Wave field inversion currently would be impractical if this was the most efficient possible scaling, because the thousands of parameters needed for realistic Earth models would then imply the need for computing an equal number of wave field simulations (whereas, computing even a few is computationally challenging). Wave field inversion would be limited to a few simplistic cases where the simulation can be computed analytically (such as in homogeneous media [Devaney, 1981]) or where Earth models can be described by just a few parameters (such as layered models [Mellman 1980]).

Adjoint Methods significantly improved the efficiency of the calculation of derivatives, because they allow the computation to be reorganized so as to scale with the number K of receivers, as contrasted to the number M of model parameters. In a typical seismic imaging problem, $K \ll M$.

Adjoint Methods came to seismology via atmospheric science, where they are used to facilitate *data assimilation* - the tuning of the forcing of global circulation models to better match observations [Hall et. al., 1982; Hall and Cacuci, 1983; Talagrand and Courtier, 1987]. Early work on seismic wave field sensitivity kernels by Marquering et al. [1998], Marquering et al. [1999], Dahlen et al. [2000] and Hung et al [2000] did not explicitly utilize Adjoint Methods (though some of their mathematical manipulations are arguably similar to them). Adjoint techniques were first introduced into wave field inversion by Tromp et al. [2005], who cite Talagrand and Courtier's [1987] paper as an inspiration. Subsequent work by Zhao et al. [2005], Van der Hilst and De Hoop [2005], Long et al. [2008], Taillandier et al. [2009], Chen et al. [2010], Xu et al. [2012] have developed and extended Adjoint Method. Early applications wave field imaging applied to seismology include Montelli et al.'s [2006] study of mantle plumes, Chen et al.'s [2007] study of the crust beneath southern California, Chauhan et al.'s [2009] study of the Sumatra subduction zones, and Zhu et al. [2012] study of the European continental mantle,.

We provide a review here of the underlying principles of the Adjoint Method.

Section 2 is devoted to a review of the key concepts of functional analysis and seismic inversion, using mathematical notation that balances compactness with familiarity. Our review of functional analysis includes linear operators and their adjoints and the attributes that make them useful to wave field inversion. The most important relationships are derived and intuitive justifications are provided for most of the rest. Adjoints of selected linear operators are derived in Appendix A.1. A simple example is used to illustrate the potential of Adjoint Methods to improve the efficiency of seismic inversion problems. Our review of inversion includes a discussion of model parameterizations, distinguishes Fréchet derivatives from ordinary partial derivatives, and identifies the cases where their respective use is appropriate. Finally, the role of the Born in calculating perturbations to the wave field is introduced and two complementary derivations are provided.

Section 3 reviews the application of the Adjoint Method of *waveform inversion*, that is, special case where the data are the displacement time series, itself, as contrasted to some quantity derived from it (such as a finite-frequency travel time). The least squares error is defined and formulas for the partial derivative of waveform and error with respect to a model parameter and their corresponding Fréchet derivative are derived. The concept of an adjoint field is developed. A direct method is used in the derivations, but the use of an implicit method based on Lagrange multipliers is explored in Appendix A.2. Section 4 applies the results of Section 3 to the simple case of a scalar wave field in a weakly heterogeneous medium with a homogeneous background slowness. The spatial patterns of the partial derivatives are illustrated and its relationship to the seismic migration method is developed.

Section 5 reviews the application of Adjoint Method to finite frequency travel times. Finite frequency travel time is defined and a perturbation technique is used to derive its partial derivative with respect to a model parameter. An Adjoint Method is then used to derive formulas for the partial derivative of error with respect to a model parameter. Section 6 the results of Section 5 areplued to the scalar wave field case. The spatial patterns of the partial derivative of error are illustrated and their interpretation as a banana-doughnut kernel is developed.

Section 7 applies the Adjoint Method to the cross-convolution measure, an error-like quantity that is used in receiver function and shear wave splitting imaging, because it is relatively insensitive to the poorly-known source time function of the teleseismic wave field. Formulas for the partial derivates are developed. Section 8 applies the results of Section 6 to the simple case of an elastic wave field in a weakly heterogeneous medium with homogeneous background slowness. The spatial patterns of the partial derivatives are illustrated and their connection to the issue of model resolution is developed.

2. Review of Concepts

2.1. Linear Operators. The word *adjoint* comes from the mathematical theory of linear operators [e.g.. Reed and Simon, 1981]. Linear operators, denoted by $\mathcal{L}$'s, include multiplication by functions, derivatives, integrals and other operations that obey the rule $\mathcal{L}(c_1 f_1 + c_2 f_2) = c_1 \mathcal{L} f_1 + c_2 \mathcal{L} f_2$ (where the $c's$ are constants and the f's are functions).

Linear operators act on functions of position and time and are themselves functions of position and time. Often, we will need to refer to several sets of position and time (e.g. of an observation, of a source) and so adopt the practice of distinguishing them with subscripts; that is, $(\mathbf{x}_\mathrm{A}, t_A)$ and $(\mathbf{x}_\mathrm{B}, t_B)$**.** Furthermore**,** we simplify expressions by abbreviating the functional dependence with a subscript; that is, $f_A \equiv f(\mathbf{x}_\mathrm{A}, t_A)$, $G_{A,B} \equiv G(\mathbf{x}_A, t_A, \mathbf{x}_B, t_B)$, etc.

The exemplary expression $f_A = \mathcal{L}_A\, u_A$ can be interpreted as generating a function f_A from a function u_A through the action of a linear operator $\mathcal{L}_A$. It is analogous to the linear algebraic equation $\mathbf{f} = \mathbf{Lu}$, where $\mathbf{u}$ and $\mathbf{f}$ are time series (vectors) of length M and where $\mathbf{L}$ is a $M \times M$ matrix. The equation $f_A = \mathcal{L}_A\, u_A$ can be thought of as the limiting case of $\mathbf{f} = \mathbf{Lu}$ when $M \to \infty$ and the time series become functions. Linear operators are important in seismology because the wave equation and its solution in terms of Green functions involve linear operators. This mathematical structure is exemplified by the scalar wave equation for an isotropic homogeneous material (which has one material parameter, the constant slowness s):

$$\mathcal{L}_A\, u_A \equiv \left(s^2 \frac{\partial^2}{\partial {t_A}^2} - {\nabla_A}^2 \right) u_A = f_A \tag{2.1a}$$

$$u_A = \mathcal{L}_A^{-1} f_A \equiv \int\limits_{t_B} \iiint\limits_{\mathbf{x}_B} F_{A,B}\, f_B\, d^3\mathbf{x}_B\ dt_B$$

$$(2.1b)$$

$$\text{with} \quad F_{A,B} \equiv \frac{\delta(t_A - t_B - T_{AB})}{4\pi R_{AB}} \quad \text{with} \quad R_{AB} \equiv |\mathbf{x}_A - \mathbf{x}_B| \quad \text{and} \quad T_{AB} \equiv sR_{AB} \tag{2.1c}$$

Here, $F_{A,B}$ is the Green function for a observer at $(\mathbf{x}_A, t_A)$ and an impulsive point source at $(\mathbf{x}_B, t_B)$ and the Dirac impulse function is denote by $\delta(.)$. The Green function integral is the inverse operator $\mathcal{L}_A^{-1}$ of $\mathcal{L}_A$ (in the sense that $\mathcal{L}_A$ generates f_A from u_A, whereas $\mathcal{L}_A^{-1}$ generates u_A from f_A.

A linear operator may need to include one or more boundary conditions in order to be fully defined and to possess an inverse. For instance, the simple first derivative equation $f_A = \mathcal{L}_A u_A = du_A/dt_A$ needs to be supplemented by the initial condition $u_A(t_A = 0) = 0$ in order for its inverse to be the integral $u_A = \mathcal{L}_A^{-1} f_A = \int_0^{t_A} f_B \, dt_B$.

The generalization to the three-component particle displacement field $\mathbf{u} = [u_x, u_y, u_z]^{\mathrm{T}}$ that is commonly used in seismology is algebraically complicated but straightforward. The equations of motion combine Newton's law, $\rho \, d\dot{u}_i/dt - \sigma_{ij,j} = f_i$ (where ρ is density and σ is stress) with Hooks' Law $\sigma_{ij} = c_{ijpq} u_{p,q}$ (where c_{ijpq} is the elastic tensor) to yield the second-order matrix differential equation $\boldsymbol{\mathcal{L}}_A \mathbf{u}_A = \mathbf{f}_A$. In the isotropic case with Lamé parameters λ and μ, the operator $\boldsymbol{\mathcal{L}}_A$ is 3×3:

$$\boldsymbol{\mathcal{L}}_A\mathbf{u} = \left(\boldsymbol{\mathcal{L}}_A^{(1)} + \boldsymbol{\mathcal{L}}_A^{(2)}\right)\mathbf{u}_A = \mathbf{f}_A = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix}_A$$

$$\boldsymbol{\mathcal{L}}_A^{(1)} = \begin{bmatrix} \rho\frac{\partial^2}{\partial t^2} - (\lambda+2\mu)\frac{\partial^2}{\partial x^2} - \mu\frac{\partial^2}{\partial y^2} - \mu\frac{\partial^2}{\partial z^2} & -(\lambda+\mu)\frac{\partial^2}{\partial x\partial y} & -(\lambda+\mu)\frac{\partial^2}{\partial x\partial z} \\ -(\lambda+\mu)\frac{\partial^2}{\partial x\partial y} & \rho\frac{\partial^2}{\partial t^2} - \mu\frac{\partial^2}{\partial x^2} - (\lambda+2\mu)\frac{\partial^2}{\partial y^2} - \mu\frac{\partial^2}{\partial z^2} & -(\lambda+\mu)\frac{\partial^2}{\partial y\partial z} \\ -(\lambda+\mu)\frac{\partial^2}{\partial x\partial z} & -(\lambda+\mu)\frac{\partial^2}{\partial y\partial z} & \rho\frac{\partial^2}{\partial t^2} - \mu\frac{\partial^2}{\partial x^2} - \mu\frac{\partial^2}{\partial y^2} - (\lambda+2\mu)\frac{\partial^2}{\partial z^2} \end{bmatrix}$$

$$\boldsymbol{\mathcal{L}}_A^{(2)} = -\begin{bmatrix} \frac{\partial(\lambda+2\mu)}{\partial x}\frac{\partial}{\partial x} + \frac{\partial\mu}{\partial y}\frac{\partial}{\partial y} + \frac{\partial\mu}{\partial z}\frac{\partial}{\partial z} & \frac{\partial\lambda}{\partial x}\frac{\partial}{\partial y} + \frac{\partial\mu}{\partial y}\frac{\partial}{\partial x} & \frac{\partial\lambda}{\partial x}\frac{\partial}{\partial z} + \frac{\partial\mu}{\partial z}\frac{\partial}{\partial x} \\ \frac{\partial\lambda}{\partial y}\frac{\partial}{\partial x} + \frac{\partial\mu}{\partial x}\frac{\partial}{\partial y} & \frac{\partial\mu}{\partial x}\frac{\partial}{\partial x} + \frac{\partial(\lambda+2\mu)}{\partial y}\frac{\partial}{\partial y} + \frac{\partial\mu}{\partial z}\frac{\partial}{\partial z} & \frac{\partial\lambda}{\partial y}\frac{\partial}{\partial z} + \frac{\partial\mu}{\partial z}\frac{\partial}{\partial y} \\ \frac{\partial\lambda}{\partial z}\frac{\partial}{\partial x} + \frac{\partial\mu}{\partial x}\frac{\partial}{\partial z} & \frac{\partial\lambda}{\partial z}\frac{\partial}{\partial y} + \frac{\partial\mu}{\partial y}\frac{\partial}{\partial z} & \frac{\partial\mu}{\partial x}\frac{\partial}{\partial x} + \frac{\partial\mu}{\partial y}\frac{\partial}{\partial y} + \frac{\partial(\lambda+2\mu)}{\partial z}\frac{\partial}{\partial z} \end{bmatrix} \qquad (2.2)$$

Here, we have written the operator $\boldsymbol{\mathcal{L}}$ as the sum of a term $\boldsymbol{\mathcal{L}}^{(1)}$ that does not contain derivatives of the material parameters and a term $\boldsymbol{\mathcal{L}}^{(2)}$ that does. We have also suppressed the subscript A on the derivatives to improve readability of the matrices.

Some authors use two coupled first-order equations, in particle velocity and strain, rather than the single second-order equation, above. The combined matrix equation is larger but algebraically simpler and more amenable to numerical integration.

2.2. The Inner Product. Central to the theory of linear operators is the concept of the inner product, which computes a number q from an arbitrary pair of functions u_A and v_A:

$$q = \int_{t_A} \iiint_{\mathbf{x}_A} u_A \, v_A \, d^3\mathbf{x}_A \; dt_A \equiv \langle u_A, v_A \rangle_A \tag{2.3}$$

The angle brackets provide a compact way of writing the inner product. The subscript A in $\langle .,. \rangle_A$ indicates that the integration is over $(\mathbf{x}_A, t_A)$. The location of the comma is significant only when its arguments are more complicated than simple functions. For instance, $\langle \mathcal{L}_A u_A, v_A \rangle_A$ implies that the linear operator $\mathcal{L}_A$ is applied to u_A but not v_A. The inner product of functions u_A and v_A is analogous to the dot product $s = \sum_i u_i v_i = \mathbf{u}^{\mathrm{T}}\mathbf{v}$ of vectors **u** and **v.** Furthermore, just as $\ell^2 = \mathbf{u}^{\mathrm{T}}\mathbf{u}$ is the squared length of the vector **u** and $d^2 = (\mathbf{u} - \mathbf{v})^{\mathrm{T}}(\mathbf{u} - \mathbf{v})$ is the squared distance between vectors **u** and **v,** so $\ell^2 = \langle u_A, u_A \rangle_A$ can be thought of as the squared length of the function u_A and $d^2 = \langle u_A - v_A, u_A - v_A \rangle_A$ can be thought of the distance between the two functions u_A and v_A. Thus, like the dot product, the inner product is very useful in quantifying sizes and distances. An important inner product in seismology is the waveform error $E_T = \langle u_A^{obs} - u_A, u_A^{obs} - u_A \rangle_A$, which defines the total error (misfit) between and observed and predicted wave fields.

In the case of a vector field, the inner product is the integral of the dot product of the fields:

$$q = \int_{t_A} \iiint_{\mathbf{x}_A} [\mathbf{u}_\mathrm{A}]^{\mathrm{T}} \, \mathbf{v}_\mathrm{A} \, d^3\mathbf{x}_A \; dt_A \equiv \langle \mathbf{u}, \mathbf{v} \rangle_A \tag{2.4}$$

2.3. The Adjoint of a Linear Operator. One or both of the arguments of an inner product can involve a linear operator $\mathcal{L}_A$ - for example, $\langle \mathcal{L}_A u_A, v_A \rangle_A$. This situation is analogous to a dot product containing a matrix **L** - for example $(\mathbf{Lu})^{\mathrm{T}}(\mathbf{v})$. In the latter case, the transposition operator can be used to “move” the matrix from one part of the dot product to the other, in the sense that $(\mathbf{Lu})^{\mathrm{T}}(\mathbf{v}) = (\mathbf{u})^{\mathrm{T}}(\mathbf{L}^{\mathrm{T}}\mathbf{v})$. The *adjoint operator*, which is denoted with the dagger symbol †, moves the linear operator from one side of the inner product to the other in a completely analogous way: $\langle \mathcal{L}_A u_A, v_A \rangle_A = \langle u_A, \mathcal{L}_A^{\dagger} v_A \rangle_A$. Just as $\mathbf{L}^{\mathrm{T}}$ is a different matrix from **L**, but constructed from it in a known way (that is, by swapping rows and columns), so the operator $\mathcal{L}_A^{\dagger}$ is different from the operator $\mathcal{L}_A$ but constructed from it in a known way (though in a way more complicated than for a matrix). Thus, far from being mysterious, the adjoint is just a $M \to \infty$ limiting case of a matrix transpose. Adjoints obey almost all of the same algebraic rules as do transposes, including:

$$\mathcal{L}_A^{\dagger\dagger} = \mathcal{L}_A \quad \text{and} \quad \left(\mathcal{L}_A^{\dagger}\right)^{-1} = \left(\mathcal{L}_A^{-1}\right)^{\dagger} \quad \text{and} \quad \left(\mathcal{L}_A^{(1)}\mathcal{L}_A^{(2)}\right)^{\dagger} = \left(\mathcal{L}_A^{(2)}\right)^{\dagger}\left(\mathcal{L}_A^{(1)}\right)^{\dagger} \tag{2.5}$$

Just as a matrix that obeys $\mathbf{L}^{\mathrm{T}} = \mathbf{L}$ or $\mathbf{L}^{\mathrm{T}} = -\mathbf{L}$ is respectively called symmetric or anti-symmetric, so an operator that obeys $\mathcal{L}_A^{\dagger} = \mathcal{L}_A$ or $\mathcal{L}_A^{\dagger} = -\mathcal{L}_A$ is respectively called *self-adjoint* or *anti-self-adjoint*. A few simple cases are (see Appendix A.1):

$\mathcal{L}u$	$\mathcal{L}^{\dagger}u$	2.6a
$c(\mathbf{x}, t)\ u$	self-adjoint	2.6b

du/dt and du/dx	anti-self-adjoint	2.6c
d^2u/dt^2 and d^2u/dx^2	self-adjoint	2.6d
$\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$	$\begin{bmatrix} \mathcal{L}_{11}^{\dagger} & \mathcal{L}_{12}^{\dagger} \\ \mathcal{L}_{21}^{\dagger} & \mathcal{L}_{22}^{\dagger} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$	2.6e
$a(t) * u$	$a(t) \star u$	2.6f
$\boldsymbol{\mathcal{L}}\mathbf{u}$ of elastic wave equation	self-adjoint	2.6g
$\langle \mathcal{F}_{B,A}, u_A \rangle_A$	$\langle \mathcal{F}_{A,B}^{\dagger}, u_A \rangle_A$	2.6g

Here $*$ signifies convolution and $\star$ signifies cross-correlation. Taking the adjoint of a first derivative reverse the sense of direction of the independent variable, since $-d/dt = d/d(-t)$ and $-d/dx = d/d(-x)$ This effect is more important for the time than for space, because the time boundary condition is usually asymmetric (the past is quiescent but the future is not), while the space boundary condition is usually symmetric (the field approaches zero as $x \to \pm\infty$). Consequently, manipulations of equations using adjoints often lead to behaviors that are "backward in time" (see Appendix A.1).

2.4. Applications of Adjuncts. Two factors combine to make Adjoint Methods especially useful in seismology. First, observations often involve a wave field u that obeys a differential equation $\mathcal{L}_A u_A = f_A$ (where f_A is a source term), so a linear operator $\mathcal{L}_A$ is associated with the problem. Second, the formulas that link the field u to observations and to observational error involve inner products.

To see why this combination of factors might be useful, consider the case where a set of N observations d_i^{obs} are related to the field u by the inner product [see Menke, Section 11.11, 2012]; that is, the predicted data is:

$$d_i = \langle h_{Ai}, u_A \rangle_A \tag{2.7}$$

Here, $h_{Ai} \equiv h_i(\mathbf{x}_A, t_A)$ are known functions and u_A is the wave field. Now suppose that we want to "tune" the source so that the observations are matched (meaning we inverting for the source f_A). A perturbation δf_A in the source causes a perturbation δu_A in the field which, in turn, causes a perturbation δd_i in the data. Because of the linearity of the system:

$$\mathcal{L}\, \delta u_A = \delta f_A \quad \text{and} \quad \delta d_i = \langle h_{Ai}, \delta u_A \rangle_A \tag{2.8}$$

Writing the solution of the differential equation as $\delta u_A = \mathcal{L}^{-1} \delta f_A$ and inserting it into the inner product yields:

$$\delta d_i = \langle h_{Ai}, \mathcal{L}^{-1} \delta f_A \rangle_A = \langle h_{Ai}, \delta u_A \rangle_A \quad \text{with} \quad \mathcal{L}\, \delta u_A = \delta f_A \tag{2.9}$$

This equation reads: to determine the perturbation δd_i in the data, solve the wave equation with a source perturbation δf_A to determine the field perturbation δu_A and then take the inner product of δu_A with the function h_{Ai}. The differential equation must be solved for every source perturbation δf_A that is considered (let's suppose that there are M of them), but once these solutions are determined, they can be applied to any number of data. Now, suppose that we manipulate the inner product:

$$\delta d_i = \langle \mathcal{L}^{\dagger -1} h_{Ai}, \delta f_A \rangle_A \equiv \langle \lambda_{Ai}, \delta f_A \rangle_A \quad \text{with} \quad \mathcal{L}^{\dagger} \lambda_{Ai} = h_{Ai} \tag{2.10}$$

Here we have introduced the *adjoint field* $\lambda_{Ai} \equiv \lambda_i(\mathbf{x}_\mathrm{A}, t_A)$ as an abbreviation for $\mathcal{L}^{\dagger -1} h_{Ai}$. This equation reads: to determine the perturbation δd_i in the data, solve the adjoint differential equation with source term h_{Ai} to determine the adjoint field λ_{Ai} and then take the inner product of λ_{Ai} with the source perturbation δf_A. The adjoint differential equation needs be solved N times (once for each datum), but once these solutions are determined, they can be applied to any number of source perturbations.

As an aside, we mention that the adjoint field plays the role of a *data kernel* G_{Ai} linking perturbations in data to perturbations in unknowns, that is $\delta d_i = \langle G_{Ai}, \delta f_A \rangle_A$ with $G_{Ai} \equiv \lambda_{Ai}$.

In many practical problems, $M \gg N$, so the adjoint formulation is preferred. The advantage is one of efficiency, only; both approaches lead to same solution. However, the value of efficiency must not be underrated, for many problems in seismology become tractable *only* because of Adjoint methods.

In seismology, this procedure can be used to determine the earthquake source, as quantified by its moment tensor density, from observed seismic waves [Kim et al., 2011]. The function h_{Ai} in Equation (2.7) is then the Dirac delta function $\delta(\mathbf{x}_R - \mathbf{x}_A)\delta(t_R - t_A)$; that is, the predicted data d_i is the field u_i observed at time t_R by a receiver at $\mathbf{x}_R$.

2.5. The Fréchet derivative. The equation $\delta d_i = \langle G_{Ai}, \delta f_A \rangle_A$ is very similar in structure to the first order perturbation rule for a set of analogous vector quantities $\Delta\mathbf{d}$ and $\Delta\mathbf{f}$:

$$\Delta d_i = \left(\mathbf{G}^{(i)}\right)^{\mathrm{T}} \Delta\mathbf{f} = \sum_j \frac{\partial d_i}{\partial f_j} \Delta f_j \quad \text{with} \quad \frac{\partial d_i}{\partial f_j} \equiv G_{ij} \tag{2.11}$$

The only differences are that the vector $\Delta\mathbf{f}$ has been replaced by the function δf_A and the summation has been replaced by integrals. Consequently, the rule $\delta d_i = \langle G_{Ai}, \delta f_A \rangle_A$ can be thought of defining a kind of derivative:

$$\delta d_i = \langle G_{Ai}, \delta f_A \rangle_A \equiv \int_{t_A} \iiint_{\mathrm{x}_\mathrm{A}} \frac{\delta d_i}{\delta f_A} \, \delta f_A \, d^3\mathbf{x}_\mathrm{A} \; dt_A \equiv \langle \frac{\delta d_i}{\delta f}, \delta f_A \rangle_A \quad \text{with} \quad \frac{\delta d_i}{\delta f_A} \equiv G_{Ai} \tag{2.12}$$

This so-called *Fréchet derivative* $\delta d_i / \delta f_A$ is distinguished from a partial derivative by the use of $\delta' s$ in place of $\partial' s$. Partial derivative Fréchet derivatives find many uses, especially because they obey the *chain rule*:

$$\frac{\delta d_i}{\delta f_A} = \langle \frac{\delta d_i}{\delta u_B}, \frac{\delta u_B}{\delta f_A} \rangle_B \tag{2.13}$$

Here u_B is an arbitrary function of space and time. The manipulation of expressions into a form that identifies a Fréchet derivative (as in the case above) is another important application of Adjoint methods.

2.6. Model Parameterization. In the case discussed above, the source perturbation δf_A is treated as the unknown. Far more common in seismology is the case where the material parameters that appear in $\mathcal{L}_A$, such as elastic constants and density, are the unknowns. An important question is whether these parameters should be described by a spatially (and possibly temporally) varying function, say $m_A \equiv m(\mathbf{x}, t)$, (as was done previously for the source) or by a set of discrete parameters $m_i, i = 1 \dots M$ that multiply a set of prescribed spatial-temporal patterns. (say $p_A^{(i)} = p_i(\mathbf{x}_A, t_A)$):

$$\delta m(\mathbf{x}_A, t_A) = \sum_{i=1}^{M} m_i \, p^{(i)}(\mathbf{x}_A, t_A) \quad \text{or} \quad \delta m_A = \sum_{i=1}^{M} m_i \, p_A^{(i)} \tag{2.14}$$

The issue is where in the solution process the transition should be made from a continuous view of the world, which is realistic but unknowable, to a discrete view, which is approximate but computable. The first approach starts with the derivation of Fréchet derivatives and converts them to partial derivatives only at last resort. The second approach uses only partial derivatives throughout. We review both approaches here, because both are used in the literature.

2.7. The Born Approximation. The differential equation $\mathcal{L}_A(m)\, u_A = f_A$ is not, in general, linear in a material parameter m, so only an approximate equation can be derived that links a perturbation is the material parameter to a perturbation in the field. This is in contrast to the case of the unknown source, in which the equation $\mathcal{L}_A \delta u_A = \delta f_A$ is exact. Here we examine the case for a single discrete parameter m_1 (for which we subsequently drop the subscript). The result can be generalized to multiple discrete parameters m_i merely by adding a substituting m_i for m. The generalization to the continuous case is somewhat more complicated and will be derived later. We compare two different approaches to deriving this equation, which, as we will discover, yield the same result.

The first approach starts with the equation $u_A = \mathcal{L}_A^{-1}\, f_A$ and differentiates it with respect to m around a point m^0

$$\left.\frac{\partial u_A}{\partial m}\right|_{m^0} = \left.\frac{\partial(\mathcal{L}_A^{-1}\, f_A)}{\partial m}\right|_{m^0} = \frac{\partial}{\partial m}\mathcal{L}_A^{-1}\,\Big|_{m^0} f_A =$$
$$-\mathcal{L}_A^{-1}|_{m^0} \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} \mathcal{L}_A^{-1}|_{m^0}\, f_A = -\mathcal{L}_A^{-1}|_{m^0} \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} u_A^0 \equiv G_A \tag{2.15}$$

Note that f_A is not a function of m, so that $\partial f_A/\partial m = 0$, that $u_A^0 = \mathcal{L}_A^{-1}|_{m^0} f_A$ is the solution to the *unperturbed* equation $\mathcal{L}_A|_{m^0}\, u_A^0 = f_A$, G is the data kernel (the partial derivative of the field with respect to a model parameter), and the derivation uses the derivative rule (Appendix A.2):

$$\frac{\partial}{\partial m}\mathcal{L}_A^{-1} = -\mathcal{L}_A^{-1}\frac{\partial \mathcal{L}_A}{\partial m}\mathcal{L}_A^{-1} \tag{2.16}$$

The second approach starts with the wave equation $\mathcal{L}_A(m)\, u_A = f_A$, represents the field $u_A = u_A^0 + \Delta u_A$ as the sum of an unperturbed part u_A^0 and a perturbation Δu_A and expands the operator $\mathcal{L}(m)$ around the point m^0, discarding terms higher than first-order:

$$\mathcal{L}_A(m) = \mathcal{L}_A|_{m^0} + \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} \Delta m \tag{2.17}$$

Inserting these representations into the wave equation and keeping only first order terms (the *Born approximation*) yields:

$$f_A = \left(\mathcal{L}_A|_{m^0} + \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} \Delta m\right)(u_A^0 + \Delta u_A\ ) = \mathcal{L}_A|_{m^0} u_A^0 + \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} u_A^0 \Delta m + \mathcal{L}_a|_{m^0}\, \Delta u_A \tag{2.18}$$

Subtracting out the unperturbed equation and rearranging yields:

$$\Delta u_A = \left(-\mathcal{L}_A^{-1}|_{m^0} \left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_{m^0} u_A^0\right) \Delta m \equiv \frac{\partial u_A}{\partial m}\ \Delta m \equiv G_A\, \Delta m \tag{2.19}$$

We can now identify the data kernel $G_A \equiv \partial u_A/\partial m$ as the factor in the parentheses and see that is it's the same formula that was derived by the first approach. That these two approaches lead to the same formula is unsurprising, since both are based on first-order approximations of the same equations.

2,8. An Exemplary Partial Derivative of an Operator. The partial derivative $\partial \mathcal{L}_A/\partial m$ may at first seem mysterious, but an example demonstrates that it is completely straightforward. Consider the special case of a scalar wave equation with slowness $s_A = s_{0A} + \delta s_A$, where the unperturbed slowness s_{0A} and the perturbation δs_A are both spatially-variable functions. We parameterize $\delta s_A = m\, p_A$, where p_A is prescribed "pattern" function and m is a scalar amplitude parameter. The linear operator in the wave equation is then:

$$\mathcal{L}_A(m) = [s_{0A} + \delta s_A]^2 \frac{\partial^2}{\partial t_A^2} - \nabla_{\mathrm{A}}^2 \approx [s_{0A}^2 + 2 s_{0A} m\, p_A] \frac{\partial^2}{\partial t_A^2} - \nabla_{\mathrm{A}}^2 \tag{2.20}$$

Taking the partial derivative with respect to m and evaluating it at $m = m^0 = 0$ yields:

$$\left.\frac{\partial \mathcal{L}_A}{\partial m}\right|_0 = 2 s_{0A} p_A \frac{\partial^2}{\partial t_A^2} \tag{2.21}$$

2.9. Relationship between a partial derivative and a Fréchet derivative. Suppose that the model is parameterized as $\delta m_A = p_A\, \Delta m$ where p_A is a prescribed spatially and temporally varying pattern and Δm is a scalar. Inserting this form of δm into the Fréchet derivative $\delta u_A = \langle \delta u_A / \delta m_B \, , \delta m_B \rangle_B$ yields:

$$\delta u_A = \langle \frac{\delta u_A}{\delta m_B}, p_B \rangle_B \ \Delta m = \frac{\partial u_A}{\partial m} \ \Delta m \quad \text{with} \quad \frac{\partial u_A}{\partial m} \equiv \langle \frac{\delta u_B}{\delta m}, p_B \rangle_B \tag{2.22}$$

Evidentially, the partial derivative can be formed by taking the inner product of the Fréchet derivative with the prescribed pattern. Alternately, suppose that the pattern is temporally- and spatially localized at B; that is $\delta m_{A,B} = m \, \delta(t_A - t_B)\delta(\mathbf{x}_A - \mathbf{x}_B)$ where m is a scalar model parameter Furthermore, suppose that this model function leads to the partial derivative is $\partial u_{A,B}/\partial m$. The effect of many such perturbations, each with its own position $\mathbf{x}_B$, time t_B, and amplitude δm_B, is the superposition (integral) of the individual ones:

$$(\delta u_A)^{\text{total}} = \langle \frac{\partial u_{A,B}}{\partial m}, \delta m_B \rangle_B \quad \text{with} \quad \frac{\delta u_A}{\delta m_B} \equiv \frac{\partial u_{A,B}}{\partial m_B} \tag{2.23}$$

Evidentially, the Fréchet derivative is just the partial derivative for a temporally- and spatially localized pattern.

3. Waveform Tomography

3.1. Definition of Error. The goal in waveform tomography is to match the predicted field u_A to the observed field u_A^{obs}, by minimizing the total error $E = \langle e_A, e_A \rangle_A$, where $e_A = u_A^{obs} - u_A$.

This optimization problem can be solved using the Gradient-Descent method, which minimizse E by iteratively perturbs an initial estimate of m. It requires the either the partial derivative $\partial E/\partial m$ or the Fréchet derivative $\delta E/\delta m$, depending upon whether the model is respectively represented by discrete parameters or continuous functions.

3.2. Partial Derivative of Error. As before, the predicted field u_A is assumed to arise through the solution of a differential equation $\mathcal{L}_A(m)\, u_A = f_A$ containing a discrete parameter m. A perturbation Δm in parameter m causes a perturbation ΔE in the total error E_T:

$$\Delta E = \left.\frac{\partial E}{\partial m}\right|_{m_0} \Delta m \quad \text{where} \quad \frac{\partial E}{\partial m} = 2\,\langle e_A^0, \left.\frac{\partial e_A}{\partial m}\right|_{m_0} \rangle = -2\,\langle e_A^0, \left.\frac{\partial u_A}{\partial m}\right|_{m_0} \rangle = -2\langle e_A^0, G_A \rangle_A \tag{3.1}$$

We simplify the notation used in subsequent equations by dropping explicit dependence on m_0. Inserting the formula for G_A yields:

$$\frac{\partial E}{\partial m} = -2\langle e_A, G_A \rangle = 2\,\langle e_A^0, \mathcal{L}_A^{-1}\, \frac{\partial \mathcal{L}_A}{\partial m} u_A^0 \rangle_A = \langle 2\,\frac{\partial \mathcal{L}_A^\dagger}{\partial m}\, \mathcal{L}_A^{-1\dagger} e_A^0, u_A^0 \rangle_A = \langle h_A, u_A^0 \rangle_A$$

$$\text{with}\;\; h_A = 2\,\frac{\partial \mathcal{L}_A^\dagger}{\partial m}\, \lambda_A \;\;\text{and}\;\; \lambda_A = \mathcal{L}_A^{-1\dagger} e_A^0 \quad \text{or} \quad \mathcal{L}_A^\dagger\; \lambda_A = e_A^0 \tag{3.2}$$

As before, we have introduced an adjoint field λ_A. The derivative $\partial E/\partial m$ is constructed as follows: First, the adjoint field λ_A is determined by solving the adjoint wave equation, which involves the adjoint operator $\mathcal{L}_A^\dagger$ and has a source term equal to the prediction error e_A^0. Second, the operator $2\,\partial \mathcal{L}_A^\dagger/\partial m$ is applied to the adjoint field to yield the function h_A. Finally, the inner product of the unperturbed field e_A^0 with the function h is computed. This process is often referred to as *correlating* h_A and ${e_A^0}^0$, since it corresponds to their zero-lag cross-correlation.

In many practical cases, we will want to consider the error E_R associated with one receiver point R:

$$E_R = E(\mathbf{x}_R) = \int [u^{obs}(\mathbf{x}_R.t_R) - u(\mathbf{x}_R.t_R)]^2 dt_R \tag{3.3}$$

We now assert that the total error E_T is the superposition of the individual errors E_R, its partial derivative is the superposition of individual partial derivatives, and the adjoint field λ_A is the superposition of individual $\lambda_{A,R}$'s:

$$E_T = \int E_R\ d^3x_R \quad \text{and} \quad \frac{\partial E_T}{\partial m} = \int \frac{\partial E_R}{\partial m}\ d^3x_R \quad \text{and} \quad \lambda_A = \int \lambda_{A,R}\ d^3x_R \tag{3.4}$$

Inserting this definition of $\lambda_{A,R}$ into the differential equation for λ_A yields:

$$\begin{gathered}\mathcal{L}_A^\dagger\ \lambda_A - e_A^0 = 0 \\ \mathcal{L}_A^\dagger\ \int \lambda_{A,R}\ d^3x_R - \int e^0(\mathbf{x}_R, t_A)\ \delta(\mathbf{x}_A - \mathbf{x}_R)\ d^3x_R\ = 0 \\ \int \left\{\mathcal{L}_A^\dagger \lambda_{A,R} - e^0(\mathbf{x}_R, t)\ \delta(\mathbf{x}_A - \mathbf{x}_R)\right\}\ d^3x_R\ = 0\end{gathered} \tag{3.5}$$

The presumption that this equation holds irrespective of the volume over which the error is defined implies that the integrand is zero, so:

$$\mathcal{L}_A^\dagger \lambda_{A,R} = e^0(\mathbf{x}_R, t)\, \delta(\mathbf{x}_A - \mathbf{x}_R) \tag{3.6}$$

Thus, each $\lambda_{A,R}$ corresponds to a point source at $\mathbf{x}_R$ with the time function of the error at that point. Similarly, if we define $h_{A,R} \equiv 2\left[\partial \mathcal{L}_A^\dagger / \partial m\right] \lambda_{A,R}$, then a procedure analogous to the one above can be used to show that:

$$\frac{\partial E_R}{\partial m} = \langle h_{A,R}, u_A^0 \rangle_A \tag{3.7}$$

A typical seismological application might involve $M \approx 10^4$ model parameters but only $K \approx 10^2$ observations points. The adjoint formulation allows all 10^4 partial derivatives (one for each model parameter) to be calculated by solving “only” $K + 1$ differential equations, one to calculate the unperturbed field u^0 and the rest to calculate the adjoint fields, of which there is one for each of the K observation points.

Physically, the adjoint field can be thought of as the scattered field, back-propagated to heterogeneities from which it *might* have originated. Mathematically, the adjoint field can be interpreted as a Lagrange multiplier associated with the constraint that the field obeys a wave equation at every point in space and time (see Appendix A.3).

3.3. Fréchet Derivative of Error. We now present a completely parallel derivation of the Fréchet derivative of the waveform error with respect to a model function m_A. The scalar field u_A satisfies partial differential equation:

$$\mathcal{L}_A(m_A)\, u_A = f_A \tag{3.8}$$

We will write both u_A and m_A in terms of background level and a perturbation:

$$u_A = u_A^0 + \delta u_A \quad and \quad m_A = m_A^0 + \delta m_A \tag{3.9}$$

Functions u_A^0, δu_A, m_A^0 and δm_A all vary with both space and time. However, in most practical cases m_A and δm_A will be constant in time. The background field u_A^0 satisfies the unperturbed equation:

$$\mathcal{L}_A(m_A^0)\, u_A^0 = f_A \tag{3.10}$$

The Fréchet derivative for the total error $E_T = \langle e_B, e_B \rangle_B$ with $e_B = u_B^{obs} - u_B$ and $e_B^0 = u_B^{obs} - u_B^0$, is derived by considering how a perturbation in the field changes the error:

$$\delta E_T = \langle e_B, e_B \rangle_B - \langle e_B^0, e_B^0 \rangle_B = \langle e_B^0 + \delta e_B, e_B^0 + \delta e_B \rangle_B - \langle e_B^0, e_B^0 \rangle_B$$

$$= \langle e_B^0, e_B^0 \rangle_B + 2\langle e_B^0, \delta e_B \rangle_B + \langle \delta e_B, \delta e_B \rangle_B - \langle e_B^0, e_B^0 \rangle_B$$

$$= \langle 2e_B, \delta e_B \rangle_B - \langle \delta u_B, \delta e_B \rangle_B \approx \langle 2e_B^0, \delta e_B \rangle_B$$

(3.11)

Note that we have discarder terms of second order in small quantities. Substituting in $\delta e_B = -\delta u_B$ yields:

$$\delta E_T = \langle -2e_B^0, \delta u_B \rangle_B$$

(3.12)

The next step is to replace δu_B in the above expression with an expression involving δm_B. We start with the Fréchet derivative of the field, which is defined by:

$$\delta u_A = \langle \left[\frac{\delta u}{\delta m}\right]_{A,B}, \delta m_B \rangle_B \equiv \langle G_{A,B}, \delta m_B \rangle_B \equiv \mathcal{G}_A \, \delta m_A$$

(3.13)

Here the inner product with Fréchet derivative $G_{A,B} \equiv [\delta u_A / \delta m_B]$ is understood to be a linear operator $\mathcal{G}_A$. Our derivation requires the Fréchet derivative of the operator $\mathcal{L}_A(m)$. It satisfies:

$$\delta \mathcal{L}_A = \langle \frac{\delta \mathcal{L}_A}{\delta m_B}, \delta m_B \rangle_B$$

(3.14)

As shown previously, this is just the partial derivative of the operator for a heterogeneity temporally- and spatially-localized at B. For example, in the case of the scalar wave equation:

$$\mathcal{L}_A(m_A) = \left\{(m_A^0)^2 \frac{\partial^2}{\partial t_A^2} - \nabla_\mathrm{A}^2\right\} + 2m_A^0 m_B \delta(\mathbf{x}_A - \mathbf{x}_B)\delta(t_A - t_B)\frac{\partial^2}{\partial t_A^2}$$

$$\frac{\delta \mathcal{L}_A}{\delta m_B} = \frac{\partial \mathcal{L}_A}{\partial m_B} = 2m_A^0 \delta(\mathbf{x}_A - \mathbf{x}_B)\delta(t_A - t_B)\frac{\partial^2}{\partial t_A^2}$$

(3.15)

The Fréchet derivative of the field is then derived by applying the Born approximation to the differential equation:

$$\mathcal{L}_A(m_A)\, u_A = f_A$$

$$\left(\mathcal{L}_A(m_A^0) + \ \delta\mathcal{L}_A(m_A)\right)(u_A^0 + \delta u_A) = f_A$$

$$\left(\mathcal{L}_A(m_A^0) + \langle \frac{\delta \mathcal{L}_A}{\delta m_B}, \delta m_B\rangle_B\right)\ (u_A^0 + \delta u_A) = f_A$$

$$\mathcal{L}_A(m_A^0)\, u_A^0 + \mathcal{L}_A(m_A^0)\, \delta u_A + \langle \frac{\delta \mathcal{L}_A}{\delta m_B}, \delta m_B\rangle_B\, u_A^0 \ \approx f_A$$

(3.16)

Subtracting out the unperturbed equation and rearranging yields:

$$\delta u_A = \langle -\mathcal{L}_A^{-1} \frac{\delta \mathcal{L}_A}{\delta m_B} u_A^0, \delta m_B\rangle_B \equiv \langle G_{A,B}, \delta m_B\rangle_B \equiv \mathcal{G}_A\, \delta m_A$$

$$G_{A,B} = -\mathcal{L}_A^{-1} \frac{\delta \mathcal{L}_A}{\delta m_B} u_A^0$$

(3.17)

The Fréchet derivative of the total error with respect to the model is obtained by substituting $\delta u_B = \mathcal{G}_B\, \delta m_B$ into the general expression for this derivative:

$$\delta E_T = \langle -2e_B^0, \mathcal{G}_B\, \delta m_B \rangle_B = \langle -2\mathcal{G}_B^\dagger e_B^0, \delta m_B \rangle_B \tag{3.18}$$

The Fréchet derivative of the total error is then:

$$\frac{\delta E_T}{\delta m_A} = -2\mathcal{G}_A^\dagger e_A^0 = \langle u_B^0, G_{B,\boldsymbol{A}}^\dagger e_B^0 \rangle_B = \langle G_{B,\boldsymbol{A}}^\dagger e_B, u_B^0 \rangle_B \tag{3.19}$$

Substituting in the formula for $G_{B,\boldsymbol{A}}^\dagger$ yields:

$$\frac{\delta E}{\delta m_A} = \langle 2\frac{\delta \mathcal{L}_B^\dagger}{\delta m_A} \mathcal{L}_B^{-1\dagger} e_B^0, u_B^0 \rangle_B = \langle g_{B,A}, u_B^0 \rangle_B$$

$$\text{with} \quad \lambda_B = \mathcal{L}_B^{-1\dagger} e_B^0 \quad \text{or} \quad \mathcal{L}_B^\dagger\ \lambda_B = e_B^0 \quad \text{and} \quad g_{B,A} = 2\left[\frac{\delta \mathcal{L}^\dagger}{\delta m}\right]_{B,A} \lambda_B \tag{3.20}$$

The quantify $\delta \mathcal{L}_B^\dagger / \delta m_A$ has the (A, B) independent variables reversed with respect to $\delta \mathcal{L}_A / \delta m_B$. However, since the Dirac function is symmetric in (A, B), the only effect is to change the independent variables in the rest of the operator from A to B.

In many practical cases, we will want to consider the error E_C associated with a receiver point R:

$$E_R \equiv \int \left[u_R^{obs} - u_R\right]^2 dt_R$$

(3.21)

The total error is then the superposition of the individual errors, its Fréchet derivative is the superposition of individual derivatives, and λ_B is a superposition of individual $\lambda_{B,R}$'s

$$E_T = \int E_R \; d^3x_R \quad \text{and} \quad \frac{\delta E_T}{\delta m_A} = \int \frac{\delta E_R}{\delta m_A} \, d^3x_R \quad \text{and} \quad \lambda_B = \int \lambda_{B,R} \; d^3x_R \tag{3.22}$$

Inserting this definition of $\lambda_{B,R}$ into the differential equation for λ_B yields:

$$\begin{gathered} {\mathcal{L}_B}^{\dagger} \; \lambda_B - e^0(\mathbf{x}_R, t) = 0 \\ \int {\mathcal{L}_B}^{\dagger} \, \lambda_{B,R} \; d^3x_R - \int e^0(\mathbf{x}_R, t) \, \delta(\mathbf{x}_B - \mathbf{x}_R) \, d^3x_R = 0 \\ \int \left\{ {\mathcal{L}_B}^{\dagger} \, \lambda_{B,R} - e^0(\mathbf{x}_R, t) \, \delta(\mathbf{x}_B - \mathbf{x}_R) \right\} d^3x_R = 0 \end{gathered} \tag{3.23}$$

The presumption that this equation holds irrespective of the volume over which the error is defined implied that the integrand is zero, so:

$${\mathcal{L}_B}^{\dagger} \, \lambda_{B,R} = e(\mathbf{x}_R, t) \, \delta(\mathbf{x}_B - \mathbf{x}_R) \tag{3.24}$$

Thus, each $\lambda_{B,R}$ corresponds to a point source at $\mathbf{x}_R$ with the time function of the error at that point. Similarly, if we define $g_{B,A,R} \equiv 2[\delta\mathcal{L}^{\dagger}/\delta m]_{B,A} \, \lambda_{B,R}$, then a procedure analogous to the one above can be used to show that:

$$\frac{\delta E_R}{\delta m_A} = \langle u_B^0, g_{B,,A,R} \rangle_B \tag{3.25}$$

These formula are very similar to the partial derivate case derived previously.

4. An Example Using the Scalar Wave Equation

4.1. The Partial Derivative of the Field With Respect to a Model Parameter. In the first part of this derivation, we pursue the strategy of explicitly calculating $u_R(m)$, where R is a receiver point and m is a scalar parameter, using the Born approximation. When then differentiate it to find the derivative $\partial u_R / \partial m|_{m=0}$, and use this derivative to infer $\partial E_T / \partial m|_{m=0}$. The advantage of this approach is that it allows terms in the formula for $\partial E_T / \partial m$ to be interpreted in terms of scattering interactions.

The scalar wave equation for an isotropic medium with constant slowness s_0 and a source that is spatially-localized at $\mathbf{x}_s$ and has time function $f(t)$ is:

$$s_0^2 \ddot{u}_R^0 - \nabla^2 u_R^0 = \delta(\mathbf{x}_R - \mathbf{x}_S)\, f(t_R) \tag{4.1}$$

It has solution:

$$u_R^0 = \frac{f(t \pm T_{RS})}{4\pi R_{RS}} \quad \text{with} \quad R_{RS} = |\mathbf{x}_R - \mathbf{x}_S| \quad \text{and} \quad T_{RS} = s_0 R_{RS} \tag{4.2}$$

Here R_{RS} is the distance between $\mathbf{x}_R$ and $\mathbf{x}_S$ and T_{RS} is the corresponding travel time. The initial condition that u^0 has a quiescent past selects the forward-in-time solution ($-$ of the $\pm$) and the condition that it has a quiescent future selects the backwards-in-time solution ($+$ of the $\pm$).

Suppose that the slowness of the medium has the form $s_R = s_0 + \delta s_R$ where s_0 is a constant background level and δs_R is small perturbation representing spatially variable heterogeneities. The quantity s_R^2, which appears in the wave equation, is approximately:

$$s_R^2 = s_0^2 \left(1 + \frac{\delta s_R}{s_0}\right)^2 \approx s_0^2 \left(1 + 2\frac{\delta s_R}{s_0}\right) = s_0^2 + 2 s_0 \delta s_R \tag{4.5}$$

The corresponding scalar field is $u_R = u_R^0 + \delta u_R$, where u_R^0 solves the constant-slowness wave equation and where δu arises because the slowness field is slightly heterogeneous. Inserting this representation into the wave equation, keeping terms only to first order, and subtracting out the homogenous equation yields the Born approximation:

$$\begin{gathered}
(s_0^2 + 2 s_0 \delta s_R)\left(\ddot{u}_R^0 + \delta \ddot{u}_R\right) - \nabla_R^2 (u_R^0 + \delta u_R) = \delta(\mathbf{x}_R - \mathbf{x}_S)\, f(t_R) \\
s_0^2 \ddot{u}_R^0 + 2 s_0 \ddot{u}_R^0 \delta s_R + s_0^2\left(\ddot{u}_R^0 + \delta \ddot{u}_R\right) - \nabla_R^2 u_R^0 - \nabla_R^2 \delta u_R = \delta(\mathbf{x}_R - \mathbf{x}_S)\, f(t_R) \\
s_0^2 \delta \ddot{u}_R - \nabla_R^2 \delta u_R = -2 s_0 \ddot{u}_R^0 \delta s_R
\end{gathered} \tag{4.6}$$

The field perturbation δu_R solves a constant-slowness wave equation with a complicated source term. Now suppose that we consider appoint-like heterogeneity of strength m localized at position $\mathbf{x}_H$:

$$\delta s_R = m\,\delta(\mathbf{x}_R - \mathbf{x}_H) \tag{4.7}$$

Substituting this expression into the Born approximation yields:

$$s_0^2 \delta\ddot{u} - \nabla^2 \delta u = -2 s_0 m\, \ddot{u}_R^0\, \delta(\mathbf{x}_R - \mathbf{x}_H) \tag{4.8}$$

This is a constant-slowness wave equation and has solution:

$$\delta u_R = -2 s_0 m\, \frac{\ddot{u}^0(\mathbf{x}_H, t_R - T_{RH})}{4\pi R_{RH}} \tag{4.9}$$

Differentiating u with respect to m yeilds:

$$\frac{\partial u_R}{\partial m} = \frac{\partial}{\partial m} \delta u_R = -2 s_0\, \frac{\ddot{u}^0(\mathbf{x}_H, t_R - T_{RH})}{4\pi R_{RH}} \tag{4.10}$$

4.2. The Partial Derivative of Error With Respect to a Model Parameter. Now suppose we have an observation u_R^{obs} for some fixed observer location R. The error E is defined as:

$$E_R = \int (e_R^0)^2\, dt_R \quad \text{with} \quad e_R^0 = u_R^{obs} - u_R^0 \tag{4.11}$$

The derivative is:

$$\frac{\partial E_R}{\partial m} = -2 \int e_R^0 \frac{\partial u_R}{\partial m} \, dt \tag{4.12}$$

Inserting Equation (4.10) yields:

$$\begin{aligned}
\frac{\partial E_R}{\partial m} &= 4 s_0 \ \frac{1}{4\pi R_{RH}} \int e_R^0 \ \ddot{u}^0(\mathbf{x}_H, t_R - T_{RH}) \, dt_R \\
&= 4 s_0 \ \frac{1}{4\pi R_{RH}} \frac{1}{4\pi R_{HS}} \int e_R^0 \ \ddot{f}(t_R - T_{RH} - T_{HS}) \, dt_R \\
&= 4 s_0 \ \frac{1}{4\pi R_{RH}} \frac{1}{4\pi R_{HS}} \int \ddot{e}_R^0 \ f(t_R - T_{RH} - T_{HS}) \, dt_R
\end{aligned} \tag{4.13}$$

The last form is derived by noting that the integral is an inner product and that the time derivative, which is self-adjoint, can be moved from f to $(u^{obs} - u)$. The source time function f is propagated outward from the source, scattered off the heterogeneity, and then propagated to the observer (Figure 1A) , where it is "correlated" (time-integrated) with the second derivative $\ddot{e}$ of the error.

We now apply the transformation of variables $t_A = t_R - T_{RH}$ to the integral in Equation (4.13). Then:

$$\frac{\partial E_R}{\partial m} = 4 s_0 \ \frac{1}{4\pi R_{RH}} \frac{1}{4\pi R_{HS}} \int \ \ddot{e}(\mathbf{x}_R, t_A + T_{RH}) \, f(t_A - T_{HS}) \, dt_A \tag{4.14}$$

In this version, the source time function is propagated forward in time from the source to the heterogeneity and the error is propagated backward in time from the observation point to the heterogeneity (Figure 1B), and the two are then correlated. We have achieved a form that is very reminiscent of formula derived using the Adjoint Method, without explicitly applying adjoint methodology. Or rather, we *have* applied adjoint methodology without recognizing that we have done so; compare with the derivation of the Green function adjoint in Equation (A.5).

4.3. Computation of $\partial E/\partial m|_{m=0}$ Using the Adjoint Method. The wave equation operator is self-adjoint, so the Adjoint field equation (see Equation 2.6) and its solution are:

$$\mathcal{L}_A^{\dagger}\,\lambda_{A,R} = \left(s_0^2\frac{\partial^2}{\partial t_A^2} - \nabla_A^2\right)\lambda_{A,R} = \; e^0(\mathbf{x}_R, t_A)\;\delta(\mathbf{x}_A - \mathbf{x}_R)$$

$$\lambda_{A,R} = \frac{e^0(t_A + T_{AR})}{4\pi R_{AR}} \qquad (4.15)$$

Here we have selected the quiescent future form of solution ($+$ of the $\pm$ in Equation 4.2). The derivative of the wave equation operator is also self-adjoint and is:

$$\frac{\partial \mathcal{L}_A}{\partial m} = \frac{\partial \mathcal{L}_A^{\dagger}}{\partial m} = 2s_0\,\delta(\mathbf{x}_A - \mathbf{x}_H)\frac{\partial^2}{\partial t_A^2} \qquad (4.16)$$

so

$$h_{A,R} = 2\frac{\partial \mathcal{L}_A^{\dagger}}{\partial m}\;\lambda_{A,R} = \; 2s_0\,\delta(\mathbf{x}_A - \mathbf{x}_H)\;\frac{\ddot{e}^0(t_A + T_{AR})}{4\pi R_{AR}}$$

(4.17)

The unperturbed field satisfies:

$$\mathcal{L}_A u_A^0 = \left(s_0^2 \frac{\partial^2}{\partial t_A^2} - \nabla_A^2 \right) = \delta(\mathbf{x}_A - \mathbf{x}_S)\, f(t_S) \quad \text{so} \quad u_A^0 = \frac{f(t_A - T_{AS})}{4\pi R_{AS}}$$

(4.18)

Here we have selected the quiescent past form of solution ($-$ of the $\pm$ in Equation 4.2). Inserting this expression into Equation (2.7) yields an expression for the derivative:

$$\frac{\partial E_R}{\partial m} = \langle h_{A,R}, u_A^0 \rangle_A = \langle 2 s_0\, \delta(\mathbf{x}_A - \mathbf{x}_H)\, \frac{\ddot{e}^0(t_A + T_{AR})}{4\pi R_{AR}}, \frac{f(t_A - T_{AS})}{4\pi R_{AS}} \rangle_A$$

$$= 4 s_0\, \frac{1}{4\pi R_{RH}} \frac{1}{4\pi R_{HS}} \int \ddot{e}^0(t_A + T_{RH})\, f(t_A - T_{HS})\, dt_A$$

(4.19)

This is the same formula that was derived in Equation (4.14) using the Born approximation. The spatial pattern of the derivative is axially-symmetric about a line drawn through source and receiver and has the form of a series of concentric ellipses of alternating sign, with foci at the source and receiver (Figure 2). The ellipses represent surfaces of equal travel time from source to heterogeneity to receiver. The amplitude of the derivative varies across the surface of an ellipse, because it depends upon the product of the source-to-heterogeneity and heterogeneity-to-receiver distances, rather than their sum.

Zhu et al. [2009] point out an interesting link between the Adjoint Method and seismic migration [Claerbout and Doherty 1972], an imaging method commonly applied in reflection seismology.

In this setting, the unperturbed field, due to a source on the Earth's surface, is down-going and the perturbed field, due to heterogeneities within the Earth that scatter waves back up to the Earth's surface, is up-going. The imaging principle of seismic migration is based on the idea that, when the perturbed field δu is back-propagated into the earth, and the unperturbed field u^0 is forward propagated into the earth, the two will arrive at a scatterer at the same time (since the unperturbed field is the source of the perturbed field). A scattered at a point $\mathbf{x}_H$ can be detected ("imaged") by correlating δu with the $\ddot{u}^0$ (the source associated with u^0). This is precisely what the adjoint formulation is doing: the unperturbed field is forward-propagated in Equation (4.15b); the perturbed field is back-propagated in Equation (4.15c) (if we assume $e \approx \delta u^{obs}$); and the two field are time-correlated at the position of the heterogeneity in Equation 4.16. Hence, migration is just using $\partial E/\partial m$ as a proxy for m^{est} (see Equation 8.4). This correspondence provides a mechanism for generalizing seismic migration to more complicated settings [Luo et al., 2013].

5. Finite Frequency Travel Time Tomography

5.1. Rationale for Finite-Frequency Measurements. Traditionally, seismic tomography has used travel times based on "picks" of the onset of motion of a seismic phase on a seismogram, either determined "by eye" by a human analyst or automatically with, say, a short term average - long term average (STA/LTA) algorithm [Coppens, 1985].. Such travel times are easy to measure on short-period seismograms but problematical at longer periods, owing to the emergent onset of the waveforms. A more suitable measurement technique for these data involves cross-correlating the observed seismic phase with a synthetic reference seismogram, because cross-correlation can accurately determine the small time difference, say τ, between two smooth pulses. However, the results of cross-correlation are dependent upon the frequency band of measurement; a phase that is observed to arrive earlier than the reference phase at one frequency may well arrive later than it at another. Consequently, finite-frequency travel times must be interpreted in the context of the frequency band at which they are measured. Finite-frequency travel time tomography is based upon a derivative $\partial\tau/\partial m$ (where m is a model parameter) than incorporates the frequency-dependent behavior of cross-correlations.

5.2. Definition of Differenial Travel Time. The differential travel time τ^0 between an observed field $u^{obs}(t) \equiv u^{obs}(\mathbf{x},t)$ and a predicted field $u^0(t) \equiv u^0(\mathbf{x},t)$ is defined as the one that maximizes the cross-correlation:

$$C(\mathbf{x},\tau;u^0) = \int u^{obs}(t-\tau)\, u^0(t)\, dt \tag{5.1}$$

Since the cross-correlation is maximum at τ^0, its first derivative is zero there:

$$\left.\frac{dC(\mathbf{x},\tau;u^0)}{d\tau}\right|_{\tau^0} = 0 \tag{5.2}$$

5.3. Perturbation in Travel Time due to a Perturbation in the Predicted Wave Field. Suppose that the predicted field is perturbed from u^0 to $u = u^0 + \delta u$. The cross-correlation is perturbed to Marquering et al. 1999]:

$$C(\mathbf{x},\tau;u) = C(\mathbf{x},\tau;u^0) + \delta C(\mathbf{x},\tau) \quad \text{with} \quad \delta C(\mathbf{x},\tau) = \int u^{obs}(t-\tau)\,\delta u(t)\,dt \tag{5.3}$$

This function has a maximum at, say, $\tau = \tau^0 + \delta\tau$. Expanding $C(\mathbf{x},\tau;u)$ in a Taylor series up to second order in small quantities yields:

$$C(\mathbf{x},\tau;u) = C(\mathbf{x},\tau;u^0)\big|_{\tau^0} + 0 + ½\left.\frac{d^2C(\mathbf{x},\tau;u^0)}{d\tau^2}\right|_{\tau^0}(\delta\tau)^2 + \delta C(\mathbf{x},\tau)\big|_{\tau^0} + \left.\frac{d\delta C(\mathbf{x},\tau)}{dt}\right|_{\tau^0}\delta\tau \tag{5.4}$$

As is shown in Equation (5.2), the second term on the r.h.s. is zero. The maximum occurs where the derivative is zero:

$$\frac{dC(\mathbf{x},\tau)}{d\tau} = 0 \approx \left.\frac{d^2C(\mathbf{x},\tau)}{d\tau^2}\right|_{\tau^0}\delta\tau + \left.\frac{d\delta C(\mathbf{x},\tau)}{d\tau}\right|_{\tau^0} \tag{5.5}$$

Solving for $\delta\tau$ yields:

$$\delta\tau \approx -\left.\frac{d\delta C(\mathbf{x},\tau)}{d\tau}\right|_{\tau^0} \Big/ \left.\frac{d^2C(\mathbf{x},\tau)}{d\tau^2}\right|_{\tau^0} \tag{5.6}$$

The numerator is:

$$\left.\frac{d\delta C}{d\tau}\right|_{\mathbf{x},\tau^0} = \left.\frac{d}{d\tau}\int u^{obs}(\mathbf{x},t-\tau)\,\delta u(\mathbf{x},t)\,dt\right|_{\tau^0} = -\int \dot{u}_B^{obs}(\mathbf{x},t_B-\tau^0)\,\delta u(\mathbf{x},t)\,dt$$

$$(5.7)$$

and the denominator is:

$$\frac{dC(\mathbf{x},\tau)}{d\tau} = \frac{d}{d\tau}\int u^{obs}(\mathbf{x},t-\tau)\,u^0(\mathbf{x},t)\,dt = -\int \dot{u}^{obs}(\mathbf{x},t-\tau)\,u^0(\mathbf{x},t)\,dt$$

$$D(\mathbf{x},\tau^0) \equiv \left.\frac{d^2C(\tau^0;u_B^0)}{d\tau^2}\right|_{\tau^0} = -\frac{d}{d\tau}\int \ddot{u}^{obs}(\mathbf{x},t-\tau)\,u^0(\mathbf{x},t)\,dt\bigg|_{\tau^0}$$

$$= \int \ddot{u}^{obs}(\mathbf{x},t+\tau^0)\,u^0(\mathbf{x},t)\,dt \tag{5.8}$$

Consequently, the perturbation in differential arrival time of an observer at $\mathbf{x}_A$ satisfies:

$$\delta\tau_A = \int\left(\frac{\dot{u}^{obs}(\mathbf{x}_A,t_B+\tau^0)}{D(\mathbf{x}_A)}\right)\delta u(\mathbf{x}_A,t_B)\,dt_B = \langle h_{A,B},\delta u_B\rangle_B \tag{5.9}$$

with

$$h_{A,B} = \frac{\dot{u}^{obs}(\mathbf{x}_A,t_B+\tau^0)}{D(\mathbf{x}_A)}\delta(\mathbf{x}_B-\mathbf{x}_A) \tag{5.10}$$

5.4. Derivative of Travel Time with Respect to a Model Parameter. According to (A.3), a perturbation Δm to a structural parameter m causes a corresponding perturbation in the field:

$$\delta u_B = -\mathcal{L}_B^{-1}\,\frac{\partial\mathcal{L}_B}{\partial m}u_B^0\,\Delta m$$

(5.11)

Inserting this expression into the formula for $\delta\tau$ yields:

$$\delta\tau_A = \langle h_{A,B}, \delta u_B \rangle_B = -\langle h_{A,B}, \mathcal{L}_B^{-1} \frac{\partial \mathcal{L}_B}{\partial m} u_B^0 \ \rangle_B \ \Delta m = -\langle \mathcal{L}_B^{-1\dagger} h_{A,B}, \frac{\partial \mathcal{L}_B}{\partial m} u_B^0 \ \rangle_B \ \Delta m$$

$$\frac{\partial \tau_A}{\partial m} = -\langle H_{A,B}, \frac{\partial \mathcal{L}_B}{\partial m} u_B^0 \ \rangle_A \quad \text{with} \quad \mathcal{L}_B^{\dagger} H_{A,B} = h_{A,B}$$

(5.12)

Note that the adjoint differential equation has a source term that is localized at the receiver point A and has a source time function proportional to $\dot{u}^{obs}(\mathbf{x}_A)$.

5.6. Fréchet Derivative. The corresponding Fréchet derivative combines $\delta\tau(\mathbf{x}_A) = \langle h_{A,B}, \delta u_B \rangle_B$ with

$$\delta u_B = \langle \left[\frac{\delta u}{\delta m}\right]_{B,C}, \delta m_C \rangle \quad \text{with} \quad G_{B,C} \equiv \left[\frac{\delta u}{\delta m}\right]_{B,C} = -\mathcal{L}_B^{-1} \left[\frac{\delta \mathcal{L}}{\delta m}\right]_{B,C} u_B^0$$

(5.13)

to yield:

$$\delta\tau(\mathbf{x}_A) = \langle h_{A,B}, \langle G_{B,C}, \delta m_C \rangle_C \rangle_B = \langle \langle h_{A,B}, G_{B,C} \rangle_B, \delta m_C \rangle_C$$

(5.14)

from whence we conclude:

$$\left[\frac{\delta\tau}{\delta m}\right]_{A,C} = \langle h_{A,B}, G_{B,C} \rangle_B = -\langle H_{A,B}, \left[\frac{\delta \mathcal{L}}{\delta m}\right]_{B,C} u_B^0 \rangle_B$$

(5.16)

with, as before,

$$\mathcal{L}_B^\dagger H_{A,B} = h_{A,B} \tag{5.17}$$

6. An Example Using the Scalar Wave Equation

6.1. Choice of the Observed Field. As in Section 4, we consider an isotropic medium with a homogeneous background slowness s_0 containing a *test* point heterogeneity of strength m located at position $\mathbf{x}_H$. This scenario allows us to address how the alignment changes as the test heterogeneity is moved to different positions relative to the source and observer The observed field is taken to be identical to the direct wave in the absence of the heterogeneity; that is, when $m = 0$. Since the $u^{obs}(\boldsymbol{x}_R, t)$ and $u^0(\boldsymbol{x}_R, t)$ already align, we can set $\tau^0 = 0$.. The differential equation is $\mathcal{L}(m)\, u = \delta(\mathbf{x} - \mathbf{x}_S)\, f(t)$, where:

$$\mathcal{L} = \mathcal{L}^\dagger = s_0^2 \frac{\partial^2}{\partial t^2} + 2 s_0 m\, \delta(\mathbf{x} - \mathbf{x}_H) \frac{\partial^2}{\partial t^2} - \nabla^2 \tag{6.1}$$

The source time function $f(t)$ is assumed to be band-limited between angular frequencies ω_1 and ω_2, e.g.:

$$f(t) = \mathrm{sinc}(\omega_1 t) - \mathrm{sinc}(\omega_2 t)$$

(6.2)

The observed field at the receiver is the direct field u^0; that is:

$$u_R^{obs} = \frac{f(t_R - T_{SR})}{4\pi R_{SR}} \tag{6.3}$$

6.2. The Partial Derivative of Travel Time With Respect to the Model. Our goal is to construct $\partial\tau/\partial m$ associated with a point heterogeneity at $\mathbf{x}_H$. First, we construct the function H_A, which involves back-propagating, via the adjoint equation, the observed field at the receiver point $\mathbf{x}_R$ to an arbitrary point $\mathbf{x}_A$:

$$\mathcal{L}^\dagger H = \frac{\dot{u}_R^{obs}}{D(\mathbf{x}_R)}\delta(\mathbf{x}-\mathbf{x}_R) \quad \text{or} \;\; H_A = -\frac{1}{D(\mathbf{x}_R)}\,\frac{\dot{f}(t_A - T_{SR} + T_{RA})}{4\pi R_{SR}\, 4\pi R_{RA}} \tag{6.4}$$

Second, we construct the function $G_{A,H}$, also for an arbitrary point $\mathbf{x}_A$:

$$G_{A,H} = -\frac{\partial\mathcal{L}}{\partial m}u^0 = -2s_0\delta(\mathbf{x}_A - \mathbf{x}_H)\frac{\partial^2}{\partial t^2}\frac{f(t - T_{SA})}{4\pi R_{SA}} = -2s_0\delta(\mathbf{x}_A - \mathbf{x}_H)\frac{\ddot{f}(t - T_{SA})}{4\pi R_{SA}} \tag{6.5}$$

Finally, we combine H_A and $G_{A,H}$ via an inner product to construct the partial derivative:

$$\left.\frac{\partial\tau}{\partial m}\right|_{\mathbf{x}_H} = \langle H_A, G_{A,H}\rangle_A = \langle -\frac{1}{D(\mathbf{x}_R)}\,\frac{\dot{f}(t_A - T_{SR} + T_{RA})}{4\pi R_{SR}\, 4\pi R_{RA}}\;\; , -2s_0\delta(\mathbf{x}_A - \mathbf{x}_T)\frac{\ddot{f}(t_A - T_{SA})}{4\pi R_{SA}}\rangle_A$$

$$= \frac{2s_0}{D(\mathbf{x}_R)} \frac{1}{4\pi R_{SR}\ 4\pi R_{RH} 4\pi R_{SH}} \int \dot{f}(t_A - T_{SR} + T_{RH})\, \ddot{f}(t_A - T_{SH})\, dt_A$$

$$= \frac{2s_0}{D(\mathbf{x}_R)} \frac{1}{4\pi R_{SR}\ 4\pi R_{RH} 4\pi R_{SH}} \int \dot{f}(t - \Delta T)\, \ddot{f}(t)\, dt \tag{6.6}$$

The last form uses the transformation of variables $t = t_A - T_{SH}$ where:

$$\Delta T = T_{SR} - T_{RH} - T_{SH} = T_{SR} - (T_{SH} + T_{HR}) \tag{6.7}$$

The quantity ΔT represents the difference in travel times between the direct *(S-R)* and scattered *(S-H-R)* paths. The quantify $D(\mathbf{x}_R)$ is given by:

$$D(\mathbf{x}_R) = \int \ddot{u}^{obs}(\mathbf{x}_R, t)\, u^0(\mathbf{x}_R, t)\, dt = \frac{1}{4\pi R_{SR}\ 4\pi R_{SR}} \int \ddot{f}(t)\, f(t)\, dt$$

$$= -\frac{1}{(4\pi R_{SR})^2} \int \left[\dot{f}(t)\right]^2 dt \tag{6.8}$$

We have used the anti-self-adjoint property of the d/dt operator to simplify the last integral.

6.3. Analysis. The derivative $\partial\tau/\partial m$ is axially symmetric about the *(S-R)* line, since R_{SR} and R_{RH} depend only on the perpendicular distance r of $\mathbf{x}_H$ from the line. Sliced perpendicular to the line, $\partial\tau/\partial m$ is “doughnut-shaped”.

The derivative $\partial\tau_A/\partial m = 0$ whenever $\Delta T = 0$. This behavior follows from d/dt being an anti-self-adjoint operator, since any quantity equal to its negative is zero:

$$\int \dot{f}(t)\,\ddot{f}(t)\;dt = -\int \ddot{f}(t)\dot{f}(t)\;dt = -\int \dot{f}(t)\,\ddot{f}(t)\;dt = 0 \qquad (6.9)$$

The time difference ΔT is zero when the test heterogeneity is between S and R and on the *(S-R)* line, so $\partial\tau/\partial m = 0$ in this case. This zero makes the 'hole' in the center of the doughnut'.

Now consider an oscillatory, band-limited source time function with a characteristic period P. Suppose we construct the elliptical volume surrounding the points $\mathbf{x}_S$ and $\mathbf{x}_R$ for which $\Delta T < P/2$. The time integral in (F.2) will have the same sign everywhere in this volume, as will $\partial\tau/\partial m$. This region defines the "banana." The banana is thinner for short periods than for long periods (Figure 4).

Moving away from the *(S-R)* line along its perpendicular, the time integral, and hence the derivative, oscillates in sign, as the $\dot{f}(t-\Delta T)$ and $\ddot{f}(t)$ factors beat against one another. The derivative also decreases in amplitude (since the factors R_{RH} and R_{SH} grow with distance). Consequently, the central banana is surrounded by a series of larger, but less intense, bananas of alternating sign.

7. Adjoint Method Applied to the Cross-Convolution Method

7.1. Definition. The cross-convolution method [Menke and Levin, 2003] is used to invert shear wave splitting and receiver function data for Earth structure [e.g. Bodin et al. 2014]. It is especially useful for extracting structural information from differences between the several components of a P or S wave because, unlike other waveform modeling approaches, it does not require knowledge of the source time function. It compares two different components u_i^{obs} and u_j^{obs} observed at the same position with their predictions u_i and u_j, using the measure:

$$\Psi(m) = \langle w_A(m), w_A(m) \rangle_A \text{ with } w_A(m) = \left(\Omega\, u_j^{obs}\right) * \left(\Omega\, u_i(m)\right) - \left(\Omega\, u_i^{obs}\right) * \left(\Omega\, u_j(m)\right) \tag{7.1}$$

Here $\Omega(t)$ is a window function that selects a particular seismic phase, such as the P wave, from the time series. The cross-convolution measure $\Psi(m)$ is a function of Earth structure, as quantified by a parameter m. Because $\Psi(m)$ scales with the amplitude of the predicted waveform, determining the model parameters by minimizing $\Phi(m) \equiv \Psi/P$, were $P(m)$ is total power, is preferable to determining them by minimizing $\Psi(m)$. The total power is given by:

$$P(m) = \langle \Omega\, \mathbf{u}(m), \Omega\, \boldsymbol{u}(m) \rangle_A \tag{7.2}$$

The partial derivative of $\Phi = \Psi/P$ with respect to a model parameter is calculated using the chain rule:

$$\left.\frac{\partial \Phi}{\partial m}\right|_{m_0} = \left.\frac{\partial \Psi P^{-1}}{\partial m}\right|_{m_0} = \left.\frac{\partial \Psi}{\partial m}\right|_{m_0} P_0^{-1} - \Psi_0 \left.\frac{\partial P}{\partial m}\right|_{m_0} P_0^{-2} \tag{7.3}$$

As we show below, the derivatives $\partial\Psi/\partial m$ and $\partial P/\partial m$ can be derived using Adjoint Methods.

G.2. The Partial Derivative of Power With Respect to Model Parameter. The $\partial P/\partial m$ derivative is:

$$\left.\frac{\partial P}{\partial m}\right|_{m_0} = 2\,\langle \Omega_A\, \mathbf{u}_A(m), \Omega\,\frac{\partial \boldsymbol{u}_A}{\partial m}\rangle_A = 2\,\langle \Omega_A^2\, \mathbf{u}_A(m), \frac{\partial \mathbf{u}_A}{\partial m}\rangle_A$$

$$= -2\,\langle \Omega^2 \mathbf{u}_A^0(m), \boldsymbol{\mathcal{L}}_A^{-1}\,\frac{\partial \boldsymbol{\mathcal{L}}_A}{\partial m}\mathbf{u}_A^0\rangle_A = -2\,\langle \frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m}\boldsymbol{\mathcal{L}}_A^{\dagger-1}\Omega_A^2\mathbf{u}_A^0, \mathbf{u}_A^0\rangle_A$$

$$\text{or}\quad \left.\frac{\partial P}{\partial m}\right|_{m_0} = -2 < \frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m}\boldsymbol{\xi}_\mathrm{A}, \mathbf{u}_A^0 >_A \quad \text{with} \quad \boldsymbol{\mathcal{L}}_A^\dagger\boldsymbol{\xi}_\mathrm{A} = \Omega_A^2\mathbf{u}_A^0$$

Here, $\boldsymbol{\xi}_A$ is the adjoint field associated with the power derivative. We consider the total power P to be the sum of the power P^R associated with individual observation points $\mathbf{x}_R$. The corresponding adjoint field $\boldsymbol{\xi}_A^R$ satisfies:

$$\boldsymbol{\mathcal{L}}_A^\dagger\boldsymbol{\xi}_A^R = \Omega_R^2\mathbf{u}^0(\mathbf{x}_R, t)\;\delta(\mathbf{x}_A - \mathbf{x}_R)$$

We now consider a point density perturbation $\rho = m\delta(\mathbf{x} - \mathbf{x}_H)$ located at $\mathbf{x}_H$. The derivative of the adjoint operator is:

$$\frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m} = \;\delta(\mathbf{x}_A - \mathbf{x}_H)\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\frac{\partial^2}{\partial t^2}$$

The power derivative is:

$$\left.\frac{\partial P}{\partial m}\right|_{m_0} = -2\int \boldsymbol{\xi}^R(\mathbf{x}_H, t_H)\cdot\;\ddot{\mathbf{u}}^0(\mathbf{x}_H, t_H)\;dt_H$$

G.3. The derivative $\partial\Psi/\partial m$. The cross-convolution function w is constructed from the predicted wave field $\mathbf{u}$ through a linear operator $\boldsymbol{\mathcal{W}}$ that is independent of the model parameter:

$$w(m) = \boldsymbol{\mathcal{W}}\,\Omega\,\mathbf{u}(m) \quad \text{with} \quad \mathbf{u} = [\cdots \quad \mathbf{u}_i, \quad \dots, \quad \mathbf{u}_j, \quad \dots] \text{ and}$$

$$\text{and} \quad \boldsymbol{\mathcal{W}} = (\mathbf{U}\,*) \quad \text{with} \quad \mathbf{U} = \left[\cdots \quad \Omega\,u_j^{obs}, \quad 0, \quad -\Omega\,u_i^{obs}, \quad \dots\right]$$

The adjoint of $\boldsymbol{\mathcal{W}}$ is the cross-correlation operator $\boldsymbol{\mathcal{W}}^{\dagger} = (\mathbf{U}^{\mathrm{T}} \star)$**.** The partial derivative of Ψ with respect to a model parameter m is:

$$\left.\frac{\partial \Psi}{\partial m}\right|_{m_0} = \frac{\partial}{\partial m}\langle w_A, w_A\rangle_A = \langle 2w_A^0, \left.\frac{\partial w_A}{\partial m}\right|_{m_0}\rangle_A$$

$$= \langle 2\boldsymbol{\mathcal{W}}_A\Omega_A\mathbf{u}_A^0, \boldsymbol{\mathcal{W}}_A\Omega_A \left.\frac{\partial \boldsymbol{u}_A}{\partial m}\right|_{m_0}\rangle_A = \langle -2\boldsymbol{\mathcal{W}}_A\Omega_A\mathbf{u}_A^0, \boldsymbol{\mathcal{W}}_A\Omega_A\boldsymbol{\mathcal{L}}_A^{-1}\frac{\partial \boldsymbol{\mathcal{L}}_A}{\partial m}\boldsymbol{u}_A^0\rangle_A$$

$$= \langle -2\frac{\partial \boldsymbol{\mathcal{L}}_A^{\dagger}}{\partial m}\boldsymbol{\mathcal{L}}_A^{\dagger-1}\Omega_A\boldsymbol{\mathcal{W}}_A^{\dagger}\boldsymbol{\mathcal{W}}_A\Omega_A\mathbf{u}_A^0\rangle_A = \langle -2\frac{\partial \boldsymbol{\mathcal{L}}_A^{\dagger}}{\partial m}\boldsymbol{\lambda}_A, \boldsymbol{u}_A^0\rangle_A$$

(7.4)

Here $\boldsymbol{\lambda}_A$ is an adjoint field that satisfies:

$$\boldsymbol{\mathcal{L}}_A^{\dagger}\,\boldsymbol{\lambda}_A = \boldsymbol{\varphi}_A \quad \text{with} \quad \boldsymbol{\varphi}_A \equiv \Omega_A\boldsymbol{\mathcal{W}}_A^{\dagger}\boldsymbol{\mathcal{W}}_A\Omega_A\mathbf{u}_A^0 = \Omega_A\left(\mathbf{U}_A^{\mathrm{T}} \star \mathbf{U}_A\right) * (\Omega_A\boldsymbol{u}_A^0) = \Omega_A\mathbf{X}_A * (\Omega_A\boldsymbol{u}_A^0)$$

$$\text{with } \mathbf{X}_A \equiv \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & ⋰ \\ \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & \left\{\Omega_A \mathrm{u}_{A,j}^{obs} \star \Omega_A u_{A,j}^{obs}\right\} & -\left\{\left(\Omega_A \mathrm{u}_{A,j}^{obs}\right) \star \left(\Omega_A u_{A,i}^{obs}\right)\right\} & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & -\left(\Omega_A \mathrm{u}_{A,i}^{obs}\right) \star \left(\Omega_A u_{A,j}^{obs}\right) & \left(\Omega_A \mathrm{u}_{A,i}^{obs}\right) \star \left(\Omega_A u_{A,i}^{obs}\right) & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & \cdots \\ ⋰ & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

(7.5)

Here $w_A^0 \equiv w_A(m_0)$. The source term of the adjoint equation involves cross-correlations of windowed components of the observed field. As with previous cases, we can view the $\boldsymbol{\lambda}_A$ as the superposition of contributions $\boldsymbol{\lambda}_A^R$ of many observation points $\mathbf{x}_R$. The adjoint equation corresponding to a single observation point is:

$$\boldsymbol{\mathcal{L}}_A^\dagger \, \boldsymbol{\lambda}_A^R = \boldsymbol{\varphi}^R(\mathbf{x}_R, t_A)\delta(\mathbf{x}_A - \mathbf{x}_R) \tag{7.6}$$

We again consider the special case of a point density heterogeneity, so that $\partial \boldsymbol{\mathcal{L}}_A^\dagger / \partial m = \delta(\mathbf{x}_A - \mathbf{x}_H)(\partial^2/\partial t^2)\,\mathbf{I}$ (where $\mathbf{I}$ is the identity matrix). The partial derivative is then:

$$\left.\frac{\partial \Psi^{\mathrm{R}}}{\partial m}\right|_{m_0} = \langle -2 \frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m} \boldsymbol{\lambda}_A^R, \mathbf{u}_A^0 \rangle_A = \langle -2\delta(\mathbf{x}_A - \mathbf{x}_H)\, \boldsymbol{\lambda}_A^R, \ddot{\mathbf{u}}_A^0 \rangle_A = -2 \int \boldsymbol{\lambda}^R(\mathbf{x}_H, t_H) \cdot \, \ddot{\mathbf{u}}(\mathbf{x}_H, t_H)\, dt_H \tag{7.7}$$

7.4. Derivatives With Respect to Lamé Parameters. Assuming a perturbation of the form $\lambda = m\,\delta(\mathbf{x}_A - \mathbf{x}_H)$ the λ derivative of the adjoint wave operator is:

$$\frac{\partial \boldsymbol{\mathcal{L}}^\dagger}{\partial m} = -\delta \begin{bmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\ \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial y^2} & \frac{\partial^2}{\partial y \partial z} \\ \frac{\partial^2}{\partial x \partial z} & \frac{\partial^2}{\partial y \partial z} & \frac{\partial^2}{\partial z^2} \end{bmatrix} - \begin{bmatrix} \frac{\partial \delta}{\partial x}\frac{\partial}{\partial x} & \frac{\partial \delta}{\partial x}\frac{\partial}{\partial y} & \frac{\partial \delta}{\partial x}\frac{\partial}{\partial z} \\ \frac{\partial \delta}{\partial y}\frac{\partial}{\partial x} & \frac{\partial \delta}{\partial y}\frac{\partial}{\partial y} & \frac{\partial \delta}{\partial y}\frac{\partial}{\partial z} \\ \frac{\partial \delta}{\partial z}\frac{\partial}{\partial x} & \frac{\partial \delta}{\partial z}\frac{\partial}{\partial y} & \frac{\partial \delta}{\partial z}\frac{\partial}{\partial z} \end{bmatrix}$$

(7.8)

Here the δ's are abbreviations for the Dirac function $\delta(\mathbf{x}_A - \mathbf{x}_H)$. Assuming a perturbation of the form $\mu = m\,\delta(\mathbf{x}_A - \mathbf{x}_H)$ the μ derivative of the adjoint wave operator is:

$$\frac{\partial \boldsymbol{\mathcal{L}}^\dagger}{\partial m} = -\delta \begin{bmatrix} 2\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} & \frac{\partial^2}{\partial x \partial y} & \frac{\partial^2}{\partial x \partial z} \\ \frac{\partial^2}{\partial x \partial y} & \mu\frac{\partial^2}{\partial x^2} + 2\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} & \frac{\partial^2}{\partial y \partial z} \\ \frac{\partial^2}{\partial x \partial z} & \frac{\partial^2}{\partial y \partial z} & \frac{\partial^2}{\partial x^2} + \mu\frac{\partial^2}{\partial y^2} + 2\frac{\partial^2}{\partial z^2} \end{bmatrix}$$

$$- \begin{bmatrix} 2\frac{\partial \delta}{\partial x}\frac{\partial}{\partial x} + \frac{\partial \delta}{\partial y}\frac{\partial}{\partial y} + \frac{\partial \delta}{\partial z}\frac{\partial}{\partial z} & \frac{\partial \delta}{\partial y}\frac{\partial}{\partial x} & \frac{\partial \delta}{\partial z}\frac{\partial}{\partial x} \\ \frac{\partial \delta}{\partial x}\frac{\partial}{\partial y} & \frac{\partial \delta}{\partial x}\frac{\partial}{\partial x} + 2\frac{\partial \delta}{\partial y}\frac{\partial}{\partial y} + \frac{\partial \delta}{\partial z}\frac{\partial}{\partial z} & \frac{\partial \delta}{\partial z}\frac{\partial}{\partial y} \\ \frac{\partial \delta}{\partial x}\frac{\partial}{\partial z} & \frac{\partial \delta}{\partial y}\frac{\partial}{\partial z} & \frac{\partial \delta}{\partial x}\frac{\partial}{\partial x} + \frac{\partial \delta}{\partial y}\frac{\partial}{\partial y} + 2\frac{\partial \delta}{\partial z}\frac{\partial}{\partial z} \end{bmatrix}$$

(7.9)

The inner products for $\partial\Psi/\partial m$ and $\partial\Psi/\partial m$ include both Dirac delta functions and their spatial derivatives. For instance, in the λ case:

$$\left.\frac{\partial \Psi}{\partial m}\right|_{m_0=0} = \langle -2\frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m}\boldsymbol{\lambda}_A, \mathbf{u}_A^0 \rangle_A$$

$$= 2\,\langle \delta(\mathbf{x}_A - \mathbf{x}_H)\frac{\partial^2 \lambda_{A,x}}{\partial x^2}, u^0_{A,x}\rangle_A + 2\,\langle \left\{\frac{\partial}{\partial x}\delta(\mathbf{x}_A - \mathbf{x}_H)\right\}\frac{\partial \lambda_{A,x}}{\partial x}, u^0_{A,x}\rangle_A + \cdots$$

$$= 2\int \frac{\partial^2 \lambda_{H,x}}{\partial x^2} u^0_{H,x} dt_H - 2\int \frac{\partial}{\partial x}\left\{\frac{\partial \lambda_{A,x}}{\partial x} u^0_{H,x}\right\} dt_H \cdots$$

$$= 2\int \frac{\partial^2 \lambda_{H,x}}{\partial x^2} u^0_{H,x} dt_H - 2\int \frac{\partial^2 \lambda_{H,x}}{\partial x^2} u^0_{H,x} dt_H - 2\int \frac{\partial \lambda_{A,x}}{\partial x}\frac{\partial u^0_{H,x}}{\partial x} dt_H \cdots$$

(7.10)

Here we have used the rule $\int f(x_A)\,\{d\delta(x_A - x_H)/dx_A\}\ dx_A = -\,df(x_H)/dx_H$. Thus, $\partial\Psi/\partial m$ and $\partial\Psi/\partial m$ involve temporal correlations between spatial gradients of both adjoint and unperturbed fields. The inner product can be succinctly written:

$$\langle -2\frac{\partial \boldsymbol{\mathcal{L}}_A^\dagger}{\partial m}\boldsymbol{\lambda}_A, \mathbf{u}_A^0\rangle = -2\sum_i\sum_j\ \int [\boldsymbol{\mathcal{D}}]_{ij}\left\{[\boldsymbol{\lambda}_H]_{\mathrm{j}}\ [\mathbf{u}_H^0]_i\right\} dt_H$$

(7.11)

Here $\boldsymbol{\mathcal{D}}$ is an operator formed from $\partial\boldsymbol{\mathcal{L}}^\dagger/\partial m$ by replacing each occurrence of δ with unity and each occurrence of $\partial\delta/\partial x_i$ with $-\partial/\partial x_i$.

8. A Cross-Convolution Example Using the Elastic Wave Equation

8.1. The elastic Green function. In an isotropic and homogeneous solid, the far-field displacement $u_{A,S}$ for an observer at $\mathbf{x}_A$ and a point force in the $\hat{\mathbf{f}}$ direction at $\mathbf{x}_S$ consists of the sum of P-wave and S-wave terms (Aki and Richards 2009):

$$u_{A,S} = \frac{s\left(t - T_{S,A}^{P}\right)}{4\pi\rho\alpha^2 R_{A,S}} \hat{\boldsymbol{\gamma}}_{\mathrm{A,S}}\left(\hat{\boldsymbol{\gamma}}_{\mathrm{A,S}} \cdot \hat{\mathbf{f}}\right) + \frac{s\left(t - T_{S,A}^{S}\right)}{4\pi\rho\beta^2 R_{A,S}} \left\{\hat{\mathbf{f}} - \hat{\boldsymbol{\gamma}}_{\mathrm{A,S}}\left(\hat{\boldsymbol{\gamma}}_{\mathrm{A,S}} \cdot \hat{\mathbf{f}}\right)\right\} \tag{8.1}$$

Here α and β are the compressional and shear wave velocities, respectively, ρ is density, $R_{S,A}$ is the distance from $\mathbf{x}_A$ to $\mathbf{x}_S$, $T_{A,S}^{P} = R_{A,S}\,/\alpha$ is the travel time of the P wave, $T_{S,A}^{S} = R_{A,S}\,/\beta$ is the travel time of the S wave, $\hat{\boldsymbol{\gamma}}_{\mathrm{A,S}} = (\mathbf{x}_A - \mathbf{x}_A)/R_{A,S}$ is the direction from source to observer, and $s(t)$ is the source time function.

We limit our discussion to P-S_v displacements from sources and receivers in the (x_1, x_3) plane. A force in the x_1-direction causes displacement:

$$\begin{bmatrix} u_x^0 \\ u_z^0 \end{bmatrix}_A = \frac{s\left(t - T_{A,S}^{P}\right)}{4\pi\rho\alpha^2 R_{A,S}} \begin{bmatrix} \cos^2\theta_{S,A} \\ \sin\theta_{S,A}\ \cos\theta_{S,A} \end{bmatrix} + \frac{s\left(t - T_{A,S}^{S}\right)}{4\pi\rho\beta^2 R_{A,S}} \begin{bmatrix} \sin^2\theta_{A,S} \\ -\sin\theta_{A,S}\ \cos\theta_{A,S} \end{bmatrix} \tag{8.2}$$

Here $\theta_{A,S}$ is the angle from the x_1-direction to the observer. A force in the x_3-direction causes displacement:

$$\begin{bmatrix} u_x^0 \\ u_z^0 \end{bmatrix}_A = \frac{s\left(t - T_{S,A}^{P}\right)}{4\pi\rho\alpha^2 R_{A,S}} \begin{bmatrix} \sin\theta_{S,A}\cos\theta_{A,S} \\ \sin^2\theta_{A,S} \end{bmatrix} + \frac{s\left(t - T_{S,A}^{S}\right)}{4\pi\rho\beta^2 R_{A,S}} \begin{bmatrix} -\sin\theta_{A,S}\ \cos\theta_{A,S} \\ \cos^2\theta_{A,S} \end{bmatrix} \tag{8.3}$$

In each case, the P and S wave particle motions are mutually perpendicular.

8.2. Derivatives of Power and Cross-Convolution Measure with Respect to Model Parameters.

We focus on the displacement field due to a force in the x_1-direction. located at $\mathbf{x}_S = 0$ and with a Gaussian source time function. The P wave observed at $\mathbf{x}_R$ consists of a leading "direct" wave followed by a secondary "scattered" wave. It is selected from the time series by multiplication by the boxcar window function $\Omega(t)$. In the absence of heterogeneity, the predicted P wave consists only of the direct wave. Heterogeneity leads to scattering, which results in scattered waves, some of which may match the observed secondary arrival. We consider a sequence of point density heterogeneities, each of strength m and located at a position $\mathbf{x}_H$. The derivative $\partial\Phi/\partial m$ quantifies whether one or more of these heterogeneities can improve the fit.

We compute $\partial\Psi/\partial m$ $\partial P/\partial m$ and $\partial \Phi/\partial m$ for a grid of $\mathbf{x}_H$'s in the (x, y) plane using both the direct and adjunct method:

The direct method computes the derivative by comparing windowed predicted and observed waves at the position of the receiver. The predicted wave is the sum of a direct wave plus a scattered wave. The former is calculated by forward-propagating P and S waves from the source to the observer. The latter is calculated by forward-propagating P and S waves to the heterogeneity, where they act as secondary sources that generate scattered P and S waves that are then propagated to the receiver. The direct and scattered waves are summed and windowed around the P wave arrival time to yield the predicted wave. It's power and cross-convolution measure are calculated and compared with those of the direct wave, providing a finite difference approximation of the derivatives.

The adjoint method computes the each derivative by comparing two time series at the position of the heterogeneity. One time series is the second derivative of source forward-propagated to the heterogeneity. The other is the adjoint wave field, which is back-propagated from an adjoint source at the receiver to the heterogeneity. The two time series are the "correlated" by time-integrating their product, yielding the derivative. Two different adjoint wave fields must be calculated, one for $\partial\Psi/\partial m$ and the other for $\partial P/\partial m$ (Figure 5).

We have verified that these two methods produce the same result. In both cases, each of the four scattering interactions (P→P, P→S, S→P and S→S) can be isolated simply by omitting a P or S wave from each of the two stages of propagation.

8.3 Resolution. Resolving power is an important one for understanding the behavior and utility of any inverse problem [Backus and Gilbert, 1968, 1971, Wiggins 1972, see also Menke 2014]. We first compute the wave field, observed at an array of receivers, associated with a point-like heterogeneity in density at $\mathbf{x}_H$; it becomes the synthetic data $\mathbf{u}^{obs}$. We then perform an approximate inversion of these data to produce an estimate of the heterogeneity. Typically, the heterogeneity spreads out in space, so it can be interpreted as a *point spread function* that quantifies resolution.

Suppose that we define a gradient vector $[\nabla\Phi]_\mathrm{i} \equiv \partial\Phi/\partial m_i$, each element of which corresponds to the derivative for a point heterogeneity at $\mathbf{x}^{(i)}$ of amplitude m_i. The steepest-descent estimate of these amplitudes is computed by moving a distance Δm, in the downhill direction, from the homogeneous $\mathbf{m}^0 = 0$ model (corresponding to Φ_0) to a heterogeneous model $\mathbf{m}^{\mathrm{est}}$ (corresponding to $\Phi \approx 0$):

$$0 - \Phi_0 \approx \nabla\Phi \cdot (\mathbf{m}^{\mathrm{est}} - 0) \quad \text{with} \quad \mathbf{m}^{\mathrm{est}} = \frac{\nabla\Phi}{|\nabla\Phi|}\Delta m$$

$$\Phi_0 \approx -\nabla\Phi \cdot \frac{\nabla\Phi}{|\nabla\Phi|}\Delta m \quad \text{so} \quad \Delta m = -\frac{\Phi_0}{|\nabla\Phi|} \quad \text{and} \quad \mathbf{m}^{\text{est}} \approx \frac{-\Phi_0}{|\nabla\Phi|^2}\nabla\Phi \tag{8.4}$$

In this approximation, the solution $\mathbf{m}^{\text{est}}$ is proportional to the gradient $\nabla\Phi$, implying that $\nabla\Phi$ can be used as an proxy for the point spread function. We examine three cases, in which the window function is chosen to include both P and S waves, or just the P wave, or just the S wave. In all three cases, $\nabla\Phi$ has sharp minimum at $\mathbf{x}_H$, implying a narrow point spread function and excellent resolution (although the horizontal resolution in the P wave case is poorer than the other two (Figure 6).

9. Conclusions

The Adjoint Methods have proven to be essential tools for imaging problems. On the practical side, they allow inversions to be organized in an extremely efficient way, allowing what might otherwise be prohibitively time-consuming calculations to be performed. On the conceptual side, they allow complex formula to be manipulated into forms in which important quantities, such as Fréchet derivatives, readily can be identified. Our review here has stressed the underlying similarity between different approaches used in the literature, including the derivation of the adjoint field equations, the use of partial or Fréchet derivatives, and the application of the method to four different types of data (wave forms, finite frequency travel times, power and cross-correlation measure).

Acknowledgements. This research was supported by the National Science Foundation under grant EAR 11-47742.

References

Aki, K. and P.G. Richards, Quantitative Seismology, Second Edition, 700 pp, ISBN-13: 978-1891389634, 2009.

Backus, G. and F. Gilbert, The resolving power of Gross Earth Data, Geophys. J.R. astr. Soc.16, 169-205, doi: 10.1190/1.1444834.coden:GPYSA70016-8033, 1968.

Backus G., and F. Gilbert,, 1970, Uniqueness in the inversion of inaccurate gross earth data, Philosophical Transactions of the Royal Society of London Series A, 266, 123–192, doi: 10.1098/rsta.1970.0005.coden:PTRMAD1364-503X, 1971.

Bodin, T., H. Yuan, and B. Romanowicz, Inversion of receiver functions without deconvolution—application to the Indian craton, Geophys. J. Int., 196(2), 1025-1033, doi:10.1093/gji/ggt431, 2014.

Chauhan, A.P.S., S. Singh, N.D. Hananto, H. Carton, F. Klingelhoefer, J.-X. Dessa, H. Permana, N.J. White, D. Graindorge, and the SumatraOBS Scientific Team, Seismic imaging of forearc backthrusts at northern Sumatra subduction zone, Geophys. J. Int. 179 : 1772-1780, doi: 10.1111/j.1365-246X.2009.04378.x, 2009.

Chen, P., L. Zhao and T.H. Jordan, Full 3D Tomography for Crustal Structure of the Los Angeles Region, 2007, Bulletin of Seismological Society of America 97, 1094-1120, doi: 10.1785/0120060222, 2007.

Chen, P., T.H. Jordan and E.-J. Lee, Perturbation Kernels for Generalized Seismological Data Functionals (GSDF), Geophysical Journal International, 183, 869-883, doi: 10.1111/j.1365-246X.2010.04758.x, 2010.

Claerbout, J.F. and S.M. Doherty, Downward continuation of moveout-corrected seismograms, Geophysics,37, 741-768. doi: 10.1190/1.1440298, 1972.

Coppens, F., First arrival picking on common-offset trace collections for automatic estimation of static corrections, Geophysical Prospecting 33, 1212-1231, DOI: 10.1111/j.1365-2478.1985.tb01360, 1985.

Dahlen, F.A., S.-H. Hung and G. Nolet, Fréchet kernels for finite-frequency traveltimes—I. Theory, Geophys. J. Int. 141, 157-174, doi: 10.1046/j.1365-246X.2000.00070.x, 2000.

Deuflhard, P., Newton Methods for Nonlinear Problems.Affine Invariance and Adaptive Algorithms. Springer Series in Computational Mathematics 35. Springer, 424pp, ISBN 3-540-21099-7, 2004.

Devaney, A.J., Inverse-scattering theory within the Rytov approximation. Optics Letters 6, 374-376. doi: 10.1364/OL.6.000374, 1981.

Hall, M.C.G., D.C. Cacuci and M.E. Schlesinger, Sensitivity Analysis of a Radiative-Convective Model by the Adjoint Method, J. Atmospheric Sci. 39, 2038-2050, DOI: 10.1175/1520-0469(1982)039<2038:SAOARC>2.0.CO;2, 1982.

Hall, M.C.G. and D.G. Cacuci, Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. J. Atmos. Sci. 40, 2537–2546, doi: 10.1175/1520-0469(1983)040<2537:PIOTAF>2.0.CO;2, 1983.

Hung, S.-H., F.A. Dahlen and G. Nolet, Fréchet kernels for finite-frequency traveltimes—II. Examples, Geophys. J. Int. 141, 175-203, doi: 10.1046/j.1365-246X.2000.00072.x, 2000.

Kim, Y., Liu, Q., and J. Tromp. Adjoint centroid-moment tensor inversions. Geophys. J. Int., 186: 264-278. DOI: 10.1111/j.1365-246X.2011.05027.x, 2011.

Long, M.D., M. de Hoop and R.D. van der Hilst, Wave-equation shear wave splitting tomography, Geophys. J. Int., 172, 311-330, DOI: 10.1111/j.1365-246X.2007.03632.x, 2008

Luo, Y., Tromp, J., Denel, B., and Calendra, H. 3D coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods. Geophysics 78, S193-S202. doi: 10.1190/geo2012-0462.1, 2013.

Marquering, H., G. Nolet G. and F.A. Dahlen, Three-dimensional waveform sensitivity kernels, Geophys. J. Int. 132, 521–534, DOI: 10.1046/j.1365-246X.1998.00426.x, 1998.

Marquering, H, FA Dahlen, G. Nolet, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, Geophys. J. Int. 137, 805-815. doi: 10.1046/j.1365-246x.1999.00837.x, 1999.

Mellman, G.R., A method of body-wave waveform inversion for the determination of earth structure, Geophys. J. Int. 62, 481-504. doi: 10.1111/j.1365-246X.1980.tb02587.x, 1980.

Menke, W. Geophysical Data Analysis: Discrete Inverse Theory, Third Edition (textbook), Academic Press (Elsevier), 330 pp, ISBN: 9780123971609, 2012.

Menke, W., Review of the Generalized Least Squares Method, Surveys in Geophysics 36, 1-25, doi: 10.1007/s10712-014-9303-1, 2014.

Menke, W. and Z. Eilon, Relationship between data smoothing and the regularization of inverse problems, Pure and Applied Geophysics 172, 2711-2726, DOI: 10.1007/s00024-015-1059-0, 2015

Menke, W. and V. Levin, The cross-convolution method for interpreting SKS splitting observations, with application to one and two-layer anisotropic earth models, Geophys. J. Int. 154, 379-392, doi: 10.1046/j.1365-246X.2003.01937.x, 2003.

Montelli, R., G. Nolet, F.A. Dahlen and G. Masters, Finite-frequency tomography reveals a variety of plumes in the mantle, Science 303, 338-343, DOI: 10.1126/science.1092485, 2006.

Reed, M. and S. Barry, Functional Analysis, Elsevier, 400 pp, ISBN 0125850506, 1981.

Snyman, J.A., Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, Springer Publishing, 257 pp, ISBN 0-387-24348-8, 2005.

Talagrand, O. and P. Courtier, Variational Assimilation of Meteorological Observations with the Adjoint Vorticity Equation. I: Theory, Quarterly Journal of the Royal Meteorological Society 113, 1311–1328 DOI: 10.1002/qj.49711347812, 1987.

Taillandier, C. M. Noble, H. Chauris and H. Calandra, First-arrival traveltime tomography based on the adjoint-state method, Geophysics 74, WCB1–WCB10, doi: 10.1190/1.3250266, 2009.

Tromp, J., C. Tape and Q. Liu, Seismic Tomography, adjoint methods, time reversal and banana-doughnut kernels, Geophys. J, Int. 160, 195-216, doi: 10.1111/j.1365-246X.2004.02453.x, 2005.

van der Hilst, R.D., and M.V. De Hoop, Banana-doughnut kernels and mantle tomography, Geophys. J. Int. 163, 956-961, doi: 10.1111/j.1365-246X.2005.02817.x, 2005.

Wiggins, R.A., The general linear inverse problem: Implications of surface waves and free oscillations for Earth structure, Rev. Geophys. Space Phys. 10, 251-285, doi: 10.1029/RG010i001p00251, 1972.

Xu, Z., P. Chen and Y. Chen, Sensitivity Kernel for the Weighted Norm of the Frequency-Dependent Phase Correlation, Pure and Applied Geophysics 170, 353-371, DOI 10.1007/s00024-012-0507-3, 2012.

Zhao, L. T.H. Jordan, KB. Olsen and P. Chen, Fréchet Kernels for Imaging Regional Earth Structure Based on Three-dimensional Reference Models, Bulletin of Seismological Society of America 95, 2066-2080, doi: 10.1785/0120050081, 2005.

Zhu, H., E. Bozdag, D. Peter, and J. Tromp. Structure of the European upper mantle revealed by adjoint tomography. Nature Geoscience 5, 493-498, doi: 10.1038/ngeo1501, 2012.

Zhu, H., Y. Luo, T. Nissen-Meyer, C. Morency and J. Tromp, Elastic imaging and time-lapse migration based on adjoint methods, Geophysics 74, WCA167-WCA177, doi: 10.1190/1.3261747, 2009.

.Appendix

A.1. Adjoints of Some Simple Operators. A function $f(\mathbf{x},t)$ is self-adjoint, since:

$$\langle fu, v\rangle = \int_t \iiint_{\mathbf{x}} \{f(\mathbf{x},t)\, u(\mathbf{x},t)\}\, v(\mathbf{x},t)\, d^3\mathbf{x}\ dt$$

$$= \int_t \iiint_{\mathbf{x}} u(\mathbf{x},t)\, \{f(\mathbf{x},t)\, v(\mathbf{x},t)\}\ d^3\mathbf{x}\ dt = \langle u, fv\rangle \tag{A.1}$$

The first derivative $\partial/\partial t$ is anti-self-adjoint, since by integration by parts:

$$\langle \frac{\partial u}{\partial t}, v\rangle = \int_t \iiint_{\mathbf{x}} \left\{\frac{\partial}{\partial t}\, u(\mathbf{x},t)\right\} v(\mathbf{x},t)\, d^3\mathbf{x}\ dt = \iiint_{\mathbf{x}} \int_{t_{min}}^{t_{max}} \left\{\frac{\partial}{\partial t}\, u(\mathbf{x},t)\right\} v(\mathbf{x},t)\, dt\ \ d^3\mathbf{x}$$

$$= \iiint_{\mathbf{x}} \int_t \{u(\mathbf{x},t)\, v(\mathbf{x},t)\}\Big|_{t_{min}}^{t_{max}} dt\ \ d^3\mathbf{x} - \iiint_{\mathbf{x}} \int_t u(\mathbf{x},t) \left\{\frac{\partial}{\partial t} v(\mathbf{x},t)\right\} dt\ \ d^3\mathbf{x} = \langle u, -\frac{\partial v}{\partial t}\rangle \tag{A.2}$$

(provided that the fields decline to zero at $t \to \pm\infty$). The second derivative $\partial^2/\partial t^2$ is self-adjoint, since:

$$\left(\frac{\partial^2}{\partial t^2}\right)^\dagger = \left(\frac{\partial}{\partial t}\frac{\partial}{\partial t}\right)^\dagger = \left(\frac{\partial}{\partial t}\right)^\dagger\left(\frac{\partial}{\partial t}\right)^\dagger = \left(-\frac{\partial}{\partial t}\right)\left(-\frac{\partial}{\partial t}\right) = \frac{\partial^2}{\partial t^2} \tag{A.3}$$

The adjoint of a Green function inner product obeys:

$$\text{If}\quad \mathcal{G}_A y_A = \langle G_{A,B}, y_B\rangle_B \quad\text{then}\quad \mathcal{G}_A^\dagger y_A = \langle G_{B,A}^\dagger, y_B\rangle_B \tag{A.4}$$

since

$$\begin{aligned}\langle \mathcal{G}_A y_A, z_A\rangle_A &= \langle\langle G_{A,B}\; y_B\rangle_B, z_A\rangle_A = \langle\langle G_{A,B}\; y_B, z_A\rangle_A\rangle_B \\ &= \langle y_B, \langle G_{A,B}^\dagger z_A\rangle_A\rangle_B = \langle y_B, \mathcal{G}_B^\dagger z_B\rangle_B\end{aligned} \tag{A.5}$$

The adjoint of a convolution is a cross-correlation:

$$\begin{aligned}\langle a * u, v\rangle &= \int_t \iiint_{\mathbf{x}} \int_\tau a(\tau)u(t-\tau)\, d\tau\; v(t)\; d^3\mathbf{x}\; dt \\ \int_{\tau'} \iiint_{\mathbf{x}} u(\tau') \int_\tau a(\tau)\; v(\tau+\tau')\;\; d\tau\; d^3\mathbf{x}\; d\tau' &= \langle u, a \star v\rangle\end{aligned} \tag{A.6}$$

Here we have employed the transformation $\tau' = t - \tau$.

The adjoint of a matrix operator is the transposed matrix of adjoints:

$$\langle \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \rangle = \langle \begin{bmatrix} \mathcal{L}_{11}a_1 + \mathcal{L}_{12}a_2 \\ \mathcal{L}_{21}a_1 + \mathcal{L}_{22}a_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \rangle$$

$$= \langle \mathcal{L}_{11}a_1, b_1 \rangle + \langle \mathcal{L}_{12}a_2, b_1 \rangle + \langle \mathcal{L}_{21}a_1, b_2 \rangle + \langle \mathcal{L}_{22}a_2, b_2 \rangle$$

$$= \langle a_1, \mathcal{L}_{11}^{\dagger}b_1 \rangle + \langle a_2, \mathcal{L}_{12}^{\dagger}b_1 \rangle + \langle a_1, \mathcal{L}_{21}^{\dagger}b_2 \rangle + \langle a_2, \mathcal{L}_{22}^{\dagger}b_2 \rangle$$

$$= \langle \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathcal{L}_{11}^{\dagger}b_1 + \mathcal{L}_{21}^{\dagger}b_2 \\ \mathcal{L}_{12}^{\dagger}b_1 + \mathcal{L}_{22}^{\dagger}a_2 \end{bmatrix} \rangle = \langle \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \begin{bmatrix} \mathcal{L}_{11}^{\dagger} & \mathcal{L}_{12}^{\dagger} \\ \mathcal{L}_{21}^{\dagger} & \mathcal{L}_{22}^{\dagger} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \rangle$$

(A.7)

The operator $\boldsymbol{\mathcal{L}}$ of the elastic wave equation is self-adjoint. Each diagonal element is self-adjoint; for instance, the (1,1) element:

$$\left(\rho \frac{\partial^2}{\partial t^2} - (\lambda + 2\mu) \frac{\partial^2}{\partial x^2} - \mu \frac{\partial^2}{\partial y^2} - \mu \frac{\partial^2}{\partial z^2} - \frac{\partial(\lambda + 2\mu)}{\partial x} \frac{\partial}{\partial x} - \frac{\partial \mu}{\partial y} \frac{\partial}{\partial y} - \frac{\partial \mu}{\partial z} \frac{\partial}{\partial z} \right)^{\dagger} =$$

$$\left(\rho \frac{\partial^2}{\partial t^2} - \frac{\partial}{\partial x} (\lambda + 2\mu) \frac{\partial}{\partial x} - \frac{\partial}{\partial y} \mu \frac{\partial}{\partial y} - \frac{\partial}{\partial z} \mu \frac{\partial}{\partial z} \right)^{\dagger}$$

$$= \rho \frac{\partial^2}{\partial t^2} - \frac{\partial}{\partial x} (\lambda + 2\mu) \frac{\partial}{\partial x} - \frac{\partial}{\partial y} \mu \frac{\partial}{\partial y} - \frac{\partial}{\partial z} \mu \frac{\partial}{\partial z}$$

(A.8)

And each pair off diagonal elements are adjoints of one another; for instance for the (1,2) and (2,1) pair:

$$\left(-(\lambda+\mu)\frac{\partial^2}{\partial x \partial y}-\frac{\partial\lambda}{\partial x}\frac{\partial}{\partial y}-\frac{\partial\mu}{\partial y}\frac{\partial}{\partial x}\right)^{\dagger}=\left(-\frac{\partial}{\partial x}\lambda\frac{\partial}{\partial y}-\frac{\partial}{\partial y}\mu\frac{\partial}{\partial x}\right)^{\dagger}=-\frac{\partial}{\partial y}\lambda\frac{\partial}{\partial x}-\frac{\partial}{\partial x}\mu\frac{\partial}{\partial y}$$

$$-(\lambda+\mu)\frac{\partial^2}{\partial x \partial z}-\frac{\partial\lambda}{\partial y}\frac{\partial}{\partial x}-\frac{\partial\mu}{\partial x}\frac{\partial}{\partial y}=-\frac{\partial}{\partial y}\lambda\frac{\partial}{\partial x}-\frac{\partial}{\partial x}\mu\frac{\partial}{\partial y}$$

(A.9)

An adjoint can have different boundary conditions than the original operator. Consider the first derivative du/dt with the initial condition $u(t=0)=0$, written as the operator $\mathcal{L}u$. It has a finite difference approximation **Lu**, where:

$$\mathcal{L} \approx \mathbf{L} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

(A.10)

The first row of **Lu** involves only the first element of **u** and is the initial condition; the subsequent rows are the first differences between adjacent elements of **u** and is the derivative. The corresponding approximation of operator $\mathcal{L}^{\dagger}$ is the transposed matrix:

$$\mathcal{L}^{\dagger} \approx \mathbf{L}^{\mathrm{T}} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

(A.11)

Then boundary condition has moved to the last row of $\mathbf{L}^{\mathrm{T}}\mathbf{u}$ has become an end condition and, as expected, the signs of the first differences have flipped.

A. 2. Derivative of the Inverse of an Operator. Perturbation theory can be used to show that, for a small number ε, the inverse of $\mathcal{L}_0 + \varepsilon\mathcal{L}_1$is [Menke and Eilon, 2015]:

$$[\mathcal{L}_0 + \varepsilon\mathcal{L}_1]^{-1} = \mathcal{L}_0^{-1} - \varepsilon\mathcal{L}_0^{-1}\mathcal{L}_1\mathcal{L}_0^{-1} + O(\varepsilon^2) \tag{A.12}$$

This expression is verified by showing that applying the operator to its inverse, and the inverse to the operator, both yield the identity operator $\mathcal{I}$:

$$[\mathcal{L}_0 + \varepsilon\mathcal{L}_1][\mathcal{L}_0^{-1} - \varepsilon\mathcal{L}_0^{-1}\mathcal{L}_1\mathcal{L}_0^{-1} + O(\varepsilon^2)] = \mathcal{I} + \varepsilon[\mathcal{L}_1\mathcal{L}_0^{-1} - \mathcal{L}_1\mathcal{L}_0^{-1}] + O(\varepsilon^2) = \mathcal{I} + O(\varepsilon^2)$$

and

$$[\mathcal{L}_0^{-1} - \varepsilon\mathcal{L}_0^{-1}\mathcal{L}_1\mathcal{L}_0^{-1} + O(\varepsilon^2)][\mathcal{L}_0 + \varepsilon\mathcal{L}_1] = \mathcal{I} + \varepsilon[\mathcal{L}_0^{-1}\mathcal{L}_1 - \mathcal{L}_0^{-1}\mathcal{L}_1] + O(\varepsilon^2) = \mathcal{I} + O(\varepsilon^2) \tag{A.13}$$

The derivative rule then follows from the definition of the derivative:

$$\frac{\partial}{\partial m}\mathcal{L}^{-1} = \lim_{\Delta m \to 0} \frac{\left[\mathcal{L} + \left[\frac{\partial \mathcal{L}}{\partial m}\right]\Delta m\right]^{-1} - \mathcal{L}^{-1}}{\Delta m} = -\mathcal{L}^{-1}\left[\frac{\partial \mathcal{L}}{\partial m}\right]\mathcal{L}^{-1} \tag{A.14}$$

A.3. The Adjoint Field as a Lagrange Multiplier. For clarity, we derive the derivative $\partial E/\partial m$ (Equation 3.2) in the discrete case where the field $\mathbf{u}(\mathbf{x}, t)$ is approximated by a discrete vector u_i with $i = 1,2,\cdots K$ that contains $\mathbf{u}(\mathbf{x}, t)$ evaluated all permutations of components, positions and times. We consider a $(K+1)$-dimensional vector space consisting of the elements of the field plus a single model parameter m (Figure A1). In this view, the elements of the field and the model parameter are all independent variables. Using Einstein notation, where repeated indices imply summation, the total error is $E = e_i e_i$ with $e_i = u_i^{obs} - u_i$. The error is independent of m, and is axially-symmetric about the line $u_i = u_i^{obs}$ (cylinder in Figure A1). The field obeys a matrix equation $C_i \equiv L_{ij}(m)\, u_j - f_j = 0$, where the $K \times K$ matric $\mathbf{L}(\mathrm{m})$ is a discrete analogue of a differential operator $\mathcal{L}(m)$ and its associated boundary conditions. The use of the abbreviation C_i highlights the sense in which each row of the matrix equation is a separate constraint applied at a different point in space and time. Since, for any given value of m, the matrix equation can be solved for a unique field $u_i(m)$, but the value of m can be freely varied, these constraint trace out a curve (grey line in Figure A.1). We want to know the gradient of the error resolved onto this curve, a quantity that we will refer to as $\partial E/\partial m|_{\mathbf{C}=0}$.

The explicit calculation of $\partial E/\partial m|_{C=0}$ in Section 3.2 starts with:

$$\frac{\partial E}{\partial m} = 2e_i \frac{\partial e_i}{\partial m} = -2e_i \frac{\partial u_i}{\partial m} = -2\mathbf{e}^{\mathrm{T}} \frac{\partial \mathbf{u}}{\partial m} \tag{A.15}$$

and substitutes in the solution $\mathbf{u} = \mathbf{L}^{-1}\mathbf{f}$. The error is then an explicit function of the model parameter and can be differentiated with respect to it:

$$\left.\frac{\partial E}{\partial m}\right|_{C=0} = -2\mathbf{e}^{\mathrm{T}} \frac{\partial}{\partial m}(\mathbf{L}^{-1}\mathbf{f}) = -2\mathbf{e}^{\mathrm{T}} \frac{\partial \mathbf{L}^{-1}}{\partial m}\mathbf{f} = 2\mathbf{e}^{\mathrm{T}}\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial m}\mathbf{L}^{-1}\mathbf{f}$$

$$= 2\mathbf{e}^{\mathrm{T}}\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial m}\mathbf{u} = 2\left(\frac{\partial \mathbf{L}^{\mathbf{T}}}{\partial m}\mathbf{L}^{-1\mathbf{T}}\mathbf{e}\right)^{\mathrm{T}}\mathbf{u}$$

(A.16)

Note that we have used the fact that $\mathbf{f}$ is not a function of m and have applied the rule $\partial\mathbf{L}^{-1}/\partial m = -\mathbf{L}^{-1}(\partial\mathbf{L}/\partial m)\mathbf{L}^{-1}$. Defining the adjoint field to be $\boldsymbol{\lambda} = \mathbf{L}^{-1\mathbf{T}}\mathbf{e}$ leads to equations:

$$\left.\frac{\partial E}{\partial m}\right|_{C=0} = 2\left(\frac{\partial\mathbf{L}^{\mathbf{T}}}{\partial m}\boldsymbol{\lambda}\right)^{\mathrm{T}}\mathbf{u} \quad \text{and} \quad \mathbf{L}^{\mathbf{T}}\boldsymbol{\lambda} = \mathbf{e}$$

(A.17)

We obtain the continuous limit by replacing vectors with functions, matrices with operators and the dot product with the inner product:

$$\left.\frac{\partial E}{\partial m}\right|_{\mathbf{C}=0} = \; 2\left\langle\frac{\partial\boldsymbol{\mathcal{L}}^{\dagger}}{\partial m}\lambda, u\right\rangle \quad \text{and} \quad \boldsymbol{\mathcal{L}}^{\dagger}\lambda = \mathrm{e}$$

(A.18)

These expressions are the same as those derived previously in Equation 3.2.

The same result can be achieved *implicitly*, using the method of Lagrange multipliers. We focus on a point m_0 along the curve $C_i = 0$. The derivative resolved onto the curve is the part of ∇E that is parallel to the curve; or equivalently, the part of ∇E that is *perpendicular* to the gradients ∇C_i of all of the constraints.

The standard way of removing the part of ∇E that is parallel to ∇C_i is to subtract from ∇E just the right amount of each ∇C_i. We start by writing:

$$\nabla E|_{\mathbf{C}=0} = \nabla E + 2\,\nabla C_i\,\lambda_i$$

(A.19)

where λ_i are a set of unknown coefficients (called *Lagrange multipliers*) and the factor of 2 has been added to simplify the subsequent derivation. The coefficients are determined by the conditions that $\nabla E|_{\mathbf{C}=0}$ is perpendicular to ∇C_i:

$$\nabla C_i \cdot \nabla E|_{\mathbf{C}=0} = 0 \tag{A.20}$$

Various derivatives are needed to perform this dot product:

$$\frac{\partial E}{\partial u_k} = 2e_i \frac{\partial e_i}{\partial u_k} = -2e_i \frac{\partial u_i}{\partial u_k} = -2e_i \delta_{ik} = -2e_k \quad \text{so} \quad \nabla_{\mathrm{u}} E = -2\mathbf{e}$$

$$\frac{\partial E}{\partial m} = 0 \quad \text{so} \quad \nabla_{\mathrm{m}} E = 0$$

$$\frac{\partial C_i}{\partial u_k} = \frac{\partial}{\partial u_k}\left(L_{ij}\, u_j - f_j\right) = L_{ij}\, \frac{\partial u_j}{\partial u_k} = L_{ij}\, \delta_{ik} = L_{ik} \quad \text{so} \quad \nabla_{\mathrm{u}} C_i = \mathbf{L}$$

$$\frac{\partial C_i}{\partial m} = \frac{\partial}{\partial m}\left(L_{ij}\, u_j - f_j\right) = \frac{\partial L_{ij}}{\partial m} u_j \quad \text{so} \quad \nabla_{\mathrm{m}} C_i = \frac{\partial \mathbf{L}}{\partial m}\mathbf{u} \tag{A.21}$$

Defining $[\boldsymbol{\lambda}]_{\mathbf{i}} \equiv \lambda_i$, $\nabla E \equiv [\nabla_{\mathrm{u}} E, \nabla_{\mathrm{m}} E]^{\mathrm{T}}$ and $\nabla C_i \equiv [\nabla_{\mathrm{u}} C_i, \nabla_{\mathrm{m}} C_i]^{\mathrm{T}}$ we have:

$$\begin{bmatrix} \nabla_{\mathrm{u}} E|_{\mathbf{C}=0} \\ \left.\dfrac{\partial E}{\partial m}\right|_{\mathbf{C}=0} \end{bmatrix} = \nabla E + 2\lambda_i \nabla C_i = \begin{bmatrix} -2\mathbf{e} \\ 0 \end{bmatrix} + 2\begin{bmatrix} \mathbf{L}^{\mathrm{T}}\boldsymbol{\lambda} \\ \boldsymbol{\lambda}^{\mathrm{T}} \dfrac{\partial \mathbf{L}}{\partial m}\mathbf{u} \end{bmatrix} \tag{A.22}$$

The coefficients $\boldsymbol{\lambda}$ are determined by that condition that the dot product $\nabla C_i \cdot \nabla E|_{\mathbf{C}=0}$ is zero:

$$\mathbf{L}(\mathbf{e}-\mathbf{L}^{\mathrm{T}}\boldsymbol{\lambda})+\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial\mathbf{L}}{\partial m}\mathbf{u}=0 \tag{A.23}$$

The choice

$$\mathbf{L}^{\mathrm{T}}\boldsymbol{\lambda}=\mathbf{e} \tag{A.24}$$

zeros the first term on the l.h.s. It also zeros the second term, since:

$$\begin{aligned}\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\boldsymbol{\lambda}^{\mathrm{T}}\frac{\partial\mathbf{L}}{\partial m}\mathbf{u}&=\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\mathbf{e}^{\mathrm{T}}\mathbf{L}^{-1}\frac{\partial\mathbf{L}}{\partial m}\mathbf{L}^{-1}\mathbf{f}=-\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\mathbf{e}^{\mathrm{T}}\frac{\partial\mathbf{L}^{-1}}{\partial m}\mathbf{f}\\&=-\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\mathbf{e}^{\mathrm{T}}\frac{\partial}{\partial m}(\mathbf{L}^{-1}\mathbf{f})=-\mathbf{u}^{\mathrm{T}}\,\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\mathbf{e}^{\mathrm{T}}\frac{\partial\mathbf{u}}{\partial m}=0\end{aligned} \tag{A.25}$$

Here we have used the rules $\boldsymbol{\lambda}^{\mathrm{T}}=\mathbf{e}^{\mathrm{T}}\mathbf{L}^{-1}$, $\mathbf{u}=\mathbf{L}^{-1}\mathbf{f}$ and $\partial\mathbf{L}^{-1}/\partial m=\mathbf{L}^{-1}(\partial\mathbf{L}/\partial m)\mathbf{L}^{-1}$. The derivative $\partial\mathbf{u}/\partial m$ is zero because, in the context of this derivation, $\mathbf{u}$ and m are independent variables. The lower part of Equation (A.20) gives:

$$\left.\frac{\partial E}{\partial m}\right|_{\mathbf{C}=0}=2\left(\frac{\partial\mathbf{L}}{\partial m}\mathbf{u}\right)^{\mathrm{T}}\boldsymbol{\lambda}=2\left(\frac{\partial\mathbf{L}^{\mathrm{T}}}{\partial m}\boldsymbol{\lambda}\right)^{\mathrm{T}}\mathbf{u} \tag{A.26}$$

This equation and Equation (A.14) are precisely the same as those derived by the explicit method. Thus the Adjoint field $\boldsymbol{\lambda}$ can be interpreted as a Lagrange multiplier that arises from the constraint that the field exactly satisfies a differential equation at every point in space and time.

Menke, Figure 1.

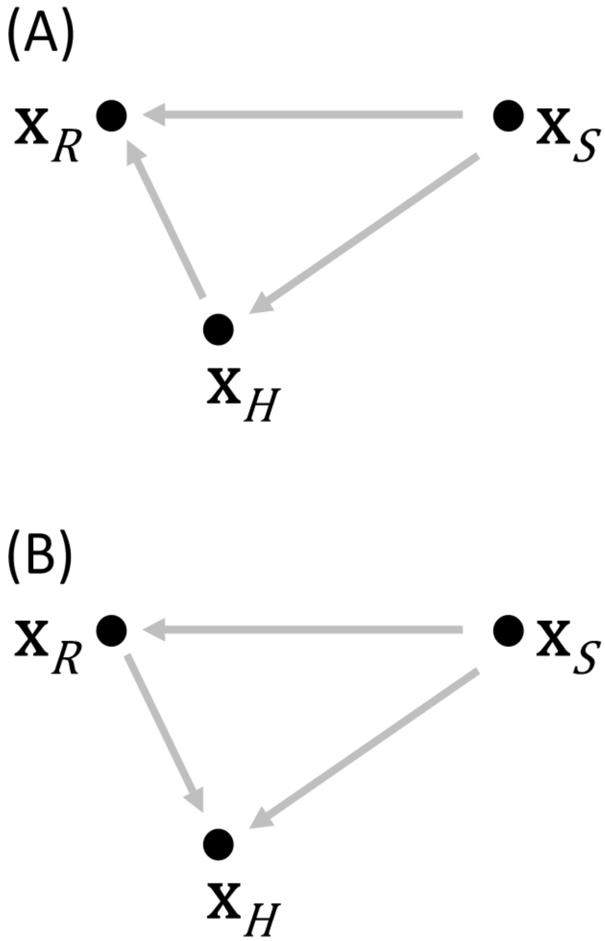

Fig. 1. (A) The Direct Method focuses on the two fields incident upon a receiver at $\mathbf{x}_R$: direct wave from the source at $\mathbf{x}_S$ that follows the SR path; and a scattered wave that has interacted with a heterogeneity at $\mathbf{x}_H$ and follows the SHR path. (B) The Adjoint Method focuses upon the fields incident upon the heterogeneity at $\mathbf{x}_H$, which includes the direct wave that follows the SH path and the adjoint field that follows the RH path. The source of the adjoint field depends upon the direct wave at the receiver, which follows the SR path.

Menke, Figure 2.

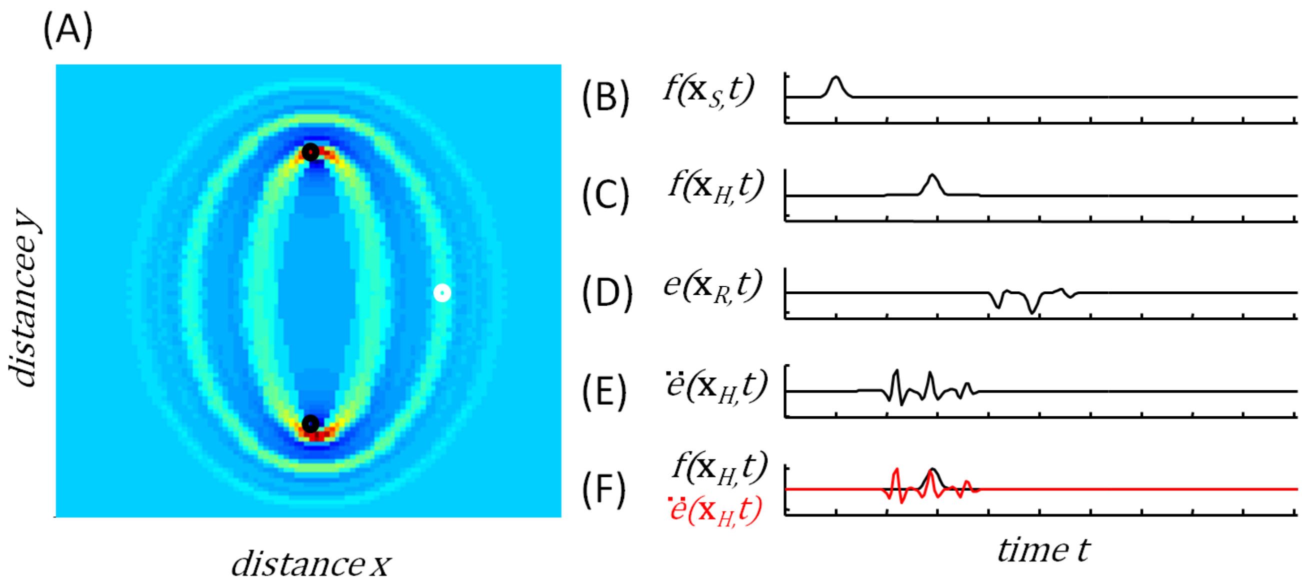

Figure 2. (A) The partial derivative $\partial E/\partial m$ (colors) for a point slowness heterogeneity in a homogeneous acoustic whole space. The amplitude of the derivatives track ellipses of equal travel time from source (lower black circle) to heterogeneity to receiver (upper black circle). (B) The source time function $(\mathbf{x}_s, t)$. (C) The source time function $f(\mathbf{x}_H, t)$ time-shifted to the receiver at $\mathbf{x}_R$. (D) The error $e(\mathbf{x}_R, t)$ at the receiver. (E) The second derivative $\ddot{e}(\mathbf{x}_H, t)$, time-shime shfted to a heterogeneity at $\mathbf{x}_H$ (white circle in Part A). (F) Comparison $\ddot{e}(\mathbf{x}_H, t)$,(red curve) and $f(\mathbf{x}_H, t)$. The overall in high-amplitudes leads to one of the elliptical bands in Part A.

Menke, Figure 3

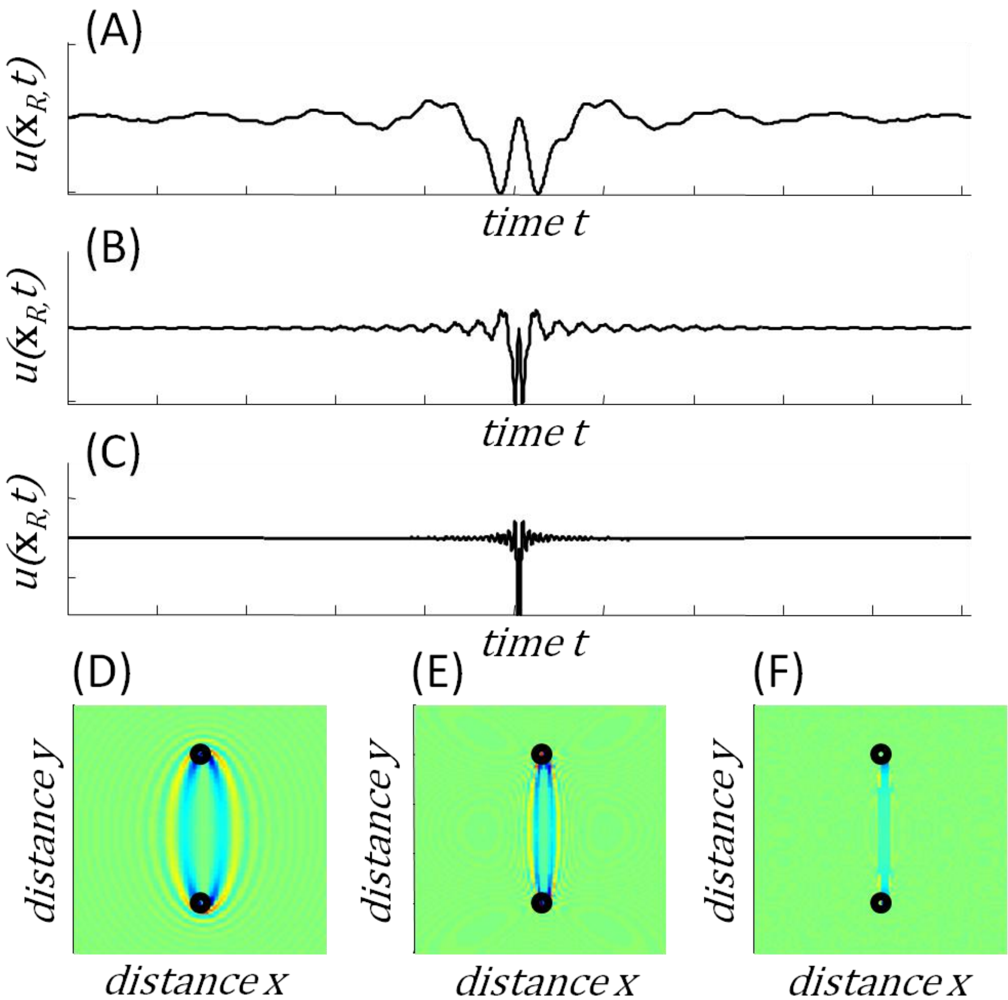

Fig. 3. Quantities associated with the banana-doughnut kernel $\partial\tau/\partial m$. (A)-(C). Three band-limited pulses $u^0(\mathbf{x}_R, t)$ originating from a source at $\mathbf{x}_R$ and observed at a receiver at $\mathbf{x}_R$ The peak frequency of these fields increases from A to C. (D)-(F) Banana-doughnut kernels (colors) for point slowness homogeneities distributed on the (x, y) plane corresponding to the pulses in Parts A-C. Note that the kernels narrow and become more linear with increasing frequency, as diffraction behavior become less importance and ray-like behavior begins to dominate.

Menke, Figure 4.

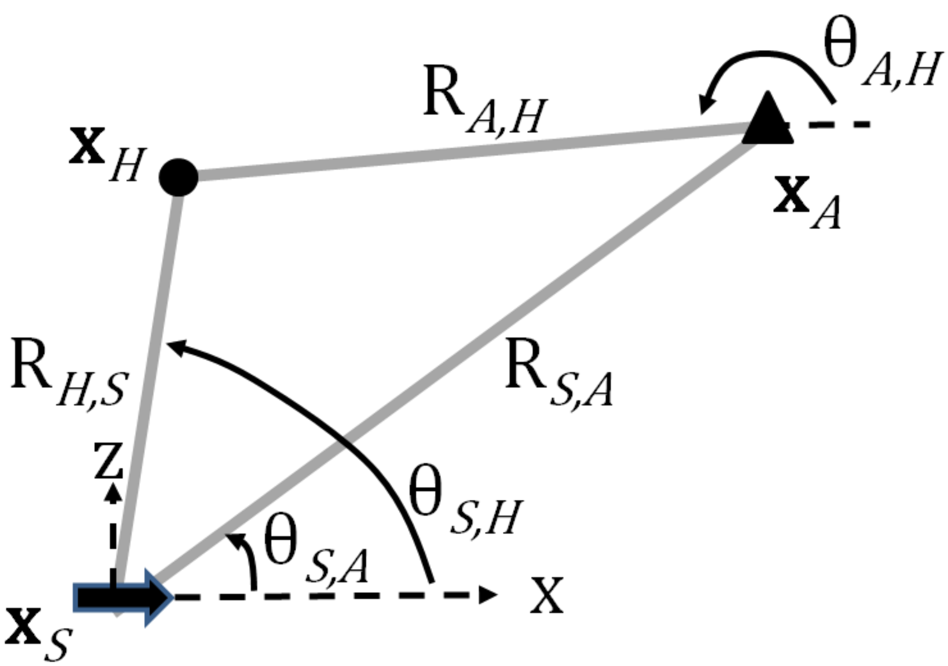

Fig. 4. Geometry used in the cross-convolution example See text for further discussion.

Menke, Figure 5.

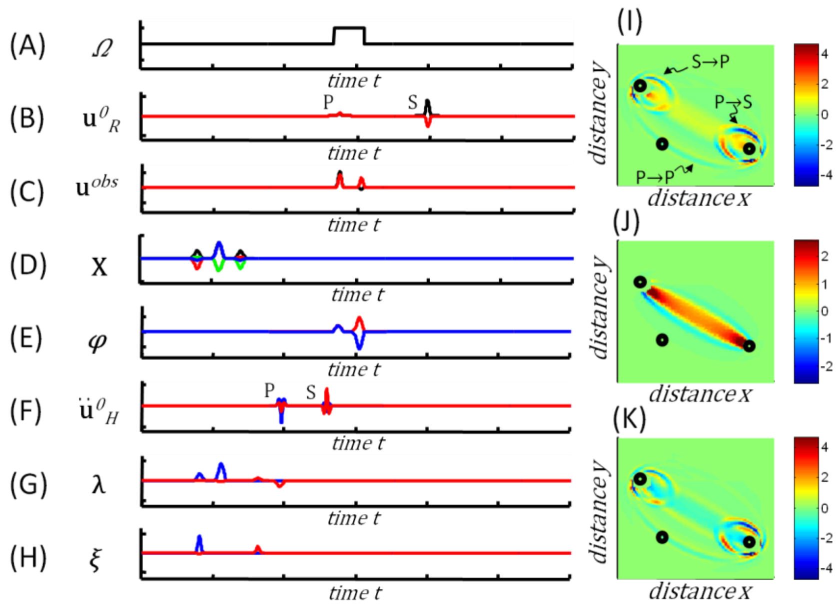

Fig. 5. Quantities associated with the cross-convolution partial derivative $\partial\Psi/\partial m$ and the power partial derivative $\partial P/\partial m$. (A) The window function Ω. (B) The horizontal (black) and vertical (red) components of the unperturbed field at the observer. (C) The horizontal (black) and vertical (red) components of the observed field at the observer, after windowing. (D) The four components of the cross-correlation function **X**. (E) The source $\boldsymbol{\varphi}$ of the adjoint field $\boldsymbol{\lambda}$ (F) The second derivative of the unperturbed field at a heterogeneity (location marked on Part I). (G) The adjoint field $\boldsymbol{\lambda}$, which is associated with the cross-correlation measure. (H) The adjoint field $\boldsymbol{\xi}$, which is associated with power. (I) The partial derivative $\partial\Psi/\partial m$ (colors) for point density homogeneities distributed on the (x, y) plane The source and receiver and the heterogeneity singled out in Parts A-H are shown (circles) and the contribution of various scattering interactions are maked. (J) Same as Part I, but for the partial derivative $\partial P/\partial m$. (J) Same as Part I, but for the partial derivative $\partial\Phi/\partial m$.

Menke, Figure 6.

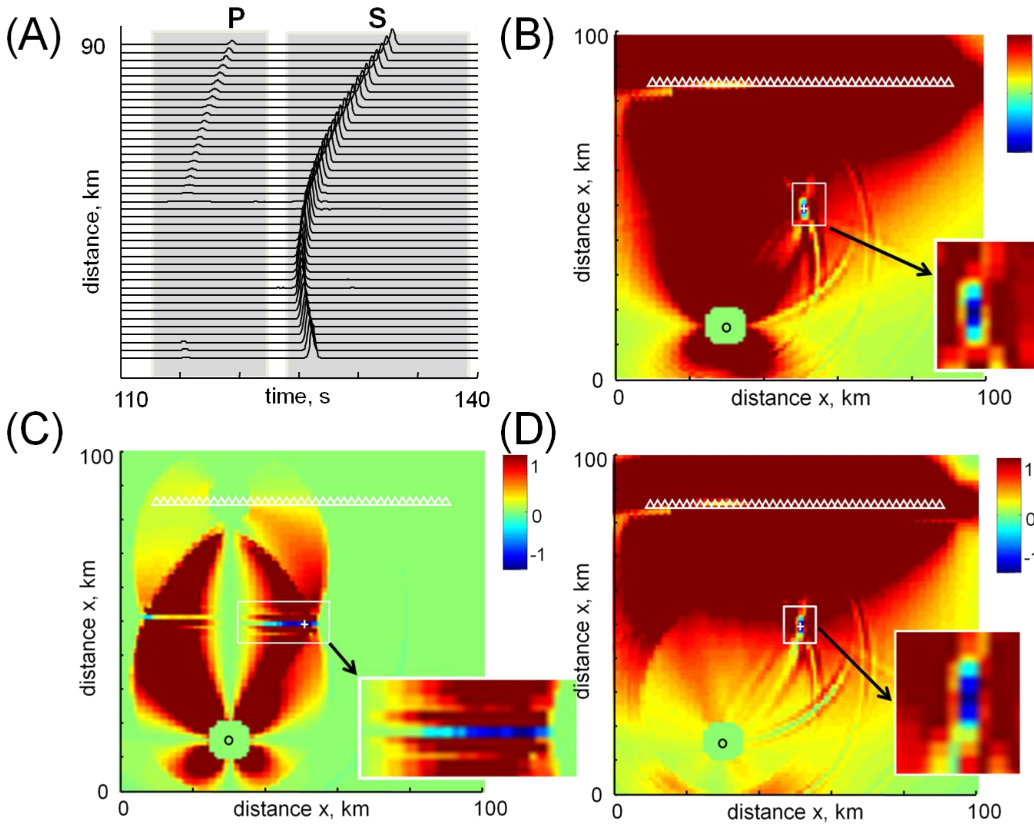

Fig.6. Quantities associated with the resolution test of the cross-convolution method. (A) The horizontal-component of the wave field (curves) is observed by a linear array of receivers and is due to a source in an elastic medium containing a “true” point heterogeneity (located in the white box in Part B).. The vertical component (not shown) was also used. P and S wave windows are shown (gray shading). (B) The partial derivative $\partial\Phi/\partial m$ (colors) for point density homogeneities distributed on the (x, y) plane. The source (circle) is at the lower left and the linear array of receivers (line of triangles) is near the top. The minimum (inset, blue) is collocated with the true heterogeneity and is spatially-localized, implying excellent resolution. (C) Same as Part B, except that the data are windowed around the P wave arrival. (C) Same as Part B, except windowed around for the S wave.

Menke, Figure 7 (really Figure A.1.)

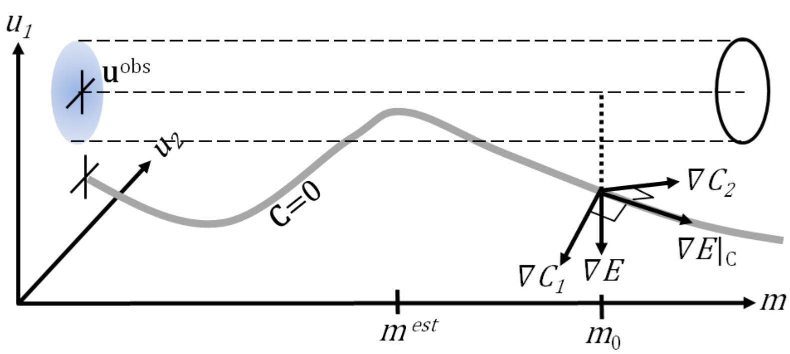

Fig. A.1. Geometrical interpretation of the process of computing the gradient of the total error subject to constraints that the field satisfies a differential equation. See text for further discussion.