Scientific
Research
Publishing

A Connection between Geometrical Spreading and the Adjoint Field in Travel Time Tomography

William Menke

Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA
Email: MENKE@LDEO.COLUMBIA.EDU

Abstract

The goal of tomography is to reconstruct a spatially-varying image function $s(x, m)$, where x is position and m is a finite-length vector of parameters. Many reconstruction methods minimize the total L_2 error $E \equiv e^{\mathrm{T}} e$, where individual errors e_i quantify misfit between predictions and observations, to quantify goodness of fit. So-called adjoint state methods allow the gradient $\partial E / \partial m_i$ to be computed extremely efficiently from an adjoint field, facilitating image reconstruction by gradient-descent methods. We examine the structure of the differential equation for the adjoint field under the ray approximation and find that it has the same form as the transport equation, whose solution involves the well-known geometrical spreading function *R*. Consequently, as *R* is routinely tabulated as part of a ray calculation, no extra work is needed to compute the adjoint field, permitting a rapid calculation of the gradient $\partial E / \partial m_i$.

Keywords

Adjoint State Method, Eikonal Equation, Geometrical Spreading, Gradient-Descent Method, Transport Equation, Travel Time Tomography

1. Introduction

Acoustic, electromagnetic and seismic waves are routinely used to probe the media through which they propagate, and especially to image the spatially-varying velocity field. A fundamental property of these waves that commonly is exploited is their travel time *T*, defined as the time between the generation of a wave at its source to its detection by a distant observer. In many cases, travel

times can be computed under the ray approximation of the exact wave equation, which is valid at high-frequencies when the scale length of heterogeneities in the medium is much larger than the wavelength of the waves. Since the 1970's, the simplicity of ray calculations has underpinned the use of travel time tomography in a variety of disciplines, including seismology [1] [2], oceanography [3], petroleum exploration [4], geotechnical engineering [5] and cosmology [6]. In some disciplines, ray-based tomography is being superseded by full wavefield methods [7] [8]; nevertheless, it remains an important part of a tomographer's toolbox on account of its computational efficiency.

Over the last several decades, the development of the so-called adjoint state method [9] has allowed tomographic imaging to be applied in cases where it was hitherto fore infeasible, because of vastly reduced computational effort. To date, this efficiency mainly has used to enable computationally-intensive forms of tomography, and especially to full wavefield tomography [7] [8]. Nevertheless, adjoint state methodology is very widely applicable. It has the potential for significantly speeding up even computationally-light problems, including ray-based tomography. The feasibility of using adjoint state methods in this form of tomography was first investigated by [10], who demonstrate its effectiveness. In this paper, we further explore it application. We study the mathematical structure of the differential equation that arises out of the adjoint state method (the equation for the so-called adjoint field) and show that it is very closely related to and in important cases identical to the transport equation of ray theory. This relationship provides an intuitive understanding of the adjoint field and suggests further ways of obtaining further computational efficiency.

Our analysis is divided into four sections: first, we review how the adjoint state method is used to streamline the computation of a critical quantity need to perform tomography; second, we review the concept of the geometrical spreading of rays and its connection to the transport equation; third, we use the adjunct state method to derive and solve the differential equation for the adjoint field; and lastly, we show that the adjoint equation is very closely related to the transport equation and that its solution can be trivially constructed when the solution to the transport equation (the geometrical spreading function) is known.

2. The Adjoint State Method for Computing the Error Derivative

The main purpose of this section is to define the error derivative, discuss its usefulness and review how the adjoint state method is used to compute it, in the special case where the unknown image is linked to the observed data via the source term in a linear differential equation.

Many types of tomography involve a set of observations $\left\{T_j^{obs}, j = 1, \cdots, N\right\}$, each associated with a spatial position $x^{(j)}$, which are related to an unknown image function $s(x)$ by the possibly-nonlinear map $T_j = g_j(s)$. Here, $x \equiv [x, y, z]^T$ are real spatial coordinates and $[.]^T$ denotes transpose. Usually, the data, spatial coordinates and image function are presumed to be real. Be-

cause no finite number of observations can define a continuous function, the image is usually approximated with a finite number of parameters; that is, as $s(\boldsymbol{x}, \boldsymbol{m})$ with $\{m_j, j = 1, \cdots, M\}$. For example, the image might be divided into voxels, each with a value m_j. A common approach to image reconstruction is to define individual errors, $e_j \equiv T_j^{obs} - T_j$ and total L_2 error $E \equiv \boldsymbol{e}^{\mathrm{T}} \boldsymbol{e}$ and then to find the $\boldsymbol{m}$ that minimizes $\Phi = E + \varepsilon^2 \|\mathcal{P}s\|$ ([11]; see also [12] [13] [14]). Here, ε is an empirically-chosen constant and $\mathcal{P}$ is a linear operator that embodies prior information, such as smoothness. Among the many optimization procedures put forward for solving this problem, several commonly-used ones, based on the linearized least-squares method (e.g. [14] [15]), employ the partial derivative $G_{jk} \equiv \partial T_j / \partial m_k$ to compute the data perturbation ΔT_j associated with the image perturbation $\Delta \boldsymbol{m}$ via $\Delta \boldsymbol{T} = \boldsymbol{G} \Delta \boldsymbol{m}$. Alternatively, other common-used procedures, based on the gradient-descent method (e.g. [16]), use the partial derivative $H_j \equiv \partial E / \partial m_j$ to compute the error perturbation ΔE associated with the image perturbation $\Delta \boldsymbol{m}$ via $\Delta E = \boldsymbol{H}^{\mathrm{T}} \Delta \boldsymbol{m}$.

The error derivative H_j can be computed from the data derivative G_{jk}:

$$H_j \equiv \frac{\partial E}{\partial m_j} = \frac{\partial}{\partial m_j} \sum_{k=1}^{N} e_k e_k = -2 \sum_{k=1}^{N} G_{kj} e_k \quad \text{or} \quad \boldsymbol{H} = -2 \boldsymbol{G}^{\mathrm{T}} \boldsymbol{e} \tag{1}$$

However, recent advances in tomography have followed the realization that H_j often can be computed without first computing G_{jk}, giving gradient-descent methods tremendous computational advantage over least-squares methods. The underlying idea of these adjoint state methods [9] [17] is to promote the error to a field $e(\boldsymbol{x})$, with the assumption that the data have been measured everywhere, so that $E = (e, e)$ and:

$$H_j \equiv \frac{\partial E}{\partial m_j} = \frac{\partial}{\partial m_j}(e, e) = 2\left(e, \frac{\partial e}{\partial m_j}\right) = -2\left(e, \frac{\partial T}{\partial m_j}\right) \tag{2}$$

Here $(.,.)$ is the inner product over spatial coordinates. Now, consider the simple case in which the data solves the linear differential equation $\mathcal{L}T = s$ (together with some appropriate boundary condition). Here, the image $s(\boldsymbol{x}, \boldsymbol{m})$ is the source term in the differential equation. By differentiating the differential equation, we obtain an expression for G_{jk}:

$$G_{jk} = \left. \frac{\partial T}{\partial m_k} \right|_{\boldsymbol{x}^{(j)}} \quad \text{with} \quad \mathcal{L} \frac{\partial T}{\partial m_k} = \frac{\partial s}{\partial m_k} \tag{3a}$$

The partial derivative of total error H_j is computed by differentiating $T = \mathcal{L}^{-1} s$ to yield $\partial T / \partial m_j = \mathcal{L}^{-1} \partial s / \partial m_j$ and inserting into (2):

$$H_j = -2\left(e, \mathcal{L}^{-1} \frac{\partial s}{\partial m_j}\right) = -2\left(\left(\mathcal{L}^{-1}\right)^{\dagger} e, \frac{\partial s}{\partial m_j}\right) = -2\left(\lambda, \frac{\partial s}{\partial m_j}\right) \tag{3b}$$

Here † denotes adjoint and $\lambda \equiv \left(\mathcal{L}^{-1}\right)^{\dagger} e = \left(\mathcal{L}^{\dagger}\right)^{-1} e$ is an adjoint field that satisfies the so-called adjoint equation $\mathcal{L}^{\dagger} \lambda = e$ (with appropriate boundary conditions). Note that the error $e(\boldsymbol{x})$ plays the role of the source term in the adjoint equation.

Cases in which the error is known everywhere are uncommon, since they imply observations within the medium, as contrasted to on its boundary. More typical are the cases where error is at discrete points $x^{(j)}$ on the boundary. These cases are handled by writing defining a partial adjoint field $\tilde{\lambda}(x)$ and error density $\tilde{e}(x)$:

$$\lambda(x) = \iiint \tilde{\lambda}(x,x') \mathrm{d}^3 x' \text{ and } e(x) = \iiint \tilde{e}(x')\delta(x-x')\mathrm{d}^3 x' \qquad (4)$$

Here, $\delta(.)$ is the Dirac impulse function. The resulting equation $\mathcal{L}^\dagger \tilde{\lambda} = \tilde{e}(x')\delta(x-x')$ is then solved only for those points $x' = x^{(j)}$ at which the error is known and adjoint field λ is taken to be the sum of the $\tilde{\lambda}s$. This procedure is equivalent to solving the original adjoint equation with error:

$$e(x) = \sum_{j=1}^{N} e_j \delta\left(x - x^{(j)}\right) \qquad (5)$$

3. The Transport Equation of Ray Theory

The main purpose of this section is to review the geometrical interpretation of the transport equation and to highlight its link to ray divergence. However, in order to provide some background for readers unfamiliar with ray theory, and to establish nomenclature, we also present an abridged derivation of the equation.

In many cases, the imaging problem involves a field $u(x,t)$ that is a function of time t as well as spatial coordinates x and that satisfies a wave equation of the form $\partial^2 u/\partial t^2 - \mathcal{L}(m)u = \delta(t-t_0)\delta(x-x^{(0)})$. Here, the differential operator $\mathcal{L}$ contains only spatial derivatives and depends on parameters m. The equation reduces to the spatial equation $-\omega^2 - \mathcal{L}(m)\hat{u} = \delta(x-x^{(0)})$ after Fourier transformation of time t to angular frequency ω, where $\hat{\ }$ denotes a transformed variable. The ray approximation is the solution to this equation in the limit $|\omega| \to \infty$, and is achieved by postulating that the solution can be written as a Laurent series of the form [18]:

$$\hat{u}(x,\omega) = A(x,\omega)\exp\{i\omega T(x)\} \text{ with } A(x,\omega) = \sum_{k=0}^{\infty}(i\omega)^{-k}A^{(k)}(x) \qquad (6)$$

Here i is the imaginary unit. The travel time function $T(x)$ represents the time needed for a fluctuation in u to propagate from $x^{(0)}$ to x, and $A(x,\omega)$ represents its amplitude. The details of the ray solution depend on $\mathcal{L}(m)$; we consider the simple (and common) case $\mathcal{L}(m) = s^2\nabla^2$, where $s(x,m)$ is a slowness function; that is, a material property that is inversely proportional to the local propagation velocity. Inserting (4) into the differential equation and equating equal powers of ω lead to the Eikonal equation for $T(x)$:

$$\nabla T \cdot \nabla T = s^2 \text{ with boundary condition } T(x^{(0)}) = 0 \qquad (7)$$

and a sequence of equations for $A^{(k)}$, the lowest order of which is the transport equation [19]:

$$2\nabla T \cdot \nabla A^{(0)} + A^{(0)}\nabla^2 T = 0 \qquad (8)$$

The unit normal to a surface of equal travel time is $\hat{t}(\boldsymbol{x}) = s^{-1}\nabla T$. A sequence of these vectors connecting surfaces of increasing travel times defines a ray; that is, a parametric curve $\boldsymbol{x}(\ell)$ with arclength ℓ and tangent $\hat{t}(\ell)$ (Figure 1(A)). The volume enclosed by a group of rays is called a ray tube. The Eikonal equation, written as two coupled first order equations in $\boldsymbol{x}(\ell)$ and $\hat{t}(\ell)$ is:

$$\frac{\mathrm{d}x}{\mathrm{d}\ell} = \hat{t} \quad \text{and} \quad \frac{\mathrm{d}\hat{t}}{\mathrm{d}\ell} = \hat{t} \times \left(s^{-1}\nabla s \times \hat{t}\right)$$

with boundary conditions

$$\boldsymbol{x}(0) = \boldsymbol{x}^{(0)} \quad \text{and} \quad \hat{t}(0) = \hat{t}^{(0)} \tag{9}$$

The ray's starting point is $\boldsymbol{x}^{(0)}$ and its take-off direction is $\hat{t}^{(0)}$. Travel time is then the path integral of the slowness along the ray, as can be seen by manipulating the formula for the directional derivative $\mathrm{d}/\mathrm{d}\ell = \hat{t}_0 \cdot \nabla$:

$$T\left(\boldsymbol{x}(\ell)\right) = \int_0^\ell \left(\hat{t} \cdot \nabla T\right)\mathrm{d}\ell' = \int_0^\ell s(\ell')\mathrm{d}\ell' \tag{10}$$

The transport equation, written in terms of $\hat{t}(\ell)$, is:

$$-\frac{\nabla \mathcal{E}}{\mathcal{E}} \cdot \hat{t} = \nabla \cdot \hat{t} \quad \text{with} \quad \mathcal{E} \equiv s\left(A^{(0)}\right)^2 \tag{11}$$
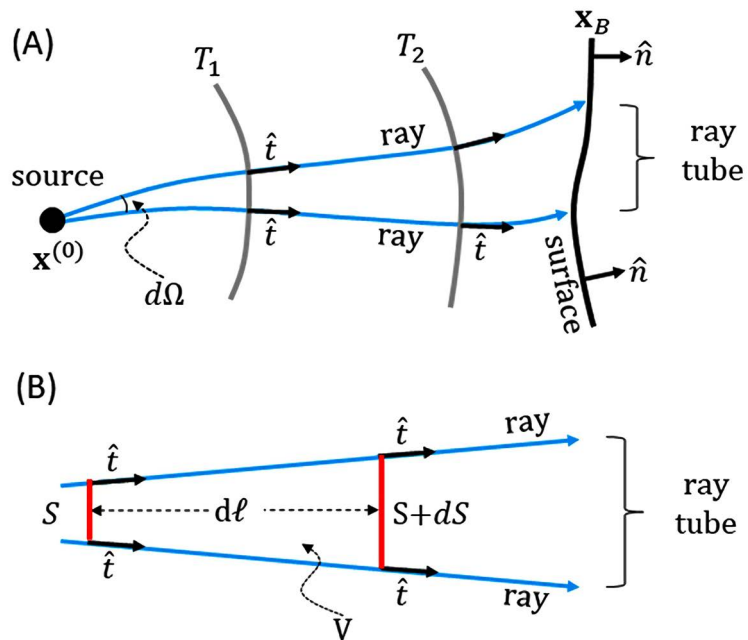

Figure 1. (A) Basic ray theory nomenclature. Wave propagates outward from a source at $\boldsymbol{x}^{(0)}$ (black circle), through the medium, to the surface $\boldsymbol{x}_S$ (with normal $\hat{n}$). Surfaces of equal travel time (wave fronts, grey curves) are labeled with their travel times T_1, T_2, etc., Normals to wave fronts define rays (blue curves) with tangents $\hat{t}$. Neighboring rays enclosing a solid angle $\mathrm{d}\Omega$ at the source define a ray tube. (B) Relationship between ray tangents $\hat{t}$ and ray tube cross-sectional area S. Gauss's theorem is applied to a small volume V along the ray tube, with the shape of a section of a cone, whose cross-sectional area S changes with arc-length ℓ and whose volume is $V = S\mathrm{d}\ell$. The tangent $\hat{t}$ is parallel to the sides of the section and normal to its ends. See text for further discussion.

The quantity $\nabla \cdot \hat{t}$ has a simple geometric interpretation, as can be seen by applying Gauss' theorem (e.g. [20]) to a volume V along a ray tube, which has the shape of a section of a cone (**Figure 1(B)**). The cross-sectional area of the ray tube increases from S on the end nearest to the source, to $S + \mathrm{d}S$ at a distance $\mathrm{d}\ell$ further away. For small volumes, the integral in Gauss' theorem is $(\nabla \cdot \hat{t})V$ where $V = S\mathrm{d}\ell$. The surface integral in Gauss' theorem has contributions only from the two ends of the cone, of $-S$ and $(S + \mathrm{d}S)$ respectively, which sum to $\mathrm{d}S$. Consequently, Gauss's theorem implies $(\nabla \cdot \hat{t}) = S^{-1}\mathrm{d}S/\mathrm{d}\ell$ and the transport equation becomes:

$$-\frac{1}{\mathcal{E}}\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}\ell} = \frac{1}{S}\frac{\mathrm{d}S}{\mathrm{d}\ell} \tag{12}$$

According to the transport equation, the fractional decrease in $\mathcal{E}$, measured along a ray, is equal to the fractional increase in area S of the ray tube. In many cases, the quantity $\mathcal{E}$ has the interpretation of the energy density, so the transport equation embodies conservation of energy. Conventionally, the area of the ray tube is written $S(\ell) = R^2(\ell)\mathrm{d}\Omega$, where $R^2(\ell)$ is the geometrical spreading function and $\mathrm{d}\Omega$ is the solid angle subtended by the ray tube at the source (e.g. [19]). Consequently, $\mathcal{E}(\ell) = cR^{-2}(\ell)$ where c is a constant. Ray-tracing algorithms that solve (9) typically tabulate both T and R (e.g. [21] [22]).

4. Adjoint Equation for Travel Time Tomography

The main purpose of this section is to derive and solve the adjoint equation needed to compute the quantity H_j, the derivative of the total travel time error with respect to a model parameter controlling the slowness of the medium. Our derivation focuses on expressing the equation in terms of quantities that vary along rays, so that it can be readily compared to the transport Equation (11). Our derivation is equivalent to, but different than, the one by [10], being a direct application of perturbation theory, as contrasted to one that employs Lagrange multipliers.

In travel time tomography, travel time observations $T(\boldsymbol{x}^{(j)})$ are considered to be the data, and the slowness $s(\boldsymbol{x})$, or rather its approximation $s(\boldsymbol{x},\boldsymbol{m})$, is the image function. In order to apply the adjunct methodology as outlined in the Introduction, the non-linear Eikonal equation must be linearized about a "background" solution. Let the slowness equal a background slowness s_0 plus a small perturbation εs_1, where ε is a small parameter, and the corresponding travel time equal a background travel time T_0 plus a small perturbation εT_1. Then to first order in ε, the Eikonal equation becomes:

$$\nabla T \cdot \nabla T = \nabla(T_0 + \varepsilon T_1) \cdot \nabla(T_0 + \varepsilon T_1) = (s_0 + \varepsilon s_1)^2 \approx s_0^2 + 2\varepsilon s_0 s_1 \tag{13}$$

Equating terms of equal order in ε yields equations for the background travel time T_0 and the perturbation in travel time T_1:

$$\nabla T_0 \cdot \nabla T_0 = s_0^2 \text{ and } \hat{t}_0 \cdot \nabla T_1 = s_1 \tag{14a,b}$$

Equation (14b) indicates that the component of ∇T_1 in the direction of the background ray direction $\hat{t}_0$ is s_1. Since s_1 plays the role the source term in the differential equation, the formulation in (3) is applicable. If we define $d\ell$ to be an increment of arc length along the unperturbed ray, then this is just an equation involving the directional derivative $d/d\ell = \hat{t}_0 \cdot \nabla$:

$$\frac{dT_1}{d\ell} = s_1 \text{ which has solution } T_1(\ell) = \int_0^{\ell} s_1(\ell')d\ell' \tag{15}$$

The perturbation in travel time is the integral of the perturbation in slowness along the unperturbed ray. We rewrite the equation for T_1 as:

$$\mathcal{L}T_1 = s_1 \text{ where } \mathcal{L} = s_0^{-1}\nabla T_0 \cdot \nabla = s_0^{-1}\begin{bmatrix} \dfrac{\partial T_0}{\partial x} & \dfrac{\partial T_0}{\partial y} & \dfrac{\partial T_0}{\partial z} \end{bmatrix}\begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix} \tag{16}$$

Using the rules $(\mathcal{L}_1\mathcal{L}_2)^\dagger = \mathcal{L}_2^{T\dagger}\mathcal{L}_1^{T\dagger}$ and $(d/dx)^\dagger = -d/dx$ (e.g., [23]) we obtain an expression for the adjoint equation:

$$\mathcal{L}^\dagger\lambda = -\begin{bmatrix} \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \end{bmatrix}\begin{bmatrix} \partial T_0/\partial x \\ \partial T_0/\partial y \\ \partial T_0/\partial z \end{bmatrix}\left(s_0^{-1}\lambda\right) = -\nabla \cdot \left[\nabla T_0\left(s_0^{-1}\lambda\right)\right] = -\nabla \cdot \left[\hat{t}_0\lambda\right] = e_0 \tag{17}$$

As is typical of first-order equations, the "left hand" boundary condition associated with $\mathcal{L}$ implies a "right hand" boundary condition for $\mathcal{L}^\dagger$ (e.g. [14]); that is, while $T = 0$ at the source $\ell = 0, \lambda = 0$ at as the end point of the ray $\ell = \ell_B$ (where it touches the boundary of the medium).

The adjoint Equation (17) can be further manipulated:

$$-\frac{\nabla\lambda}{\lambda}\cdot\hat{t}_0 - \frac{e_0}{\lambda} = \nabla\cdot\hat{t}_0 \text{ or } \frac{d\lambda}{d\ell} + P(\ell)\lambda = Q(\ell)$$

$$\text{with } P(\ell) = \frac{1}{S}\frac{dS}{d\ell} \text{ and } Q(\ell) = -e_0 \tag{18}$$

The formal solution to (17) is well-known (e.g. [24]):

$$\lambda(\ell) = \frac{\left(C + v(\ell)\right)}{\mu(\ell)} \text{ with } \mu(\ell) = \exp\left\{\int_0^{\ell} P(\ell')d\ell'\right\} \text{ and } v(\ell) = \int_0^{\ell}\mu(\ell')Q(\ell')\ell' \tag{19}$$

Here the constant C is chosen to enforce the boundary condition $\lambda(\ell_B) = 0$.

5. Analysis of the Role of the Geometrical Spreading

The main purpose of this section is show that the solution to the adjoint equation can be constructed from the geometrical spreading function, and to interpret this result.

In any region in which $e_0 = 0$, the adjoint Equation (18) has the same form as the transport Equation (12). Since the error $e(\boldsymbol{x})$ is rarely known within the medium, but rather only on its boundary $\boldsymbol{x}_B$, this restriction is satisfied by all commonly-encountered cases. As we will show below, the similarity of form provides considerable insight into the behavior of the adjoint field λ.

Ray divergence enters into the adjoint equation through the $\nabla\cdot\hat{t}_0$ term. In

order to highlight its contribution, we first examine a solution in which this term is zero. Consider a plane wave propagating in the z-direction through a homogenous layer with $0 \leq z \leq z_B$ (Figure 2). The background travel time is $T_0 = s_0 z$, the ray direction is $\hat{t} = s_0^{-1} \nabla T_0 = \hat{z}$ and $\ell = z$. The plane wave satisfies the background Eikonal equation (14a),since $\nabla T_0 \cdot \nabla T_0 = s_0^2 (\hat{z} \cdot \hat{z}) = s_0^2$. Since the rays of a plane wave do not diverge, $\nabla \cdot \hat{t}_0 = 0$.

Now consider the case where the background slowness is everywhere too small by an amount b, so that the background error $e_0 = T^{obs} - T_0$ grows linearly with distance z, that is, $e_0(x, y, z) = bz$. We will assume that this error is known only on the boundary $z = z_B$. Following (5), the adjoint equation is $\mathrm{d}\lambda/\mathrm{d}z = -b z_B \delta(z - z_B)$. Because of the Dirac impulse function, the boundary condition for λ requires some scrutiny. We will consider that the error is defined just below the boundary, at $\equiv z_B - \epsilon^2$, where $\epsilon^2 \ll z_B$. In order to satisfy both the boundary condition of $\lambda(z_B) = 0$ and the adjoint equation, the solution must be discontinuous at z_B^-; and in the immediate vicinity of z_B^- must be $\lambda(z) = b z_B^- H(z_B^- - z)$. Effectively, the boundary condition is $\lambda(z_S) = e_0(x, y, z_S)$. The solution of the adjoint equation is $\lambda(z) = b z_B$; note that it does not depend upon z.

Now consider a slowness perturbation in the form of a very thin rectangular prism, centered at z_H, of thickness D, and having sides at x_1 and $x_2 = x_1 + L$, and y_1 and $y_2 = y_1 + L$ (so that its volume is DL^2). Since the prism is very thin, it can be approximated as a Dirac impulse function in depth z:

$$\varepsilon s_1(x, y, z) = m_1 W(x, x_1, x_2) W(y, y_1, y_2) D \delta(z - z_H)$$
$$\text{with } W(x, x_1, x_2) \equiv H(x - x_1) H(x_2 - x) \tag{20}$$

Here, $H(.)$ is the Heaviside function, which is unity when its argument is positive and zero otherwise. The partial derivative of total error is:

$$H_1 = -2\left(\frac{\mathrm{d}(\varepsilon s_1)}{\mathrm{d}m_1}, \lambda \right)$$
$$= -2D \iiint W(x, x_1, x_2) W(x, y_1, y_2) \delta(z - z_H) b z_B \mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{21}$$
$$= -2b z_B D L^2$$

An expected, $H_1 < 0$, since increasing m_1 lowers the error. Also as expected, H_1 is proportional to the area L^2 of the prism, since the larger its area, the larger the region to which the slowness perturbation is applied. Interestingly, H_1 is independent of the position z_H of the prism; that is, the prism can be moved up or down without affecting the error. As we will show below, this insensitivity to position is due to the absence of ray divergence in this plane wave case.

We now consider a spherical wave propagating in the r-direction in through a homogenous sphere with $0 \leq r \leq r_B$ (Figure 3), described by spherical polar coordinates (r, θ, φ). The background travel time is $T_0 = s_0 r$, the ray direction is $\hat{t} = s_0^{-1} \nabla T_0 = \hat{r}$ and $\ell = r$. The spherical wave satisfies the background Eikonal Equation (14a), since $\nabla T_0 \cdot \nabla T_0 = s_0^2 (\hat{r} \cdot \hat{r}) = s_0^2$. The area of a ray tube is
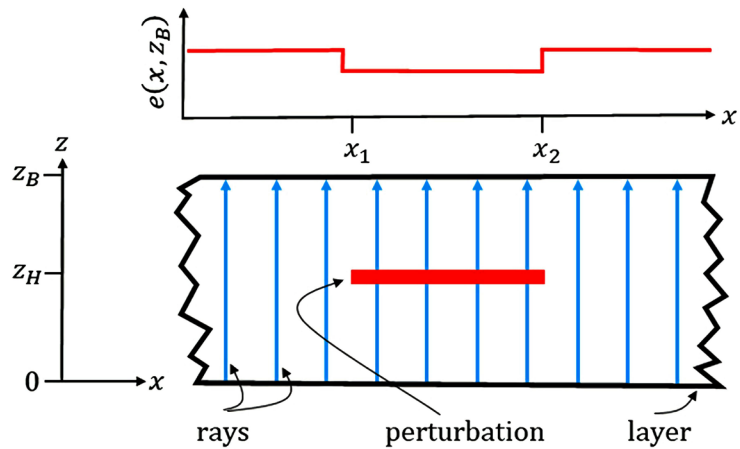
Figure 2. Rays (blue) of a plane wave cross a layer with bottom and top surfaces at $z = 0$ and $z = z_B$, respectively. A prismatic slowness perturbation (red rectangle) is placed within the layer, with left and right edges at $x = x_1$ and $x = x_2$, respectively. The travel time error $e(x, z_B)$, measured on the upper surface (top plot), is reduced in the region $x_1 < x < x_2$ where the rays project the prism. Because the rays do not diverge, the size of this region is independent of the depth of the perturbation.
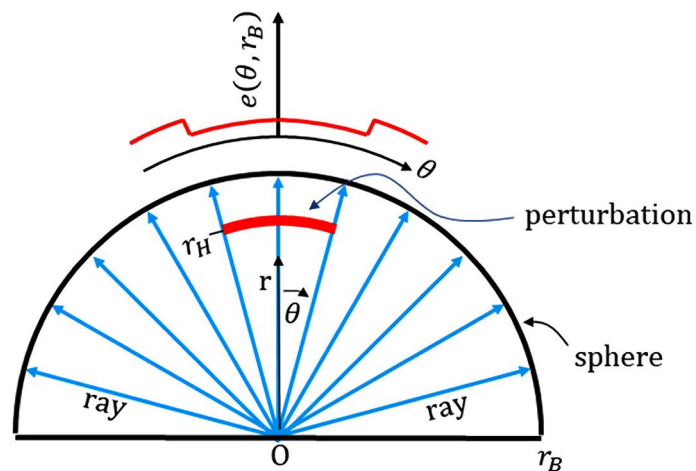

Figure 3. Rays (blue) of a spherical wave start at a source at the center of the sphere at $r = 0$ and propagate outward through the sphere to its surface at $r = r_B$. A slowness perturbation with the shape of a spherical cap (red cap) is placed within the sphere at radius r_H, with left and right edges at polar angle $-\theta_H$ and $+\theta_H$, respectively. The travel time error $e(\theta, r_B)$, measured on the upper surface, is reduced in the region where the perturbation is projected by the rays (graph at top).

$S = r^2 \mathrm{d}\Omega$, from whence we conclude that the geometrical spreading function is $R^2(r) = r^2$ and the ray divergence is $\nabla \cdot \hat{t}_0 = S^{-1} \mathrm{d}S/\mathrm{d}\ell = 2/r$. As in the plane wave case, the background slowness is everywhere too small by an amount b, leading to a background error $e_0(r, \theta, \varphi) = br$. We will assume that this error is known only on the boundary $r = r_B$. The adjoint Equation (18) reduces to:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}r} + \frac{2}{r}\lambda = 0 \quad \text{with boundary condition } \lambda(r_B, \theta, \varphi) = br_H \tag{22}$$

The solution is $\lambda(r,\theta,\varphi) = (br_B)(r_B^2/r^2)$. As is asserted in the Introduction, the solution to this transport-like equation is related to the geometrical spreading function by $\lambda(r,\theta,\varphi) \propto R^{-2}$.

Now consider a slowness perturbation in the form of a very thin spherical cap of fixed thickness D, centered at r_H and $\varphi_H = 0$ and subtending a variable polar angle area θ_H such that its area is fixed as $L^2 = 2\pi r_H^2(1 - \cos\theta_H)$:

$$\varepsilon s_1(x,y,z) = m_1 H(\theta_H - \theta) D\delta(r - r_H) \tag{23}$$

For a position r_H away from the origin where a spherical cap of thickness D and area L^2 is possible, the partial derivative of total error is:

$$\begin{aligned}
H_1 &= -2\left(\frac{d(\varepsilon s_1)}{dm_1}, \lambda\right) \\
&= -2br_B^3 D \iiint \left[H(\theta_H - \theta)\delta(r - r_H)\right]\left[r^{-2}\right] r^2 \sin\theta \, dr \, d\theta \, d\varphi \\
&= -2br_B^3 D \iint H(\theta_H - \theta)\sin\theta \, d\theta \, d\varphi \int \delta(r - r_H) \, dr \\
&= \left(-2bDr_B^3\right) 2\pi(1 - \cos\theta_H)\frac{r_H^2}{r_H^2} = \frac{\left(-2br_B^3 D\right)L^2}{r_H^2} \\
&= -2br_B DL^2 \left(\frac{r_B}{r_H}\right)^2 = -2br_B DL^2 \frac{R^2(r_B)}{R^2(r_H)}
\end{aligned} \tag{24}$$

The spherical wave solution (24) differs from the plane wave solution (21) by a factor that involves the ratio of geometric spreading functions, $R(r_B)/R(r_H)$, evaluated at the heterogeneity and the surface. The area, on the surface of the sphere, subtended by the prism decreases with its radius r_H, decreasing the error $e(r_S, \theta, \varphi)$ over wider region (Figure 4). This example illustrates the importance of geometric spreading on the amplitude of the adjoint field and on the effectiveness of a given perturbation to reduce the error E. Given several perturbations of equal size, the most effective is one whose projection on the boundary, by rays interacting with it, is the largest.

Although the adjoint field is singular at the source (ray starting point) $r = 0$, the partial derivative H_k is finite there, as can be seen by considering a spherical heterogeneity of radius r_H centered on the origin of the form $\varepsilon s_1(x,y,z) = m_1 H(r_H - r)$:

$$\begin{aligned}
H_1 &= -2\left(\frac{d(\varepsilon s_1)}{dm_1}, \lambda\right) = -2br_B^3 \iiint \left[H(r_H - r)\right]\left[r^{-2}\right] dr \, r \, d\theta \, r \sin\theta \, d\varphi \\
&= -2br_B^3 \iint \sin\theta \, d\theta \, d\varphi \int_0^{r_H} dr = \left(-2br_B^3\right)(4\pi)r_H = -8\pi br_B^3 r_H
\end{aligned} \tag{25}$$

when the background slowness $s_0(x)$ is spatially varying, the rays have a complicated spatial pattern and the background error $e_0(x_B)$, measured on the boundary x_B, is spatially varying. Suppose that the medium has a surface x_B with outward pointing normal $\hat{n}_B(x_B)$. A ray connecting an interior point x to x_B can be labeled by x_B. Then, $x_B(x)$ means the point on a boundary at which a ray passing through x ends, and arc-length $\ell(x, x_B)$ means the distance at x along a ray that ends at $x_B(x)$. Similarly, the geometrical spreading
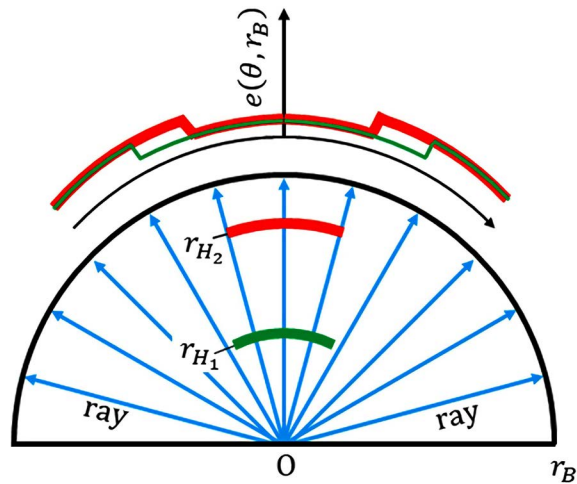
Figure 4. Rays (blue) of a spherical wave, as in **Figure 3**. One of two alternate slowness perturbations (green and red caps) are placed within the sphere, at radii r_{H_1} and r_{H_2}, respectively, with $r_{H_1} < r_{H_2}$. These caps have equal area L^2 and equal thickness D. The travel time error $e(\theta, r_B)$, measured on the surface of the sphere, is reduced in the region where the perturbation is projected by the rays (green and red curves in top plot). The reduction in error in this region is the same in both cases, because the thicknesses of the perturbations are equal. However, because the rays diverge, the size affected region is larger for the perturbation at r_{H_1}.

function can be written as $R(x, x_B)$; that is, the geometrical spreading function at x associated with the ray that ends at x_B. Then, the adjoint field is then:

$$\lambda(x) = \frac{e_0(x_B)}{\hat{t}(x_B) \cdot \hat{n}_B(x_B)} \frac{R^2(x_B, x_B)}{R^2(x, x_B)} \tag{26}$$

Here, the dot product between the ray tangent and surface normal is introduced to account for the increased surface area intersected by the ray tube, in the case (unlike the examples, above) where the ray tube obliquely impinges upon the boundary. Now, suppose that slowness perturbation is represented with voxels, where voxel k has volume V_k, amplitude m_k, and centroid position $x^{(k)}$. When the adjoint field varies slowly compared to the length scale of a voxel (a requirement that excludes the source point) the error derivative is:

$$H_k = -2\left(\frac{d(\varepsilon s_1)}{dm_k}, \lambda\right) \approx -2V_k \frac{e_0(x_B)}{\hat{t}(x_B) \cdot \hat{n}_B(x_B)} \frac{R^2(x_B, x_B)}{R^2(x^{(k)}, x_B)} \tag{27}$$

Here x_B is the end point of the ray passing through $x^{(k)}$. This result emphasizes the link between the geometrical spreading function R and the partial derivative of total error E. (When the voxel is close to, or overlaps the origin, H_k is still well-defined and finite, but the inner product in (27) must be computed appropriately).

6. Conclusion

The key result in this paper is the demonstration that the adjoint equation in

ray-based travel time tomography has the same form as the well-known transport equation for ray theoretical amplitudes. Consequently, the spatial variation of the adjoint field λ is completely controlled by the geometrical spreading function R. This result provides an intuitive understanding of the primary factor controlling the size of the partial derivative $H_j = \partial E / \partial m$ of total L_2 error E with respect to the slowness m_j of a voxel. The partial derivative H_j is large when ray divergence causes the projection of the voxel on the measurement surface to be large. Since this result provides an explicit formula for λ in terms of R, it enables H_j to be calculated without resorting to the numerical solution of the adjoint equation. Only an inner product needs to be calculated, and in the case of a voxel parameterization of the slowness image, it can be calculated trivially.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1]   Aki, K., Christoffersson, A. and Husebye, E.S. (1976) Three-Dimensional Seismic Structure of the Lithosphere under Montana Lasa. *Bulletin of the Seismological Society of America*, **66**, 501-524.

[2]   Menke, W. (1977) Lateral in Homogeneities in P Velocity under the Tarbella Array of the Lesser Himalayas of Pakistan. *Bulletin of the Seismological Society of America*, **67**, 725-734.

[3]   Munk, W., Worcester, P. and Wunsch, C. (1995) Ocean Acoustic Tomography. Cambridge University Press, Cambridge, 433 p.

[4]   Justice, J.H., Vassiliou, A.A., Singh, S., Logel, J.D., Hansen, P.A., Hall, B.R., Hurt, P.R. and Solanki, J.J. (1989) Acoustic Tomography for Monitoring Enhanced Oil Recovery. *Leading Edge*, **8**, 12-19. https://doi.org/10.1190/1.1439605

[5]   Santamarina, J.C. (1994) An Introduction to Geotomography. In: Woods, R.D., Ed., *Geophysical Characterization of Sites*, Oxford & IBH Publishing Co., New Delhi, 35-44.

[6]   Müller, T. (2014) GeoViS—Relativistic Ray Tracing in Four-Dimensional Spacetimes. *Computer Physics Communications*, **185**, 2301-2308. https://doi.org/10.1016/j.cpc.2014.04.013

[7]   Nolet, G. (1987) Seismic Wave Propagation and Seismic Tomography. In: Nolet, G., Ed., *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics*, Springer, New York, 1-23.

[8]   Dahlen, F., Hung, S.-H. and Nolet, G. (2002) Fréchet Kernels for Finite-Frequency Travel Times—I. Theory. *Geophysical Journal International*, **141**, 157-174. https://doi.org/10.1046/j.1365-246X.2000.00070.x

[9]   Hall, M.C.G., Cacuci, D.G. and Schlesinger, M.E. (1982) Sensitivity Analysis of a

Radiative-Convective Model by the Adjoint Method. *Journal of the Atmospheric Sciences*, **29**, 2038-2050.
https://doi.org/10.1175/1520-0469(1982)039<2038:SAOARC>2.0.CO;2

[10] Bin Waheed, U., Flagg, G. and Yarman, C.E. (2016) First-Arrival Traveltime Tomography for Anisotropic Media Using the Adjoint-State Method. *Geophysics*, **81**, R147-R155. https://doi.org/10.1190/geo2015-0463.1

[11] Levenberg, K. (1944) A Method for the Solution of Certain Non-Linear Problems in Least-Squares. *Quarterly of Applied Mathematics*, **2**, 164-168.
https://doi.org/10.1090/qam/10666

[12] Engl, H. (1993) Regularization Methods for the Stable Solution of Inverse Problems. *Surveys in Mathematics for Industry*, **3**, 71-143.

[13] Menke, W. and Eilon, Z. (2015) Relationship between Data Smoothing and the Regularization of Inverse Problems. *Pure and Applied Geophysics*, **172**, 2711-2726.
https://doi.org/10.1007/s00024-015-1059-0

[14] Menke, W. (2018) Geophysical Data Analysis: Discrete Inverse Theory. Fourth Edition, Elsevier, Amsterdam, 350 p.

[15] Lawson, C.L. and Hanson, R.J. (1995) Solving Least Squares Problems. Prentice-Hall, Englewoods Cliffs, New Jersey, 337 p.

[16] Snyman, J.A. and Wilke, D.N. (2018) Practical Mathematical Optimization—Basic Optimization Theory and Gradient-Based Algorithms. Springer Optimization and Its Applications, Second Edition, Springer, New York.

[17] Tromp, J., Tape, C. and Liu, Q. (2005) Seismic Tomography, Adjoint Methods, Time Reversal and Banana-Doughnut Kernels. *Geophysical Journal International*, **160**, 195-216. https://doi.org/10.1111/j.1365-246X.2004.02453.x

[18] Cerveny, V. (2001) Seismic Ray Theory. Cambridge University Press, Cambridge, 722 p.

[19] Aki, K. and Richards, P.G. (2009) Quantitative Seismology. Second Edition, University Science Books, Mill Valley, 742 p.

[20] Colley, S.J. (2012) Vector Calculus. Fourth Edition, Pearson, New York, 603 p.

[21] Phillips, W.S. and Fehler, M.C. (1991) Traveltime Tomography: A Comparison of Popular Methods. *Geophysics*, **56**, 1522-1692. https://doi.org/10.1190/1.1442974

[22] Menke, W. (2005) Case Studies of Seismic Tomography and Earthquake Location in a Regional Context. In: Levander, A. and Nolet, G., Eds., *Seismic Earth*: *Array Analysis of Broadband Seismograms*, Geophysical Monograph Series 157, American Geophysical Union, Washington DC, 7-36. https://doi.org/10.1029/157GM02

[23] Lanczos, C., 1961. Linear Differential Operators. Van Nostrand-Reinhold (London, U.K.), 580 pp., ISBN: 978-1-614-27302-8.

[24] Adkins, W.A. and Davidson, M.G. (2012) Ordinary Differential Equations. Springer, New York, 799 p. https://doi.org/10.1007/978-1-4614-3618-8_1