Inside Out: Inverse Problems
MSRI Publications
Volume **47**, 2003

Microlocal Analysis of Seismic Inverse Scattering

MAARTEN V. DE HOOP

Abstract. We review applications of microlocal analysis (MA) to reflection seismology. In this inverse method one attempts to estimate the index of refraction of waves in the earth from seismic data measured at the Earth's surface. Seismic imaging creates images of the Earth's upper crust using seismic waves generated by artificial sources and recorded into extensive arrays of sensors (geophones or hydrophones). The technology is based on a complex, and rapidly evolving, mathematical theory that employs advanced solutions to a wave equation as tools to solve approximately the general seismic inverse problem, with complications introduced by the heterogeneity and anisotropy of the Earth's crust. We describe several important developments using MA to generate these wave-solutions by manipulating the wavefields directly on their phase space. We also consider some recent applications of MA to global seismology.

1. Introduction

Microlocal analysis plays an increasingly important role in seismology, particularly in the imaging and inversion of seismic data. Here we consider imaging and inversion via the generalized Radon transform (GRT), concentrating on advances since the work of Beylkin [9], applying the work of Guillemin [51] and Taylor [98]. It is the aim of this exposition to connect microlocal analysis with seismology in the context of inverse scattering. The analysis of a related problem, the X-ray transform (see Greenleaf and Uhlmann [48; 49]) also contributes to the further understanding of the GRT in seismology. Microlocal analysis and the general theory of Fourier integral operators are described in the books by Hörmander [60; 61; 62], Duistermaat [43], and Treves [101; 102].

The author thanks The Mathematical Sciences Research Institute for partial support, through NSF grant DMS-9810361. He also thanks the members of the Mathematical Sciences Research Institute, and in particular Gunther Uhlmann, for providing a very stimulating environment during the Inverse Problems program in Fall 2001. The author thanks John Stockwell for his many suggestions to improve this survey.

Exploration seismology versus global seismology. Exploration seismology concerns the investigation of sedimentary structures in the upper crust of the earth, whereas global seismology concerns the investigation of the entire earth. In exploration reflection seismology, the following processes are amenable to the application of microlocal analysis: dip and azimuth moveout and seismic wavefield continuation, (map) migration, imaging, amplitude versus scattering angles analysis, inversion and resolution analysis, and migration velocity analysis. (Some of these processes also find application in acoustic emission and sonic borehole imaging.) In global seismology, the primary phases amenable to the application of microlocal analysis are earthquake generated body waves, both those that interact with the main transitions in the deep earth (such as the core mantle boundary) and those used in transmission wave-equation tomography. Also, spectral asymptotics applies to the study of free oscillations of the earth; this subject is beyond the scope of this survey. Here we discuss primarily exploration seismology, and we provide a brief outlook on global seismology.

Inverse scattering in seismology, in principle, yields an estimate of a distribution representing the elastic stiffness tensor in the earth. This tensor appears as coefficients in the elastic wave equation. From a geoscientist's perspective, however, stiffness is a manifestion of geodynamical processes such as mantle convection, magneto-hydrodynamics of the outer core, deformation and subduction of the continental crust, and sedimentary processes, with their own underlying mathematical models. Thus the inverse scattering problem becomes one of seismic waves coupled to one of these dynamical processes.

Caustics. The importance of microlocal techniques becomes apparent if caustics are formed in propagating wavefields in the earth. Caustics arise due to the heterogeneity and anisotropy of the elastic properties of the subsurface. It may be a matter of scale, though, whether or not some of the anisotropy originates from heterogeneity. Caustics form progressively in heterogeneous media, but may be intrinsic to the presence of anisotropy. Caustics due to heterogeneity are ubiquitous. For example, in models with small, smooth, random fluctuations in wave speed, which vary on a length scale large compared to a wave length but small compared with the propagation distance, caustic formation will occur with probability one; see White *et al.* [110].

Historical perspective. Some of the notions developed in microlocal analysis appear to have been independently discovered in seismology. Most notably, Hagedoorn [53] invented a purely graphical method for seismic imaging that is recognizable as a Fourier integral operator and its canonical relation. Rieber [85] and later Riabinkin [84] determined and exploited the "slopes" in addition to arrival times of the seismic events, which relates directly to the wavefront set of the data, to unravel complexities in the wavefield. Stolt [93] carefully used the notion of *migration dip* in imaging, which aids in the reconstruction of the wavefront set of the subsurface's stiffness. In *map migration* [108; 52; 109] (for

the state of the art, see [66]) an injectivity condition is assumed that appears in Guillemin's [51] Bolker condition in the treatment of the generalized Radon transform.

1A. Exploration seismology. In a seismic experiment one generates elastic waves in the earth using active sources at the earth's surface. The waves that return to the surface of the earth are observed; see Figure 1 (in fact, sources and receivers are not always on the surface of the earth; this case is also considered). The problem is to reconstruct the elastic properties of the subsurface from the data thus obtained.

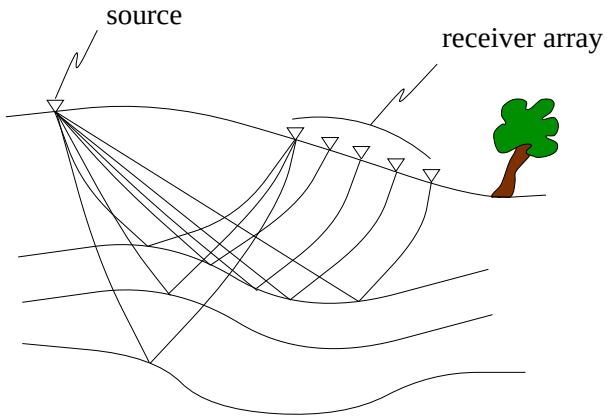


Figure 1. Schematic depth section showing reflection ray paths; a seismic experiment.

Hagedoorn's early approach to imaging of seismic reflection data can be summarized and illustrated as follows. If we restrict the seismic experiment to the acquisition of coinciding sources and receivers (*zero offset*) then the inverse scattering problem is formally determined. Figure 2(b) shows an earth model with a single reflecting surface; Figure 2(a) is a display of reflection data (seismic "traces") that would be produced at 81 locations at zero depth above the reflector. The source was a regularized delta function generating a pulse. For a pulse to travel down and up 1 km in upper medium with a speed of 2 km/s takes one second. The specular reflections occur at those points where the characteristic (ray) from the source/receiver location is normal to the reflector. The reflection can be mapped into the reflector as follows. On Figure 2(a) choose a source/receiver location (at the surface) and draw an isochrone curve (a circle for constant velocity) through any event on the corresponding seismic trace. This is illustrated in Figure 2 (bottom) for all seismic traces in Figure 2(a). The envelope of the isochrones delineates the reflector surface.

The Hagedoorn-derived methods being based totally on geometrical optics did not explicitly consider amplitudes. Amplitudes were considered by the inverse scattering based methods that followed. The approach presented in this

survey originates with the work of Beylkin [9; 10; 11] and other authors (see the references below), applying microlocal analysis to the seismic inverse problem. Beylkin [10] considered the seismic inverse scattering problem in acoustic media with constant density. He modeled the data using the Born approximation, wherein the scattering is linearized in the medium coefficients. The medium perturbation $\delta c(x)$ acts as a distribution of scatterers superimposed on a smooth background medium $c(x)$. Given the background medium $c(x)$ an operator was given to reconstruct $\delta c(x)$ microlocally from an n-dimensional subset of the data (from data that constitute a function of n variables). Beylkin-derived methods excluded caustics, assumed scalar wave propagation, and isotropy.

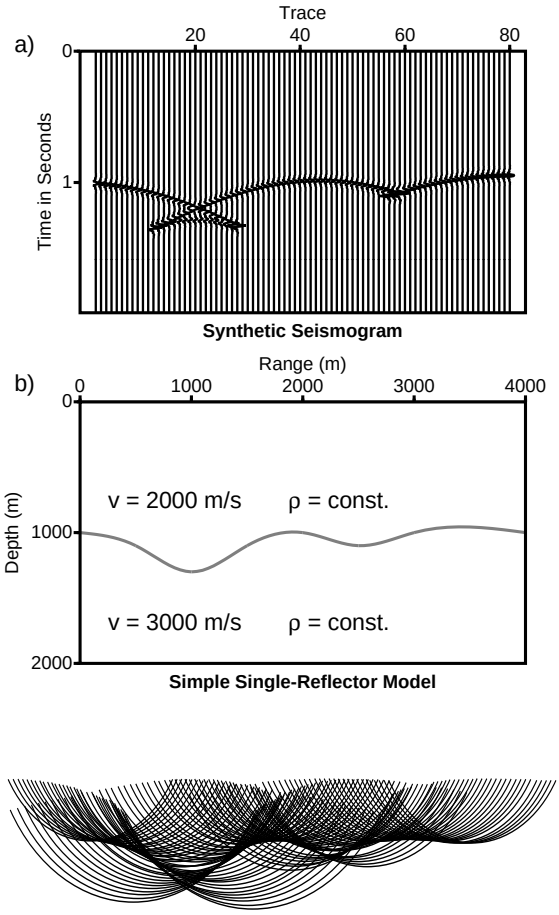


Figure 2. Top: zero-offset (coinciding source/receiver) reflection. Middle: reflector, where $v =$ acoustic wave speed, $\rho =$ density. Bottom: compass construction. After [18].

Present perspective on modeling and inversion. Data that are redundant in the sense that they are a function of more than n variables (*multiple offset* data) can be seen as a family of n-dimensional datasets, where each n-dimensional subset in the family has a fixed value of some coordinate, which we refer to as e. The result of the inversion, some manifestation of a reflectivity $r(x, e)$, should not depend on e. This is the criterion that must be used to estimate the background medium from the data; see for instance Symes [95]. Under the assumptions made by Beylkin [10], there exists microlocally an invertible map, transforming seismic data to a reflectivity function $r(x, e)$, of which the singular part should not depend on e. We consider such a transformation in a general framework that allows the presence of caustics and in anisotropic elastic rather than isotropic acoustic media. The treatment of elastic waves is based upon the decoupling of the hyperbolic system into n scalar equations (see Taylor [98], Ivrii [67], Dencker [39]) after which Fourier integral operator techniques are invoked. Each scalar equation governs the propagation of a particular mode, such as qP and $\mathrm{S}_{1,\ldots,n-1}$.

It is common to distinguish two ways of modeling reflection data. In the first way, we assume the Born approximation. This approximation is essentially a linearization, wherein the medium parameters are written as the sum of a background medium and a perturbation that is assumed to be small and localized. It is assumed that the background is smooth and that the perturbation contains the singularities of the medium. In the second, it is assumed that the medium consists of different regions separated by smooth interfaces. The medium parameters are assumed to be smooth on each region, and smoothly extendible across each interface, but they vary discontinuously at an interface. Such interfaces are the *seismic reflectors*. We discuss how to model the high-frequency part of the data using Fourier integral operators, following the approach of Taylor [98], obtaining a generalization of the Kirchhoff approximation.

Subject to certain geometrical assumptions, including the Bolker condition [51], the multi-modal data can be written as an invertible Fourier integral operator, H_{MN} say, acting on a *reflectivity* distribution, $r_{MN}(x, e)$, that is a function of subsurface position x and the additional variable e, essentially parametrizing the scattering angle and scattering azimuths between an incoming and outgoing characteristic. The position of the singularities of $r_{MN}(x, e)$ does not depend on e. In the Kirchhoff approximation for elastic media, the function $r_{MN}(x, e)$ equals to highest order $R_{MN}(x, e)\, \|\partial z_n/\partial x\|\, \delta(z_n(x))$, where $R_{MN}(x, e)$ is the appropriately normalized reflection coefficient for the pair of elastic modes (M, N), and $\|\partial z_n/\partial x\|\, \delta(z_n(x))$ is the singular function of the interface given as a level set. For the Born approximation $r_{MN}(x, e)$ is given by pseudodifferential operators that take into account the radiation patterns acting on the medium perturbation. The coordinate e is *a priori* defined only on the coisotropic submanifold $\mathcal{L} \subset T^*Y \setminus 0$ — where Y represents the acquisition manifold to which the scattered wavefield (δG) is restricted via an operator $\mathcal{R}$ — that contains the wavefront set of the data. To construct an invertible Fourier integral operator from data

to the function $r_{MN}(x,e)$, the coordinate e has to be defined on an open part of $T^*Y\setminus 0$; in Stolk and De Hoop [91] an extension is constructed of the coordinate function e from $\mathcal{L}$ to an open neighborhood of $\mathcal{L}$ in $T^*Y\setminus 0$ (which is not unique).

For the Born approximation, an inverse is also obtained in the least-squares sense via the parametrix of a normal operator, N_{MN}. The normal operator and its regularized inverse, $\langle N_{MN}^{-1}\rangle$ render a coupled spatial-parameter resolution analysis of the seismic experiment. The normal operator also provides a means to analyze nonmicrolocal contributions occurring if the Bolker condition is violated and replaced by a weaker condition leading to a characterization of artifacts.

When the data are redundant there is in addition a criterion to determine if the medium above the interface (the background medium in the Born approximation) is correctly chosen. The position of the singularities of the function $r_{MN}(x,e)$, obtained by acting with H_{MN}^{-1} on the data, should not depend on e. There exist pseudodifferential operators, W_{MN}, that, if the medium above the interface is correctly chosen, annihilate the data. This allows one to carry out an extension of differential semblance optimization in elastic media with caustics.

As mentioned, the wavefront set of the data is contained, under certain conditions, in a coisotropic submanifold of the acquisition cotangent bundle. It reveals a structure of characteristic strips. Restricting in the imaging operator the seismic data to a common coordinate value on these strips, yields a generalized Radon transform [10; 34; 38] that maps the reflection data into a seismic image. (Under certain conditions this generalized Radon transform is a Fourier integral operator.) Collecting these seismic images from the points on the characteristic strips corresponding to available data results in the set of so-called common-image-point gathers. In the presence of caustics, a filter needs to be designed and applied prior to extracting a trace from each of the common-image-point gathers in the set, to form a model image of the singular component of the medium.

From this image, we model seismic data that correspond to a different coordinate value on the characteristic strips. The result of this procedure is a composition of Fourier integral operators yielding seismic wavefield continuation, be it in the single scattering approximation. Relevant examples of seismic wavefield continuation are the *transformation to zero offset* [1] and the *transformation to common (prescribed) azimuth* [13]. The distribution kernel of transformation to zero offset is called dip moveout; the distribution kernel of transformation to common azimuth is called azimuth moveout.

Table 1 summarizes the operators that will be introduced in this survey.

Synopsis of publications. Many publications exist about high-frequency methods to invert seismic data in acoustic media. These methods date back to Hagedoorn [53]; from a seismic perspective, it has taken twenty years and more to develop the basic analysis of them [87; 31; 94; 75; 86]. From a mathematical perspective, the analysis started with the reconstruction of the singular compo-

Section		
[2,3]	modeling (FIO)	$[\delta G]$
[3]	acquisition (FIO)	$[F = \mathcal{R}\delta G]$
	imaging (FIO)	$[F^*]$
[4,5]	normal (Ψdo + nonmicrolocal operator)	$[N = F^*F]$
	resolution (Ψdo)	$[\langle N^{-1}\rangle N]$
	inversion (Least Squares, FIO)	$[\langle N^{-1}\rangle F^*]$
[6,7]	extended modeling (invertible FIO)	$[H, H^{-1}]$
[8]	annihilator (Ψdo)	$[W]$
[9]	generalized Radon transform (FIO)	$[L]$
[10]	"continuation" (FIO)	$[\mathcal{R}^c F\langle N^{-1}\rangle L]$

Table 1. Operators we will discuss. FIO stands for Fourier integral operator and Ψdo for pseudodifferential operator.

nent of the medium coefficients in the Born approximation, without caustics, by Beylkin [10]. Bleistein [17] discussed the case of a smooth jump using Beylkin's results. Rakesh [82] showed that the modeling operator in the Born approximation is a Fourier integral operator in the presence of caustics. Hansen [56] analyzed the inversion in an acoustic medium with multipathing for both the Born approximation and the case of a smooth jump. Ten Kroode *et al.* [99] extended the work of Hansen. Guillemin [51] discussed the Bolker condition in the context of generalized Radon transforms, that ensures invertibility of the modeling operator in the least-squares sense (see, e.g., De Hoop and Brandsberg–Dahl [33] and Stolk and De Hoop [91]). Stolk [89] simplified the analysis considering a case when the Bolker condition is violated. Nolan and Symes [80] discussed the imaging and inversion of seismic data with different (restricted) acquisition geometries.

The mathematical treatment of systems of equations, such as the elastic equations, in the high-frequency approximation has been given by Taylor [98]. This fundamental paper also discusses the interface problem. Beylkin and Burridge [12] discussed the imaging of seismic data in the Born approximation in isotropic elastic media, under a no-caustics assumption. De Hoop and Bleistein [32] discussed the imaging and inversion in general anisotropic elastic media, using a Kirchhoff-type approximation. In the presence of caustics, the foundations of this approximation were given by Stolk and De Hoop [91]. The generalized Radon transform in elastic media was developed in De Hoop *et al.* [34; 38]. The Born approximation for seismic data with maximal acquisition geometry in anisotropic elastic media allowing for multipathing was discussed and analyzed by De Hoop and Brandsberg–Dahl [33] and Stolk and De Hoop [91].

We mention two alternative (finite-frequency) but related approaches to inverse scattering of seismic data: The optimization approach (e.g., Tarantola [96; 97], De Hoop and De Hoop [35]), which falls into the category of *reverse time migration*, and the wavefield decomposition/double-square-root equation approach

(Claerbout [30], De Hoop *et al.* [36] and Stolk and De Hoop [92]), which falls into the category of *downward continuation migration*.

1B. Sedimentary environment. The medium parameters, stiffness c_{ijkl} and density ρ appear as the coefficients in the hyperbolic system of partial differential equations. Their properties, such as symmetry, are constrained by the types of rocks occurring in the subsurface and their microstructure as well as the ambient state of stress. Sedimentary rocks of interest include shales, sandstones, and carbonates (limestone, chalk, marlstone and dolomite).

Microscopically shales have anisotropic properties owing to the orientation of mineral grains; see Figure 3. Macroscopically, shales exhibit anisotropy due to the orientation of laminations owing to bedding or crossbedding. Their characteristic properties have been measured ultrasonically in the laboratory [68]. Their typical symmetry is hexagonal, with the restriction that triplication of the shear wave does not occur on the symmetry axis. Starting from their microstructure, shales were modeled mathematically by Hornby [63; 64] using contact theory [106; 107] on the one hand, and a combination of self consistent [58; 24] and differential effective medium approximations on the other. In view of their strongly anisotropic permeability, shales can be effective as seals over hydrocarbon reservoirs.

Reservoir rocks must be porous and permeable. These often consist of sandstones or carbonates. The porosity and permeability may result from intergrain voids or from fractures or from a combination of the two; see Figure 4. A sandstone is porous, and presumably filled with a fluid-gas mixture. Though sandstone by itself often may be assumed to be isotropic, fracturing breaks this symmetry typically to orthotropic. A commonly applied mathematical model for crack-induced anisotropy can be found in Hudson [65]. Even though the scattering theory presented here is valid in the "high-frequency" regime, the length

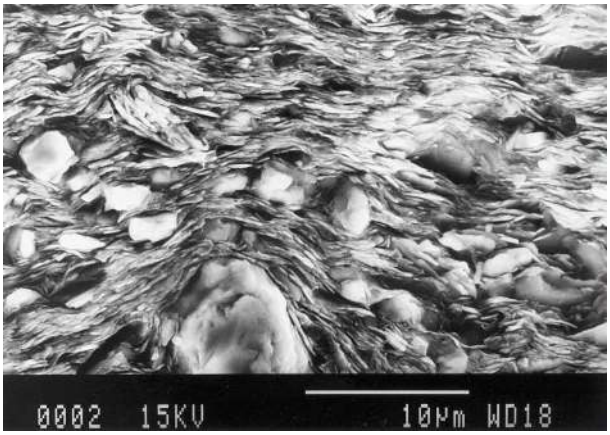


Figure 3. SEM picture of a typical shale (from Hornby *et al.* [64]).

scale of the microstructure of rocks is vastly smaller than the dominant seismic wavelengths. Effectively, the fluid-filled poro-elastic medium behaves like an (an)elastic solid. Some of the most established linear mathematical models for poro-elasticity, in particular those concerning a porous rock saturated with a fluid, were developed by Gassmann [47], Biot [14; 15; 16], Brown and Korringa [23], and Berryman and Milton [8]. Biot's equations were derived from microstructure using a homogenization approach by Burridge and Keller [26]. A theory that replaces the fluid in the pores by a gas-fluid mixture was developed by Batzle and Wang [7].

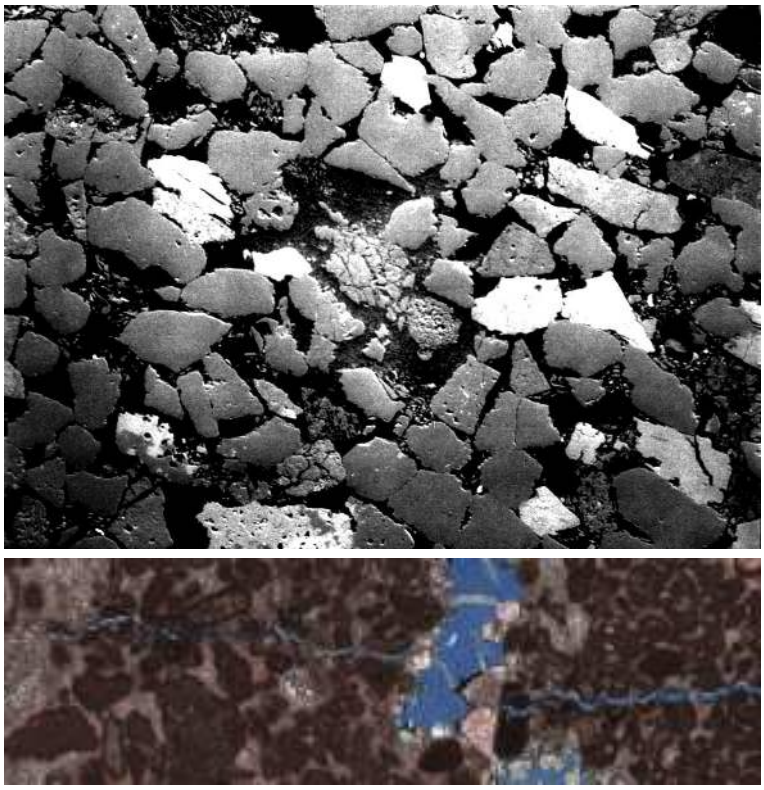


Figure 4. SEM picture of a sandstone (top) with a crack (light against darker background, running across the bottom image).

Seismic waves scatter at singularities in the elastic properties of the subsurface. These singularities are typically attributable to geological transitions (interfaces), unconformities, faults, as well as the interior structure of a formation such as one consisting of sand channels.

1C. Notation. Propagation of seismic waves occurs in the Earth's interior. The aim of seismic inverse scattering is to obtain information about selected target regions within the interior. The target is contained within an open set

$X \subset \mathbb{R}^n$. In practice $n = 2$ or 3, but we leave it unspecified. In exploration geophysics the subsurface refers to the shallow interior of the Earth. Subsurface position is denoted by x. Sources and receivers will be contained in open subsets O_s, O_r respectively, of the boundary ∂X of X. Their position is denoted by $\tilde{x}, \hat{x}$, respectively. Measurement of data takes place during a time interval $(0, T)$. The set of $(\hat{x}, \tilde{x}, t)$ for which data are taken is called the acquisition manifold Y; we assume that coordinates y' on Y are given. We assume that the particle displacement of the waves is measured for point sources at $\tilde{x}, t = 0$ with all its components, both at the source and at the receiver. Thus we assume that (after preprocessing) the data match the Green's function $G_{il}(\hat{x}, \tilde{x}, t)$, for $(\hat{x}, \tilde{x}, t) \in Y$.

We refer to the codimension of the set of $Y \subset \partial X \times \partial X \times (0, T)$ as the codimension of the acquisition manifold, and we denote it by c. Owing to the practicalities of data acquisition limits exist on the dimension of the set Y, which is expressed by c. For example, in marine data acquisition the receivers may lie along a line behind the source, in which case we have $n = 3, c = 1$, $\partial X = \{x \in \mathbb{R}^n : x_3 = 0\}$, $Y = \{(\hat{x}, \tilde{x}, t) \in \mathbb{R}^3 \times \mathbb{R}^3 \times (0, T) \,|\, \hat{x}_3 = \tilde{x}_3 = \hat{x}_2 - \tilde{x}_2 = 0\}$. We call such acquisition geometries common azimuth. Thus the data are a function of $2n - 1 - c$ variables. However, from such data we aim to determine a function of n variables; hence the data have redundancy with dimension $n - 1 - c$. The inverse problem is thus formally overdetermined.

The material presented in this survey has been published in the following papers: De Hoop *et al.* [34], De Hoop and Bleistein [32], Burridge *et al.* [25], De Hoop *et al.* [38], De Hoop and Brandsberg–Dahl [33], Stolk [89], Stolk and De Hoop [91], and Stolk [90].

2. Propagation of Elastic Waves in Smoothly Varying Media

Seismic wave amplitudes are sufficiently small such that the linearized theory of infinitesimal deformation applies. When combined with the equation of motion, this yields the elastic wave equation

$$\left(\rho\,\delta_{il}\frac{\partial^2}{\partial t^2} - \frac{\partial}{\partial x_j}c_{ijkl}\frac{\partial}{\partial x_k}\right)(\text{displacement})_l = (\text{volume force density})_i. \quad (2\text{–}1)$$

Here $\rho(x)$ is the volume density of mass and $c_{ijkl}(x)$ is the elastic stiffness tensor, with $i, j, k, l = 1, \ldots, n$.

2A. Decoupling the modes. In general, the elastic wave equation supports different wave types (modes). Seismologists easily identify the individual modes on seismograms. It is advantageous in the formulation of inverse scattering to trace the individual modes. Decoupling of the modes is accomplished by the diagonalization of system (2–1). To diagonalize this system, it is convenient to remove the x-dependent coefficient ρ multiplying the time derivative. Thus we

introduce the equivalent system

$$P_{il}u_l = f_i, \tag{2–2}$$

where

$$u_l = \sqrt{\rho}(\text{displacement})_l, \qquad f_i = \frac{1}{\sqrt{\rho}}(\text{volume force density})_i, \tag{2–3}$$

and

$$P_{il} = \delta_{il}\frac{\partial^2}{\partial t^2} - \frac{\partial}{\partial x_j}\frac{c_{ijkl}}{\rho}\frac{\partial}{\partial x_k} + \text{l.o.t.} \tag{2–4}$$

is the partial differential operator. Here we use the assumption that ρ is smooth and bounded away from zero. Both systems (2–1) and (2–2) are real, time reversal invariant, and their solutions satisfy reciprocity.

We describe how the system (2–2) can be decoupled by transforming it with appropriate pseudodifferential operators; see Taylor [98], Ivrii [67] and Dencker [39]. The goal is to transform the operator P_{il} by conjugation with a matrix-valued pseudodifferential operator $Q(x,D)_{iM}$, $D = D_x = -\mathrm{i}\frac{\partial}{\partial x}$, to an operator that is of diagonal form, modulo a regularizing part,

$$Q(x,D)^{-1}_{Mi}\, P_{il}(x,D,D_t)\, Q(x,D)_{lN} = \operatorname{diag}(P_M(x,D,D_t)\,;\, M = 1,\dots,n)_{MN}, \tag{2–5}$$

where $D_t = -\mathrm{i}\frac{\partial}{\partial t}$. The indices M, N denote the mode of propagation, and refer to qP and $\mathrm{S}_{1,\dots,n-1}$ wave propagation. In fact, for the construction of Fourier integral operator solutions developed in the scalar wave case, it is sufficient to transform the partial differential operator to block-diagonal form, where each of the blocks $P_M(x,D,D_t)$has scalar principal part (proportional to the identity matrix). In this case we will use the indices M, N to denote the block, and we will omit indices for the components within each block. Let

$$u_M = Q(x,D)^{-1}_{Mi}u_i, \qquad f_M = Q(x,D)^{-1}_{Mi}f_i. \tag{2–6}$$

The system (2–2) is then equivalent to the uncoupled equations

$$P_M(x,D,D_t)u_M = f_M. \tag{2–7}$$

Since the time derivative in P_{il} is already in diagonal form, it remains only to diagonalize its spatial part,

$$A_{il}(x,D) = -\frac{\partial}{\partial x_j}\frac{c_{ijkl}}{\rho}\frac{\partial}{\partial x_k} + \text{l.o.t.}$$

The goal becomes finding Q_{iM} and A_M such that (2–5) is valid with P_{il}, P_M replaced by A_{il}, A_M. The operator P_M is now

$$P_M(x,D,D_t) = \frac{\partial^2}{\partial t^2} + A_M(x,D).$$

Because of the properties of stiffness related to (i) the conservation of angular momentum, (ii) the properties of the strain-energy function, and (iii) the

positivity of strain energy, subject to the adiabatic and isothermal conditions, the principal symbol $A^{\rm prin}_{il}(x,\xi)$ of $A_{il}(x,D)$ is a positive symmetric matrix. Hence, it can be diagonalized by an orthogonal matrix. On the level of principal symbols, composition of pseudodifferential operators reduces to multiplication. Therefore, we let $Q^{\rm prin}_{iM}(x,\xi)$ be this orthogonal matrix, and we let $A^{\rm prin}_{M}(x,\xi)$ be the eigenvalues of $A^{\rm prin}_{il}(x,\xi)$, so that

$$Q^{\rm prin}_{Mi}(x,\xi)^{-1}A^{\rm prin}_{il}(x,\xi)Q^{\rm prin}_{lN}(x,\xi)=\mathrm{diag}(A^{\rm prin}_{M}(x,\xi))_{MN}. \tag{2–8}$$

(The eigenvalue-eigenvector system is sometimes referred to as the system of Christoffel equations.) The principal symbol $Q^{\rm prin}_{iM}(x,\xi)$ is the matrix that has as its columns the orthonormalized polarization vectors associated with the modes of propagation.

If the multiplicities of the eigenvalues $A^{\rm prin}_{M}(x,\xi)$ are constant, then the principal symbol $Q^{\rm prin}_{iM}(x,\xi)$ depends smoothly on (x,ξ) and microlocally equation (2–8) carries over to an operator equation. Taylor [98] has shown that if this condition is satisfied, then decoupling can be accomplished to all orders, where each block corresponds to a different eigenvalue. In fact, he proved the following slightly more general result.

Lemma 2.1 (Taylor). *Suppose the pseudodifferential operator $Q_{iM}(x,D)$ of order 0 is such that*

$$Q(x,D)^{-1}_{Mi}A(x,D)_{il}Q(x,D)_{lN}=\begin{pmatrix}A_{(1)}(x,D) & 0\\ 0 & A_{(2)}(x,D)\end{pmatrix}_{MN}+a(x,D)_{MN},$$

where the symbols $A_{(1)}(x,\xi)$ and $A_{(2)}(x,\xi)$ are homogeneous of order two and $a(x,\xi)_{MN}$ is polyhomogeneous of order one. Suppose the spectra of $A_{(1)}(x,\xi)$ and $A_{(2)}(x,\xi)$ are disjoint on a conic neighborhood of some $(x_0,\xi_0)\in T^*X\setminus 0$. Then by modifying Q with lower order terms the system can be transformed such that

$$a(x,D)_{MN}=\begin{pmatrix}a_{(1)}(x,D) & 0\\ 0 & a_{(2)}(x,D)\end{pmatrix}_{MN}+\text{smoothing remainder},$$

microlocally around (x_0,ξ_0).

This implies that if the multiplicity of a particular eigenvalue $A^{\rm prin}_{M}(x,\xi)$ is constant, then the system can be transformed such that the part related to this eigenvalue decouples from the rest of the system, modulo a smoothing remainder. In this survey we will assume that at least some of the modes decouple (microlocally). This is stated as Assumption 1 below.

We now give an alternative characterization of the quantities $A^{\rm prin}_{M}(x,\xi)$ and $Q^{\rm prin}_{iM}(x,\xi)$. The values $\tau=\pm\sqrt{A^{\rm prin}_{M}(x,\xi)}$ are precisely the solutions to the equation

$$\det P^{\rm prin}_{il}(x,\xi,\tau)=0. \tag{2–9}$$

The multiplicity of $A_M^{\text{prin}}(x,\xi)$ is equal to the multiplicity of the corresponding root of (2–9). The columns of $Q_{iM}^{\text{prin}}(x,\xi)$ satisfy

$$Q_{iM}^{\text{prin}} \in \ker P_{il}^{\text{prin}}\big(x,\xi,\sqrt{A_M^{\text{prin}}(x,\xi)}\,\big).$$

Because $P_{il}^{\text{prin}}(x,\xi,\tau)$ is homogeneous in (ξ,τ), one may choose to use the *slowness vector* $-\tau^{-1}\xi$ instead of the cotangent or wave vector ξ in the calculations. The set of $-\tau^{-1}\xi$ such that (2–9) holds is called the slowness surface, which can be easily visualized. A (section of the) slowness surface for the case of a transversely isotropic medium in $n=3$ dimensions is given in Figure 5(a).

The slowness surface consists of n sheets each corresponding to a mode of propagation. The innermost sheet is convex and is associated with the qP wave. The other sheets need not be convex. The multiplicity of the eigenvalues changes at the points (directions) where the different sheets intersect. In seismology it is quite common to use a parametrization of the slowness surface that differs from the stiffness tensor that directly controls the geometry (shape) of the different sheets. Examples are the Lamé parameters for isotropic media and the Thomsen parameters [100] for transversely isotropic media. For a general insight into such parametrizations; see Tsvankin [103].

The second-order equations (2–7) clearly describe the decoupling of the original system into different elastic modes. These equations inherit the symmetries of the original system, such as time-reversal invariance and reciprocity. Time-reversal invariance follows because the operators $Q_{iM}(x,D), A_M(x,D)$ can be chosen in such a way that $Q_{iM}(x,\xi) = -\overline{Q_{iM}(x,-\xi)}$, $A_M(x,\xi) = \overline{A_M(x,\xi)}$. Then Q_{iM}, A_M are real-valued. Reciprocity for the causal Green's function $G_{ij}(x,x_0,t-t_0)$ means that $G_{ij}(x,x_0,t-t_0) = G_{ji}(x_0,x,t-t_0)$. Such a relationship also holds (modulo smoothing operators) for the Green's function $G_M(x,x_0,t-t_0)$ associated with (2–7). This follows because the transpose operator $Q(x,D)^t_{Mi}$ (obtained by interchanging x,x_0 and i,M in the distribution kernel $Q_{iM}(x,x_0)$ of $Q_{iM}(x,D)$) is also a pseudodifferential operator, with principal symbol $Q^{\text{prin}}(x,\xi)^t_{Mi}$. As noted before for the principal symbol, it follows from the fact that $A^t_{ij} = A_{ij}$ that we can choose Q orthogonal, which is to say, such that $Q(x,D)_{iM}Q(x,D)^t_{Mj} = \delta_{ij}$. From the fact that

$$G_M(x,x_0,t-t_0) = Q(x,D)^{-1}_{Mi}G_{ij}(x,x_0,t-t_0)Q(x_0,D_{x_0})_{jM}$$

it then follows that microlocally G_M is reciprocal, i.e.,

$$G_M(x,x_0,t-t_0) = G_M(x_0,x,t-t_0) \quad \text{modulo smoothing operators.}$$

Up to principal symbols, the equation above represents rotations at the receiver (i) and the source (j) side. In seismology this is referred to as the Alford rotation [3].

REMARK 2.2. We already observed that if an eigenvalue $A_M^{\text{prin}}(x,\xi)$ has constant multiplicity $m_M > 1$ say, then u_M is an m_M-dimensional vector and (2–7) is a

$m_M \times m_M$ system, with scalar principal symbol. For such a system a microlocal solution can be constructed in the same way as for scalar systems. In this case all kinematic quantities, such as bicharacteristics, phase functions, and canonical relations depend only on M. Other quantities such as u_M and $Q_{iM}(x,D)$ will have multiple components. The Green's function G_M and its amplitude $\mathcal{A}_M$, to be introduced just before (2–20), are then $m_M \times m_M$ matrices. To simplify notation we do not take this into account explicitly.

2B. The Green's function. To evaluate the Green's function we use the first-order system for u_M that is equivalent to (2–7),

$$\frac{\partial}{\partial t}\begin{pmatrix} u_M \\ \partial u_M/\partial t \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -A_M(x,D) & 0 \end{pmatrix}\begin{pmatrix} u_M \\ \partial u_M/\partial t \end{pmatrix} + \begin{pmatrix} 0 \\ f_M \end{pmatrix}. \tag{2–10}$$

This system can be decoupled also. Let $B_M(x,D) = \sqrt{A_M(x,D)}$, which is a pseudodifferential operator of order 1 that exists because $A_M(x,D)$ is positive definite. The principal symbol of $B_M(x,D)$ is given by $B_M^{\mathrm{prin}}(x,\xi) = \sqrt{A_M^{\mathrm{prin}}(x,\xi)}$. We find then that (2–10) is equivalent to the two first-order equations

$$\left(\frac{\partial}{\partial t} \pm \mathrm{i}B_M(x,D)\right) u_{M,\pm} = f_{M,\pm} \tag{2–11}$$

under the transformations

$$\begin{aligned} u_{M,\pm} &= \tfrac{1}{2}u_M \pm \tfrac{1}{2}\mathrm{i}B_M(x,D)^{-1}\frac{\partial u_M}{\partial t}, \\ f_{M,\pm} &= \pm\tfrac{1}{2}\mathrm{i}B_M(x,D)^{-1}f_M. \end{aligned} \tag{2–12}$$

We construct operators $G_{M,\pm}$ with Lagrangian distribution kernel $G_{M,\pm}(x,x_0,t)$ that solve the initial value problem for (2–11). Then using Duhamel's principle we find that

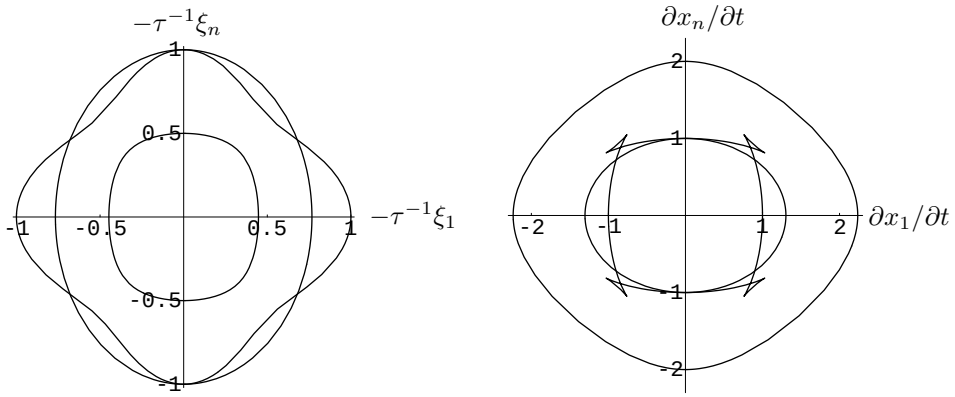
$$u_{M,\pm}(x,t) = \int_0^t \int_X G_{M,\pm}(x,x_0,t-t_0) f_{M,\pm}(x_0,t_0)\,\mathrm{d}x_0\,\mathrm{d}t_0$$

solves (2–11). It follows from (2–12) that the Green's function for the second-order decoupled equation is given by

$$\begin{aligned} &G_M(x,x_0,t) \\ &\quad = \tfrac{1}{2}\mathrm{i}G_{M,+}(x,x_0,t)B_M(x_0,D_{x_0})^{-1} - \tfrac{1}{2}\mathrm{i}G_{M,-}(x,x_0,t)B_M(x_0,D_{x_0})^{-1}. \end{aligned} \tag{2–13}$$

The operators $G_{M,\pm}$ are Fourier integral operators. Their construction is well known; see for example Duistermaat [43], Chapter 5. Singularities are propagated along the bicharacteristics, that are determined by Hamilton's equations generated by the principal symbol (factor i divided out) $\tau \pm B_M^{\mathrm{prin}}(x,\xi)$ of (2–11),

$$\begin{aligned} \frac{\partial x}{\partial \lambda} &= \pm\frac{\partial}{\partial \xi}B_M^{\mathrm{prin}}(x,\xi), & \frac{\partial t}{\partial \lambda} &= 1, \\ \frac{\partial \xi}{\partial \lambda} &= \mp\frac{\partial}{\partial x}B_M^{\mathrm{prin}}(x,\xi), & \frac{\partial \tau}{\partial \lambda} &= 0. \end{aligned} \tag{2–14}$$



(a) Cotangent: $B_M^{\text{prin}}(x,\tau^{-1}\xi)=1$ (b) Tangent: $v_M=\partial B_M^{\text{prin}}/\partial\xi\big|_{B_M^{\text{prin}}(x,\tau^{-1}\xi)=1}$

Figure 5. (a) Section of a slowness surface (the characteristic surface) for a transversely isotropic medium in $n=3$ dimensions. (b) Set of velocities associated to the slowness surface in a). Note the caustics that occur due to the fact that one of the (shear wave) sheets is not convex.

Solving these equations is what seismologists call *ray tracing* [27]. The solution may be parametrized by t. We denote the solution of (2–14) with the + sign and initial values x_0, ξ_0 by $(x_M(x_0,\xi_0,t),\xi_M(x_0,\xi_0,t))$. The solution with the $-$ sign is found upon reversing the time direction; in other words, it is given by $(x_M(x_0,\xi_0,-t),\xi_M(x_0,\xi_0,-t))$. For the later analysis we also use the direction $\alpha = \|\xi_0\|^{-1}\xi_0$ and τ combined to replace ξ_0 in the initial values of the bicharacteristic solution: $\xi_0 = \boldsymbol{\xi}_0(x_0,\alpha,\tau)$.

The first equality in (2–14) represents the velocity $\partial x/\partial t$ of the bicharacteristic identified as the group velocity. Because B_M^{prin} is homogeneous in ξ and Euler's relation, $\langle \xi, \partial_\xi B_M^{\text{prin}}\rangle = B_M^{\text{prin}} = \mp\tau$ it follows directly that the group velocity is orthogonal to the slowness surface. Solving (2–14) reveals the formation of caustics. Caustics may form progressively in the presence of heterogeneities, or instantaneously in the presence of anisotropy even in the absence of heterogeneity. An example of the latter is shown in Figure 5(b).

A complete view of the propagation of singularities is provided by the canonical relation of the operator $G_{M,\pm}$, given by

$$C_{M,\pm} = \{(x_M(x_0,\xi_0,\pm t),t,\xi_M(x_0,\xi_0,\pm t),\mp B_{M,\pm}(x_0,\xi_0);x_0,\xi_0)\}. \qquad (2\text{–}15)$$

A convenient choice of phase function is described in Maslov and Fedoriuk [73]. They state that one can always use a subset of the cotangent vector components as phase variables. Let us choose coordinates for $C_{M,+}$ of the form

$$(x_I, x_0, \xi_J, \tau), \qquad (2\text{–}16)$$

where $I \cup J$ is a partition of $\{1,\ldots,n\}$. It follows from Theorem 4.21 in Maslov and Fedoriuk [73] that there is a function $S_{M,+}(x_I,x_0,\xi_J,\tau)$, such that locally

$C_{M,+}$ is given by

$$\begin{aligned} x_J &= -\frac{\partial S_{M,+}}{\partial \xi_J}, \qquad & t &= -\frac{\partial S_{M,+}}{\partial \tau}, \\ \xi_I &= \frac{\partial S_{M,+}}{\partial x_I}, \qquad & \xi_0 &= -\frac{\partial S_{M,+}}{\partial x_0}. \end{aligned} \tag{2–17}$$

Here we take into account the fact that $C_{M,+}$ is a canonical relation, which introduces a minus sign for ξ_0. A nondegenerate phase function for $C_{M,+}$ is then found to be

$$\phi_{M,+}(x, x_0, t, \xi_J, \tau) = S_{M,+}(x_I, x_0, \xi_J, \tau) + \langle \xi_J, x_J \rangle + \tau t. \tag{2–18}$$

In case $J = \varnothing$, the generating function $S_{M,+}$ reduces to frequency, τ, times the negative of travel time.

On the other hand, the canonical relation $C_{M,-}$ is given by

$$C_{M,-} = \big\{(x, t, -\xi, -\tau; x_0, -\xi_0) : (x, t, \xi, \tau; x_0, \xi_0) \in C_{M,+}\big\}.$$

Thus a phase function for $C_{M,-}$ is

$$\phi_{M,-}(x, x_0, t, \xi_J, \tau) = -\phi_{M,+}(x, x_0, t, -\xi_J, -\tau).$$

We may define the canonical relation for G_M as $C_M = C_{M,+} \cup C_{M,-}$and a phase function $\phi_M = \phi_{M,-}$if $\tau > 0$, $\phi_M = \phi_{M,+}$ if $\tau < 0$.

We have to assume that the decoupling is valid microlocally around the bicharacteristic. In that case Theorem 5.1.2 of Duistermaat [43] implies that the operator $G_{M,\pm}$ is microlocally a Fourier integral operator of order $-\frac{1}{4}$. Hence, microlocally we have an expression for $G_{M,\pm}$ in the form of an oscillatory integral

$$G_{M,\pm}(x, x_0, t) = (2\pi)^{-(|J|+1)/2-(2n+1)/4} \int \mathcal{A}_{M,\pm}(x_I, x_0, \xi_J, \tau) \mathrm{e}^{\mathrm{i}\phi_{M,\pm}(x,x_0,t,\xi_J,\tau)} \, \mathrm{d}\xi_J \, \mathrm{d}\tau. \tag{2–19}$$

The factors of (2π) in front of the integral are according to the convention of Treves [102] and Hörmander [62]. In the special case of $J = \varnothing$ and considering the amplitude $\mathcal{A}_{M,\pm}$ up to leading order, the integral reduces to the leading order term of the Debye series expansion in geometrical optics.

The amplitude $\mathcal{A}_{M,\pm}(x_I, x_0, \xi_J, \tau)$ satisfies a transport equation along the bicharacteristics $(x_M(x_0, \xi_0, \pm t), \xi_M(x_0, \xi_0, \pm t))$. Properties of amplitudes are described for instance in Treves [102], Section 8.4. The amplitude is an element of $M_{C_M} \otimes \Omega^{1/2}(C_M)$, the tensor product of the Keller–Maslov bundle M_{C_M} and the half-densities on the canonical relation C_M. If the subprincipal part of $A_M(x, D)$ is a matrix, then the amplitude is also a matrix; see Remark 2.2. The Keller–Maslov bundle gives a factor i^k, where k is an index, which we will absorb in the amplitude. The index keeps track of the passage through caustics.

It is possible to choose a Maslov phase function with a different set of phase variables, for instance $\xi_{\tilde{J}}$ (and not τ), where $\tilde{I} \cup \tilde{J}$ is a partition of $\{1, \ldots, n\}$ and

$C_{M,\pm}$ is parametrized by $(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})$. In that case the transformed amplitude $\tilde{\mathcal{A}}_{M,\pm}(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})$ contains a Jacobian factor to the power one half, i.e.

$$\left|\tilde{\mathcal{A}}_{M,\pm}(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})\right| = |\mathcal{A}_{M,\pm}(x_I, x_0, \xi_J, \tau)| \left|\frac{\partial(x_I, x_0, \xi_J, \tau)}{\partial(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})}\right|^{1/2}, \qquad (2\text{–}20)$$

where in the Jacobian both sets of variables are coordinates on $C_{M,\pm}$.

We calculate the left-hand side of (2–20). For this purpose, consider the Green's function $G_{M,\pm}(x, x_0, t - t_0)$ with t and $t_0 = 0$ fixed. This function can be be viewed as the kernel of an invertible Fourier integral operator, mapping the displacement at $t = 0$, $u|_{t=0} \in \mathcal{E}'(X)$ to the displacement at t, $u|_t \in \mathcal{D}'(X)$, with phase $\tilde{\phi}_{M,\pm}(x, x_0, t, \xi_{\tilde{J}})$ and amplitude $\tilde{\mathcal{A}}_{M,\pm}(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})$. To highest order the energy at time t is given by

$$\int |B_M(x, D)u_{M,\pm}(x, t)|^2 \, \mathrm{d}x.$$

Conservation of this quantity is reflected by the relation

$$G_{M,\pm}(t)^* B_M(x, D)^* B_M(x, D) G_{M,\pm}(t) = B_{M,\pm}(x_0, D_{x_0})^* B_{M,\pm}(x_0, D_{x_0}),$$

where the left-hand side denotes a composition of Fourier integral operators and * denotes the adjoint. Since the left-hand side is a product of invertible Fourier integral operators, we can use the theory of Section 8.6 in Treves [102]. We find that to highest order

$$\left|(2\pi)^{-1/4}\tilde{\mathcal{A}}_{M,\pm}(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})\right|^2 = \left|\det \frac{\partial \xi_0}{\partial(x_{\tilde{I}}, \xi_{\tilde{J}})}\right| \left|\frac{B_M(x_0, \xi_0)}{B_M(x, \xi)}\right|^2.$$

The value of $B_M(x, \xi)$ equals the frequency τ and is conserved along the bicharacteristic. Recall that (x_0, ξ_0, t) are valid coordinates for $C_{M,\pm}$ (cf. (2–15)). The Jacobian $|\partial(x_0, \xi_0, t)/\partial(x_I, x_0, t, \xi_J)|$ is equal to the factor $|\det \partial\xi_0/\partial(x_I, \xi_J)|$, the reciprocal of which describes the geometrical spreading. It follows that to highest order

$$\left|\tilde{\mathcal{A}}_{M,\pm}(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})\right| = (2\pi)^{1/4} \left|\det \frac{\partial(x_0, \xi_0, t)}{\partial(x_{\tilde{I}}, x_0, t, \xi_{\tilde{J}})}\right|^{1/2}. \qquad (2\text{–}21)$$

From (2–20) it now follows that

$$|\mathcal{A}_{M,\pm}(x_I, x_0, \xi_J, \tau)| = (2\pi)^{1/4} \left|\det \frac{\partial(x_0, \xi_0, t)}{\partial(x_I, x_0, \xi_J, \tau)}\right|^{1/2}. \qquad (2\text{–}22)$$

We give our result about the Green's function for (2–7), collecting the results of this section, and using equations (2–12) and (2–22) to obtain a statement about the amplitude. We will assume that microlocally around the relevant bicharacteristics the decoupling is valid. Let $\mathrm{Char}(P_M)$ be the characteristic set of $P_M(x, D, D_t)$ given by $\{(x, t, \xi, \tau) : P_M(x, \xi, \tau) = 0\}$. The Green's function is such that precisely the singularities of f_M at $\mathrm{Char}(P_M)$ propagate (see Hörmander [61], Theorem 23.2.9). Thus we have

ASSUMPTION 1. *On a neighborhood of the bicharacteristic the multiplicity of the eigenvalue* $A_M^{\mathrm{prin}}(x,\xi)$ *in* (2–8) *is constant.*

LEMMA 2.3. *Suppose that for the bicharacteristics through* $\mathrm{WF}(f_M)\cap\mathrm{Char}(P_M)$ *Assumption* 1 *is satisfied. Then* u_M *is given microlocally, away from* $\mathrm{WF}(f_M)$*, by*

$$u_M(x,t)=\int G_M(x,x_0,t-t_0)f_M(x_0,t_0)\,\mathrm{d}x_0\,\mathrm{d}t_0, \tag{2–23}$$

where $G_M(x,x_0,t)$ *is the kernel of a Fourier integral operator with canonical relation* C_M *and order* $-1\frac{1}{4}$*, mapping functions of* x_0 *to functions of* (x,t)*. It can be written as*

$$G_M(x,x_0,t)=(2\pi)^{-(|J|+1)/2-(2n+1)/4}\int\mathcal{A}_M(x_I,x_0,\xi_J,\tau)\mathrm{e}^{\mathrm{i}\phi_M(x,x_0,t,\xi_J,\tau)}\,\mathrm{d}\xi_J\,\mathrm{d}\tau. \tag{2–24}$$

For the amplitude $\mathcal{A}_M(x_I,x_0,\xi_J,\tau)$ *we have, to highest order,*

$$|\mathcal{A}_M(x_I,x_0,\xi_J,\tau)|=(2\pi)^{1/4}\tfrac{1}{2}|\tau|^{-1}\left|\det\frac{\partial(x_0,\xi_0,t)}{\partial(x_I,x_0,\xi_J,\tau)}\right|^{1/2}. \tag{2–25}$$

The elastic system for generic elastic media has been investigated by Braam and Duistermaat [20]. The set of singular points is generically of codimension three (thus one lower than one would expect naively), and is of conical form in the neighborhood of the singular point. Braam and Duistermaat give a normal form for such systems and investigate the behavior of its associated bicharacteristics and polarization spaces. In this case the system cannot be decoupled. However, in a generic elastic medium there cannot be an open set of bicharacteristics that pass through a singular point, because the singular points form a set of codimension 3. In this sense the set of bicharacteristics that is to be excluded is small. An analysis of conical refraction has been carried out by Melrose and Uhlmann [74] and Uhlmann [104].

When the elastic tensor (for $n=3$) has symmetries it is determined by less than 21 coefficients. (The classification and analysis of the characteristic sets of such media can be found in the book by Musgrave [77].) In this case the singularities can be of different types. For example, in some classes of media, such as transversely isotropic media, the determinant decomposes into smooth factors. Then the multiplicities of the eigenvalues $A_M^{\mathrm{prin}}(x,\xi)$ can vary on a larger (codimension 2) subset of $T^*X\setminus 0$. Because the bicharacteristics are curves on a codimension 1 surface, Assumption 1 can be violated on an open set of bicharacteristics.

In seismology, representations of the type (2–24) have been used by Chapman and Drummond [29] and Kendall and Thomson [69].

REMARK 2.4. In isotropic media, when Assumption 1 is certainly satisfied, the Hopf–Rinow theorem guarantees that any two points in the domain probed by the waves can be connected by at least one characteristic in each mode of

propagation. In anisotropic media, in general, this is no longer true because there will not be an associated smooth Riemannian metric.

2C. Sources. In exploration seismology, a vibrator source is modeled as a point body force that is of the form

$$f_i(x_0, t_0) = d_i \delta(x_0 - s)\mathcal{W}(t_0),$$

acting at s with signature $\mathcal{W}$; the direction $d \in S^{n-1}$. Typically, $\mathcal{W}(t_0)$ is viewed as a regularization of $\delta(t_0)$.

An earthquake in global seismology is modeled with a symmetric moment tensor M_{ij} of rank 2; then

$$f_i(x_0, t_0) = -M_{ij}\partial_{x_{0,j}}\delta(x_0 - s)\, H(t_0 - t_s),$$

where s denotes the hypocentral location and t_s the origin time. We set $t_s = 0$. In $n = 3$ dimensions, depending on the eigenvalues $\lambda_{1,2,3}$ of M, pure shear faults ($\lambda_1 = -\lambda_3$, $\lambda_2 = 0$), pure tension cracks ($\lambda_1 \neq 0$, $\lambda_2 = \lambda_3 = 0$), explosive sources ($\lambda_1 = \lambda_2 = \lambda_3 \neq 0$), or compensated linear dipoles ($\lambda_1 \neq 0$, $\lambda_2 = \lambda_3 = -\frac{1}{2}\lambda_1$) can be simulated. Substituting the body force representation into (2–23) together with (2–6) leads to the product of operators $G_M Q(x_0, D_{x_0})^{-1}_{Mi} M_{ij}\tau^{-1}\partial/\partial x_{0,j}$. This product is a Fourier integral operator with the same phase as G_M, and amplitude that to highest order (by integrating by parts in x_0) equals the product $\mathcal{A}_M(x_I, x_0, \xi_J, \tau)Q(x_0, \boldsymbol{\xi}_0)^{-1}_{Mi} M_{ij}\,\mathrm{i}\tau^{-1}\boldsymbol{\xi}_{0,j}$, where $\boldsymbol{\xi}_0 = \xi_0(x_I, x_0, \xi_J, \tau)$.

3. High-Frequency Born Modeling and Imaging

Modeling under the Born approximation can be obtained as the leading term in a forward scattering series, while inversion in this approximation can be likewise obtained from the inverse scattering series; see Moses [76] and Razavy [83]. These series originate from a contrast formulation in the medium coefficients assuming a background. This formulation leads to the Lipmann–Schwinger–Dyson equation for the scattered field. We choose the background to be smooth and the contrast to be singular, viewing the contrast as a perturbation from the background. This is important in its own right, and it will also be a motivation for our approach to the model with smooth jumps introduced in the Kirchhoff approximation.

Microlocal analysis of the Born approximation has been discussed by a number of authors. In the absence of caustics, for the acoustic case, see Beylkin [10]; for the isotropic elastic case, see Beylkin and Burridge [12]; for the anisotropic elastic case, see De Hoop, Spencer and Burridge [38]. In the acoustic case, allowing for multipathing (caustics), see Rakesh [82] and Hansen [56]. For the acoustic problem with (non)maximal acquisition geometry, see Nolan and Symes [80]. For the elastic case with maximal acquisition geometry, see De Hoop and Brandsberg–Dahl [33] and Stolk and De Hoop [91].

3A. Scattering: perturbation of the Green's function. In the contrast formulation the total value of the medium parameters c_{ijkl}, ρ is written as the sum of a smooth background constituent $\rho(x), c_{ijkl}(x)$ plus a singular perturbation $\delta\rho, \delta c_{ijkl}$, namely

$$c_{ijkl} + \delta c_{ijkl}, \qquad \rho + \delta\rho.$$

This decomposition induces a perturbation of P_{il} (cf. (2–4)),

$$\delta P_{il} = \delta_{il}\frac{\delta\rho}{\rho}\frac{\partial^2}{\partial t^2} - \frac{\partial}{\partial x_j}\frac{\delta c_{ijkl}}{\rho}\frac{\partial}{\partial x_k}.$$

We denote the causal Green's operator associated with (2–2) by G_{il} and its distribution kernel by $G_{il}(x, x_0, t - t_0)$. The first-order perturbation $\delta G_{il}$of G_{il} is derived by demanding that the leading-order term in $(P_{ij} + \delta P_{ij})(G_{jk} + \delta G_{jk})$ vanishes. This results in the representation

$$\delta G_{il}(\hat{x}, \tilde{x}, t) = -\int_0^t \int_X G_{ij}(\hat{x}, x_0, t-t_0) \underbrace{\delta P_{jk}(x_0, D_{x_0}, D_{t_0})\, G_{kl}(x_0, \tilde{x}, t_0)}_{-\text{linearized contrast source}} \,\mathrm{d}x_0\,\mathrm{d}t_0, \tag{3–1}$$

which is the Born approximation. Here $\tilde{x}$ denotes a source location, $\hat{x}$ a receiver location, and x_0 a scattering point. Because the background model is smooth the operator δG_{il} contains only the single scattered field.

We apply the decoupling given by equation (2–6). Omitting the factors $Q_{iM}(\hat{x}, D_{\hat{x}}), Q(\tilde{x}, D_{\tilde{x}})^{-1}_{Nl}$ at the beginning and end of the product, we obtain an expression for the perturbation of the Green's function $\delta G_{MN}(\hat{x}, \tilde{x}, t)$for the pair of modes M (scattered) and N (incident):

$$\begin{aligned}\delta G_{MN}(\hat{x}, \tilde{x}, t) = &-\int_0^t \int_X G_M(\hat{x}, x_0, t - t_0) Q(x_0, D_{x_0})^{-1}_{Mi} \\ &\times \Big(\delta_{il}\frac{\partial}{\partial t_0}\frac{\delta\rho}{\rho}\frac{\partial}{\partial t_0} - \frac{\partial}{\partial x_{0,j}}\frac{\delta c_{ijkl}}{\rho}\frac{\partial}{\partial x_{0,k}}\Big) Q(x_0, D_{x_0})_{lN} G_N(x_0, \tilde{x}, t_0)\,\mathrm{d}x_0\,\mathrm{d}t_0.\end{aligned} \tag{3–2}$$

Microlocally we can write G_M as in (2–24), with appropriate substitutions for its arguments. For G_N we use in addition the reciprocity relation $G_N(x_0, \tilde{x}, t_0) = G_N(\tilde{x}, x_0, t_0)$. The product of operators

$$G_M Q(x_0, D_{x_0})^{-1}_{Mi}\frac{\partial}{\partial x_{0,j}}$$

is a Fourier integral operator with the same phase as G_M, and amplitude that to highest order equals the product $\mathcal{A}_M(\hat{x}_{\hat{I}}, x_0, \hat{\xi}_{\hat{J}}, \tau) Q(x_0, \hat{\xi}_0)^{-1}_{Mi}\mathrm{i}\hat{\xi}_{0,j}$, where $\hat{\xi}_0 = \xi_0(\hat{x}_{\hat{I}}, x_0, \hat{\xi}_{\hat{J}}, \tau)$.

Assuming that the medium perturbation vanishes around $\hat{x}$ and $\tilde{x}$ a cut-off is introduced for t_0 near 0 and t. In the resulting expression one of the two frequency variables $\hat{\tau}, \tilde{\tau}$ can now be eliminated using the integral over t_0 (see for instance Duistermaat [43], Section 2.3). In this case the result can be obtained readily by noting that the integral over t_0 can be extended to the

whole of $\mathbb{R}$ (the phase is not stationary for t_0 outside $[0,t]$), and then using that $\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}t_0(\hat\tau-\tilde\tau)}\,\mathrm{d}t_0 = 2\pi\delta(\hat\tau-\tilde\tau)$. The resulting formula for δG_{MN} is, modulo lower-order terms in the amplitude,

$$\delta G_{MN}(\hat x,\tilde x,t) = (2\pi)^{-(3n+1)/4-(|\hat J|+|\tilde J|+1)/2}\int \mathcal{B}_{MN}(\hat x_{\hat I},\hat\xi_{\hat\jmath},\tilde x_{\tilde I},\tilde\xi_{\tilde\jmath},x_0,\tau)$$
$$\times\left(w_{MN;ijkl}(\hat x_{\hat I},\tilde x_{\tilde I},x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau)\frac{\delta c_{ijkl}(x_0)}{\rho(x_0)} + w_{MN;0}(\hat x_{\hat I},\tilde x_{\tilde I},x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau)\frac{\delta\rho(x_0)}{\rho(x_0)}\right)$$
$$\times\,\mathrm{e}^{\mathrm{i}\Phi_{MN}(\hat x,\tilde x,t,x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau)}\,\mathrm{d}x_0\,\mathrm{d}\hat\xi_{\hat\jmath}\,\mathrm{d}\tilde\xi_{\tilde\jmath}\,\mathrm{d}\tau. \tag{3–3}$$

Here (see (2–18) for the construction of ϕ_M,ϕ_N),

$$\Phi_{MN}(\hat x,\tilde x,t,x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau) = \phi_M(\hat x,x_0,t,\hat\xi_{\hat\jmath},\tau)+\phi_N(\tilde x,x_0,t,\tilde\xi_{\tilde\jmath},\tau)-\tau t. \tag{3–4}$$

The amplitude factors $\mathcal{B}_{MN}$ are given by

$$\mathcal{B}_{MN}(\hat x_{\hat I},\tilde x_{\tilde I},x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau) = (2\pi)^{-(n-1)/4}\,\mathcal{A}_M(\hat x_{\hat I},x_0,\hat\xi_{\hat\jmath},\tau)\,\mathcal{A}_N(\tilde x_{\tilde I},x_0,\tilde\xi_{\tilde\jmath},\tau). \tag{3–5}$$

We will refer to the factors $w_{MN;ijkl}, w_{MN;0}$ as the radiation patterns. These are given by

$$w_{MN;ijkl}(\hat x_{\hat I},\tilde x_{\tilde I},x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau) = Q_{iM}(x_0,\hat\xi_0)Q_{lN}(x_0,\tilde\xi_0)\,\hat\xi_{0,j}\tilde\xi_{0,k},$$
$$w_{MN;0}(\hat x_{\hat I},\tilde x_{\tilde I},x_0,\hat\xi_{\hat\jmath},\tilde\xi_{\tilde\jmath},\tau) = -Q_{iM}(x_0,\hat\xi_0)Q_{iN}(x_0,\tilde\xi_0)\,\tau^2,$$

where $\hat\xi_0 = \xi_0(\hat x_{\hat I},x_0,\hat\xi_{\hat\jmath},\tau)$, $\tilde\xi_0 = \xi_0(\tilde x_{\tilde I},x_0,\tilde\xi_{\tilde\jmath},\tau)$. The scattering is depicted in Figure 6. We illustrate a couple of radiation patterns in Figure 7.

We investigate the map

$$\left(\frac{\delta c_{ijkl}}{\rho},\frac{\delta\rho}{\rho}\right)\mapsto \delta G_{MN}(\hat x,\tilde x,t)$$

induced by (3–3). We use the notation C_{ϕ_M} to indicate the subset of the global canonical relation C_M that is associated with a phase function ϕ_M; cf. (2–15).

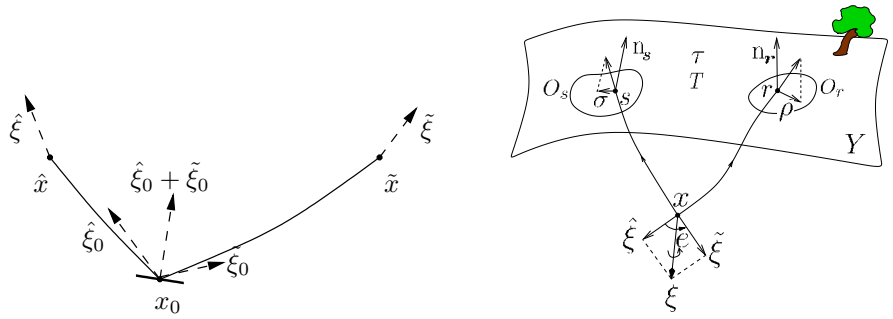


Figure 6. The scattering cotangent vectors ($\Lambda_{0,MN}$, left) and parametrization of Λ_{MN} (right).

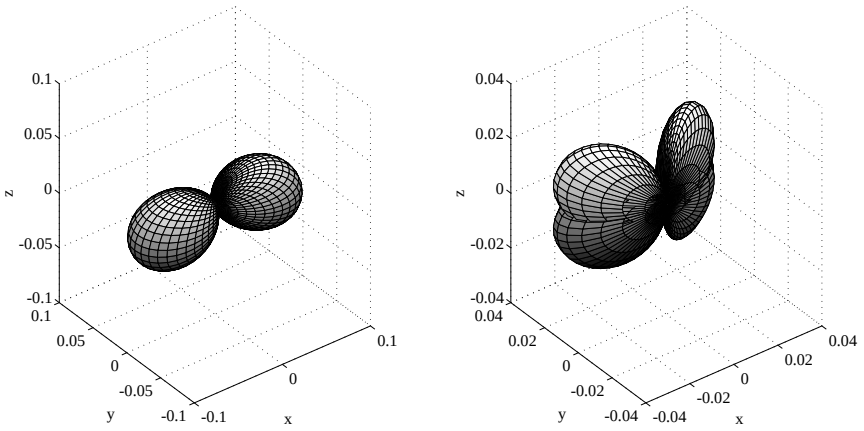


Figure 7. Radiation patterns ($n = 3$) for δc_{1313} (left) and δc_{1112} (right) for qP-qP scattering in an isotropic background. Magnitude is given as as a function of scattering angle and azimuth; cf. (4–4).

LEMMA 3.1. *Assume that if* $(\hat{x}, \hat{t}, \hat{\xi}, \tau; x_0, \hat{\xi}_0) \in C_{\phi_M}$, $(\tilde{x}, \tilde{t}, \tilde{\xi}, \tau; x_0, \tilde{\xi}_0) \in C_{\phi_N}$ *then* $\hat{\xi}_0 + \tilde{\xi}_0 \neq 0$. *Then the map*

$$\Big(\frac{\delta c_{ijkl}}{\rho}, \frac{\delta \rho}{\rho}\Big) \mapsto \delta G_{MN}(\hat{x}, \tilde{x}, t)$$

of (3–3) *is a Fourier integral operator* $\mathcal{E}'(X) \to \mathcal{D}'(X \times X \times (0,T))$. *Its canonical relation is*

$$\Lambda_{0,MN} = \big\{(\hat{x}, \tilde{x}, \hat{t} + \tilde{t}, \hat{\xi}, \tilde{\xi}, \tau; x_0, \hat{\xi}_0 + \tilde{\xi}_0) : (\hat{x}, \hat{t}, \hat{\xi}, \tau; x_0, \hat{\xi}_0) \in C_{\phi_M}, (\tilde{x}, \tilde{t}, \tilde{\xi}, \tau; x_0, \tilde{\xi}_0) \in C_{\phi_N}\big\}. \quad (3\text{–}6)$$

PROOF. We show that $\Phi_{MN}(\hat{x}_{\hat{I}}, \tilde{x}_{\tilde{I}}, t, x_0, \hat{\xi}_{\hat{J}}, \tilde{\xi}_{\tilde{J}}, \tau)$ is a nondegenerate phase function. The derivatives with respect to the phase variables are

$$\frac{\partial \Phi_{MN}}{\partial \tau} = -\hat{t}(\hat{x}_{\hat{I}}, x_0, \hat{\xi}_{\hat{J}}, \tau) - \tilde{t}(\tilde{x}_{\tilde{I}}, x_0, \tilde{\xi}_{\tilde{J}}, \tau) + t,$$

$$\frac{\partial \Phi_{MN}}{\partial \hat{\xi}_{\hat{J}}} = -\hat{x}_{\hat{J}}(\hat{x}_{\hat{I}}, x_0, \hat{\xi}_{\hat{J}}, \tau) + \hat{x}_{\hat{J}}, \qquad \frac{\partial \Phi_{MN}}{\partial \tilde{\xi}_{\tilde{J}}} = -\tilde{x}_{\tilde{J}}(\tilde{x}_{\tilde{I}}, x_0, \tilde{\xi}_{\tilde{J}}, \tau) + \tilde{x}_{\tilde{J}},$$

where $\hat{x}_{\hat{J}}(\hat{x}_{\hat{I}}, x_0, \hat{\xi}_{\hat{J}}, \tau), \tilde{x}_{\tilde{J}}(\tilde{x}_{\tilde{I}}, x_0, \tilde{\xi}_{\tilde{J}}, \tau)$ are as defined in (2–17), for the receiver side and the source side respectively. The derivatives of these expressions with respect to the variables $(\hat{x}_{\hat{J}}, \tilde{x}_{\tilde{J}}, t)$ are linearly independent, so Φ_{MN} is nondegenerate. From expression (3–4) it follows that the canonical relation of this operator is given by (3–6). By the hypothesis the canonical relation contains no elements with $\hat{\xi}_0 + \tilde{\xi}_0 = 0$, hence it is continuous as a map $\mathcal{E}'(X) \to \mathcal{D}'(X \times X \times (0,T))$. □

The condition in Lemma 3.1 is violated if and only if $M = N$ and there exists a "direct" bicharacteristic from $\tilde{x}, \tilde{\xi}$ to $\hat{x}, -\hat{\xi}$. From the symmetry of

the bicharacteristic under the transformation $\xi \to -\xi, t \to -t$ it follows that indeed in this case the condition is violated. On the other hand, we have $B_M(x_0, \hat{\xi}_0) = B_N(x_0, \tilde{\xi}_0) = \pm\tau$. If $\hat{\xi}_0 = -\tilde{\xi}_0$, then we must have $M = N$, because $B_M(x_0, \hat{\xi}_0) = B_M(x_0, -\hat{\xi}_0)$ and the condition that the eigenvalues in (2–8) are different for different modes. If $M = N$ and $\hat{\xi}_0 = -\tilde{\xi}_0$ then we have the mentioned direct bicharacteristic.

3B. Restriction: acquisition. Data are measurements of the scattered wave field which we relate here to the Green's function perturbation in (3–2). These data are assumed to be representable by $\delta G_{MN}(\hat{x}, \tilde{x}, t)$ for $(\hat{x}, \tilde{x}, t)$ in some acquisition manifold, which contains the source and receiver points and time. To make this explicit, let $y \mapsto (\hat{x}(y), \tilde{x}(y), t(y))$ be a coordinate transformation, such that $y = (y', y'')$ and the acquisition manifold Y is given by $y'' = 0$. Assume that the dimension of y'' is $2 + c$, where c is the codimension of the geometry (the 2 enforces "remote sensing"). Then the data are modeled by

$$\delta G_{MN}(\hat{x}(y', 0), \tilde{x}(y', 0), t(y', 0)). \tag{3–7}$$

It follows that the map $(\frac{\delta c_{ijkl}}{\rho}, \frac{\delta \rho}{\rho})$ to the data may be seen as the compose of the map of Lemma 3.1 with the restriction operator to $y'' = 0$. The restriction operator that maps a function $f(y)$ to $f(y', 0)$ is a Fourier integral operator with canonical relation $\Lambda_r = \{(y', \eta'; (y', y''), (\eta', \eta'')) \in T^*Y \times T^*\check{Y} : y'' = 0\}$, where $\check{Y} = X \times X \times (0, T)$. The composition of the canonical relations $\Lambda_{0,MN}$ and Λ_r is well defined if the intersection of $\Lambda_r \times \Lambda_{0,MN}$ with $T^*Y \setminus 0 \times \mathrm{diag}(T^*\check{Y} \setminus 0) \times T^*X \setminus 0$ is transversal [43]. In this case we must have that the intersection of $\Lambda_{0,MN}$ with the manifold $y'' = 0$ is transversal. For the later analysis, the source and receiver points s, r are defined through $(s, r, t) = (\hat{x}(y', 0), \tilde{x}(y', 0), t(y', 0))$; instead, we will shortcut the coordinate transformation and identify

$$y' = (s, r, t) \quad \text{and} \quad \eta' = (\sigma, \rho, \tau),$$

see Figure 6. The tangential slownesses then follow as

$$p^s = \tau^{-1}\sigma, \qquad p^r = \tau^{-1}\rho. \tag{3–8}$$

Let's repeat our assumptions:

Assumption 2. *There are no elements $(y', 0, \eta', \eta'') \in T^*Y \setminus 0$ such that there is a direct bicharacteristic from $(\hat{x}(y', 0), \hat{\xi}(y', 0, \eta', \eta''))$ to $(\tilde{x}(y', 0), -\tilde{\xi}(y', 0, \eta', \eta''))$ with arrival time $t(y', 0)$.*

Assumption 3. *The intersection of $\Lambda_{0,MN}$ with the manifold $y'' = 0$ is transversal, that is,*

$$\frac{\partial y''}{\partial(x_0, \hat{\xi}_0, \tilde{\xi}_0, \hat{t}, \tilde{t})} \text{ has maximal rank.} \tag{3–9}$$

In the following theorem we parametrize (3–6) by $(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)$ using the parametrization of C_{ϕ_M} given by (2–15). Thus we let $\tau=\mp B_M(x_0,\hat\xi_0)$ and

$$\begin{aligned}\hat x &= x_M(x_0,\hat\xi_0,\pm\hat t), \qquad & \tilde x &= x_N(x_0,\tilde\xi_0,\pm\tilde t),\\ \hat\xi &= \xi_M(x_0,\hat\xi_0,\pm\hat t), \qquad & \tilde\xi &= \xi_N(x_0,\tilde\xi_0,\pm\tilde t).\end{aligned}$$

We suppose that $\big(y'(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t),\ \eta'(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)\big)$ is obtained by transforming $(\hat x,\tilde x,\hat t+\tilde t,\hat\xi,\tilde\xi,\tau)$ to (y,η) coordinates.

THEOREM 3.2. [91] *If Assumptions* 2 *and* 3 *are satisfied, the operator* $F_{MN;ijkl}$ *(resp.* $F_{MN;0}$*) that maps the medium perturbation* $\delta c_{ijkl}/\rho$ *(resp.* $\delta\rho/\rho$*) to the data as a function of* y' (3–7) *is microlocally a Fourier integral operator with canonical relation*

$$\begin{aligned}\Lambda_{MN}=\big\{&(y'(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t),\eta'(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t);x_0,\hat\xi_0+\tilde\xi_0):\\ &B_M(x_0,\hat\xi_0)=B_N(x_0,\tilde\xi_0)=\pm\tau, y''(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)=0\big\}.\end{aligned}\tag{3–10}$$

The order is $(n-1+c)/4$*. The amplitude is given to highest order (in coordinates* (y'_I,η'_J,x_0) *for* Λ_{MN}*, where* I,J *is a partition of* $\{1,\dots,2n-1-c\}$*) by the products* $\mathcal{B}_{MN}(y'_I,\eta'_J,x_0)w_{MN;ijkl}(y'_I,\eta'_J,x_0)$ *and* $\mathcal{B}_{MN}(y'_I,\eta'_J,x_0)w_{MN;0}(y'_I,\eta'_J,x_0)$ *respectively, where*

$$\begin{aligned}\big|\mathcal{B}_{MN}(y'_I,\eta'_J,x_0)\big| &= \tfrac14\tau^{-2}(2\pi)^{-(n+1+c)/4}\\ &\times\left|\det\frac{\partial(\hat x,\tilde x,t)}{\partial y}\right|^{-1/2}\left|\det\frac{\partial(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)}{\partial(x_0,y'_I,y'',\eta'_J,\Delta\tau)}\right|^{1/2}_{\Delta\tau=0,y''=0}.\end{aligned}\tag{3–11}$$

Here we define $\Delta\tau=\hat\tau-\tilde\tau$*, so that the first constraint in* (3–10) *reads* $\Delta\tau=0$*. The map* $(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)\mapsto(x_0,y'_I,y'',\eta'_J,\Delta\tau)$ *is bijective.*

PROOF. The first statement has been argued above. The order of the operator is given by

$$\chi+\frac{K}{2}-\frac{\dim X+\dim Y'}{4},$$

where χ is the degree of homogeneity of the amplitude and K is the number of phase variables. The factors $\{w_{MN;ijkl},w_{MN;0}\}$ are homogeneous of order 2 in the ξ and τ variables; the degree of homogeneity of the factor $\mathcal{B}_{MN}$ follows from (2–22). We find

$$\begin{aligned}\text{order } F_{MN;ijkl} &= 2+\big(-2-\tfrac12(|\hat J|+|\tilde J|+2)+n\big)+\tfrac12\big(|\hat J|+|\tilde J|+1\big)-\tfrac14(3n-1-c)\\ &= \tfrac14(n-1+c).\end{aligned}$$

We calculate now the amplitude of the Fourier integral operator in Lemma 3.1. The factor $w_{MN;ijkl}$ is simply multiplicative. Suppose we choose coordinates on $\Lambda_{0,MN}$ to be $(\hat x_{\hat I},\hat\xi_{\hat J},\tilde x_{\tilde I},\tilde\xi_{\tilde J},\hat\tau,\tilde\tau,x_0)$, with ultimately $\hat\tau=\tilde\tau$ and define $\tau=$

$(\hat\tau+\tilde\tau)/2$, $\Delta\tau=\hat\tau-\tilde\tau$. Using (2–25) and (3–5) we find that the amplitude $\mathcal{B}_{MN}(x_0,\hat x_{\hat I},\hat\xi_{\hat\jmath},\tilde x_{\tilde I},\tilde\xi_{\tilde\jmath},\tau)$ is given by

$$\left|\mathcal{B}_{MN}(\hat x_{\hat I},\hat\xi_{\hat\jmath},\tilde x_{\tilde I},\tilde\xi_{\tilde\jmath},\tau,x_0)\right|=\tfrac14\tau^{-2}(2\pi)^{-(n-1)/4}\left|\det\frac{\partial(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)}{\partial(\hat x_{\hat I},\hat\xi_{\hat\jmath},\tilde x_{\tilde I},\tilde\xi_{\tilde\jmath},\tau,x_0,\Delta\tau)}\right|^{1/2}.$$

The transformation from $(\hat x,\tilde x,t)$ to y coordinates in Fourier integral (3–7) induces an additional factor $\left|\det\partial(\hat x,\tilde x,t)/\partial y\right|^{-1/2}$ (note that for the Fourier integral operators it would be more natural to transform as a half-density). The amplitude transforms as a half-density on the canonical relation, and we obtain the factor

$$\left|\det\frac{\partial(y_I',y'',\eta_J')}{\partial(\hat x_{\hat I},\hat\xi_{\hat\jmath},\tilde x_{\tilde I},\tilde\xi_{\tilde\jmath},\tau)}\right|^{1/2}.$$

The additional factor $(2\pi)^{-(2+c)/4}$ arises from the normalization. We find (3–11). □

Natural coordinates for the canonical relation are given by $(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)$ such that $B_M(x_0,\hat\xi_0)-B_N(x_0,\tilde\xi_0)=0, y''(x_0,\hat\xi_0,\tilde\xi_0,\hat t,\tilde t)=0$. There is a natural density directly associated with this set, the quotient density. The Jacobian in (3–11) reveals that the amplitude factor $|\mathcal{B}_{MN}(y_I',\eta_J',x_0)|$ is in fact given by the associated half-density times $\frac14\tau^{-2}(2\pi)^{-(n+1+c)/4}\left|\partial(\hat x,\tilde x,t)/\partial y\right|^{-1/2}$.

REMARK 3.3. If $c=0$ and there are no rays tangent to the acquisition manifold, that is, if

$$\operatorname{rank}\frac{\partial y''}{\partial(\hat t,\tilde t)}=2, \tag{3–12}$$

then a convenient way to parametrize the canonical relation is found using the phase directions $\hat\alpha=\hat\xi_0/\|\hat\xi_0\|$, $\tilde\alpha=\tilde\xi_0/\|\tilde\xi_0\|\in S^{n-1}$ and the frequency τ. See also Figure 6 (right).

3C. Imaging. We collect the medium perturbations into the column matrix

$$g_\alpha=\left(\frac{\delta c_{ijkl}}{\rho},\frac{\delta\rho}{\rho}\right).$$

The Born approximate forward operator $(F_{MN;ijkl},F_{MN;0})$ (cf. Theorem 3.2) is then represented by $F_{MN;\alpha}$.

The imaging operator is the adjoint $F^*_{MN;\alpha}$ of $F_{MN;\alpha}$. If $F_{MN;\alpha}$ is a Fourier integral operator then the imaging operator is a Fourier integral operator also. In a geometrical context, seismologists view imaging as "moving" the singular support of the data restricted to a given offset (offset is the difference between the source s and receiver r points) to the singular support of the medium perturbation. Seismologists refer to this process as *migration*.

4. Linearized (Born) Inversion

4A. Imaging-inversion: the least-squares approach. The standard procedure to deal with the fact that the seismic inverse problem is overdetermined is to use the method of least squares. Let us consider data from a single pair of modes (M, N). The normal operator $N_{MN;\alpha\beta}$ is defined as the compose of $F_{MN;\beta}$ and its adjoint $F^*_{MN;\alpha}$,

$$N_{MN;\alpha\beta} = F^*_{MN;\alpha} F_{MN;\beta} \tag{4–1}$$

(no summation over M, N). If $N_{MN;\alpha\beta}$ is invertible (as a matrix-valued operator with indices $\alpha\beta$), then

$$F^{-1}_{MN;\alpha} = (N_{MN})^{-1}_{\alpha\beta} F^*_{MN;\beta} \tag{4–2}$$

(no summation over M, N) is a left inverse of $F_{MN;\alpha}$ that is optimal in the sense of least squares[1].

The properties of the composition in (4–1) depend on those of Λ_{MN}. Let π_Y, π_Xbe the projection mappings of Λ_{MN} to $T^*Y \setminus 0$, $T^*X \setminus 0$ respectively. We show that under the following assumption $N_{MN;\alpha\beta}$ is a pseudodifferential operator, so that the problem of inverting $N_{MN;\alpha\beta}$ reduces to a finite-dimensional problem for each $(x, \xi) \in \pi_X(\Lambda_{MN})$.

ASSUMPTION 4. (*Guillemin* [51]) *The projection π_Y of Λ_{MN} on $T^*Y \setminus 0$ is an embedding.*

This assumption is also known as the Bolker condition. Because Λ_{MN} is a canonical relation that projects submersively (under Assumption 1) on the subsurface variables (x, ξ), the projection of (3–10) on $T^*Y \setminus 0$ is immersive [62, Lemma 25.3.6 and (25.3.4)]. Therefore only the injectivity in the assumption need be verified [99]. In fact, it is precisely the injectivity condition that has been tacitly assumed in what seismologists call *map migration*.

Note that the "proper" part of Assumption 4 is always satisfied: The definition of proper is that the pre-image of a compact set is a compact set. Let us assume we have a compact subset of $T^*Y \setminus 0$. The pre-image consists of elements of Λ_{MN} corresponding to those "points" where the source and receiver rays intersect. The set of these points can be written as a set on which some continuous function vanishes. Therefore this set is closed. It is also bounded, and hence it is compact.

Assumption 4 implies that the image of π_Y is a submanifold, $\mathcal{L}$ say, of $T^*Y \setminus 0$. Using that Λ_{MN} is a canonical relation we have

[1]Equation (4–1) is for the case where one minimizes the difference with the data δG_{MN} in L^2 norm $\|\delta G_{MN} - F_{MN;\alpha} g_\alpha\|$. It can easily be adapted to the case where one minimizes a Sobolev norm of different order, or a weighted L^2 norm. This would introduce extra factors in the amplitude.

LEMMA 4.1. [91] *The projection π_Y of Λ_{MN} on $T^*Y \setminus 0$ is an immersion if and only if the projection π_X of Λ_{MN} on $T^*X \setminus 0$ is a submersion. In this case the image of π_Y is locally a coisotropic submanifold of $T^*Y \setminus 0$.*

As a consequence of immersivity implied by the embedding in Assumption 4, we can use $(x, \xi) \in T^*X \setminus 0$ as the first $2n$ (local) coordinates on Λ_{MN}. In addition, we need to parametrize the subsets of the canonical relation given by $(x, \xi) =$ constant; we denote such parameters by e. The new parametrization of Λ_{MN} is (identifying x_0 in (3–10) with x)

$$\Lambda_{MN} = \{(y'(x, \xi, e), \eta'(x, \xi, e); (x, \xi))\}. \tag{4–3}$$

The results do not depend on the precise definition of e. As noted before, if the variables $(\hat{t}, \tilde{t})$ can be solved from the second constraint in (3–10) (cf. equation (3–12)), then Λ_{MN} can be parametrized using $(x, \hat{\alpha}, \tilde{\alpha}, \tau)$, where $(\hat{\alpha}, \tilde{\alpha})$ are phase directions. In that case (x, ξ, e) should be related by a coordinate transformation to $(x, \hat{\alpha}, \tilde{\alpha}, \tau)$. In isotropic media with $M = n$ (where $\|\hat{\xi}_0\| = \|\tilde{\xi}_0\|$) a suitable choice is the pair scattering angle/azimuth, given by [38]

$$\boldsymbol{e}(x, \hat{\alpha}, \tilde{\alpha}) = \left(\underbrace{\arccos(\hat{\alpha} \cdot \tilde{\alpha})}_{\theta}, \ \frac{-\hat{\alpha} + \tilde{\alpha}}{2 \sin(\arccos(\hat{\alpha} \cdot \tilde{\alpha})/2)} \right) \in (0, \pi) \times S^{n-2}. \tag{4–4}$$

The azimuth, the second component, defines together with ξ the plane spanned by $(\hat{\alpha}, \tilde{\alpha})$. It is not very difficult to show that in elastic media the scattering angle (the first component) can be used as a coordinate when the slowness sheets are convex, but not always when one of the slowness sheets fails to be convex. In the canonical relation Λ_{MN} the range of ξ values at x such that y' yields a data point (after the application of mentioned pseudodifferential cutoff) controls the spatial resolution of the reconstruction of $g_\alpha(x)$ (we refer to this as illumination or insonification). The construction of this range is illustrated in Figure 8 and dates back to Ewald [45]; see also Devaney [40].

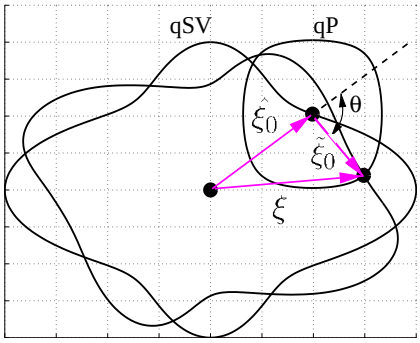


Figure 8. Generalization of the Ewald sphere. Here M = qSV, N = qP in a hexagonal medium; e (here θ) and τ are fixed.

REMARK 4.2. We show that the first part of Assumption 4 implies that

$$\frac{\partial B_M}{\partial \xi}(x,\hat{\xi}_0)+\frac{\partial B_N}{\partial \xi}(x,\tilde{\xi}_0)\neq 0;$$

in other words, the group velocities at the scattering point do not add up to 0. We have seen in Theorem 3.2 that Λ_{MN} may be parametrized by $(x,\hat{\xi}_0,\tilde{\xi}_0,\hat{t},\tilde{t})$, where $(\hat{\xi}_0,\tilde{\xi}_0)$ are such that

$$B_M(x_0,\hat{\xi}_0)=B_N(x_0,\tilde{\xi}_0)=\pm\tau$$

(and we have the additional constraint $y''(x_0,\hat{\xi}_0,\tilde{\xi}_0,\hat{t},\tilde{t})=0$). The projection π_X is given by $(x,\hat{\xi}_0+\tilde{\xi}_0)$. Consider tangent vectors to Λ_{MN} given by vectors $v_{\hat{\xi}_0}, v_{\tilde{\xi}_0}$. These must satisfy

$$v_{\hat{\xi}_0}\cdot\frac{\partial B_M}{\partial \xi}(x,\hat{\xi}_0)=v_{\tilde{\xi}_0}\cdot\frac{\partial B_N}{\partial \xi}(x,\tilde{\xi}_0)=\pm v_\tau. \tag{4–5}$$

Thus, if $\dfrac{\partial B_M}{\partial \xi}(x,\hat{\xi}_0)=-\dfrac{\partial B_N}{\partial \xi}(x,\tilde{\xi}_0)$, (4–5) implies that $(v_{\hat{\xi}_0}+v_{\tilde{\xi}_0})\cdot\dfrac{\partial B_M}{\partial \xi}(x,\hat{\xi}_0)$ vanishes, implying that the projection of Λ_{MN} on $T^*X\setminus 0$ is not submersive.

If $c=0$, and $\operatorname{rank}\partial y''/\partial(\hat{t},\tilde{t})=2$ (no tangent rays), then the constraint $y''=0$ may be used to solve for the parameters $\hat{t},\tilde{t}$ and (4–5) is the only condition on $(\hat{\xi}_0,\tilde{\xi}_0)$. In that case $(\partial B_M/\partial\xi)(x,\hat{\xi}_0)\neq-(\partial B_N/\partial\xi)(x,\tilde{\xi}_0)$ implies that the projection is submersive. The solutions to $y''(x_0,\hat{\xi}_0,\tilde{\xi}_0,\hat{t},\tilde{t})=0$ in (3–10) are then denoted as

$$\tilde{t}=T_N(x_0,\tilde{\alpha}),\quad \hat{t}=T_M(x_0,\hat{\alpha}),\quad \text{while}\ \ \tilde{t}+\hat{t}=T_{MN}(x_0,\hat{\alpha},\tilde{\alpha}).$$

For later notational convenience, we write the associated solutions for source and receiver points in y' as (s_N,r_M,T_{MN}) and their cotangent vectors in η' as (σ_N,ρ_M,τ) with

$$s_N(x_0,\tilde{\alpha}),\ \sigma_N(x_0,\tilde{\alpha},\tau)=\tau p_N^s,\ r_M(x_0,\hat{\alpha}),\ \rho_M(x_0,\hat{\alpha},\tau)=\tau p_M^r,$$

where p^s and p^r are defined in (3–8). In other cases, the set of $(\hat{\xi}_0,\tilde{\xi}_0)$ is in general a smaller subset of $T^*_xX\setminus 0\times T^*_xX\setminus 0$.

When constructing the compose (4–1) there is a subtlety that we have to take into account, namely that the linearized forward operator is only *microlocally* a Fourier integral operator. To make it globally a Fourier integral operator, we apply a pseudodifferential cutoff $\psi(y',D_{y'})$ (which seismologists would call a *tapered mute*) with compact support. Due to the fact that an embedding is proper, the forward operator is then a finite sum of local Fourier integral operators.

THEOREM 4.3. [91] *Let $\psi(y',D_{y'})$ be a pseudodifferential cutoff with conically compact support in $T^*Y\setminus 0$, such that for the set*

$$\big\{(y',\eta';x_0,\xi_0)\in\Lambda_{MN}:(y',\eta')\in\operatorname{supp}\psi\big\} \tag{4–6}$$

Assumptions 2, 3, *and* 4 *are satisfied. Then*

$$F^*_{MN;\beta}\psi(y',D_{y'})^*\psi(y',D_{y'})F_{MN;\alpha} \tag{4–7}$$

is a pseudodifferential operator of order $n-1$, *with principal symbol*

$$\begin{aligned} N_{MN;\beta\alpha}(x,\xi) = {} & \tfrac{1}{16}(2\pi)^{-n} \\ & \times \int |\psi(y'(x,\xi,e),\eta'(x,\xi,e))|^2\tau^{-4}\overline{w_{MN;\beta}(x,\xi,e)}w_{MN;\alpha}(x,\xi,e) \\ & \times \left|\det\frac{\partial(\hat{x},\tilde{x},t)}{\partial y}\right|^{-1}\left|\det\frac{\partial(x,\hat{\xi}_0,\tilde{\xi}_0,\hat{t},\tilde{t})}{\partial(x,\xi,e,y'',\Delta\tau)}\right|_{\substack{\Delta\tau=0\\ y''=0}} \mathrm{d}e, \end{aligned} \tag{4–8}$$

where $\tau=\tau(x,\xi,e)$.

Proof. We use the clean intersection calculus for Fourier integral operators (see Treves [102], for example) to show that (4–7) is a Fourier integral operator. The canonical relation of $F^*_{MN;\alpha}$ is given by

$$\Lambda^*_{MN} = \big\{(x,\xi;y',\eta') : (y',\eta';x,\xi)\in\Lambda_{MN}\big\}.$$

Let $L=\Lambda^*_{MN}\times\Lambda_{MN}$ and $M=T^*X\setminus 0\times\operatorname{diag}(T^*Y\setminus 0)\times T^*X\setminus 0$. We have to show that the intersection $L\cap M$ is clean, that is,

$$L\cap M \text{ is a manifold,} \tag{4–9}$$

$$TL\cap TM = T(L\cap M). \tag{4–10}$$

It follows from Assumption 4 (injectivity) that $L\cap M$ must be given by

$$L\cap M = \{(x,\xi,y',\eta',y',\eta',x,\xi) : (y',\eta';x,\xi)\in\Lambda_{MN}\}. \tag{4–11}$$

Because Λ_{MN} is a manifold this set satisfies (4–9). The property (4–10) follows from the assumption that the map π_Y is immersive. The excess is given by

$$\begin{aligned} \mathsf{e} &= \dim(L\cap M) - (\dim L + \dim M - \dim T^*X\setminus 0\times T^*Y\setminus 0\times T^*Y\setminus 0\times T^*X\setminus 0) \\ &= n-1-c. \end{aligned}$$

Taking into account the pseudodifferential cutoff $\psi(y',D_{y'})$, it follows that (4–7) is a Fourier integral operator. The canonical relation $\Lambda^*_{MN}\circ\Lambda_{MN}$ of $F^*_{MN;\beta}\psi^*\psi F_{MN;\alpha}$ is contained in the diagonal of $T^*X\setminus 0\times T^*X\setminus 0$, so it is a pseudodifferential operator. The order is given by $2\operatorname{order}F_{MN;\alpha}+\mathsf{e}/2 = n-1$ (note that the codimension c drops out).

We write

$$\psi(y',D_{y'})^*\psi(y',D_{y'}) = \sum_i \chi^{(i)}(y',D_{y'}),$$

where the symbols $\chi^{(i)}(y',\eta')$ have small enough support, so that the distribution kernel of $\chi^{(i)}(y',D_{y'})F_{MN;\alpha}$ can be written as the oscillatory integral

$$\chi^{(i)}(y',D_{y'})\mathcal{F}_{MN;\alpha}(y',x) = (2\pi)^{-(3n-1-c)/4-|J|/2}\int \chi^{(i)}(y'_I,\eta'_J,x) \\ \times \mathcal{B}_{MN}(y'_I,\eta'_J,x)w_{MN;\alpha}(y'_I,\eta'_J,x)\mathrm{e}^{\mathrm{i}(S^{(i)}_{MN}(y'_I,x,\eta'_J)+\langle\eta'_J,y'_J\rangle)}\,\mathrm{d}\eta'_J,$$

where $\psi^{(i)}(y'_I,\eta'_J,x) = \psi^{(i)}(y'_I,y'_J(y'_I,\eta'_J,x),\eta'_I(y'_I,\eta'_J,x),\eta'_J)$. We have written $\Phi^{(i)}_{MN}(y',x,\eta'_J) = S^{(i)}_{MN}(y'_I,x,\eta'_J) + \langle\eta'_J,y'_J\rangle$, (cf. (2–18) and (3–4)). We do not indicate the dependence of J on i explicitly. The distribution kernel of the normal operator is then given by a sum of terms

$$\int(\overline{\psi(y',D_{y'})\mathcal{F}_{MN;\beta}(y',x)})(\psi(y',D_{y'})\mathcal{F}_{MN;\alpha}(y',x_0))\,\mathrm{d}y' \\ = (2\pi)^{-(3n-1-c)/2-|J|}\sum_i\int\chi^{(i)}(y'_I,\eta'_{0,J},x_0) \\ \times\overline{\mathcal{B}_{MN}(y'_I,\eta'_J,x)}\mathcal{B}_{MN}(y'_I,\eta'_{0,J},x_0)\overline{w_{MN;\beta}(y'_I,\eta'_J,x)}w_{MN;\alpha}(y'_I,\eta'_{0,J},x_0) \\ \times\mathrm{e}^{\mathrm{i}(S^{(i)}_{MN}(y'_I,x_0,\eta'_{0,J})-S^{(i)}_{MN}(y'_I,x,\eta'_J)+\langle\eta'_{0,J},y'_J\rangle-\langle\eta'_J,y'_J\rangle)}\,\mathrm{d}\eta'_{0,J}\,\mathrm{d}\eta'_J\,\mathrm{d}y'.$$

We now apply the method of stationary phase to integrate out the variables $(y'_J,\eta'_{0,J})$. For the remaining variables we use the Taylor expansion,

$$S^{(i)}_{MN}(y'_I,x_0,\eta'_J) - S^{(i)}_{MN}(y'_I,x,\eta'_J) = \langle x-x_0,\xi(y'_I,\eta'_J,x_0)\rangle + O(|x-x_0|^2).$$

Thus we find (to highest order)

$$(2\pi)^{-(3n-1-c)/2}\sum_i\int\chi^{(i)}(y'_I,\eta'_J,x)^2|\mathcal{B}_{MN}(y'_I,\eta'_J,x)|^2 \\ \times\overline{w_{MN;\beta}(y'_I,\eta'_J,x)}w_{MN;\alpha}(y'_I,\eta'_J,x)\mathrm{e}^{\mathrm{i}\langle x-x_0,\xi(y'_I,\eta'_J,x_0)\rangle}\,\mathrm{d}\eta'_J\,\mathrm{d}y'_I.$$

We now change variables $(x,y'_I,\eta'_J)\to(x,\xi,e)$, and use (3–11). We sum over i and arrive at

$$\mathcal{N}_{MN;\beta\alpha}(x,x_0) = \tfrac{1}{16}(2\pi)^{-2n} \\ \times\int|\psi(y'(x,\xi,e),\eta'(x,\xi,e))|^2\tau^{-4}\overline{w_{MN;\beta}(x,\xi,e)}w_{MN;\alpha}(x,\xi,e) \\ \times\left|\det\frac{\partial(\hat{x},\tilde{x},t)}{\partial y}\right|^{-1}\left|\det\frac{\partial(x,\hat{\xi}_0,\tilde{\xi}_0,\hat{t},\tilde{t})}{\partial(x,\xi,e,y'',\Delta\tau)}\right|_{\substack{\Delta\tau=0\\ y''=0}}\mathrm{e}^{\mathrm{i}\langle x-x_0,\xi\rangle}\,\mathrm{d}\xi\,\mathrm{d}e.$$

It follows that the principal symbol of $N_{MN;\beta\alpha}$ is given by (4–8). □

In the inverse, $(N_{MN})^{-1}_{\alpha\beta}F^*_{MN;\beta}$, seismologists distinguish the action of the parametrix of the normal operator from the imaging operator: They refer to the first action as *amplitude versus angles* (AVA) inversion, when e is given by (4–4).

REMARK 4.4. So far, we have discussed the inversion of data from one pair of modes (M, N). Often data will be available for some subset S of all possible pairs of modes. Define the normal operator for this case as

$$N_{\alpha\beta} = \sum_{(M,N)\in \mathrm{S}} F^*_{MN;\alpha} F_{MN;\beta} = \sum_{(M,N)\in \mathrm{S}} N_{MN;\alpha\beta}.$$

If all the $N_{MN;\alpha\beta}$ are pseudodifferential operators, so is $N_{\alpha\beta}$. A left inverse is now given by $N^{-1}_{\alpha\beta} F^*_\beta$, where F^*_β is the vector of Fourier integral operators containing the $F^*_{MN;\beta}$, $(M, N) \in \mathrm{S}$.

4B. Parameter resolution. In practice, the parametrix $(N_{MN})^{-1}_{\alpha\beta}$ is replaced by a regularized inverse, $\langle (N_{MN})^{-1}\rangle_{\alpha\beta}$ say. Following the Backus–Gilbert approach [4], we subject the symbol matrix $N_{MN;\alpha\beta}(x, \xi)$ for given (x, ξ) and mode pair MN to a singular value decomposition and invoke thresholding. The thresholding yields the regularization and limits the set of parameters that can be resolved. This is apparent in the symbol resolution matrix,

$$\langle (N_{MN})^{-1}\rangle (x, \xi) N_{MN}(x, \xi),$$

namely through its deviation from the identity matrix: One can read off the linear combinations of parameters that can be resolved. An example is shown in Figure 9, where the background was assumed to be isotropic and $M = N =$ P. (There we employ the Voigt notation, C_{IJ}, that replaces the tensor notation c_{ijkl}.)

Seismologists rarely attempt to reconstruct the stiffness tensor components directly. The resolution analysis defines a hierarchy of linear combinations of stiffness tensor components that can actually be extracted from the data.

5. Phantom Images, Artifacts of Type I

We investigate some of the consequences of a violation of the Bolker condition. To this end, we reconsider the composition $F^*_{MN;\beta} F_{MN;\alpha}$ defining the normal operator for a mode pair MN. The canonical relation associated with the imaging operator $F^*_{MN;\beta}$ is given by (cf. (3–10))

$$\Lambda^*_{MN} = \{(x, \xi; y', \eta') : (y', \eta'; x, \xi) \in \Lambda_{MN}\}.$$

Let $L = \Lambda^*_{MN} \times \Lambda_{MN}$ and $M = T^*X \setminus 0 \times \mathrm{diag}(T^*Y \setminus 0) \times T^*X \setminus 0$ as before. The compose $\Lambda^*_{MN} \circ \Lambda_{MN}$ is given by the projection $L \cap M$ on $T^*X \setminus 0 \times T^*X \setminus 0$, namely

$$\Lambda^*_{MN} \circ \Lambda_{MN} = \{(z, \zeta; x, \xi) : (y', \eta'; z, \zeta) \in \Lambda_{MN},\ (y', \eta'; x, \xi) \in \Lambda_{MN}\}.$$

If the intersection $L \cap M$ is clean (cf. (4–9)-(4–10)) the compose $\Lambda^*_{MN} \circ \Lambda_{MN}$ is again a canonical relation [62, Theorem 21.2.14]. (Such result was obtained for transversal intersection by Hörmander [59] and refined to clean intersection by Duistermaat and Guillemin [42].) The intersection $L \cap M$ is clean precisely if L

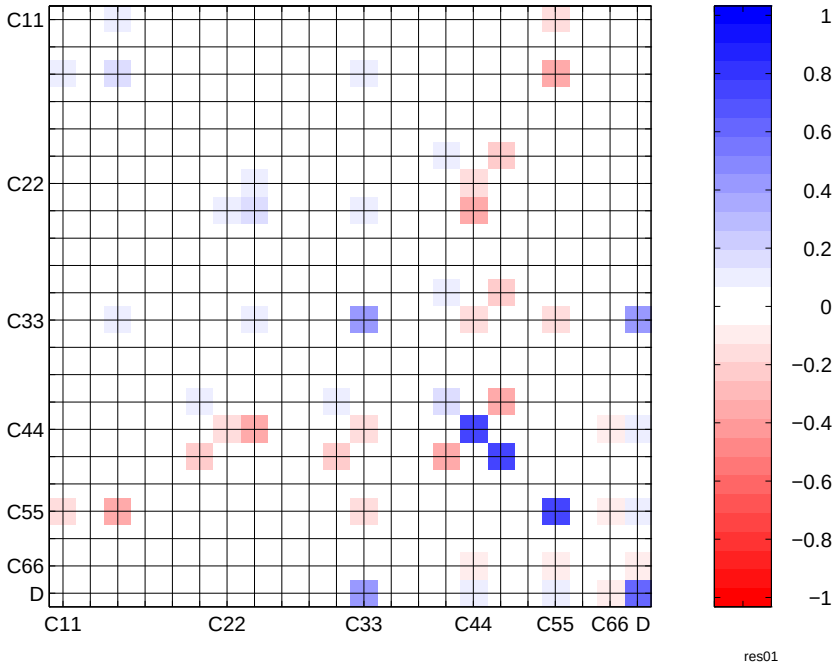


Figure 9. Normal operator symbol resolution matrix; $n = 3$, ξ points in the 3 direction, eigenvalues with a magnitude of less than 0.1 times the primary eigenvalue are zeroed (the matrix has rank 4).

and M intersect transversally in a submanifold of $T^*X\backslash 0 \times T^*Y\backslash 0 \times T^*Y\backslash 0 \times T^*X\backslash 0$.

5A. The composition. Employing the parametrization of Λ_{MN} given in Remark 3.3 and below (4–3), the intersection solves the equation

$$y'(x,\hat\alpha,\tilde\alpha,\tau) = y'(z,\hat\beta,\tilde\beta,\tau), \qquad \eta'(x,\hat\alpha,\tilde\alpha,\tau) = \eta'(z,\hat\beta,\tilde\beta,\tau). \tag{5–1}$$

(Here we have already substituted the immediate equality of frequencies, τ.) These equations describe the following geometry:

(i) x and z lie on the ray in mode M determined by (r,ρ); at x this ray has phase or cotangent direction $\hat\alpha$ and at z this ray has phase or cotangent direction $\hat\beta$;

(ii) x and z lie on the ray in mode N determined by (s,σ); at x this ray has (phase) cotangent direction $\tilde\alpha$ and at z this ray has (phase) cotangent direction $\tilde\beta$;

(iii) since

$$\underbrace{T_M(x,\hat\alpha) + T_N(x,\tilde\alpha)}_{T_{MN}(x,\hat\alpha,\tilde\alpha)} = T_M(z,\hat\beta) + T_N(z,\tilde\beta),$$

if $z \neq x$, $T_M(z,\hat\beta) > T_M(x,\hat\alpha)$ implies $T_N(x,\tilde\alpha) > T_N(z,\tilde\beta)$ and vice versa, in which case the ray originating at r reaches x prior to reaching z while the ray originating at s reaches z prior to reaching x. Because $T_M(z,\hat\beta) - T_M(x,\hat\alpha) = T_N(x,\tilde\alpha) - T_N(z,\tilde\beta)$, the rays originating at z with initial phase directions $\hat\beta$ and $-\tilde\beta$ intersect in x at the same time.

From these geometrical observations, it follows that (5–1) can be recast in the form, for some t_{MN}:

$$x_M(z,\hat\beta,\tau,t_{MN}) - x_N(z,\tilde\beta,\tau,-t_{MN}) = 0, \tag{5–2}$$

and

$$x = x_M(z,\hat\beta,\tau,t_{MN}), \tag{5–3}$$

$$\hat\alpha = \alpha_M(z,\hat\beta,\tau,t_{MN}) \quad := \frac{\xi_M(z,\hat\beta,\tau,t_{MN})}{\big\|\xi_M(z,\hat\beta,\tau,t_{MN})\big\|}, \tag{5–4}$$

$$\tilde\alpha = \alpha_N(z,\tilde\beta,\tau,-t_{MN}) := \frac{\xi_N(z,\tilde\beta,\tau,-t_{MN})}{\big\|\xi_N(z,\tilde\beta,\tau,-t_{MN})\big\|}, \tag{5–5}$$

so that the compose $\Lambda^*_{MN}\circ\Lambda_{MN}$ is given by the set

$$\big\{(z,\zeta;x,\xi) : \text{(5–2)–(5–3) are satisfied, with } \xi = \hat{\boldsymbol\xi}_0(x,\hat\alpha,\tau) + \tilde{\boldsymbol\xi}_0(x,\tilde\alpha,\tau),\ \zeta = \hat{\boldsymbol\xi}_0(z,\hat\beta,\tau) + \tilde{\boldsymbol\xi}_0(z,\tilde\beta,\tau)\big\}. \tag{5–6}$$

Compare (3–10). In these equations we used the solution representation of the Hamilton system in (2–14). So far, the composition has not been subjected to any assumptions.

5B. Transversal intersection ($c = 0$). In case $t_{MN} = 0$ then the solution to equations (5–3)-(5–3) is found to be $(z,\hat\beta,\tilde\beta) = (x,\hat\alpha,\tilde\alpha)$. In the context of the discussion in the previous subsection, we call these solutions the *reciprocal-ray* solutions: the ray originating at s and terminating at r coincides with the ray originating at r and terminating at s. These are the only solutions under the Bolker condition as discussed in Theorem 4.3. It was noted that then L and M intersect cleanly with excess $\mathsf{e} = n - 1 - c$.

In the more general situation when there are solutions with $t_{MN} \neq 0$ (that correspond with the violation of travel time injectivity [99]) the normal operator will attain a contribution that is in general not microlocal (corresponding to *nonreciprocal-ray* solutions). To ensure that this contribution can be represented by a Fourier integral operator, we invoke the condition that L and M still intersect transversally (cleanly with excess 0). This typically occurs in configurations with waveguiding behavior. In seismological terms, then the diffraction surfaces associated with x and z have higher-order contact simultaneously in "common shot" and in "common receiver" data gathers.

Upon linearizing (5–1) the transversal intersection condition can be formulated as the condition that (for the notation, see Remark 4.2)

$$\mathsf{M}=\begin{pmatrix} \frac{\partial r_M}{\partial x} & \frac{\partial r_M}{\partial \hat\alpha} & 0 & \frac{\partial r_M}{\partial z} & \frac{\partial r_M}{\partial \hat\beta} & 0\\ \frac{\partial p^r_M}{\partial x} & \frac{\partial p^r_M}{\partial \hat\alpha} & 0 & \frac{\partial p^r_M}{\partial z} & \frac{\partial p^r_M}{\partial \hat\beta} & 0\\ \frac{\partial s_N}{\partial x} & 0 & \frac{\partial s_N}{\partial \tilde\alpha} & \frac{\partial s_N}{\partial z} & 0 & \frac{\partial s_N}{\partial \tilde\beta}\\ \frac{\partial p^s_N}{\partial x} & 0 & \frac{\partial p^s_N}{\partial \tilde\alpha} & \frac{\partial p^s_N}{\partial z} & 0 & \frac{\partial p^s_N}{\partial \tilde\beta}\\ \frac{\partial T_{MN}}{\partial x} & \frac{\partial T_M}{\partial \hat\alpha} & \frac{\partial T_N}{\partial \tilde\alpha} & \frac{\partial T_{MN}}{\partial z} & \frac{\partial T_M}{\partial \hat\beta} & \frac{\partial T_N}{\partial \tilde\beta} \end{pmatrix} \text{ has maximal rank.}$$

Following the simplification of (5–1) to (5–2), in [89] it is shown that this rank is maximal if and only if the matrix C defined by

$$\begin{pmatrix} \frac{\partial x_M}{\partial z}\Big|_{(z,\hat\beta,t)} - \frac{\partial x_N}{\partial z}\Big|_{(z,\tilde\beta,-t)} & \frac{\partial x_M}{\partial \hat\beta}\Big|_{(z,\hat\beta,t)} & - \frac{\partial x_N}{\partial \tilde\beta}\Big|_{(z,\tilde\beta,-t)} & \frac{\partial x_M}{\partial t}\Big|_{(z,\hat\beta,t)} + \frac{\partial x_N}{\partial t}\Big|_{(z,\tilde\beta,-t)} \end{pmatrix}$$

has maximal rank.

The derivation follows. In view of observation (iii) below (5–1), the map

$$(z,\hat\beta)\to(r_M(z,\hat\beta),p^r_M(z,\hat\beta),T_M(z,\hat\beta))$$

is equal to the composition of maps

$$(z,\hat\beta)\to(x_M(z,\hat\beta,\tau,t),\alpha_M(z,\hat\beta,\tau,t))$$

$$(x,\hat\alpha)\to(r_M(x,\hat\alpha),p^r_M(x,\hat\alpha),T_M(x,\hat\alpha)+t).$$

Using the chain rule, it follows that

$$\frac{\partial(r_M,p^r_M,T_M)}{\partial(z,\hat\beta)}\Big|_{(z,\hat\beta)} = \frac{\partial(r_M,p^r_M,T_M)}{\partial(x,\hat\alpha)}\Big|_{(x,\hat\alpha)} \frac{\partial(x_M,\alpha_M)}{\partial(z,\hat\beta)}\Big|_{(z,\hat\beta,t)}. \tag{5–7}$$

In a similar fashion, we obtain

$$\frac{\partial(s_N,p^s_N,T_N)}{\partial(z,\tilde\beta)}\Big|_{(z,\tilde\beta)} = \frac{\partial(s_N,p^s_N,T_N)}{\partial(x,\tilde\alpha)}\Big|_{(x,\tilde\alpha)} \frac{\partial(x_N,\alpha_N)}{\partial(z,\tilde\beta)}\Big|_{(z,\tilde\beta,-t)}. \tag{5–8}$$

Using these relations, the matrix M can then be factorized as

$$\mathsf{M}=\mathsf{AS},$$

in which

$$\mathsf{A} = \begin{pmatrix} \dfrac{\partial(r_M, p_M^r)}{\partial(x,\hat\alpha)} & 0 \\ 0 & \dfrac{\partial(s_N, p_N^s)}{\partial(x,\tilde\alpha)} \\ \dfrac{\partial T_M}{\partial(x,\hat\alpha)} & \dfrac{\partial T_N}{\partial(x,\tilde\alpha)} \end{pmatrix}$$

is a $(4n-3)\times(4n-2)$ matrix, contains derivatives of the mappings from the subsurface to the acquisition manifold, and has full rank since

$$\left.\frac{\partial(r_M, p_M^r, T_M)}{\partial(x,\hat\alpha)}\right|_{(x,\hat\alpha)} \quad \text{and} \quad \left.\frac{\partial(s_N, p_N^s, T_N)}{\partial(x,\tilde\alpha)}\right|_{(x,\tilde\alpha)}$$

are invertible in view of Liouville's theorem, while

$$\mathsf{S} = \begin{pmatrix} I_n & 0 & 0 & \left.\dfrac{\partial x_M}{\partial z}\right|_{(z,\hat\beta,t)} & \left.\dfrac{\partial x_M}{\partial \hat\beta}\right|_{(z,\hat\beta,t)} & 0 \\ 0 & I_{n-1} & 0 & \left.\dfrac{\partial \alpha_M}{\partial z}\right|_{(z,\hat\beta,t)} & \left.\dfrac{\partial \alpha_M}{\partial \hat\beta}\right|_{(z,\hat\beta,t)} & 0 \\ I_n & 0 & 0 & \left.\dfrac{\partial x_N}{\partial z}\right|_{(z,\tilde\beta,-t)} & 0 & \left.\dfrac{\partial x_N}{\partial \tilde\beta}\right|_{(z,\tilde\beta,-t)} \\ 0 & 0 & I_{n-1} & \left.\dfrac{\partial \alpha_N}{\partial z}\right|_{(z,\tilde\beta,-t)} & 0 & \left.\dfrac{\partial \alpha_N}{\partial \tilde\beta}\right|_{(z,\tilde\beta,-t)} \end{pmatrix}.$$

Following the factorization, we have

$$\operatorname{rank}\mathsf{M} = \operatorname{rank}\mathsf{S} - \dim(\operatorname{range}\mathsf{S} \cap \ker\mathsf{A}). \tag{5–9}$$

The kernel of A follows from the observation (see [33, Appendix A]) that

$$\left.\frac{\partial(r_M, p_M)}{\partial(x,\hat\alpha)}\right|_{(x,\hat\alpha)} \begin{pmatrix} v_M(x,\hat\alpha) \\ 0 \end{pmatrix} = 0, \quad \left.\frac{\partial T_M}{\partial(x,\hat\alpha)}\right|_{(x,\hat\alpha)} \begin{pmatrix} v_M(x,\hat\alpha) \\ 0 \end{pmatrix} = -1;$$

compare Figure 5, where $v_M(x,\hat\alpha) := v_M(x, \tau^{-1}\xi_M(x,\hat\alpha))$. A similar observation holds for (s_N, p_N, T_N) with $-v_N(x,\tilde\alpha)$ replacing $v_M(x,\hat\alpha)$. We conclude that

$$\ker\mathsf{A} = \operatorname{span}\{v_M(x,\hat\alpha), 0, -v_N(x,\tilde\alpha), 0\}.$$

To the range of S the following reasoning applies. Writing $V = [V_{x,\hat{}}, V_{\hat\alpha}, V_{x,\tilde{}}, V_{\tilde\alpha}]$ for an element of $\operatorname{range}\mathsf{S}$ we find, upon subtracting rows, from the structure of the matrix S that

$$V_{x,\hat{}} - V_{x,\tilde{}} \in \operatorname{range}\mathsf{C}'$$

where

$$\mathsf{C}' = \begin{pmatrix} \left.\dfrac{\partial x_M}{\partial z}\right|_{(z,\hat\beta,t)} - \left.\dfrac{\partial x_N}{\partial z}\right|_{(z,\tilde\beta,-t)} & \left.\dfrac{\partial x_M}{\partial \hat\beta}\right|_{(z,\hat\beta,t)} & - \left.\dfrac{\partial x_N}{\partial \tilde\beta}\right|_{(z,\tilde\beta,-t)} \end{pmatrix}.$$

Accounting for the identity submatrices in S, it then follows that $\operatorname{rank}\mathsf{S} = 3n - 2 + \operatorname{rank}\mathsf{C}'$. Following the subtraction of rows, we find that

$$\dim(\operatorname{range}\mathsf{S}\cap\ker\mathsf{A}) = \dim(\operatorname{range}\mathsf{C}'\cap\underbrace{\operatorname{span}\{v_M(x,\hat\alpha)+v_N(x,\tilde\alpha)\}}_{\text{one-dimensional}}). \tag{5–10}$$

If $\operatorname{span}\{v_M(x,\hat\alpha)+v_N(x,\tilde\alpha)\}\subset\operatorname{range}\mathsf{C}'$, $\operatorname{rank}\mathsf{C}=\operatorname{rank}\mathsf{C}'$; otherwise $\operatorname{rank}\mathsf{C}=\operatorname{rank}\mathsf{C}'+1$, i.e.

$$\operatorname{rank}\mathsf{C} = \operatorname{rank}\mathsf{C}' + 1 - \dim(\operatorname{range}\mathsf{C}'\cap\operatorname{span}\{v_M(x,\hat\alpha)+v_N(x,\tilde\alpha)\}). \tag{5–11}$$

Combining (5–10), (5–11) with (5–9) yields

$$\operatorname{rank}\mathsf{M} = \operatorname{rank}\mathsf{C} + 3n - 3.$$

Hence, $\operatorname{rank}\mathsf{M}$ is maximal if and only if $\operatorname{rank}\mathsf{C}$ is maximal.

5C. Nonmicrolocal contribution to the normal operator ($c = 0$). The projection π_Y of Λ_{MN} on $T^*Y\setminus 0$ is immersive. (This follows from Lemma 4.1 and the fact that the projection π_X of Λ_{MN} on $T^*X\setminus 0$ is submersive under Assumption 1.) Hence, for any $(y',\eta';x,\xi)\in\Lambda_{MN}$ there is a small neighborhood $\Lambda_0\subset\Lambda_{MN}$ such that $\pi_Y:\Lambda_0\to T^*Y\setminus 0$ is an embedding. Then

$$(\Lambda_0^*\times\Lambda_0)\cap M = \{(x,\xi,y',\eta';y',\eta',x,\xi) : (x,\xi,y',\eta')\in\Lambda_0\}.$$

Applying this argument for arbitrary $(y',\eta';x,\xi)$ in the canonical relation, we can find subsets $\Lambda_k\subset\Lambda_{MN}$, $k\in K$ for some set K, such that $\pi_Y : \Lambda_k\to T^*Y\setminus 0$ is an embedding. Let

$$L_{\mathrm{mloc}} := \bigcup_{k\in K}\Lambda_k^*\times\Lambda_k,$$

then $L_{\mathrm{mloc}}\cap M$ consists of points corresponding with the *reciprocal* intersection solutions. For each $k\in K$ the intersection

$$(\Lambda_k^*\times\Lambda_k)\cap M$$

is transversal in the submanifold $T^*X\setminus 0\times\operatorname{diag}(\pi_Y(\Lambda_k))\times T^*X\setminus 0$.

Complementary to the microlocal part, $L_{\mathrm{nmloc}} = L\setminus L_{\mathrm{mloc}}$, $L_{\mathrm{nmloc}}\cap M$ consists of points corresponding with the *nonreciprocal* intersection solutions. This intersection is transversal if and only if the following assumption is satisfied:

Assumption 5 (Stolk). *The matrix* C *given by*

$$\left(\left.\frac{\partial x_M}{\partial z}\right|_{(z,\hat\beta,t)} - \left.\frac{\partial x_N}{\partial z}\right|_{(z,\tilde\beta,-t)} \quad \left.\frac{\partial x_M}{\partial\hat\beta}\right|_{(z,\hat\beta,t)} \quad - \left.\frac{\partial x_N}{\partial\tilde\beta}\right|_{(z,\tilde\beta,-t)} \quad \left.\frac{\partial x_M}{\partial t}\right|_{(z,\hat\beta,t)} + \left.\frac{\partial x_N}{\partial t}\right|_{(z,\tilde\beta,-t)}\right)$$

has maximal rank.

If $\Lambda_{MN;c} \subset \Lambda_{MN}$ is a conically compact subset of Λ_{MN}, and we replace L by $L_c = \Lambda^*_{MN;c} \times \Lambda_{MN;c}$, the microlocal and the nonmicrolocal parts of $L_c \cap M$ are both conically compact. It follows that in the parameters $(x, \hat{\alpha}, \tilde{\alpha}, \tau)$ the microlocal and nonmicrolocal parts are separated.

If $\Lambda_{MN;c}$ is a compact subset of Λ_{MN}, there is a finite collection $\{\Lambda_k\}_{k\in\check{K}}$ such that $\Lambda_{MN;c} \subset \bigcup_{k\in\check{K}} \Lambda_k$ and $\Lambda^*_k \circ \Lambda_l$ is either diagonal, empty, or accounts for a nonmicrolocal contributrion. Upon applying appropriate pseudodifferential cutoffs, we obtain $\psi(y', D_{y'})F_{MN;\alpha} = F_{MN;\alpha;k}$ where $F_{MN;\alpha;k}$ has canonical relation Λ_k; in view of the previous observation, we can then write $F_{MN;\alpha}$ as a finite sum of $F_{MN;\alpha;k}$.

To each of the compositions, $F^*_{MN;\alpha;k}F_{MN;\alpha;l}$, the calculus of Fourier integral operators applies. Their orders follow to be $(n-1)/2 + \mathsf{e}/2$. With Assumption 5, the excess of the nonmicrolocal contributions is zero whence their contributions to the normal operator have order $\frac{n-1}{2}$ (while for the microlocal part the excess was found to be $\mathsf{e} = n-1$).

The nonmicrolocal contribution to the normal operator is at most as singular as its pseudodifferential contribution:

THEOREM 5.1 (STOLK). *The operator $F_{MN;\alpha} : H_s \to H_{s-(n-1)/2}$ is continuous, and hence $N_{MN;\alpha\beta} : H_s \to H_{s-n+1}$ is continuous.*

PROOF. The modeling operator $F_{MN;\alpha}$ can be written as the finite sum

$$\sum_{k\in\check{K}} F_{MN;\alpha;k}$$

by "partitioning" its canonical relation: $\Lambda_k \subset \Lambda_{MN}$ is sufficiently small such that it satisfies the Bolker condition. Then $(F_{MN;\alpha;k})^* F_{MN;\beta;k}$ is pseudodifferential of order $n-1$ and hence continuous as a mapping $H_s \to H_{s-(n-1)}$. We find that $F_{MN;\alpha;k}$ is continuous as a mapping $H_s \to H_{s-(n-1)/2}$.

The nonmicrolocal contribution to the normal operator is thus associated with off-diagonal terms of the form $(F_{MN;\alpha;k})^* F_{MN;\beta;l}$. These can be simply estimated from the fact that $\|F_{MN;\alpha;k}u - F_{MN;\beta;l}u\|^2 \geq 0$, namely,

$$2\langle F_{MN;\alpha;k}u, F_{MN;\beta;l}u\rangle \leq \|F_{MN;\alpha;k}u\|^2 + \|F_{MN;\beta;l}u\|^2.$$

Since $F_{MN;\alpha;k}$ is continuous as a mapping $H_s \to H_{s-(n-1)/2}$, it follows that $(F_{MN;\alpha;k})^* F_{MN;\beta;l}$ is continuous as a mapping $H_s \to H_{s-(n-1)}$. □

As a consequence, this theorem implies that the normal operator may not be invertible in the sense that $(N_{MN;\mathrm{mloc}})^{-1}_{\alpha\beta} F^*_{MN;\beta}\, F_{MN;\gamma}$ (cf. equation (4–2)) attains a nonmicrolocal contribution as singular as the identity. The nonmicrolocal contribution, in this case, generates a phantom image, which we also call an artifact of type I; see Figure 10.

The transversal intersection condition for nonreciprocal ray pairs was investigated in [99; 33]; the simplified analysis presented here can be found in [89].

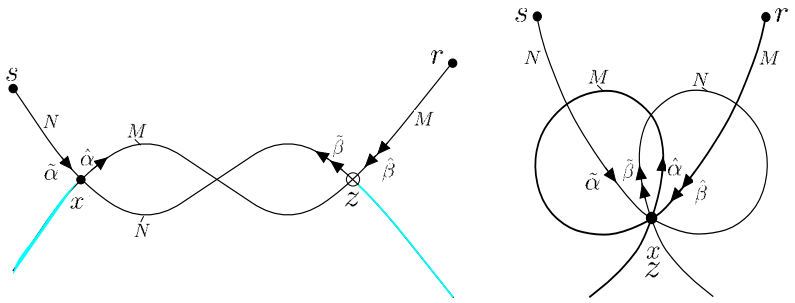


Figure 10. Nonmicrolocal contribution - $z \neq x$ (phantom separated in position; left) and $\zeta \neq \xi$ (phantom separated in orientation; right).

5D. The case $c \geq 1$. Artifacts of type I occur, subject to Assumption 5, in acquisition geometries with $c = 0$ though not generically. Two cases are illustrated in Figure 10. The situation $c \geq 1$ is more of a concern. In principle, neither Assumption 4 nor Assumption 5 are necessarily satisfied. In particular, if the acquisition geometry renders a codimension $c \geq 1$, assumptions of this type may be violated. The precise characterization of the Lagrangian of the normal operator, other than the diagonal, will then yield the artifacts (or phantom images). It is expected that then the calculus of singular Fourier integral operators due to Greenleaf and Uhlmann [48; 50; 49] still applies.

6. Symplectic Geometry of Seismic Data

The wavefront set of the modeled data is not arbitrary. This is a consequence of the fact that data consist of multiple experiments designed to provide a degree of redundancy, which we explain here. Under the Born approximation, subject to the restriction to the acquisition manifold, the singular part of the medium parameters is a function of n variables, while the data are a function of $2n-1-c$ variables. This redundancy is exploited in the parameter reconstruction, and is important in the reconstruction of the background medium (or the medium above the interface in the case of a smooth jump) as well.

We consider again the canonical relation Λ_{MN}. Suppose Assumption 4 is satisfied. In this section, denote by Ω the map

$$\Omega : (x,\xi,e) \mapsto (y'(x,\xi,e),\eta'(x,\xi,e)) : T^*X \setminus 0 \times E \to T^*Y \setminus 0$$

introduced above (4–3); seismologists refer to it as *map demigration*. This map conserves the symplectic form of $T^*X \setminus 0$. That is, if $w_{x_i} = \partial(y',\eta')/\partial x_i$ and similarly for w_{ξ_i}, w_{e_i}, we have

$$\begin{aligned}
&\sigma_Y(w_{x_i}, w_{x_j}) = \sigma_Y(w_{\xi_i}, w_{\xi_j}) = 0,\\
&\sigma_Y(w_{\xi_i}, w_{x_j}) = \delta_{ij},\\
&\sigma_Y(w_{e_i}, w_{x_j}) = \sigma_Y(w_{e_i}, w_{\xi_j}) = \sigma_Y(w_{e_i}, w_{e_j}) = 0.
\end{aligned} \tag{6–1}$$

The (x, ξ, e) are *symplectic coordinates* on the projection of Λ_{MN} on $T^*Y \setminus 0$, which is a subset $\mathcal{L}$ of $T^*Y \setminus 0$.

The image $\mathcal{L}$ of the map Ω is coisotropic, as noted in Lemma 4.1. The sets $(x, \xi) = \text{const.}$ are the isotropic fibers of the fibration of Hörmander [61], Theorem 21.2.6; see also Theorem 21.2.4. Duistermaat [43] calls them characteristic strips (see Theorem 3.6.2). We have sketched the situation in Figure 11. The wavefront set of the data is contained in $\mathcal{L}$ and is a union of fibers.

Using the following result we can extend the coordinates (x, ξ, e) to symplectic coordinates on an open neighborhood of $\mathcal{L}$.

LEMMA 6.1. [91] *Let $\mathcal{L}$ be an embedded coisotropic submanifold of $T^*Y \setminus 0$, with coordinates (x, ξ, e) such that* (6–1) *holds. Denote $\mathcal{L} \ni (y', \eta') = \Omega(x, \xi, e)$. We can find a homogeneous canonical map G from an open part of $T^*(X \times E) \setminus 0$ to an open neighborhood of $\mathcal{L}$ in $T^*Y \setminus 0$, such that $G(x, e, \xi, \varepsilon = 0) = \Omega(x, \xi, e)$.*

PROOF. The e_i can be viewed as (coordinate) functions on $\mathcal{L}$. We first extend them to functions on the whole $T^*Y \setminus 0$ such that the Poisson brackets $\{e_i, e_j\}$ satisfy

$$\{e_i, e_j\} = 0, \qquad 1 \leq i, j \leq m - n, \tag{6–2}$$

where $m = \dim Y = 2n - c - 1$. This can be done successively for $e_1, \ldots, e_{m-n}$ by the method that we describe in the sequel, see Treves [102, Chapter 7, proof of Theorem 3.3] or Duistermaat [43, proof of Theorem 3.5.6]. Suppose we have extended $e_1, \ldots, e_l$, we extend e_{l+1}. In order to satisfy (6–2) e_{l+1} has to be a solution u of

$$H_{e_i} u = 0, \qquad 1 \leq i \leq l,$$

where H_{e_i} is the Hamilton field associated with the function e_i, with initial condition on some manifold transversal to the H_{e_i}. For any $(y', \eta') \in \mathcal{L}$ the covectors $\mathrm{d}e_i$, $1 \leq i \leq l$ restricted to $T_{(y', \eta')}\mathcal{L}$ are linearly independent, so the H_{e_i} are transversal to $\mathcal{L}$ and they are linearly independent modulo $\mathcal{L}$. So we can give the initial condition $u|_{\mathcal{L}} = e_{l+1}$ and even prescribe u on a larger manifold, which leads to nonuniqueness of the extensions e_i.

We now have $m - n$ commuting vectorfields H_{e_i} that are transversal to $\mathcal{L}$ and linearly independent on some open neighborhood of $\mathcal{L}$. The Hamilton systems with parameters ε_i read

$$\frac{\partial y'_j}{\partial \varepsilon_i} = \frac{\partial e_i}{\partial \eta'_j}(y', \eta'), \qquad \frac{\partial \eta'_j}{\partial \varepsilon_i} = -\frac{\partial e_i}{\partial y'_j}(y', \eta'), \qquad 1 \leq i, j \leq m - n.$$

Let $G(x, e, \xi, \varepsilon)$ be the solution for (y', η') of the Hamilton systems combined with initial value $(y', \eta') = \Omega(x, \xi, e)$ with *flowout parameters* ε. This gives a diffeomorphic map from a neigborhood of the set $\varepsilon = 0$ in $T^*(X \times E) \setminus 0$ to a neighborhood of $\mathcal{L}$ in $T^*Y \setminus 0$. One can check from the Hamilton systems that this map is homogeneous.

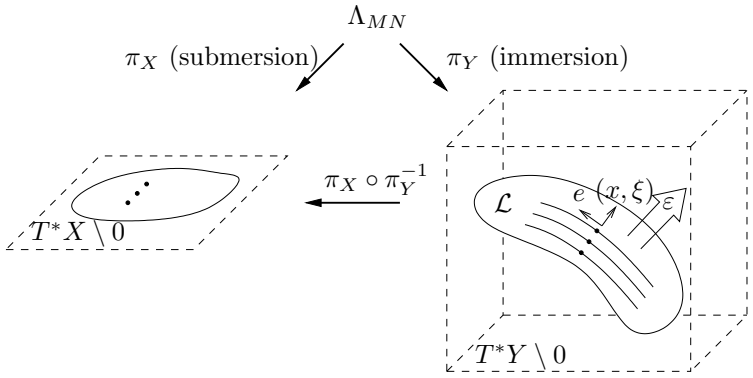


Figure 11. Visualization of the symplectic structure of Λ_{MN} (cone structure omitted).

It remains to check the commutation relations. The relations (6–1) are valid for any ε, because the Hamilton flow conserves the symplectic form on $T^*Y \setminus 0$. The commutation relations for $\partial(y',\eta')/\partial\varepsilon_i$ follow, using that $\partial(y',\eta')/\partial\varepsilon_i = H_{e_i}$. □

Let M_{MN} be the canonical relation associated with the map G we just constructed, i.e. $M_{MN} = \{(G(x,e,\xi,\varepsilon);x,e,\xi,\varepsilon)\}$. We now construct a Maslov-type phase function for M_{MN} that is directly related to a phase function for Λ_{MN}. Suppose (y'_I,η'_J,x) are suitable coordinates for Λ_{MN} ($\varepsilon=0$). For ε small, the constant-ε subset of M_{MN} allows the same set of coordinates, thus we can use coordinates $(y'_I,\eta'_J,x,\varepsilon)$ on M_{MN}. Now there is (see Theorem 4.21 in Maslov and Fedoriuk [73]) a function $S_{MN}(y'_I,x,\eta'_J,\varepsilon)$ such that M_{MN} is given by

$$y'_J = -\frac{\partial S_{MN}}{\partial \eta'_J}, \qquad \eta'_I = \frac{\partial S_{MN}}{\partial y'_I}, \qquad \xi = -\frac{\partial S_{MN}}{\partial x}, \qquad e = \frac{\partial S_{MN}}{\partial \varepsilon}.$$

Thus a phase function for M_{MN} is given by

$$\Psi_{MN}(y',x,e,\eta'_J,\varepsilon) = S_{MN}(y'_I,x,\eta'_J,\varepsilon) + \langle \eta'_J, y'_J\rangle - \langle \varepsilon, e\rangle. \tag{6–3}$$

A Maslov-type phase function for Λ_{MN} then follows as

$$\Phi_{MN}(y',x,\eta'_J) = \Psi_{MN}\Big(y',\, x,\, \frac{\partial S_{MN}}{\partial\varepsilon}\Big|_{\varepsilon=0},\, \eta'_J,\, 0\Big) = S_{MN}(y'_I,\eta'_J,x,0) + \langle \eta'_J, y'_J\rangle.$$

The variable ε also plays a role in the formulation of the generalized Radon transform in Section 9.

7. Modeling and Inversion under the Kirchhoff Approximation

Another way to model the subsurface is to assume that it consists of different regions (*layers*) separated by smooth interfaces. The medium parameters, stiffness c_{ijkl} and density ρ, are assumed to vary smoothly on each region, and smoothly extendible across each interface, but they vary discontinuously at an

interface. Here we model the reflection of waves at a smooth interface between two such layers with smoothly varying medium parameters.

7A. Reflection at an interface: Microlocal analysis of the "Kirchhoff" approximation. The amplitude of the scattered waves is determined essentially by the reflection coefficients, and, implicitly, also by the curvature of the reflecting interface. Expressions for these coefficients are well known for the case of two constant coefficient media separated by a plane interface (see e.g. Aki and Richards [2], Chapter 5). In the case of smoothly varying media the reflection coefficients determine the scattering in the high-frequency limit. For a treatment of reflection and transmission of waves using microlocal analysis, see Taylor [98]; for the acoustic case, see also Hansen [56].

Mathematically the reflection and transmission of waves is formulated as a boundary value problem. The displacement u_l must satisfy the elastic wave equation under given initial conditions. In addition the displacement and the normal traction must be continuous at the interface. Denote the normal to the interface by ν. The

$$P_{il}u_l = f_i \text{ away from the interface}, \qquad u_l = 0 \text{ for } t < 0, \tag{7–1}$$

must hold, while

$$\begin{aligned} \rho^{-1/2}u_l &\text{ is continuous at the interface,} \\ \nu_j c_{ijkl}\frac{\partial}{\partial x_k}(\rho^{-1/2}u_l) &\text{ is continuous at the interface.} \end{aligned} \tag{7–2}$$

Here we have the factors ρ because of our normalization (2–3). We assume that the source vanishes on a neighborhood of the interface. That this is a well-posed problem can be shown using energy estimates; see, for instance, Lions and Magenes [71], Section 3.8.

The solutions to the partial differential equation with $f = 0$ follow from the theory discussed in Section 2. The singularities are propagated along the bicharacteristics, curves in $T^*(X \times \mathbb{R}) \setminus 0$, given by

$$(x_M(x_0, \xi_0, \pm t), t, \xi_M(x_0, \xi_0, \pm t), \mp B_M(x_0, \xi_0)).$$

This is the bicharacteristic associated with the $M, \pm$ constituent of the solution; see Section 2. We define a bicharacteristic to be incoming if its direction is from inside a layer towards the interface for increasing time. We define a bicharacteristic to be outgoing if its direction is away from the interface into a layer for increasing time.

Assume that the incoming bicharacteristic stays inside a layer from $t = 0$ until it hits the interface, then the solution along such a bicharacteristic is determined completely by the partial differential equation and the initial condition. On the other hand, the solution along the outgoing bicharacteristics is not determined by the partial differential equation and the initial condition. We show that

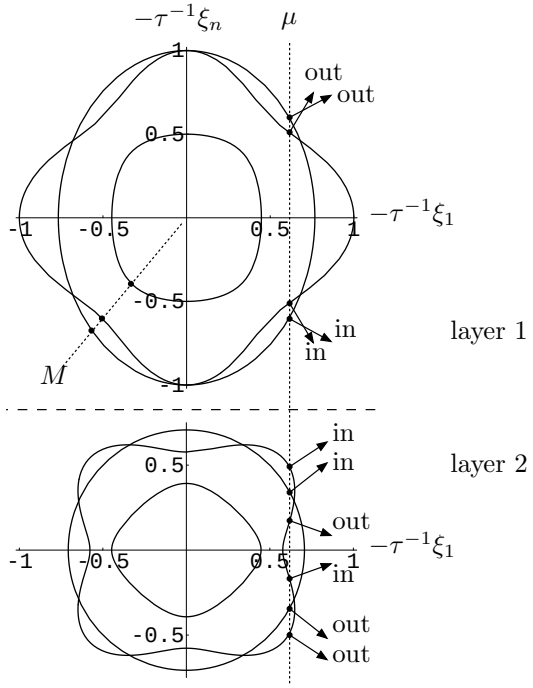


Figure 12. 2-dimensional section of an $n = 3$-dimensional slowness surfaces at some point of the interface, for the medium on both sides of the interface. The slownesses of the modes that interact (i.e. reflect and transmit into each other) are the intersection points with a line that is parallel to the normal of the interface. The group velocity, which is normal to the slowness surface, determines whether the mode is incoming or outgoing.

the solution along the outgoing bicharacteristics is determined by the partial differential equation and the interface conditions in (7–2).

We consider the consequences of the interface conditions. Assume for the moment that the interface is located at $x_n = 0$. We denote $x' = (x_1, \ldots, x_{n-1}), x = (x', x_n)$ and similarly, $\xi = (\xi', \xi_n)$. The wavefront set of the restriction of u_l to $x_n = 0$ satisfies

$$\mathrm{WF}(u_l|_{x_n=0}) = \big\{(x', t, \xi', \tau) : \text{there is } \xi_n \text{ with } (x', 0, t, \xi', \xi_n, \tau) \in \mathrm{WF}(u_l)\big\}.$$

It follows that a solution traveling along a bicharacteristic that intersects the boundary at some point $(x', 0, t)$ interacts with any other such solution as long as the associated values for ξ', τ in their wavefront sets coincide (Snell's law). This is depicted in Figure 12.

Depending upon the boundary coordinate x' and the "tangential" slowness $-\tau^{-1}\xi'$, the number of interacting bicharacteristics may vary. For large values of $-\tau^{-1}\xi'$ there will be neither incoming nor outgoing modes; for small values there are n incoming and n outgoing modes. The situation where the vertical

line in Figure 12 is tangent to the slowness surface corresponds to rays tangent to the interface. Such rays are associated with head-waves and are not treated in our analysis. Equation (2–9) implies that the incoming and the outgoing modes correspond to the real solutions ξ_n of

$$\det P_{il}(x', 0, \xi', \xi_n, \tau) = 0.$$

This equation has $2n$ real or complex conjugated roots. The complex roots correspond to evanescent wave constituents. To number the roots we use an index μ.

In the following theorem we show that if none of the rays involved is tangent, there exists a pseudodifferential operator-type relation between the different modes restricted to the surface $x_n = 0$; we calculate its principal symbol in the proof. Let $x \mapsto z(x) : \mathbb{R}^n \to \mathbb{R}^n$ be a coordinate transformation such that the interface is given by $z_n = 0$. The corresponding cotangent vector is denoted by ζ, and satisfies

$$\zeta_i(\xi) = \left(\left(\frac{\partial z}{\partial x} \right)^{-1} \right)^{t_{ij}} \xi_j;$$

moreover the z form coordinates on a manifold Z.

ASSUMPTION 6. *There are no rays tangent to the interface $z_n = 0$ microlocally at (z', t, ζ', τ).*

THEOREM 7.1. *Suppose the roots τ of* (2–9) *have constant multiplicity and Assumption* 6 *is valid microlocally on some neighborhood in $T^*(Z' \times \mathbb{R}) \setminus 0$. Let $u^{\mathrm{in}}_{N(\nu)}$ be microlocal constituents of a solution describing the "incoming" modes, and suppose $G_{M(\mu)}$ refers to an "outgoing" Green's function* (2–19). *Microlocally, the single reflected/transmitted constituent of the solution is given by*

$$u_{M(\mu)}(x,t) = \int_{z_n=0} G_{M(\mu)}(x, x(z), t - t_0) 2\mathrm{i} D_{t_0} \big(R_{\mu\nu}(z, D_{z'}, D_{t_0}) u^{\mathrm{in}}_{N(\nu)}(x(z), t_0) \big) \times \left| \det \frac{\partial x}{\partial z} \right| \left\| \frac{\partial z_n}{\partial x} \right\| \mathrm{d}z' \, \mathrm{d}t_0, \quad (7\text{–}3)$$

where $R_{\mu\nu}(z, D_{z'}, D_t)$ is a pseudodifferential operator of order 0.

In the proof we derive the explicit form of $R^{\mathrm{prin}}_{\mu\nu}(z, \zeta', \tau)$; see Remark 7.2. The $|\det \partial x/\partial z| \, \|\partial z_n/\partial x\| \, \mathrm{d}z'$ integration is the Euclidean surface integral over the surface $z_n = 0$.

PROOF. (See [91].) For the moment we assume $z(x) = x$, i.e. that we have a reflector at $x_n = 0$, and smooth coefficients on either side. We show that at the interface there is a relation of the type

$$u^{\mathrm{out}}_{M(\mu)}(x', 0, t) = R^0_{\mu\nu}(x', 0, D', D_t) u^{\mathrm{in}}_{N(\nu)}. \quad (7\text{–}4)$$

We use the notation $c_{jk;il} = c_{ijkl}$ and also $(c_{jk})_{il} = c_{ijkl}$ [111]. The partial differential equation (2–1) reads in this notation

$$\Big(\rho\delta_{il}\frac{\partial^2}{\partial t^2} - c_{jk;il}\frac{\partial^2}{\partial x_j \partial x_k}\Big)(\rho^{-1/2}u_l) + \text{l.o.t.} = 0.$$

This equation can be rewritten as a first-order system in the variable x_n for the vector V_a of length $2n$ that contains both the displacement and the normal traction (normal to the surface $x_n = \text{const.}$)

$$V_a = \begin{pmatrix} \rho^{-1/2}u_i \\ c_{nk;il}\frac{\partial(\rho^{-1/2}u_l)}{\partial x_k} \end{pmatrix}, \qquad i = 1, \ldots, n \tag{7–5}$$

in preparation for the boundary value problem (7–1), (7–2). Here a is an index in $\{1, \ldots, 2n\}$. The first-order system then is

$$\frac{\partial V_a}{\partial x_n} = \mathrm{i}C_{ab}(x, D', D_t)V_b,$$

where C_{ab} is a matrix partial differential operator given to highest order by

$$C_{ab}(x, D', D_t) = -\mathrm{i}\begin{pmatrix} -\sum_{q=1}^{n-1}\sum_{j=1}^{n}(c_{nn})^{-1}_{ij}c_{nq;jl}\frac{\partial}{\partial x_q} & (c_{nn})^{-1}_{il} \\ -\sum_{p,q=1}^{n-1} b_{pq;il}\frac{\partial^2}{\partial x_p \partial x_q} + \rho\delta_{il}\frac{\partial^2}{\partial t^2} & -\sum_{p=1}^{n-1}\frac{\partial}{\partial x_p}c_{pn;ij}(c_{nn})^{-1}_{jl} \end{pmatrix}_{ab}.$$

Here $b_{pq;il} = c_{pq;il} - \sum_{j,k=1}^{n} c_{pn;ij}(c_{nn})^{-1}_{jk}c_{nq;kl}$ (in this equation we indicate the summations explicitly because the summations over p, q are $1, \ldots, n-1$, whereas $j \in \{1, \ldots, n\}$).

We next decouple this first-order system microlocally, in a way similar to the decoupling in Section 2A. We want to find scalar pseudodifferential operators $C_\mu(x, D', D_t)$ and a matrix pseudodifferential operator $L_{a\mu}(x, D', D_t)$ such that

$$C_{ab}(x, D', D_t) = L_{a\mu}(x, D', D_t)\,\mathrm{diag}(C_\mu(x, D', D_t))_{\mu\nu}\,L^{-1}_{\nu b}(x, D', D_t).$$

The principal symbols $C^{\mathrm{prin}}_\mu(x, \xi', \tau)$ are the solutions for ξ_n of

$$\det P^{\mathrm{prin}}_{il}(x, (\xi', \xi_n), \tau) = 0. \tag{7–6}$$

In fact, it suffices for the transformed operator (the matrix $\mathrm{diag}\, C_\mu(x, D', D_t)_{\mu\nu}$) to be block-diagonal: a block for each different real root of (7–6), a block with eigenvalues with positive imaginary part, and a block with eigenvalues with negative imaginary part. (This has also been discussed by Taylor [98].) By hypothesis of the theorem this transformation can be realized. This is because when varying ξ', τ, the multiplicity of a real eigenvalue only changes when the multiplicity of the corresponding root of (2–9) changes, or when two real eigenvalues become

complex. The number of complex eigenvalues with positive or negative imaginary part changes only when two real eigenvalues become complex or vice versa. The latter case occurs only when there are tangent rays, and is hence excluded.

The $2n \times 2n$ principal symbol $L^{\mathrm{prin}}_{a\mu}$ (the columns appropriately normalized) is given by

$$L^{\mathrm{prin}}_{a\mu}(x,\xi',\tau) = \begin{pmatrix} Q^{\mathrm{prin}}_{iM(\mu)}(x,(\xi',C^{\mathrm{prin}}_{\mu}(x,\xi',\tau))) \\ c_{in;kl}(-\mathrm{i}(\xi',C^{\mathrm{prin}}_{\mu}(x,\xi',\tau))_k) Q^{\mathrm{prin}}_{lM(\mu)}(x,(\xi',C^{\mathrm{prin}}_{\mu}(x,\xi',\tau))) \end{pmatrix}_{a\mu}.$$

(The polarization vector $Q_{iM}(x,\xi)$ can also be defined for complex ξ). We define $V_\mu = L(x,D',D_t)^{-1}_{\mu a} V_a$. (The index mapping $\mu \mapsto M(\mu)$ assigns the appropriate mode to the normal component of the wave vector).

If the principal symbol of $C_\mu(x,\xi',\tau)$ is real, the decoupled equation for mode μ is of hyperbolic type. It corresponds to an outgoing wave or to an incoming wave, depending on the direction of the corresponding ray. If the principal symbol of $C_\mu(x,\xi',\tau)$ is complex, the decoupled operator for mode μ is of elliptic type. Depending on the sign of the imaginary part it corresponds to a mode that grows in the n-direction, a backward parabolic equation, or one that decays, a forward parabolic equation. The growing mode has to be absent, in view of energy considerations; see also Hörmander [62], Section 20.1.

The matrix $L_{a\mu}$ is fixed up to normalization of its columns. For the elliptic modes ($\operatorname{Im} C^{\mathrm{prin}}_{\mu}(x,\xi',\tau) \neq 0$) the normalization is unimportant. For the hyperbolic modes the normalization can be such that the vector $V_\mu = L(x,D',D_t)^{-1}_{\mu a} V_a$ agrees microlocally with the corresponding mode $u_{M,\pm}$ defined in Section 2. To see this, assume V_μ refers to the same mode as $u_{M,\pm}$. In that case there is an invertible pseudodifferential operator $\psi(x,D,D_t)$ of order 0 such that $V_\mu = \psi u_{M,\pm}$. Now we can define $V_{\mu,\mathrm{new}} = \psi^{-1} V_{\mu,\mathrm{old}}$. Because ψ may depend on ξ_n, this factor cannot directly be absorbed in L. However, since $V_{\mu,\mathrm{old}}$ satisfies a first-order hyperbolic equation the dependence on ξ_n can be eliminated and the factor ψ^{-1} can be absorbed in L.

In proving this let the in-modes be the modes for which the amplitude is known, which are of the incoming hyperbolic and the growing elliptic type. Denote by $L^{(1)}_{a\mu}, L^{(2)}_{a\mu}$ the matrix $L_{a\mu}$ on either side of the interface. We define the $2n \times 2n$ matrix L^{in} such that it contains the columns related to incoming modes extracted from $L^{(1)}_{a\mu}, L^{(2)}_{a\mu}$, i.e.

$$L^{\mathrm{in}}_{a\mu} = \begin{pmatrix} L^{(1),\mathrm{in}} & -L^{(2),\mathrm{in}} \end{pmatrix}_{a\mu},$$

and define $L^{\mathrm{out}}_{a\mu}$ similarly (so, here, μ is slightly different). The interface conditions (7–2) now read

$$L^{\mathrm{out}}_{a\mu} V^{\mathrm{out}}_\mu + L^{\mathrm{in}}_{a\mu} V^{\mathrm{in}}_\mu = 0.$$

If we set $R^0_{\mu\nu} = -(L^{\mathrm{out}})^{-1}_{\mu a} L^{\mathrm{in}}_{a\nu}$ (for the question whether the inverse exists; see Remark 7.2 after the proof) then the part referring to the hyperbolic modes gives Equation (7–4).

The $u_M^{\rm out}$ are determined at the interface by (7–4). Describing their propagation away from the interface is a (microlocal) initial value problem similar to the problem for $G_{M,\pm}$ above, where now the x_n variable plays the role of time. The solution is again a Fourier integral operator, with canonical relation generated by the bicharacteristics. It follows that we can use $\phi_{M,\pm}(x,t-t_0,x_0,\xi_J,\tau)$ as phase function (taking care that $n\notin J$). The amplitude $\mathcal{A}_{M,\pm}(x_I,x_0,\xi_J,\tau)$ satisfies the transport equation as before. However, the restriction of the Fourier integral operator to the "initial surface" $x_n=0$ is a pseudodifferential operator that is not necessarily the identity. Let us assume

$$u_M^{\rm out}(x,t)=\int_{x_{0,n}=0} G_{M,\pm}(x,(x_0',0),t-t_0)\psi(x,D_{x_0'},D_{t_0})u_M^{\rm out}(x_0',0,t_0)\,{\rm d}x_0'\,{\rm d}t_0, \tag{7–7}$$

where $\psi(x,D',D_t)$ is to be found such that the restriction of this representation to $x_n=0$ is the identity. The $\pm$ sign is chosen such that $G_{M,\pm}$ is the outgoing mode. We can use again Section 8.6 of Treves [102] to find that the principal symbol of this pseudodifferential operator should be

$$\psi(x,\xi',\tau)=\left|\frac{\partial B_M}{\partial\xi_n}(x,\xi',C_\mu^{\rm prin}(\xi',\tau))\right|=\left|\frac{\partial x_{M,n}}{\partial t}(x,\xi',C_\mu^{\rm prin}(\xi',\tau),0)\right|, \tag{7–8}$$

i.e., the normal component of the velocity of the ray, the group velocity.

We now replace $G_{M,\pm}$ by (the relevant part of) G_M, using the equality $G_M=\frac{1}{2}{\rm i}G_{M,+}B_M(x,D)^{-1}-\frac{1}{2}{\rm i}G_{M,-}B_M(x,D)^{-1}$. Taking this and the relation $B_M^{\rm prin}(x,\xi)=\mp\tau$ into account, we have now obtained (7–3) for the case that $z=x$ (no coordinate transformation).

We argue that (7–3) is also true when $z(x)$ is a general coordinate transformation. This follows from transforming the equations (7–1), (7–2) to z coordinates. In general, to highest order, the symbol of (pseudo)differential operators transforms as $\psi^{\rm transf}(z,\zeta,\tau)=\psi\big(x(z),(\partial z/\partial x)^{\rm t}\zeta,\tau\big)$. Tracing the steps of the proof we find the following equivalent of (7–4)

$$u_{M(\mu)}^{\rm out}(x(z',0),t)=R^0_{\mu\nu}(z',0,D_{z'},D_t)u_{N(\nu)}^{\rm in}(x(z',0),t). \tag{7–9}$$

When the interface is at $z_n=0$ we can obtain (7–7) in z coordinates instead of x coordinates. Transforming G_M,u_M back to x coordinates we find that, for x away from the interface,

$$u_M(x)=\int_{z_n=0} G_M(x,x(z),t-t_0)\left|\frac{\partial z_{M,n}}{\partial t}(z,D_{z'},D_{t_0})\right| \times u_M^{\rm out}(x(z),t_0)\left|\det\frac{\partial x}{\partial z}\right|{\rm d}z'\,{\rm d}t_0.$$

Here $\left|\dfrac{\partial z_{M,n}}{\partial t}(z, D_{z'}, D_t)\right|$ is the transformed version of (7–8). Expression (7–3) follows, with

$$R_{\mu\nu}(z, \zeta', \tau) = \left|\frac{\partial z_{M,n}}{\partial t}(z, \zeta', \tau)\right| \left\|\frac{\partial z_n}{\partial x}\right\|^{-1} R^0_{\mu\nu}(z, \zeta', \tau). \qquad \square$$

REMARK 7.2. The principal symbol $R^{0,\mathrm{prin}}_{\mu\nu}(z, \zeta', \tau)$ that occurs in the proof is simply the flux-normalized reflection coefficient for the amplitudes. The principal symbol $R^{\mathrm{prin}}_{\mu\nu}(z, \zeta', \tau)$ is obtained by multiplying $R^{0,\mathrm{prin}}_{\mu\nu}$ with the normal component of the velocity of the ray, given (for $z(x) = x$) by (7–8). The reflection coefficients satisfy unitary relations; see Chapman [28] and Kennett [70] (the appendix to Chapter 5). These follow essentially from conservation of energy. It follows that the matrix of reflection coefficients is well defined and in particular that the inverse of $L^{\mathrm{out}}_{a\mu}$ exists. Chapman [28] also gives a direct proof of the reciprocity relations for the reflection coefficients.

REMARK 7.3. We have shown that the reflected/transmitted wave is given by a composition of Fourier integral operators acting on the source. In the proof of Theorem 7.1, one can recognize the elements of the derivation of the Kirchhoff(-Helmholtz) approximate scattering theory for scalar waves [18]. In the case of multiple reflections or transmissions (for instance in a medium consisting of a number of smooth domains separated by smooth interfaces) this is also the case (cf. Frazer and Sen [46]). It follows that microlocally the solution operator describing the reflected solutions is itself a Fourier integral operator, where the canonical relation is given by the generalized bicharacteristics (i.e. the reflected and transmitted bicharacteristics) and the amplitude is essentially the product of the ray amplitudes and the reflection/transmission coefficients. The integration over z' accounts for the effects associated with the curvature of the interface.

7B. Modeling: Kirchhoff versus Born. In this subsection we match the expression for the data modeled using the smooth jump (Kirchhoff) approximation to the expressions for the Born modeled data we obtained in Section 3. The smooth medium above the interface plays the role of the background medium in the Born approximation.

From Theorem 7.1 it follows that reflection of an incident N-mode with covector $\tilde{\xi}_0$ into a scattered M-mode with covector $\hat{\xi}_0$ can take place if the frequencies are equal and $\hat{\xi}_0 + \tilde{\xi}_0$ is normal to the interface. In other words, $\hat{\xi}_0 + \tilde{\xi}_0$ must be in the wavefront set of the singular function of the interface, $\delta(z_n(x))$. Given $\tilde{\xi}_0, \hat{\xi}_0$ one can identify $\mu(M), \nu(N)$, and define (at least to highest order) the reflection coefficient as a function of $(x, \hat{\xi}_0, \tilde{\xi}_0)$, $R^{\mathrm{prin}}_{MN}(x, \hat{\xi}_0, \tilde{\xi}_0) = R^{\mathrm{prin}}_{\mu(M),\nu(N)}(z'(x), \zeta'(\tilde{\xi}_0), \tau)$. This factor can now be viewed as a function either of coordinates (y'_I, x, η'_J) or of coordinates (x, ξ, e) on Λ_{MN} (strictly speaking only defined for x in the interface, and ξ normal to the interface). To highest order, R_{MN} does not depend on $\|\xi\|$ and is simply a function of (x, e). We obtain the

following result, which is a generalization of the Kirchhoff approximation. The normalization factor $\|\partial z_n/\partial x\|$ of the δ-function is such that integral

$$\int (\cdots) \left\| \frac{\partial z_n}{\partial x} \right\| \delta(z_n(x))\, \mathrm{d}x$$

is an integral over the surface $z_n = 0$ with Euclidean surface measure in x coordinates.

THEOREM 7.4. *Suppose Assumptions* 1, 6, 2, *and* 3 *are satisfied microlocally for the relevant part of the data. Let* $\Phi_{MN}(y', x, \eta'_J), \mathcal{B}_{MN}(y'_I, x, \eta'_J)$ *be the phase and amplitude as in Theorem* 3.2, *but here for the smooth medium above the interface. Then the data modeled with the smooth jump interface model is given microlocally by*

$$G^{\mathrm{refl}}_{MN}(y') = (2\pi)^{-|J|/2-(3n-1-c)/4} \int \big(\mathcal{B}_{MN}(y'_I, x, \eta'_J) 2\mathrm{i}\tau(\eta') R_{MN}(y'_I, x, \eta'_J) + \text{l.o.t.} \big) \times \mathrm{e}^{\mathrm{i}\Phi_{MN}(y', x, \eta'_J)} \left\| \frac{\partial z_n}{\partial x} \right\| \delta(z_n(x))\, \mathrm{d}\eta'_J\, \mathrm{d}x; \quad (7\text{–}10)$$

in other words, by a Fourier integral operator with canonical relation Λ_{MN} *and order* $(n-1+c)/4 - 1$ *acting on the distribution* $\|\partial z_n/\partial x\| \delta(z_n(x))$.

PROOF. We write the distribution kernel of the reflected data (7–3) in a form similar to (3–3). First recall the reciprocal expression for the Green's function (2–24),

$$G_N(x(z), \tilde{x}, t_0) = (2\pi)^{-(|\tilde{J}|+1)/2-(2n+1)/4} \times \int \mathcal{A}_N(\tilde{x}_{\tilde{I}}, x(z), \tilde{\xi}_{\tilde{J}}, \tau) \mathrm{e}^{\mathrm{i}\phi_N(\tilde{x}, x(z), t_0, \tilde{\xi}_{\tilde{J}}, \tau)}\, \mathrm{d}\tilde{\xi}_{\tilde{J}}\, \mathrm{d}\tau.$$

By using Theorem 7.1, and doing an integration over a t and a τ variable one finds that the Green's function for the reflected part is given by

$$G^{\mathrm{refl}}_{MN}(\hat{x}, \tilde{x}, t) = (2\pi)^{-(|\hat{J}|+|\tilde{J}|+1)/2-n} \times \int_{z_n=0} \big(2\mathrm{i}\tau \mathcal{A}_M(\hat{x}_{\hat{I}}, x(z), \hat{\xi}_{\hat{J}}, \tau) \mathcal{A}_N(\tilde{x}_{\tilde{I}}, x(z), \tilde{\xi}_{\tilde{J}}, \tau) R_{\mu(M)\nu(N)}(z, \zeta', \tau) + \text{l.o.t.} \big) \times \mathrm{e}^{\mathrm{i}\Phi_{MN}(\hat{x}, \tilde{x}, t, x(z), \hat{\xi}_{\hat{J}}, \tilde{\xi}_{\tilde{J}}, \tau)} \left| \det \frac{\partial x}{\partial z} \right| \left\| \frac{\partial z_n}{\partial x} \right\| \mathrm{d}\hat{\xi}_{\hat{J}}\, \mathrm{d}\tilde{\xi}_{\tilde{J}}\, \mathrm{d}\tau\, \mathrm{d}z',$$

where ζ' depends on $(x(z), \tilde{\xi}_0)$ (the indices μ, ν for the reflection coefficients have been explained in Section 7A). The integration $\int(\cdots)\mathrm{d}z'$ is now replaced by $\int(\cdots)\delta(z_n)\mathrm{d}z$. The latter can be transformed back to an integral over x. Thus we obtain

$$(2\pi)^{-(|\hat{J}|+|\tilde{J}|+1)/2-n}$$
$$\times \int \big(2\mathrm{i}\tau \mathcal{A}_M(\hat{x}_{\hat{I}}, x, \hat{\xi}_{\hat{J}}, \tau)\mathcal{A}_N(\tilde{x}_{\tilde{I}}, x, \tilde{\xi}_{\tilde{J}}, \tau) R_{\mu(M)\nu(N)}(z(x), \zeta'(\tilde{\xi}_J, x), \tau) + \text{l.o.t.}\big)$$
$$\times \mathrm{e}^{\mathrm{i}\Phi_{MN}(\hat{x},\tilde{x},t,x,\hat{\xi}_{\hat{J}},\tilde{\xi}_{\tilde{J}},\tau)} \left\| \frac{\partial z_n}{\partial x} \right\| \delta(z_n(x))\, \mathrm{d}\hat{\xi}_{\hat{J}}\, \mathrm{d}\tilde{\xi}_{\tilde{J}}\, \mathrm{d}\tau\, \mathrm{d}x.$$

This formula is similar to (3–3), except for the fact that the amplitude is different and $\delta c_{ijkl}(x)/\rho(x)$, $\delta\rho(x)/\rho(x)$ is replaced by the δ-function $\|\partial z_n/\partial x\|\delta(z_n(x))$. Also the factors $w_{MN;ijkl}, w_{MN;0}$ depend only on the background medium, while $R_{\mu(M)\nu(N)}$ depends on the total medium. The phase function Φ_{MN} now comes from the smooth medium above the reflector.

The data are modeled by $G^{\text{refl}}_{MN}(\hat{x}, \tilde{x}, t)$ with $(\hat{x}, \tilde{x}, t)$ in the acquisition manifold, as is explained in the text below Lemma 3.1. We follow the approach of Section 3, and do a coordinate transformation $(\hat{x}, \tilde{x}, t) \mapsto (y', y'')$, such that the acquisition manifold is given by $y'' = 0$. It follows that under Assumptions 2 and 3 the data are obtained as the image of a Fourier integral operator acting on $\|\partial z_n/\partial x\|\delta(z_n(x))$ and that it is given by (7–10). □

We now construct the reflectivity function and the operator that maps it to seismic data. This is done by applying the results of Section 6 to the Kirchhoff modeling formula (7–10), and its equivalent under the Born approximation (3–3).

THEOREM 7.5. [91] *Suppose that microlocally Assumptions* 1, 6, 2, 3, *and* 4 *are satisfied. Let* $H_{MN} : \mathcal{E}'(X \times E) \to \mathcal{D}'(Y)$ *be the Fourier integral operator with canonical relation given by the extended map* $(x, \xi, e, \varepsilon) \mapsto (y', \eta')$ *constructed in Section* 6, *and with amplitude to highest order given by*

$$(2\pi)^{n/2}(2\mathrm{i}\tau)\mathcal{B}_{MN}(y'_I, x, \eta'_J, \varepsilon),$$

such that $\mathcal{B}_{MN}(\varepsilon = 0)$ *is as given in Theorem* 3.2. *Then the data, in both Born and Kirchhoff approximations, are given by* H_{MN} *acting on a distribution* $r_{MN}(x, e)$ *of the form*

$$r_{MN}(x, e) = (\text{pseudo})(x, D_x, e)(\text{distribution})(x). \tag{7–11}$$

For the Kirchhoff approximation this distribution equals $\|\partial z_n/\partial x\|\delta(z_n(x))$, *while the principal symbol of the pseudodifferential operator equals* $R_{MN}(x, e)$, *so to highest order* $r_{MN}(x, e) = R_{MN}(x, e)\|\partial z_n/\partial x\|\delta(z_n(x))$. *For the Born approximation the function* $r_{MN}(x, e)$ *is given by a pseudodifferential operator acting on*

$$\Big(\frac{\delta c_{ijkl}}{\rho}, \frac{\delta\rho}{\rho}\Big)_\alpha,$$

with principal symbol $(2\mathrm{i}\tau(x, \xi, e))^{-1} w_{MN;\alpha}(x, \xi, e)$; *see* (3–5).

PROOF. We do the proof for the Kirchhoff approximation using (7–10); for the Born approximation the proof is similar. Because Assumption 4 is satisfied, the projection π_Y of Λ_{MN} into $T^*Y \setminus 0$ is an embedding, and the image is a

coisotropic submanifold of $T^*Y\setminus 0$. Therefore we can apply Lemma 6.1. Formula (6–3) implies that the phase factor $\mathrm{e}^{\mathrm{i}\Phi_{MN}}$ can be written in the form

$$\begin{aligned}\mathrm{e}^{\mathrm{i}\Phi_{MN}(y'_I,x,\eta'_J)} &= \mathrm{e}^{\mathrm{i}(S_{MN}(y'_I,x,\eta'_J,0)+\langle y'_J,\eta'_J\rangle)}\\ &= (2\pi)^{-(n-1-c)}\int \mathrm{e}^{\mathrm{i}(S_{MN}(y'_I,x,\eta'_J,\varepsilon)+\langle y'_J,\eta'_J\rangle-\langle e,\varepsilon\rangle)}\,\mathrm{d}\varepsilon\,\mathrm{d}e;\end{aligned}$$

we define

$$\Psi_{MN}(y',x,e,\eta'_J,\varepsilon) = S_{MN}(y'_I,x,\eta'_J,\varepsilon)+\langle y'_J,\eta'_J\rangle-\langle e,\varepsilon\rangle.$$

Thus the number of phase variables is increased by making use of a stationary phase argument. Let $\mathcal{B}_{MN}(y'_I,x,\eta'_J,\varepsilon)$ be as described. Then we obtain

$$\begin{aligned}G^{\mathrm{refl}}_{MN}(y') = (2\pi)^{-(|J|+n-1-c)/2-(2n-1-c)/2}&\\ \times\int\big((2\pi)^{n/2}2\mathrm{i}\tau(\eta')\mathcal{B}_{MN}(y'_I,x,\eta'_J,\varepsilon)R_{MN}(x,e)+\text{l.o.t.}\big)&\\ \times\mathrm{e}^{\mathrm{i}\Psi_{MN}(y',x,e,\eta'_J,\varepsilon)}\left\|\frac{\partial z_n}{\partial x}\right\|\delta(z_n(x))\,\mathrm{d}\eta'_J\,\mathrm{d}\varepsilon\,\mathrm{d}x\,\mathrm{d}e.&\end{aligned}$$

In this formula, the data are represented as a Fourier integral operator acting on $\|\partial z_n/\partial x\|\delta(z_n(x))$ considered as a function of (x,e). Multiplying by H^{-1}_{MN} gives a pseudodifferential operator of the form described acting on $\|\partial z_n/\partial x\|\delta(z_n(x))$. Thus the result follows. □

Figure 13 illustrates the relation between R_{MN} and the Born approximation.

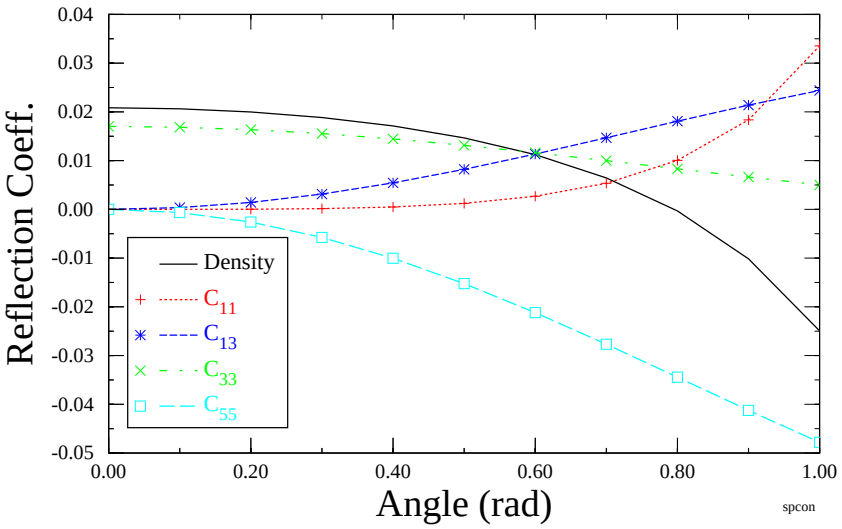


Figure 13. Constituent reflection coefficients linearized in $(\delta c_{ijkl}/\rho,\ \delta\rho/\rho)_\alpha$ for incident qP and scattered qP in a transversely isotropic medium: The relation between the Born and Kirchhoff approximations.

7C. Inversion. The operator H_{MN} is invertible. A choice of phase function and amplitude for its inverse is given by (see Chapter 8 of Treves [102])

$$-\Psi_{MN}(y', x, e, \eta'_J, \varepsilon), \qquad \mathcal{B}_{MN}(y'_I, x, \eta'_J, \varepsilon)^{-1} \left|\det \frac{\partial(y', \eta')}{\partial(y'_I, x, \eta'_J, \varepsilon)}\right|,$$

respectively. Thus microlocally an explicit expression for $r_{MN}(x, e)$ in terms of the data is given by

$$r_{MN}(x,e) = \int \mathcal{B}_{MN}(y'_I, x, \eta'_J, \varepsilon)^{-1} \left|\det \frac{\partial(y', \eta')}{\partial(y'_I, x, \eta'_J, \varepsilon)}\right| \times \mathrm{e}^{-\mathrm{i}\,\Psi_{MN}(y', x, e, \eta'_J, \varepsilon)} d_{MN}(y')\, \mathrm{d}\eta'_J\, \mathrm{d}\varepsilon\, \mathrm{d}y'.$$

Because the function $r_{MN}(x, e)$ is to highest order equal to the product of reflection coefficient and the singular function of the reflector surface, this reconstruction of the function $r_{MN}(x, e)$ leads to the following result for Kirchhoff data.

COROLLARY 7.6. *Suppose that the medium above the reflector is given, and that it satisfies Assumptions* 1, 6, 2, 3, *and* 4. *Then one can reconstruct the position of the interface and the angle dependent reflection coefficient* $R_{\mu\nu}(x, e)$ *on the interface.*

REMARK 7.7. In amplitude-versus-angles analysis it is the reflection coefficient obtained from the inversion that is directly subjected to interpretation. Depending on the (medium) coefficients symmetries, the reflection coefficient may be expanded into trigonometric functions of scattering angle and azimuth (e). (The foundation of such expansions dates back to the work of Shuey [88].) The leading order coefficients in such expansion provide information about the P and S phase velocities just below the reflector.

Such information may be used to infer additional properties of the subsurface. For example, the ratio of P and S phase velocities is nearly independent of pore pressure (or effective stress) and hence renders information about the fluid compressibility (and hence fluid mixture) in a porous formation. The S phase velocity as such will reveal primarily the pore pressure. This separation of information is most easily recovered by crossplotting the P and S phase impedances. The analysis is of key importance in time-lapse seismology with application for example to reservoir monitoring.

8. Annihilators of Seismic Data

The results of the previous subsections may be exploited in the problem of estimating the smooth background medium (or, in the Kirchhoff approximation, the smooth medium parameters above/in between the interfaces). If $n-1-c > 0$ there is a redundancy in the data through the variable e. The redundancy in the data manifests itself as a redundancy in images from these data. If the smooth

medium parameters (above the interface) are correct, then applying the operator H_{MN}^{-1} of Theorem 7.5 to the data results in a reflectivity function $r_{MN}(x,e)$, such that the position of the singularities does not depend on e. Thus we obtain multiple images of the reflectivity parametrized by e such that their singular supports (in x) should agree. This can be used as a criterion to assess the accuracy of the choice of the background medium.

One way to measure this agreement is by taking a derivative with respect to e. We develop this strategy further. The derivative $\partial r_{MN}(x,e)/\partial e$ is one order less singular if $r_{MN}(x,e)$ depends smoothly on e, as in (7–11), than when it does not (for instance a δ function versus its derivative in the Kirchhoff case). Taking also the factor in front of the δ function of r_{MN} into account — see (7–11 — we obtain that to the highest two orders

$$\Big(R_{MN}(x,e)\frac{\partial}{\partial e} - \frac{\partial R_{MN}^{\mathrm{prin}}}{\partial e}(x,e)\Big) r_{MN}(x,e) = 0. \tag{8–1}$$

If $R_{MN}(x,e)$ is nonzero then the lower order terms can be chosen such that this equation is valid to all orders.

Conjugating the differential operator of (8–1) with the invertible Fourier integral operator H_{MN}, we obtain a pseudodifferential operator on $\mathcal{D}'(Y)$. Thus we obtain the following corollary of Theorem 7.5

COROLLARY 8.1. *Let the pseudodifferential operators $W_{MN}(y',D_{y'})$be given by*

$$W_{MN}(y',D_{y'}) = H_{MN}\Big(R_{MN}(x,e)\frac{\partial}{\partial e} - \frac{\partial R_{MN}}{\partial e}(x,e)\Big) H_{MN}^{-1}.$$

Then for Kirchhoff data $d_{MN}(y')$ we have to the highest two orders

$$W_{MN}(y',D_{y'})d_{MN}(y') = 0. \tag{8–2}$$

For values of e where $R_{MN}(x,e) \neq 0$, the operator $W_{MN}(y',D_{y'})$ can be chosen such that (8–2) *is valid to all orders.*

We refer to W_{MN} as an annihilator of the data.

REMARK 8.2. In principle, the annihilators $W_{MN}(y',D_{y'})$ can be used to quantify the agreement between the data and the background medium. Symes [95] discusses such criteria for acoustic media using differentiation with respect to the offset coordinate. Upon introducing the seminorm $\|W_{MN}d_{MN}\|$, an error criterion is obtained and an optimization scheme can be derived to minimize the action of the annihilators and to obtain an improved estimate of the background medium coefficients. Such an optimization scheme requires a "gradient" computation combined with a regularization procedure. Here the "gradient" is viewed as a derivative along a curve in the space of background media. We have adapted the Tikhonov regularization that penalizes roughness in the background medium. The gradient computation makes use of ray perturbation theory. The procedure is known to seismologists as *migration velocity analysis*. An example is presented in Section 9E.

9. The Generalized Radon Transform

Honoring the symplectic geometry presented in Section 6, we extract a generalized Radon transform from the modeling and inversion Fourier integral operators developed in Sections 4 and 7. The generalized Radon transform becomes the basis of a processing procedure. It is developed through several intermediate steps by selecting relevant neighborhoods of the canonical relation Λ_{MN}. These steps enable the use of traveltime in the phase function while restricting the imaging operator to a given value of e.

9A. Diffraction stack. To describe the kernel of the operator $F_{MN;\alpha}$ as an oscillatory integral on a neighborhood of the point on Λ_{MN} parametrized by $(x_0, \hat{\alpha}, \tilde{\alpha}, \tau)$, the minimum number of phase variables is given by the corank of the projection

$$D\pi : T\Lambda_{MN} \to T(T^*Y \times T^*X)$$

at $(x_0, \hat{\alpha}, \tilde{\alpha}, \tau)$, which is here given by

$$\text{corank } D\pi = 1 + \text{corank } \frac{\partial s_N}{\partial \tilde{\alpha}}(x, \tilde{\alpha}) + \text{corank } \frac{\partial r_M}{\partial \hat{\alpha}}(x, \hat{\alpha}).$$

This corank is > 1 when s or r is in a caustic point relative to x. Let

$$\Lambda'_{MN} = \Lambda_{MN} \setminus \big(\text{closed neighborhood of } \{\lambda \in \Lambda_{MN} : \text{corank } D\pi > 1\}\big). \quad (9\text{–}1)$$

The subset Λ'_{MN} can be described by phase functions of the *traveltime form*

$$\tau\,(t - T^{(m)}_{MN}(x, s, r))$$

with the only phase variable τ and where $T^{(m)}_{MN}$ is the value of the time variable in (3–10); see Remark 4.2. The index m labels the branches of the multi-valued traveltime function. Thus the set $\{T^{(m)}_{MN}\}_{m\in M}$ describes the canonical relation (3–10) except for a neighborhood of the subset of the canonical relation where mentioned projection is degenerate. Each $T^{(m)}_{MN}$ is defined on a subset $D^{(m)}$ of $X \times O_s \times O_r$ (possibly dependent on MN). We define $F^{(m)}_{MN;\alpha}$ to be a contribution to $F_{MN;\alpha}$ with phase function given by $\tau(t - T^{(m)}_{MN})$, and symbol $\mathcal{A}^{(m)}_{MN}$ in a suitable class such that on a subset Λ'_{MN} of the canonical relation where the projection is nondegenerate, $F_{MN;\alpha}$ is given microlocally by $\sum_{m\in M} F^{(m)}_{MN;\alpha}$.

Remark 9.1. With the phase function $\tau(t - T^{(m)}_{MN})$, the application of the imaging or inversion operator on the data, up to leading order, amounts to a "diffraction stack", which is for the contribution at the image point x, an integration of the analytic extension of the data over (s, r) subjecting the time to $t = T^{(m)}_{MN}(x, s, r)$. Let s, r be restricted to a hyperplane (the earth's surface) and introduce the half offset H and midpoint m such that $s = m - H$ and $r = m + H$. Assume that $M = N$. Upon composing $T^{(m)}_{NN}$ with this map, we obtain $t^{(m)}_N(x, m, H) = T^{(m)}_{NN}(x, m - H, m + H)$. Let $t^{(m)}_0 = t^{(m)}_N(x, m, H = 0)$ denote the zero-offset traveltime. It is common practice to write $H = h\varpi$ with

$\varpi \in S^{n-2}$. Since $\mathrm{d}t_N^{(m)}/\mathrm{d}h|_{h=0} = 0$, expanding $t_N^{(m)}$ into a Taylor series, and squaring the result leads to the expansion

$$(t_N^{(m)})^2 = (t_0^{(m)})^2 + (2h)^2 \underbrace{\langle \varpi, U\varpi \rangle}_{V^2_{\mathrm{nmo}}(\varpi)} + \dots,$$

in which $U_{ij} = t_0^{(m)} \partial_{H_i} \partial_{H_j} t_N^{(m)}|_{H=0}/4$ is an $(n-1)\times(n-1)$ matrix, which defines the *normal moveout velocity* along a common midpoint line in the direction ϖ [103]. This hyperboloid was the original shape used in the diffraction stack that was organized in (m, h, ϖ) rather than (s, r). In the generalized Radon transform we can think of $\xi/\|\xi\|$ replacing m and e replacing H.

9B. Common image point gathers in scattering angles. In (4–4) we introduced the scattering angles. Here we introduce in addition the *migration dip* ν, defined as the direction of ξ in Λ_{MN} (cf. (3–10)),

$$\nu_{MN}(x, \hat{\alpha}, \tilde{\alpha}) = \frac{\xi}{\|\xi\|} \in S^{n-1}. \tag{9–2}$$

On $D^{(m)}$ there is a map $(x, \hat{\alpha}, \tilde{\alpha}) \mapsto (x, s, r)$. We define $\boldsymbol{e}_{MN}^{(m)} = \boldsymbol{e}_{MN}^{(m)}(x, s, r)$ as the composition of $\boldsymbol{e}_{MN}$ with the inverse of this map. Likewise, we define $\nu_{MN}^{(m)}$. Note that also $T_{MN}^{(m)}$ is the composition of T_{MN} with mentioned inverse, and that we can introduce $w_{MN;\alpha}^{(m)}$ in a similar manner.

In preparation of the generalized Radon transform (GRT) we define the angles imaging operator, $\check{L}$, via a restriction in $F_{MN;\beta}^*$ of the mapping $\boldsymbol{e}_{MN}^{(m)}$ to a prescribed value e; that is, the distribution kernel of each contribution $(F_{MN;\beta}^{(m)})^*$ is multiplied by $\delta(e - \boldsymbol{e}_{MN}^{(m)}(x, s, r))$. (This restriction transfers over to the construction of the left inverse in (4–2).) Invoking the Fourier representation of this δ, the kernel of L follows as (cf. Section 6)

$$\check{L}(x, e, r, s, t) = \sum_{m\in M} (2\pi)^{-(n-1)} \int \overline{\mathcal{A}_{MN}^{(m)}(x, s, r, \tau) w_{MN;\beta}^{(m)}(x, s, r, \tau)} \times \mathrm{e}^{\mathrm{i}\Phi_{MN}^{(m)}(x,e,s,r,t,\varepsilon,\tau)}\, \mathrm{d}\tau \mathrm{d}\varepsilon, \tag{9–3}$$

where $\mathcal{A}_{MN}^{(m)} = \mathcal{B}_{MN}$ with $|J| = 1$ (and $\eta_J' = \tau$) is a symbol for the m-th contribution to $F_{MN;\beta}$, supported on $D^{(m)}$, and

$$\Phi_{MN}^{(m)}(x, e, s, r, t, \varepsilon, \tau) = \tau(T_{MN}^{(m)}(x, s, r) - t) + \langle \varepsilon, e - \boldsymbol{e}_{MN}^{(m)}(x, s, r)\rangle.$$

As before, ε is the cotangent vector corresponding to e.

Let $\psi_L = \psi_L(D_s, D_r, D_t)$ be a pseudodifferential cutoff such that $\psi_L(\sigma, \rho, \tau) = 0$ on a conic neighborhood of $\tau = 0$. Then $\psi_L \check{L}$ is a Fourier integral operator [90] with canonical relation

$$\Lambda_{\check{L}} = \bigcup_{m\in M} \big\{(x, \boldsymbol{e}_{MN}^{(m)}, \xi_{MN}^{(m)}, \varepsilon; s, r, T_{MN}^{(m)}, \sigma_N, \rho_M^{(m)}, \tau) : (x, s, r) \in D^{(m)}, \varepsilon \in \mathbb{R}^{n-1}, \tau \in \mathbb{R}\setminus 0\big\} \subset T^*(X\times E)\setminus 0 \times T^*Y\setminus 0, \tag{9–4}$$

where $\boldsymbol{e}_{MN}^{(m)}$ and $T_{MN}^{(m)}$ are functions of x, s, r and

$$\xi_{MN}^{(m)} = \xi_{MN}^{(m)}(x,s,r,\tau,\varepsilon) = \partial_x \Phi_{MN}^{(m)} = \tau \partial_x T_{MN}^{(m)}(x,s,r) - \langle \varepsilon, \partial_x \boldsymbol{e}_{MN}^{(m)}(x,s,r)\rangle, \tag{9–5}$$

while similar expressions hold for $\sigma_N^{(m)}$ and $\rho_M^{(m)}$.

Effectively, for each x we select a different subset of the data. This is fundamentally different from the common offset Kirchhoff integral approach which amounts to a straightforward restriction in the acquisition manifold.

9C. Artifacts of type II. With the choice (4–4) for $\boldsymbol{e}$, the following assumption is implied. However, for other choices of $\boldsymbol{e}$ it needs to be verified.

ASSUMPTION 7. *Consider the mapping*

$$\Xi : \Lambda_{MN} \to T^*X \setminus 0 \times E, \;\; \lambda(x,\hat{\alpha},\tilde{\alpha},\tau) \mapsto (x,\xi,e), \;\; \textit{with } \xi = \|\xi\|\nu_{MN}.$$

Composing this mapping with the inverse of the mentioned map $(x,\hat{\alpha},\tilde{\alpha}) \mapsto (x,s,r)$, *yields per branch* m *a mapping* $\Xi^{(m)}$ *from* (x,s,r,τ) *to an element of* $T^*X \setminus 0 \times E$. $\Xi^{(m)}$ *is locally diffeomorphic, i.e.*

$$\operatorname{rank} \frac{\partial(\xi_{MN}^{(m)}, \boldsymbol{e}_{MN}^{(m)})}{\partial(s,r,\tau)}\bigg|_{\varepsilon=0} \quad \textit{is maximal, at given } x \textit{ and branch } m.$$

Let d_{MN} be the Born modeled data in accordance with Theorem 3.2. To reveal any artifacts generated by $\check{L}$, i.e. singularities in $\check{L}d_{MN}$ at positions not corresponding to an element of $\mathrm{WF}(g_\alpha)$, we consider the composition $\check{L}\, F_{MN;\alpha}$. With Assumption 7 this composition is equal to the sum of a smooth e-family of pseudodifferential operators and, in general, a non-microlocal operator. The wavefront set of the non-microlocal operator contains no elements with $\varepsilon = 0$ [90, Theorem 6.1]. The origin of contributions from $\varepsilon \neq 0$ is illustrated in Figure 14. A filter needs to be applied to remove contributions from $|\varepsilon| \geq \varepsilon_0 > 0$; we define the generalized Radon transform L as the Fourier integral operator with canonical relation U_L — also denoted as Λ_L — to be a *neighborhood* of $\Lambda_{\check{L}} \cap \{\varepsilon = 0\}$, which derives from the left inverse (4–2).

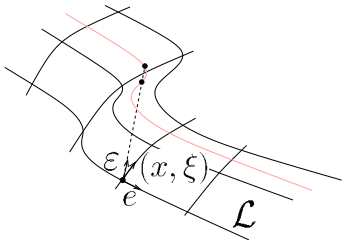


Figure 14. The origin of artifacts generated by the generalized Radon transform. (Inside the $T^*Y \setminus 0$ box of Figure 11.)

The artifacts in the compose of the canonical relation of $\check{L}$ with that of $F_{MN;\alpha}$ can be evaluated by solving the system of equations

$$r = r_M(x, \hat{\alpha}), \tag{9–6}$$

$$s = s_N(x, \tilde{\alpha}), \tag{9–7}$$

$$T^{(m)}_{MN}(z, s, r) = T_M(x, \hat{\alpha}) + T_N(x, \tilde{\alpha}), \tag{9–8}$$

$$\rho^{(m)}_M(z, s, r, \tau, \varepsilon) = -\tau p^r_M(x, \hat{\alpha}), \tag{9–9}$$

$$\sigma^{(m)}_N(z, s, r, \tau, \varepsilon) = -\tau p^s_N(x, \tilde{\alpha}). \tag{9–10}$$

(The frequency is preserved.) Equations (9–6)–(9–8) imply that the image point z must lie on the isochrone determined by (x, s, r). Equations (9–9) and (9–10) enforce a match of slopes (apparent in the appropriate "slant stacks") in the measurement process:

$$-\tau\partial_r T^{(m)}_{MN}(z, s, r) + \big\langle \varepsilon, \partial_r \boldsymbol{e}^{(m)}_{MN}(z, s, r)\big\rangle = -\tau p^r_M(x, \hat{\alpha}), \tag{9–11}$$

$$-\tau\partial_s T^{(m)}_{MN}(z, s, r) + \big\langle \varepsilon, \partial_s \boldsymbol{e}^{(m)}_{MN}(z, s, r)\big\rangle = -\tau p^s_N(x, \tilde{\alpha}). \tag{9–12}$$

For $\varepsilon \neq 0$ the take-off angles of the pairs of rays at (r, s) may be distinct. Equations (9–11) and (9–12) imply the matrix compatibility relation (upon eliminating ε/τ)

$$\begin{aligned}\big(\partial_r \boldsymbol{e}^{(m)}_{MN}(z, s, r)\big)^{-1}\big(p^r_M(x, \hat{\alpha}) - \partial_r T^{(m)}_{MN}(z, s, r)\big)\\ = \big(\partial_s \boldsymbol{e}^{(m)}_{MN}(z, s, r)\big)^{-1}\big(p^s_N(x, \tilde{\alpha}) - \partial_s T^{(m)}_{MN}(z, s, r)\big).\end{aligned} \tag{9–13}$$

Those summarize the geometrical composition equations determining the artifacts: For each $(x, \hat{\alpha}, \tilde{\alpha}) \in K$ solve the $3n - 2$ equations (9–6)–(9–8), (9–13) for the $3n - 2$ unknowns (z, s, r). (From (9–11) we then obtain ε/τ hence ε.)

Remark 9.2. The generalized Radon transform reconstructs a distribution in $\mathcal{E}'(X)$ smoothly indexed by $e \in E$. Thus, we can carry out the composition $(N_{MN})^{-1}_{\alpha\beta} L$ (no summation over M, N) as in (4–2) to yield the generalized Radon transform *inversion* [33]. Likewise, we can carry out a composition with the modeling operator $F_{MN;\alpha}$ (or $H_{MN;\alpha}$).

9D. Filters. In general, filters need to be applied to the common image point gathers to remove the artifacts of type II. At the same, the generalized Radon transform is a transformation of data as a function of the $(2n - 1)$ variables (s, r, t) to a distribution of the $(2n - 1)$ variables (x, e). After the removal of artifacts, the alignment in the e directions ($\varepsilon = 0$) of the singular support of this distribution represented in the common image point gathers simplifies the task of denoising the final image of the medium perturbation. Treating the artifacts as noise as well, a joint approach based upon non-adaptive wavelet thresholding [41]

applies; an analysis of subbands will then aid in the suppression of the artifacts, associated with $\varepsilon \neq 0$.

The removal of the illumination effects (which can be written as the action of a pseudodifferential operator on g_α) is also amenable to the use of wavelets. A possible approach is matching pursuit [72].

9E. An example: Estimating the background model. To get to an example, a discretization both of c_{ijkl}, ρ (background) and of $\delta c_{ijkl}, \delta\rho$ (singular perturbation) has to be chosen. We have represented the background by *cubic splines* the smoothness of which aid in the numerical computation of geometrical spreading.

In Figure 16 we illustrate the performance of an optimization minimizing the annihilators developed with the aid of the generalized Radon transform. In the example, the medium is isotropic and hence we have only two parameters i.e. the P-wave and S-wave speeds. We reconstruct a smooth Gaussian lens (in P-wave speed) in six iterations (the intermediate model on the left and some ray geometry originating at a scattering point on the reflector on the right), starting from an initial model that did not contain a lens whence in the initial model no caustics were formed. Note the alignment in the final common image point gather reflecting the reduncancy in the data. In the final model caustics do occur, as illustrated in Figure 15.

9F. An example: reconstructing the singular perturbation. From a geological point of view, it is attractive to represent the singular perturbation by *fractional splines*.

We apply the generalized Radon transform to multicomponent ocean bottom seismic data acquired over the Valhall field in the Norwegian sector of the North Sea to obtain common image point gathers representing $r_{MN}(x, e)$. In this region, it is believed that the presence of gas in the overburden yields lenses that cause caustics to form. An isotropic elastic background model was obtained;

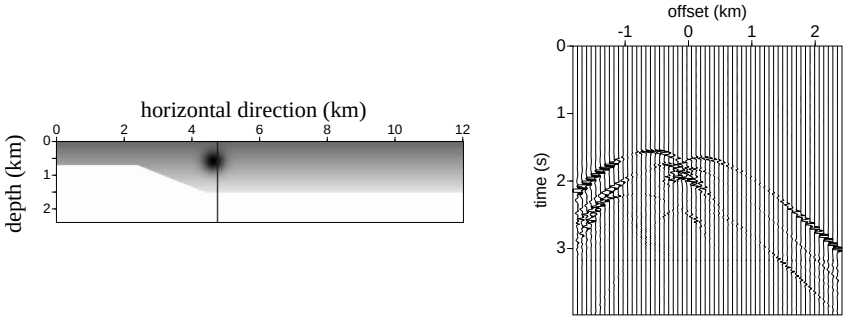


Figure 15. The Gaussian lens model with reflecting surface (left) and the modeled data (vertical component) for a given source position, in an experiment of the type in Figure 1 (right).

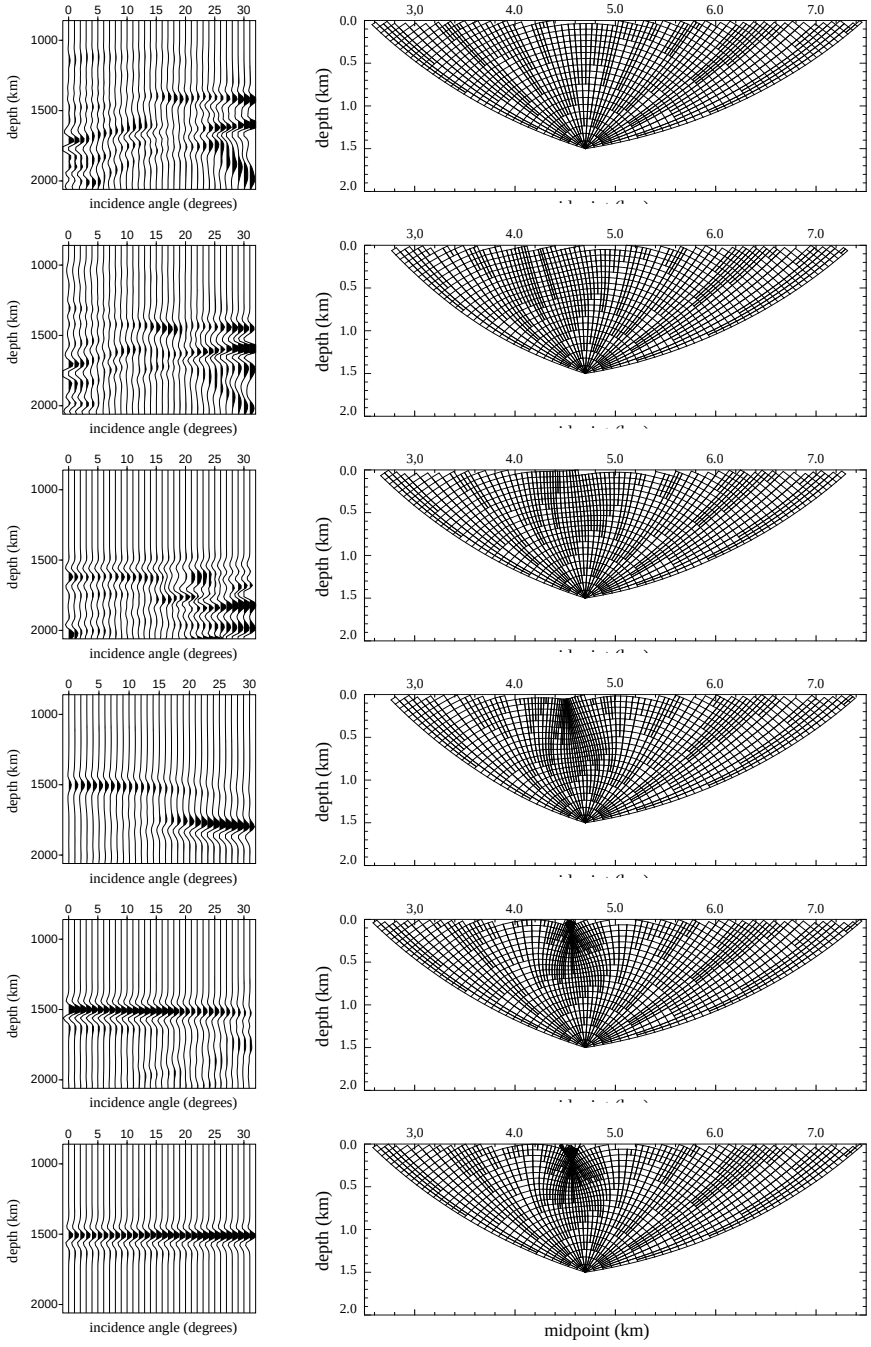


Figure 16. Successive (P, P) common image point gathers (left) and characteristics associated with a particular Green's function originating at the reflector (right) in the process of minimizing the annihilator. A smooth Gaussian lens is being reconstructed (from [22]).

the P-wave speed is shown in Figure 17 (top). Note the presence of the lenses. We will illustrate the data, the action of the generalized Radon transform, the common image point gathers, and images of particular medium parameter combinations in the slice depicted in Figure 17 (bottom). The common image point gathers will be restricted to x lying on the black vertical line and e being the scattering angle. A fan of characteristics originating at the (dark) reservoir layer is shown in Figure 18, which illustrates the formation of caustics in G_M ($M = \mathrm{P}$).

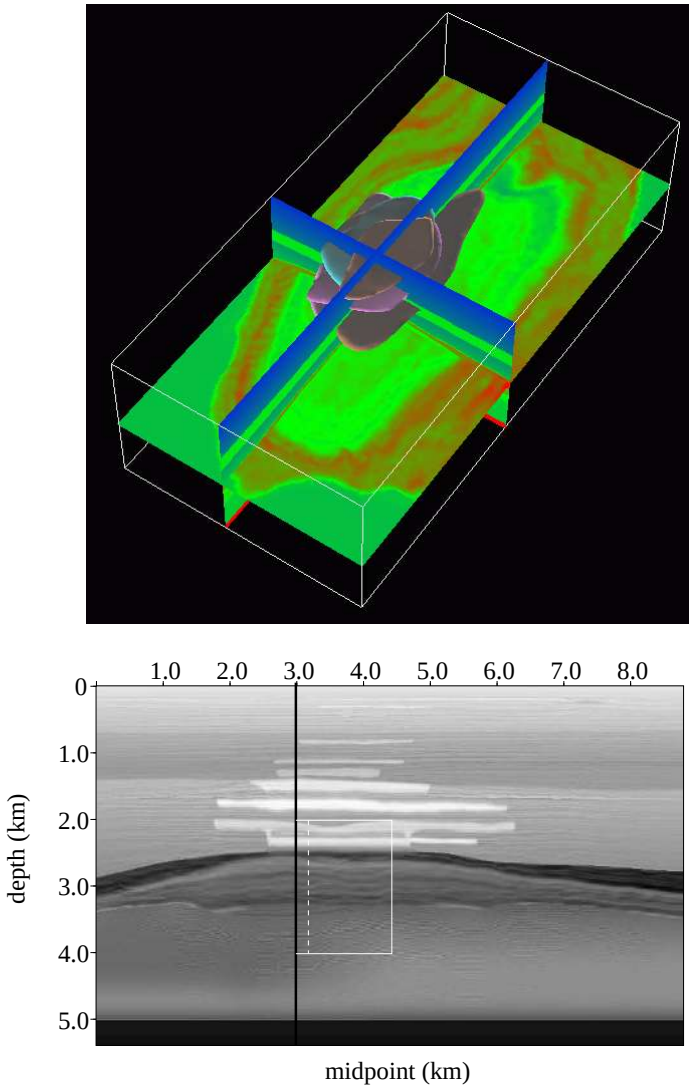


Figure 17. The c_P model in perspective (top) and a slice (bottom). Note the stack of lenses that contain gas. The dark layer represents the reservoir.

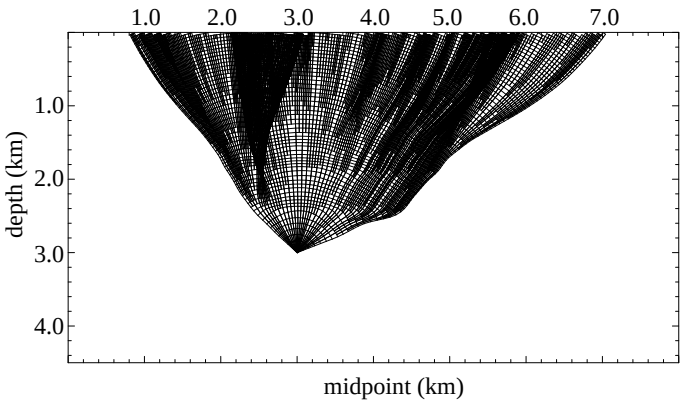


Figure 18. Characteristics in the model of Figure 17 (bottom).

The horizontal component of the data, for given source position, is shown in Figure 19. These illustrations can be thought of as regularizations of the data and medium perturbation distributions.

For the pairs (P,P) and (P,S) the common image point gathers are shown in Figure 20 (left, middle). Observe the alignment of a sequence of (singular) supports at the arrow with scattering angle (here converted to incidence angle) at reservoir depth. The $r_{MN}(x,e)$ for (N,M) =(P,S) corresponding with the

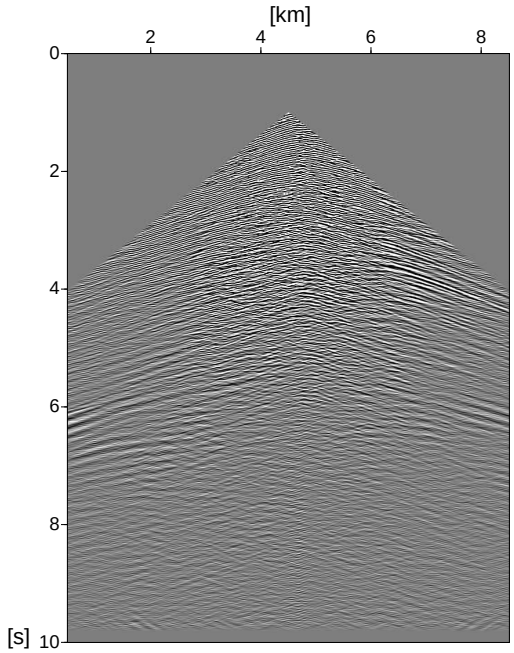


Figure 19. Data: fixed source position, horizontal component (parallel to acquisition surface).

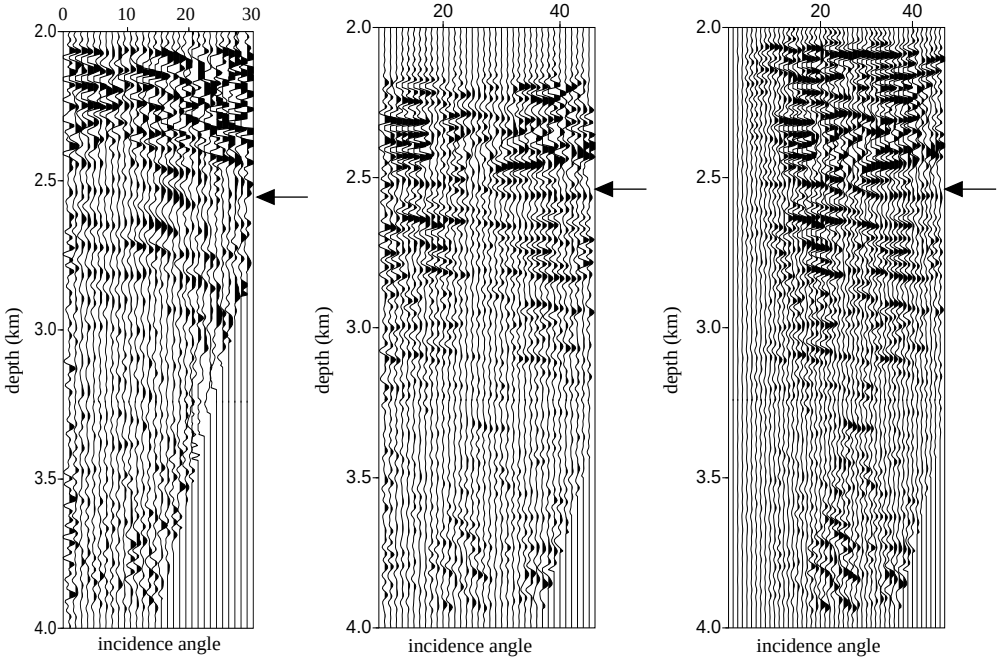


Figure 20. Common image point gathers with x on the black line in Figure 17 (bottom). In the framework of the Born approximation for (P, P) (left) and (P, S) (middle) and in the framework of the Kirchhoff approximation for (P, S) (right); from [21].

middle common image point gather is shown on the right. Note the change of sign of the amplitude at the key reflector, which is an indication of the presence of anisotropy.

In Figure 21, the images of P- and S-phase impedances inside the white box of Figure 17 (bottom) are presented, and compared with a standard seismic image from (P, S) in Figure 22. The use of these images combined was addressed in Remark 7.7.

10. Wavefield "Continuation"

In general, $F_{MN;\alpha}F^*_{MN;\beta}$ cannot be a pseudodifferential operator. However, exploiting carefully the redundancies in the data, L^*L (note that L is an imaging operator itself) attains, under certain conditions, pseudodifferential properties. This observation is at the basis of seismic wavefield "continuation".

In this section, we use a simplified notation: We suppress the subscripts MN in the operators F (originally $F_{MN;\alpha}$) and N (originally $N_{MN;\alpha\beta}$). Also, we

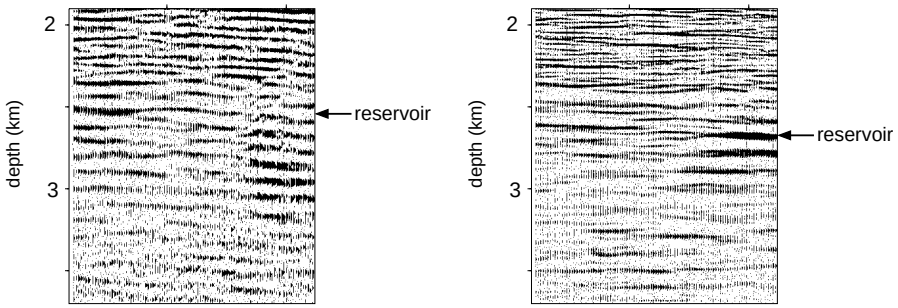


Figure 21. High resolution images of impedance revealing sedimentary layers and faults, from: (P, P) scattered waves (left) and (P, S) scattered waves (right).

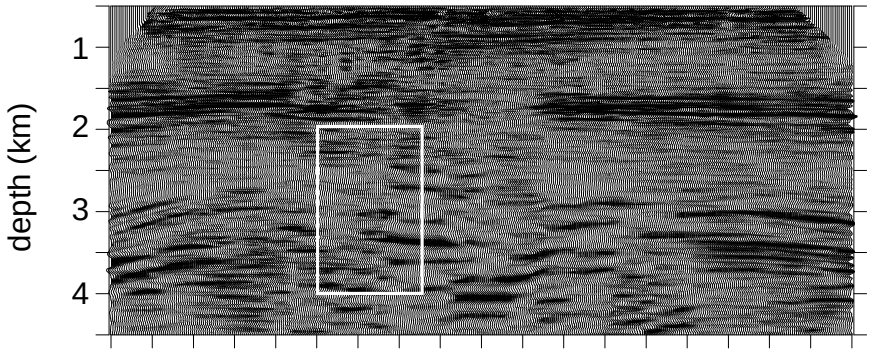


Figure 22. A (P, P) image from hydrophone data obtained with standard seismic processing (transversely isotropic) corresponding with Figure 17 (bottom). The white box corresponds with the image of Figure 21 (left).

re-introduce y replacing y' to enable with the same notation a further restriction to acquisition submanifolds.

10A. Modeling restricted to an acquisition submanifold. Single reflection seismic wavefield continuation aims at generating from reflection data — through the canonical relation (3–10) — associated with $T^*X \setminus 0 \times E_{\rm i}$, in which $E_{\rm i}$ is an $(n-1)$-dimensional open neighborhood of e say, reflection data associated with $T^*X \setminus 0 \times E_{\rm o}$, in which $E_{\rm o} \supset E_{\rm i}$. Such continuation, within the acquisition manifold Y, is accomplished through the composition of Fourier integral operators generating an intermediate image of δc_{ijkl} (resp. $\delta\rho$). In the previous section, we analyzed a Fourier integral operator, the generalized Radon transform, that generates (linear combinations of) δc_{ijkl} (resp. $\delta\rho$) from data on $T^*X \setminus 0 \times E_{\rm i}$. In this section we consider, once data are modeled from δc_{ijkl} (resp. $\delta\rho$) as in Theorem 3.2, the restriction to an acquisition submanifold parametrized by (x, ξ, e) through the canonical relation (3–10), such that $e \in E_{\rm o}$. In the following subsections, the restriction, modeling and generalized Radon transform imaging

operators will be composed to yield the continuation. In this composition, the background coefficients (ρ, c_{ijkl}) are used, but, naturally, δc_{ijkl} (resp. $\delta\rho$) does not appear. The continuation is illustrated in Figure 23.

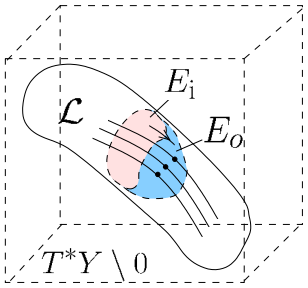


Figure 23. Continuation and characteristic strips. (Inside the $T^*Y \setminus 0$ box of Figure 11.)

Our starting point is the situation where $c = 0$. We consider the restriction to an acquisition submanifold by increasing c to $\tilde{c}$, say. This submanifold is written as $Y^{\tilde{c}} = \Sigma^{\tilde{c}} \times (0,T)$, with $\Sigma^{\tilde{c}} \xrightarrow{\imath} O_s \times O_r$ representing an embedded manifold of codimension $c = \tilde{c} \geq 0$. We reconsider Assumption 3. Let

$$(y'_1, \ldots, y'_{2n-2-c}, \underbrace{y'_{2n-1-c}}_{t}, y''_{2n-c}, \ldots, y''_{2n-1})$$

denote a local coordinate system on Y such that Σ^c is given by $(y''_{2n-c}, \ldots, y''_{2n-1})$ $= (0, \ldots, 0)$ locally. In the applications under consideration, we order the coordinates such that $y'_{2n-1-c} = t$.

Assumption 8. *The intersection of Λ_{MN} with the manifold $y'' = 0$ is transversal, i.e.*

$$\frac{\partial y''}{\partial(x, \hat{\alpha}, \tilde{\alpha}, \tau)} \quad \textit{has maximal rank.}$$

Applying [43, Thm. 4.2.2] to the pair F and the restriction $\mathcal{R} = \mathcal{R}^c$ from $O_s \times O_r \to \Sigma^c$ with Assumption 8 implies that $\mathcal{R}^c F$ is a Fourier integral operator of order $(n-1+c)/4$ with canonical relation

$$\begin{aligned}\Lambda^c_{MN} = \big\{&(y', t, \eta', \tau; x, \xi) : \exists\ y', y'', \eta', \eta'' \text{ such that} \\ &(y', y'', \eta', \eta''; x, \xi) \in \Lambda_{MN} \text{ and } y'' = 0\big\} \subset\ T^*Y^c \setminus 0 \times T^*X \setminus 0. \quad (10\text{–}1)\end{aligned}$$

(Here $\Lambda^0_{MN} = \Lambda_{MN}$.) We mention two examples: Zero offset, where $c = n-1$ and $\Sigma^c := \Sigma^{\mathrm{ZO}} \subset \mathrm{diag}(\partial X)$ (subject to the $n-1$ constraints $r = s$ when $\arccos(\hat{\alpha} \cdot \tilde{\alpha}) = 0$ and e_{o} at x follows from (4–4)), and common azimuth (CA), where $c = 1$ and $\Sigma^c := \Sigma^{\mathrm{CA}}$ subject to one constraint typically of the form that the $(n-1)$st coordinate in $r - s$ is set to zero, while $E_{\mathrm{o}} \ni e$ at x follows from the mapping $\boldsymbol{e}^{(m)}$. We set $Y^{\mathrm{ZO}} = \Sigma^{\mathrm{ZO}} \times (0,T)$ and $Y^{\mathrm{CA}} = \Sigma^{\mathrm{CA}} \times (0,T)$.

The restriction to acquisition submanifolds is placed in the context of inversion in [80].

10B. Continuation. We analyze the continuation of multiple finite-offset seismic data — in the absence of knowledge about the singular medium perturbation. The compose FL is a well-defined operator $\mathcal{D}'(Y) \to \mathcal{D}'(Y)$. Its wavefront set is contained in the composition of the wavefront sets of F and L [43, Thm. 1.3.7], hence in the composition of canonical relations,

$$\begin{aligned}\Lambda_{MN}\circ\Lambda'_L = \big\{&(s_2,r_2,t_2,\sigma_2,\rho_2,\tau_2;s_1,r_1,t_1,\sigma_1,\rho_1,\tau_1) : \exists\, x,\xi,\varepsilon \text{ such that}\\ &(s_2,r_2,t_2,\sigma_2,\rho_2,\tau_2;x,\xi)\in\Lambda_{MN} \text{ and } (x,e,\xi,\varepsilon;s_1,r_1,t_1,\sigma_1,\rho_1,\tau_1)\in\Lambda_L\big\}\\ &\subset\ T^*Y\setminus 0\times T^*Y\setminus 0, \quad (10\text{–}2)\end{aligned}$$

with $\Lambda'_L = \{(x,\xi;s,r,t,\sigma,\rho,\tau) : \exists x,\xi,\varepsilon \text{ such that } (x,e,\xi,\varepsilon;s,r,t,\sigma,\rho,\tau)\in\Lambda_L\}$. (To avoid the introduction of Λ'_L, we could consider the composition of H_{MN} with L; see Theorem 7.5.) Whether the compose is a Fourier integral operator is yet to be investigated.

Using the parametrizations of Λ_{MN} in Remark 3.3 and Λ_L in (9–4), the compose (10–2) can be evaluated through solving a system of equations, the first n being trivial fixing the scattering point $x_0 = x$, the second n equating the cotangent vectors

$$\underbrace{\tau_2\partial_x T_{MN}(x_0,\hat\alpha,\tilde\alpha)}_{\boldsymbol{\xi}(x,\hat\alpha,\tilde\alpha,\tau_2)} = \underbrace{\tau_1\partial_x T^{(m)}_{MN}(x,s,r) - \langle\varepsilon,\partial_x \boldsymbol{e}^{(m)}_{MN}(x,s,r)\rangle}_{\boldsymbol{\xi}^{(m)}_{MN}(x,s,r,\tau_1,\varepsilon)}. \qquad (10\text{–}3)$$

Given $\boldsymbol{e}(x,\hat\alpha,\tilde\alpha) = e$ ($n-1$ constraints) these constitute n equations with the $2n-1$ unknowns $(\hat\alpha,\tilde\alpha,\tau_2)$. (On $D^{(m)}$ the constraints on e can be invoked on s,r instead, namely, via the inverse of the map $(x,\hat\alpha,\tilde\alpha)\mapsto(x,s,r)$ as before.)

LEMMA 10.1. *With Assumptions* 4 *and* 7 *the composition FL yields a smooth family of Fourier integral operators parametrized by e. Their canonical relations are given by*

$$\Lambda_C = \Lambda_{MN}\circ U_L = \{(s_2,r_2,t_2,\sigma_2,\rho_2,\tau_2;s_1,r_1,t_1,\sigma_1,\rho_1,\tau_1)\}$$

parametrized by $(x_0,\hat\alpha,s_1,\tau_1,\varepsilon)$, where upon substituting $x = x_0$ and once r_1 is obtained from s_1 through the value e of $\boldsymbol{e}^{(m)}_{MN}$ (which mapping is defined below equation (4–4)*), $(s_1,r_1,t_1,\sigma_1,\rho_1)$ are given in* (9–4)*, and, given $(\hat\alpha,\varepsilon)$, $(s_2,r_2,t_2,\sigma_2,\rho_2)$ are given in Theorem* 3.2 *in which $(\tilde\alpha,\tau_2)$ are obtained by solving* (10–3).

PROOF. First we extend the operator F to act on distributions in $\mathcal{E}'(X\times E)$ by assuming that the action does not depend on $e\in E$. The calculus of Fourier integral operators gives sufficient conditions that the composition of two Fourier integral operators, here F and L, is again a Fourier integral operator. The essential condition is that the composition of canonical relations is transversal,

i.e. that $\mathcal{L} = \Lambda_{MN} \times U_L$ and $\mathcal{M} = T^*Y \setminus 0 \times \mathrm{diag}(T^*(X \times E) \setminus 0) \times T^*Y \setminus 0$ intersect transversally. We have

$$\begin{array}{ccccc} & & \Lambda_{MN} & & U_L & & \\ & \swarrow & & \searrow & \swarrow & \searrow & \\ T^*Y \setminus 0 & & & T^*X \setminus 0\ (\times E) & & T^*Y \setminus 0 \end{array} \tag{10–4}$$

where the inner two projections are submersions.

On the other hand, in a neighborhood of a point in Λ_{MN} given by (9–4), Λ_{MN} can be parametrized as in Λ'_{MN}. Using this parametrization one finds that the composition of Λ_{MN} and Λ_L is transversal if and only if the matrix

$$\frac{\partial}{\partial(s, r, \hat{\alpha}, \tilde{\alpha}, \tau_2, \varepsilon, \tau_1)} \Big(\boldsymbol{\xi}(x, \hat{\alpha}, \tilde{\alpha}, \tau_2) - \boldsymbol{\xi}^{(m)}_{MN}(x, s, r, \tau_1, \varepsilon) \Big)$$

has maximal rank (cf. (10–3)). This follows, for example, just from the $\boldsymbol{\xi}$ contribution in view of the submersivity of the projection $\pi_X : \Lambda_{MN} \to T^*X \setminus 0$. However, it follows also from the $\boldsymbol{\xi}^{(m)}_{MN}$ contribution: Parametrizing Λ_L by (x, ξ, ε) and restricting Λ_L to U_L further to $\varepsilon = 0$, results in a parametrization in terms of (x, ξ) (with the artifacts filtered out). Then $\boldsymbol{\xi}^{(m)}_{MN}$ becomes ξ and it follows that the composition of Λ_{MN} and U_L is transversal if and only if

$$\text{rank } \frac{\partial}{\partial(\xi, \hat{\alpha}, \tilde{\alpha}, \tau_2)} \big(\boldsymbol{\xi}(x, \hat{\alpha}, \tilde{\alpha}, \tau_2) - \xi \big) \quad \text{is maximal.}$$

This is indeed the case. □

Subjecting the operator F in the composition to the constraint that $\boldsymbol{e}$ (cf. (4–4)) attains a prescribed value, the parameter $\hat{\alpha}$ in the lemma will be eliminated.

Remark 10.2. Following seismological convention, we have used the terminology wavefield continuation. In fact, this is continuation in the context of continuation theorems also. We consider the continuation of the wavefield in the acquisition manifold from $T^*X \setminus 0 \times E_{\mathrm{i}}$ to $T^*X \setminus 0 \times E_{\mathrm{o}}$. This continuation is unique in the sense that $FLd = 0$ implies $F^*FLd = 0$ and, since $F^*F = N$ is strictly elliptic and pseudodifferential, then $Ld = 0$ so that the image of $(\delta c_{ijkl}, \delta\rho)$ vanishes. In the single scattering approximation this implies that $d = F(\delta c_{ijkl}, \delta\rho) = 0$, all modulo smoothing contributions.

Remark 10.3. The subject of data regularization is the transformation of measured reflection data, sampled in accordance with the actual acquisition, to data associated with a regular sampling of the acquisition manifold Y. In our approach the operator $\mathcal{R}^c F \int_{E_{\mathrm{i}}} \mathrm{d}e \langle N^{-1} \rangle L$ replaces the forward interpolation operator in the usual regularization procedures.

10C. Transformation to common azimuth: Azimuth MoveOut. Azimuth MoveOut [13] (AMO) is the process following composing $\mathcal{R}^1 = \mathcal{R}^1_{\rm CA}$ restricting Y to $Y^{\rm CA}$ with modeling operator F with the imaging generalized Radon transform L centered at a given value of e (conventionally for given value of offset $r-s$); the sing supp of the Lagrangian-distribution kernel of the resulting operator is what seismologists call the *AMO impulse response*. The composition $F\,L$ has been addressed in Lemma 10.1. The general restriction has been addressed in Section 10A. Here we combine these results in the following

THEOREM 10.4. *With Assumptions* 4, 7 *and* 8 *with* $Y^c = Y^{\rm CA}$, *the composition* $\mathcal{R}^1_{\rm CA} F\,L$ *yields a smooth family of Fourier integral operators parametrized by* e. *The resulting operator is called Azimuth MoveOut*.

The following Bolker-like condition ensures that the restriction to common azimuth is "image preserving". Let $\Lambda^{\rm CA}_{MN}$ denote the canonical relation of $\mathcal{R}^1_{\rm CA}F$ in accordance with the analysis of Section 10A,

ASSUMPTION 9. *The projection*

$$\pi_{Y^{\rm CA}} \;:\; \Lambda^{\rm CA}_{MN} \to T^*Y^{\rm CA} \setminus 0$$

is an embedding.

This assumption is most easily verified whether an element in $T^*Y^{\rm CA}\setminus 0$ uniquely determines an element in $T^*X \setminus 0$ smoothly given the background medium. Using "all" the data (when available), integration over the $(n-1)$ dimensional e removes the artifacts under the Bolker condition, Assumption 4: We obtain the transformation to common azimuth (TCA)

COROLLARY 10.5. *Let* $\langle N^{-1}\rangle$ *denote the regularized inverse of the normal operator in Theorem* 4.3. *With Assumptions* 2, 3, 4 *and* 8 (*with* $\Sigma^c = \Sigma^{\rm CA}$), *the composition* $\mathcal{R}^1_{\rm CA}F\langle N^{-1}\rangle F^* = \int {\rm d}e \mathcal{R}^1_{\rm CA}F\langle N^{-1}\rangle L$ *is a Fourier integral operator*, $\mathcal{D}'(Y) \to \mathcal{D}'(Y^{\rm CA})$. *With Assumption* 9 *the reduced dataset generates the same image as the original dataset*.

The proof follows that of Theorem 4.3 closely (see [91, Theorem 4.5]).

11. Sampling Canonical Relations: Quasi-Monte Carlo Integration Methods

The canonical relations of the modeling and imaging operators can be optimally sampled through their parametrizations. Here we consider the parametrization in $(x,\hat\alpha,\tilde\alpha,\tau)$ and the parametrization in (x,ξ,e). In seismic experiments, time is evenly sampled and hence, as far as τ is concerned, Nyquist's criterion (Shannon's law) applies. As far as $(\hat\alpha,\tilde\alpha)$ is concerned, we discuss a sampling approach based upon quasi-Monte Carlo methods (see De Hoop and Spencer [37]).

For quasi-Monte Carlo methods we refer here to one original paper by Niederreiter [78]; for a comprehensive treatment of their foundations, see Hammersley and Handscomb [55]. We also apply more recent work by Wözniakowski [112].

11A. The notion of discrepancy. The basic idea underlying Monte Carlo integration is straightforward. The example often quoted in numerical textbooks (e.g., Press *et al.* [81]) is that of determining the volume of a general region E contained in the s-dimensional unit hypercube I^s with $I = [0, 1)$. If N points are chosen at random over the unit hypercube then the volume of $E \subset I^s$, $V(E)$ say, is given by

$$V(E) \approx \frac{1}{N} \sum_{i=1}^{N} c_E(x_i), \tag{11–1}$$

where $c_E(x_i)$ is the characteristic function that takes the value 0 if the point x_i is outside E, 1 otherwise. In other words, the volume is computed by simply counting the number of points in I^s that fall within E and dividing by the total number of points.

Likewise, the integral $I_E[f]$ of an integrable function f over E is approximated by the mean

$$I_E[f] = \int_{I^s} f(x) c_E(x)\, \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) c_E(x_i). \tag{11–2}$$

From the central limit theorem it can be deduced that the integration error arising from using Eq.(11–2) is Gaussian-like distributed and its expected value is $\mathcal{O}(N^{-1/2})$. The attractive feature of this result is that the order of the error is independent of the dimension s of the problem and, hence, Monte Carlo integration methods become increasingly favorable for higher-dimensional problems.

Monte Carlo methods work as well as they do, because randomly chosen points in s dimensions sample the s-dimensional unit hypercube I^s "fairly". This property is not uniquely confined to purely random numbers. Any fair, deterministic, distribution will suffice and may be superior. Fairness is here defined based on a deterministic measure of the difference between an estimate of the volume of a region K resulting from the use of the N point samples $x_i, i = 1, 2, \ldots, N$ and the true volume, $V(K)$ say. We assume that K is a Cartesian region contained in I^s, i.e., $K = K_1 \times K_2 \times \cdots \times K_s$ with $K_n = [0, k^n)$, $0 \le k^n \le 1$, $n = 1, 2, \ldots, s$. Let

$$R_N(k; x_1, x_2, \ldots, x_N) = \left| \frac{1}{N} \sum_{i=1}^{N} c_K(x_i) - V(K) \right|, \quad k = \left\{ k^1, k^2, \ldots, k^s \right\}. \tag{11–3}$$

Given $x_i, i = 1, 2, \ldots, N$, then the *discrepancy* D_N is defined as the supremum,

$$D_N = \sup_{k \in I^s} R_N(k; x_1, x_2, \ldots, x_N),$$

or root-mean-square average or any equivalent measure, of R_N over all $K \subset I^s$ i.e. $k \in I^s$. For example, Wözniakowski [112] uses the L^2 discrepancy T_N of a set of points $x_i, i = 1, 2, \ldots, N$,

$$T_N^2 = \int_{I^s} R_N^2(k; x_1, x_2, \ldots, x_N)\, \mathrm{d}k.$$

Finite sets of points with low discrepancies provide valid approximations to a uniform distribution of points.

It is possible to express the error bounds on the integral of f over E in terms of the discrepancy of the point set $x_i, i = 1, 2, \ldots, N$ in I^s and, for example, the Hardy–Krause variation of f on I^s. An error bound can be obtained in which the influence of the regularity of the integrand has been separated from the influence of the uniformity of the distribution of nodes. Hence, the desirability of sampling based on a set of points or nodes with low discrepancy, to give an accurate estimate of the integral.

There exist two approaches to low-discrepancy sets:

(i) Given N, find N points in I^s with small discrepancy D_N (*low-discrepancy point set*).
(ii) Find a set of N points in I^s, such that the first M points of the sequence show low discrepancy D_M for any $M \leq N$ (*low-discrepancy sequence*).

Point sets of dimension s can be derived from sequences of dimension $s-1$ by a method described by Neiderreiter [79]: For $s \geq 2$, let

$$x_i' = \{x_i^1, \ldots, x_i^{s-1}\} \in I^{s-1}, \qquad i = 1, 2, \ldots, N \tag{11–4}$$

be a low-discrepancy sequence in I^{s-1}. Let D_M' be the discrepancy of the first $M \leq N$ terms of the sequence. Then, for given N, put

$$x_i = \left\{ \frac{i-1}{N}, x_i' \right\} \in I^s, \quad i = 1, 2, \ldots, N. \tag{11–5}$$

Niederreiter [79] has shown that the discrepancy D_N of these points satisfies

$$N\, D_N \leq \max_{1 \leq M \leq N} M\, D_M' + 1, \tag{11–6}$$

so that they form a low-discrepancy point set.

11B. Halton sequences and Hammersley point sets. The discrepancy in L^2 and other norms has been extensively studied, and relations with number theory have been established. Halton [54] was the first to demonstrate that it is possible to construct a sequence of points with discrepancy of order $D_N = \mathcal{O}(N^{-1}[\log N]^s)$. These sequences are now known as Halton sequences. If x_i' is a point in such a Halton sequence of N points in I^{s-1}, then

$$x_i = \left\{ \frac{i-1}{N}, x_i' \right\} \quad \text{or} \quad \mathbf{1} - \left\{ \frac{i-1}{N}, x_i' \right\}, \quad i = 1, 2, \ldots, N$$

is a Hammersley point set in s-dimensional space, with discrepancy $D_N = \mathcal{O}(N^{-1}[\log N]^{s-1})$ (see Eq.(11–6)) . In fact, upon "shifting" the Hammersley points, the L^2 discrepancy can be minimized to yield $T_N = \mathcal{O}(N^{-1}[\log N]^{(s-1)/2})$, which result is optimal (Wözniakowski [112]). The convergence rate of the summation replacing the integration is better for the Hammersley point set than for a set of randomly distributed points.

Several methods for constructing Halton sequences and Hammersley point sets are referred to in Niederreiter [79]. All those methods rely on expansions of integers, to different bases for each of the s coordinates. We will give the construction involving expansions of integers in prime number bases; this construction has been reviewed and exploited by Wözniakowski [112].

Consider the creation of the j'th point in a Halton sequence in an $(s-1)$-dimensional space. Let the first $s-1$ prime numbers be denoted $p_1, p_2, \ldots, p_{s-1}$. Expand the integer j as a power series in each of these $s-1$ prime numbers:

$$\forall_{m\in\{1,\ldots,s-1\}} : \; j = \sum_{\mu=0}^{\lceil \log_{p_m} j \rceil} a_\mu \, (p_m)^\mu, \quad a_\mu \in \{0, \ldots, p_m - 1\}, \tag{11–7}$$

where $\lceil \cdot \rceil$ denotes the integral part. By reversing the order of the digits in j, we can uniquely construct a fraction lying between 0 and 1, namely, the radical inverse

$$\phi_{p_m}(j) = \sum_{\mu=0}^{\lceil \log_{p_m} j \rceil} a_\mu \, (p_m)^{-\mu-1}, \tag{11–8}$$

which can then be used to assemble a Halton sequence,

$$y'_j = \left\{ \phi_{p_1}(j), \phi_{p_2}(j), \ldots, \phi_{p_{s-1}}(j) \right\}. \tag{11–9}$$

Consider the first M terms, $j = 1, 2, \ldots, M$, with $M = (p_1 p_2 \cdots p_{s-1})^h$, $h \in \mathbb{N}$ and extend this sequence *periodically* as

$$y'_{j+M} = y'_j. \tag{11–10}$$

Choosing $h = \lceil \log_2 N \rceil + 1$, $N \geq 2$, leads to the desired discrepancy estimate $\mathcal{O}(N^{-1}[\log N]^{s-1})$ for the first N points of the set. The "shifted" Hammersley point set for dimension s is then given by

$$x_i = \mathbf{1} - z_i, \quad i = 1, 2, \ldots, N, \tag{11–11}$$

where

$$z_i = \left\{ \frac{i - 1 + \eta}{N}, \; y'_i \right\}, \quad \eta \in \mathbb{R} \;\text{ such that }\; 0 \leq i - 1 + \eta < N. \tag{11–12}$$

For $\eta = 0$, expressions (11–11) and (11–12) define the original Hammersley point set. The constant shift η was used by Wözniakowski [112] in order to minimize the L^2 discrepancy, though an explicit expression for the optimal shift has not been found, hence its value must be determined by experiment.

11C. Spherical geometry: S^{n-1} and $S^{n-1} \times S^{n-1}$. For the use of quasi-Monte Carlo methods in the application of Fourier integral operators, we must design an algorithm for integration over double spheres rather than over rectangular domains. In this framework, it is appropriate to consider the construction of low-discrepancy sets over the sphere S^{n-1}. In the application of the generalized Radon transform (for a given value of e) we encounter an integration over $\nu \in S^{n-1}$; in the application of the (imaging-)inversion operator we encounter an integration over $(\hat{\alpha}, \tilde{\alpha}) \in S^{n-1} \times S^{n-1}$ or over $(\nu, e) \in S^{n-1} \times S^{n-1}$. We will discuss and illustrate the case $n = 3$.

The integration over double spheres $S^2 \times S^2$ can be written as the double integral over two spheres defined by the unit vectors $\hat{\alpha}$ and $\tilde{\alpha}$. We consider these spheres separately. The horizontal projection of S^2 onto the tangent cylinder along the equator is an area preserving map; thus we may choose a point on the cylinder and obtain a corresponding point on the sphere. We shall draw $(\hat{\vartheta}, \hat{\varphi})$ from a Hammersley point set in the rectangle $[-\pi, \pi] \times [-1, 1]$ and assign the point $\hat{\alpha} \in S^2$ in accordance with

$$\big(\sqrt{1-\hat{\varphi}^2}\cos\hat{\vartheta},\ \sqrt{1-\hat{\varphi}^2}\sin\hat{\vartheta},\ \hat{\varphi}\big).$$

(This mapping implies in the case of randomly chosen points a uniform distribution with respect to the natural area measure on S^2.) We apply this procedure also for $\tilde{\alpha}$. Once the sampling in $(\hat{\alpha}, \tilde{\alpha})$ is accomplished, we deduce the sampling in (ν, e) using (4–4) and (9–2).

12. An Outlook on Global Seismology

The primary phases amenable to the application of microlocal analysis are earthquake generated short-period body waves; see, for example, the work of Bostock *et al.* [19]. It was by ray methods that the depth to the core mantle boundary was first estimated, and the existence of an inner core was recognized. Indeed, the phases that interact with the inner core boundary (ICB) and core mantle boundary (CMB) can be modeled with microlocal techniques. In the crust, the Moho discontinuity can be thought of as a conormal distribution reflections off which are detected and interpreted. There are also transitions of a different nature in the deep earth. We mention the ones associated with anomalously large velocity gradients (around 400 km and 600 km depths).

In the analysis of discontinuities in the mantle transition zone, Ps conversions from teleseismic body waves are processed using a delay-and-sum approach [105]. This approach can be mathematically justified using the linearized inversion formulation of Section 4 in a planarly layered background medium. Studies of lithosperic and upper mantle structure, that account for anisotropic elasticity, provide constraints on continental dynamics and evolution.

Here we focus our final discussion on inverse scattering at the CMB and selective neighborhoods. Most seismological research of the CMB region has

been based on the scattering [6; 5] and diffraction [57] of relatively high frequency body waves. The dynamic wave group of PKKP fits in the framework of the presented inverse scattering theory. These are waves that propagate through the mantle, refract into the core, bounce back from the underside of the CMB, and refract into the mantle again on their way to the receivers at earth's surface (see Figure 24, which illustrates the formation of caustics).

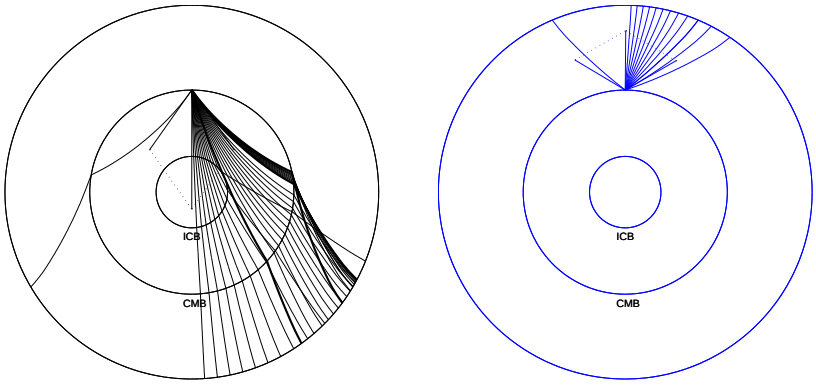


Figure 24. An illustration of characteristics associated with PKKP showing the formation of caustics (left) and characteristics associated with PcP (right).

PKKP is most readily detected in vertical-component short-period records. Often the analysis is restricted to high-frequency data to avoid contamination by long arc surface wave propagation. Other members of this wavegroup are SKKP, PKKS, and SKKS. This multiplicity can provide data redundancy for the study of the CMB near the underside reflections. PKKP and SKKS (and SKKP and PKKS in between them) can best be observed at epicentral distances up to some 100 degrees, and in a time window of some 10 minutes before the arrival of P′P′ (PKIKPPKIKP, i.e., a wave that passes through the core and reflects at earth's surface instead of at the underside of the CMB). However, due to inner core attenuation and the small reflection coefficient at near-vertical incidence, the PKKPdf arrivals are typically very weak [44]; hence, most data comes from the caustics.

12A. The core mantle boundary and its vicinity. The core mantle boundary (CMB) is located at about 2880 km depth, where the temperature is about 4000 K and the pressure is about 135 GPa, and marks one of the most dramatic changes in composition and physical properties in our planet. The CMB separates the solid mantle silicates from the liquid iron-alloy in the outer core. From mantle to core, the increase in density is about 4500 kg m^{-3} (compared to the 2700 kg m^{-3} difference between air and crustal rock near the earth's surface), the increase in temperature is about 1000-1500 K, the shear wave speed drops from some 7.2 km sec^{-1} in the lowermost mantle to zero in the liquid outer core, and

the compressional wavespeed drops from 13.7 km s^{-1} to 8 km s^{-1}. The CMB also separates two vastly different dynamic regimes. Across it the viscosity drops at least 10 and perhaps as much as 20 orders of magnitude. Mantle convection in the stiff mantle silicates, driven by thermal buoyancy, is slow, a few cm yr^{-1}; in contrast, thermal and compositional buoyancy drives turbulent flow in the liquid core at several mm s^{-1}, that is, some 6 orders of magnitude faster than in the mantle. Another consequence of the enormous viscosity contrast is that the mantle can support lateral variations in density, temperature, and elastic properties, whereas the core is usually considered homogeneous.

12B. Heterogeneity in the outermost core. On the core side, the PKKP underside reflections at high latitude straddle the intersection of the virtual "tangential cylinder", which is an essential feature of the magneto-hydrodynamics of the outer core related to the generation of the earth's magnetic field (e.g. [113; 114]). These reflections allow us to investigate if there are any changes in the character of the inside of the CMB associated with the topology of outer core flow. Of particular interest is the search for any evidence of heterogeneity in the outermost core. Core flow is partly driven by compositional buoyancy, and it is possible that "puddles" of iron that is enriched in light elements form either in topographic highs of the CMB or in certain locations relative to the virtual tangential cylinder or Taylor columns in the outer core.

References

[1] DMO processing. Geophysics Reprint Series, 1995.

[2] K. Aki and P. G. Richards. *Quantitative seismology: theory and methods*, volume 1. Freeman, San Francisco, 1980.

[3] R. M. Alford. Shear data in the presence of azimuthal anisotropy. In *Expanded Abstracts*, pages 476–479. Society of Exploration Geophysicists, 1986.

[4] G. Backus and F. Gilbert. The resolving power of gross earth data. *Geophys. J. Roy. Astr. Soc.*, 16:169–205, 1968.

[5] K. Bataille and S. M. Flatté. Inhomogeneities near the core mantle boundary inferred from short-period scattered pkp waves recorded at the global digital seismograph network. *J. Geophys. Res.*, 93:15057–15064, 1988.

[6] K. Bataille and F. Lund. Strong scattering of short-period seismic waves by the core mantle boundary and the p-diffracted wave. *Geophys. Res. Lett.*, 23:2413–2416, 1996.

[7] M. Batzle and Z. Wang. Seismic properties of pore fluids. *Geophysics*, 57:1396–1408, 1992.

[8] J. G. Berryman and G. W. Milton. Exact results for generalized gassmann's equations in composite porous media with two constituents. *Geophysics*, 56:1950–1960, 1991.

[9] G. Beylkin. The inversion problem and applications of the generalized Radon transform. *Comm. Pure Appl. Math.*, XXXVII:579–599, 1984.

[10] G. Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *J. of Math. Phys.*, 26:99–108, 1985.

[11] G. Beylkin. Reconstructing discontinuities in multidimensional inverse scattering problems: smooth errors vs small errors. *Applied Optics*, 24:4086–4088, 1985.

[12] G. Beylkin and R. Burridge. Linearized inverse scattering problems in acoustics and elasticity. *Wave Motion*, 12:15–52, 1990.

[13] B. Biondi, S. Fomel, and N. Chemingui. Azimuth moveout for 3-D prestack imaging. *Geophysics*, 63:574–588, 1998.

[14] M. A. Biot. Theory of propagation of elastic waves in a fluid-saturated porous solid. i. low-frequency range. *J. Acoust. Soc. Am.*, 28:168–178, 1956.

[15] M. A. Biot. Theory of propagation of elastic waves in a fluid-saturated porous solid. ii. higher frequency range. *J. Acoust. Soc. Am.*, 28:179–191, 1956.

[16] M. A. Biot. Mechanics of deformation and acoustic propagation in porous media. *J. Appl. Phys.*, 23:1482–2498, 1962.

[17] N. Bleistein. On imaging of reflectors in the earth. *Geophysics*, 52(6):931–942, 1987.

[18] N. Bleistein, J. K. Cohen, and J. W. Jr. Stockwell. *Mathematics of multidimensional seismic imaging, migration and inversion*. Springer-Verlag, New York, 2000.

[19] M. G. Bostock, S. Rondenay, and J. Shragge. Multiparameter two-dimensional inversion of scattered teleseismic body waves. *J. Geophys. Res.*, 106:30771–30782, 2001.

[20] P. J. Braam and J. J. Duistermaat. Normal forms of real symmetric systems with multiplicity. *Indag. Mathem.*, 4:197–232, 1993.

[21] S. Brandsberg-Dahl, M. V. De Hoop, and B. Ursin. Focusing in dip and ava compensation on scattering-angle/azimuth gathers. *Geophysics*, 68:232–254, 2003. in print.

[22] S. Brandsberg-Dahl, B. Ursin, and M. V. De Hoop. Seismic velocity analysis in the scattering-angle/azimuth domain. *Geophys. Prosp.*, in print (July 2003 issue). submitted.

[23] R. J. S. Brown and J. Korringa. On the dependence of the elastic properties of a porous rock on the compressibility of a pore fluid. *Geophysics*, 40:608–616, 1975.

[24] B. Budiansky. On the elastic moduli of some heterogeneous materials. *J. Mech. Phys. Solids*, 13:223–227, 1965.

[25] R. Burridge, M. V. De Hoop, D. Miller, and C. Spencer. Multiparameter inversion in anisotropic media. *Geophys. J. Int.*, 134:757–777, 1998.

[26] R. Burridge and J. B. Keller. Poroelasticity equations derived from microstructure. *J. Acoust. Soc. Am.*, 70:1140–1146, 1981.

[27] V. Červený. *Seismic ray theory*. Cambridge University Press, Cambridge, 2001.

[28] C. H. Chapman. Reflection/transmission coefficient reciprocities in anisotropic media. *Geophysical Journal International*, 116:498–501, 1994.

[29] C. H. Chapman and R. Drummond. Body-wave seismograms in inhomogeneous media using maslov asymptotic theory. *Bull. Seism. Soc. Am.*, 72:277–317, 1982.

[30] J. F. Claerbout. *Fundamentals of Geophysical Data Processing*. McGraw-Hill, 1976.

[31] R. W. Clayton and R. H. Stolt. A born-wkbj inversion method for acoustic reflection data. *Geophysics*, 46:1559–1567, 1981.

[32] M. V. De Hoop and N. Bleistein. Generalized Radon transform inversions for reflectivity in anisotropic elastic media. *Inverse Problems*, 13:669–690, 1997.

[33] M. V. De Hoop and S. Brandsberg-Dahl. Maslov asymptotic extension of generalized Radon transform inversion in anisotropic elastic media:a least-squares approach. *Inverse Problems*, 16:519–562, 2000.

[34] M. V. De Hoop, R. Burridge, C. Spencer, and D. Miller. Generalized Radon transform amplitude versus angle (grt/ava) migration/inversion in anisotropic media. In *Proc SPIE 2301*, pages 15–27. SPIE, 1994.

[35] M. V. de Hoop and A. T. de Hoop. Wavefield reciprocity and optimization in remote sensing. *Proc. R. Soc. Lond. A (Mathematical, Physical and Engineering Sciences)*, 456:641–682, 2000.

[36] M. V. de Hoop, J. H. Le Rousseau, and B. Biondi. Symplectic structure of wave-equation imaging: A path-integral approach based upon the double-square-root equation. *Geoph. J. Int.*, 153:52–74, 2003. submitted.

[37] M. V. De Hoop and C. Spencer. Quasi monte-carlo integration over $s^2 \times s^2$ for migration $\times$ inversion. *Inverse Problems*, 12:219–239, 1996.

[38] M. V. De Hoop, C. Spencer, and R. Burridge. The resolving power of seismic amplitude data: An anisotropic inversion/migration approach. *Geophysics*, 64:852–873, 1999.

[39] N. Dencker. On the propagation of polarization sets for systems of real principal type. *Journal of Functional Analysis*, 46:351–372, 1982.

[40] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. Geosc. Remote Sens.*, GE-22:3–13, 1984.

[41] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[42] H. H. Duistermaat and V. Guillemin. The spectrum of positive elliptic operators and periodic bicharacteristics. *Inv. Math.*, 29:39–79, 1975.

[43] J. J. Duistermaat. *Fourier integral operators*. Birkhäuser, Boston, 1996.

[44] P. S. Earle and P. N. Shearer. Observations of high-frequency scattered energy associated with the core phase pkkp. *Geophys. Res. Lett.*, 25:405–408, 1998.

[45] P. P. Ewald. Zur begründung der kristalloptik. *Annalen der Physik, IV*, 49:1–38, 1916.

[46] N. L. Frazer and M. K. Sen. Kirchhoff-helmholtz reflection seismograms in a laterally inhomogeneous multi-layered medium – i. theory. *Geophys. J. R. Astr. Soc.*, 80:121–147, 1985.

[47] F. Gassmann. Uber die elastizität poröser medien. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 96:1–23, 1951.

[48] A. Greenleaf and G. Uhlmann. Nonlocal inversion formulas for the x-ray transform. *Duke Math. J.*, 58:205–240, 1989.

[49] A. Greenleaf and G. Uhlmann. Composition of some singular fourier integral operators and estimates for restricted x-ray transforms. *Ann. Inst. Fourier*, 40:443–446, 1990.

[50] A. Greenleaf and G. Uhlmann. Estimates for singular Radon transforms and pseudodifferential operators with singular symbols. *J. Func. Anal.*, 89:202–232, 1990.

[51] V. Guillemin. *Pseudodifferential operators and applications (Notre Dame, Ind., 1984)*, chapter On some results of Gel'fand in integral geometry, pages 149–155. Amer. Math. Soc., Providence, RI, 1985.

[52] A. G. Haas and J. R. Viallix. Krigeage applied to geophysics. *Geophysical Prospecting*, 24:49–69, 1976.

[53] J. G. Hagedoorn. A process of seismic reflection interpretation. *Geophysical Prospecting*, 2:85–127, 1954.

[54] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numeric. Math.*, 2:85–90, 1960.

[55] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Methuen, London, 1964.

[56] S. Hansen. Solution of a hyperbolic inverse problem by linearization. *Communications in Partial Differential Equations*, 16:291–309, 1991.

[57] D. V. Helmberger, E. J. Garnero, and X.-D. Ding. Modeling 2-d structure at the core-mantle boundary. *J. Geophys. Res.*, 101:13963–13972, 1996.

[58] R. Hill. A self-consistent mechanics of composite matericals. *J. Mech. Phys. Solids*, 13:213–222, 1965.

[59] L. Hörmander. Fourier integral operators. i. *Acta Math.*, 127:79–183, 1971.

[60] L. Hörmander. *The analysis of linear partial differential operators*, volume I. Springer-Verlag, Berlin, 1983.

[61] L. Hörmander. *The analysis of linear partial differential operators*, volume III. Springer-Verlag, Berlin, 1985.

[62] L. Hörmander. *The analysis of linear partial differential operators*, volume IV. Springer-Verlag, Berlin, 1985.

[63] B. E. Hornby. *The elastic properties of shales*. PhD thesis, University of Cambridge, Cambridge, 1994.

[64] B. E. Hornby, L. M. Schwartz, and J. A. Hudson. Anisotropic effective medium modeling of the elastic properties of shales. *Geophysics*, 59:1570–1583, 1994.

[65] J. A. Hudson. Overall properties of a cracked solid. *Math. Proc. Camb. Phil. Soc.*, 88:371–384, 1980.

[66] E. Iversen, H. Gjøystdal, and J. O. Hansen. Prestack map migration as an engine for parameter estimation in ti media. In *70th Ann. Mtg. Soc. Explor. Geoph., Expanded Abstracts*, pages 1004–1007, 2000.

[67] V. Ya. Ivrii. Wave fronts of solutions of symmetric pseudodifferential systems. *Siberian Mathematical Journal*, 20:390–405, 1979.

[68] L. E. A. Jones and H. F. Wang. Ultrasonic velocities in cretaceous shales from the williston basin. *Geophysics*, 46:288–297, 1981.

[69] J. M. Kendall and C. J. Thomson. Maslov ray summation, pseudo-caustics, lagrangian equivalence and transient seismic waveforms. *Geoph. J. Int.*, 113:186–214, 1993.

[70] B. L. N. Kennett. *Seismic Wave Propagation in Stratified Media*. Cambridge University Press, Cambridge, 1983.

[71] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*, volume 1. Springer-Verlag, Berlin, 1972.

[72] S. Mallat. *A wavelet tour of signal processing*. Academic Press, San Diego, 1997.

[73] V. P. Maslov and M. V. Fedoriuk. *Semi-classical approximation in quantum mechanics*. Reidel Publishing Company, 1981.

[74] R. Melrose and G. Uhlmann. Microlocal structure of involutive conical refraction. *Duke Math. J.*, 46:571–582, 1979.

[75] D. Miller, M. Oristaglio, and G. Beylkin. A new slant on seismic imaging: migration and integral geometry. *Geophysics*, 52(6):943–964, 1987.

[76] H. E. Moses. Calculation of scattering potential from reflection coefficients. *Phys. Rev.*, 102:559–567, 1956.

[77] M. J. P. Musgrave. *Crystal acoustics*. Holden-Day, 1970.

[78] H. Niederreiter. Quasi-monte carlo methods and pseudo-random numbers. *Bull. Am. Math. Soc.*, 84:957–1041, 1978.

[79] H. Niederreiter. *Numerical Integration III. International series of numerical mathematics Vol* **85**, chapter Quasi-Monte Carlo methods for multidimensional numerical integration. Birkhauser-Verlag, Berlin, 1988.

[80] C. J. Nolan and W. W. Symes. Global solution of a linearized inverse problem for the wave equation. *Communications in Partial Differential Equations*, 22(5-6):919–952, 1997.

[81] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in FORTRAN, The art of scientific computing*. Cambridge University Press, 1994.

[82] Rakesh. A linearised inverse problem for the wave equation. *Comm. in Part. Diff. Eqs.*, 13:573–601, 1988.

[83] M. Razavy. Determination of the wave velocity in an inhomogeneous medium from reflection data. *J. Acoust. Soc. Am.*, 58:956–963, 1975.

[84] L. A. Riabinkin. Fundamentals of resolving power of controlled directional reception (cdr) of seismic waves. In Gardner G. H. F. and Lu L., editors, *Slant stack processing*, number 14 in Geophysics Reprint Series, pages 36–60. Society of Exploration Geophysicists, Tulsa, 1957.

[85] F. Rieber. A new reflection system with controlled directional sensitivity. *Geophysics*, I(1):97–106, 1936.

[86] J. Schleicher, M. Tygel, and P. Hubral. 3-D true-amplitude finite-offset migration. *Geophysics*, 58:1112–1126, 1993.

[87] W. A. Schneider. Integral formulation for migration in two and three dimensions. *Geophysics*, 43:49–76, 1978.

[88] R. T. Shuey. A simplication of the zoeppritz equations. *Geophysics*, 50:609–614, 1985.

[89] C. C. Stolk. Microlocal analysis of a seismic linearized inverse problem. *Wave Motion*, 32:267–290, 2000.

[90] C. C. Stolk. Microlocal analysis of the scattering angle transform. Preprint, Dept. of Computational and Applied Mathematics, Rice University, www.caam.rice.edu/~cstolk/angle.ps, July 2001.

[91] C. C. Stolk and M. V. De Hoop. Microlocal analysis of seismic inverse scattering in anisotropic, elastic media. *Comm. Pure Appl. Math.*, 55:261–301, 2002.

[92] C. C. Stolk and M. V. De Hoop. Seismic inverse scattering in the "wave equation" approach. *SIAM J. Math. Anal.*, 2002, submitted; available at http://www.msri.org/publications/preprints/online/2001-047.html.

[93] R. H. Stolt. Migration by fourier transform. *Geophysics*, 43:23–48, 1978.

[94] R. H. Stolt and A. B. Weglein. Migration and inversion of seismic data. *Geophysics*, 50:2456–2472, 1985.

[95] W. W. Symes. A differential semblance algorithm for the inverse problem of reflection seismology. *Comput. Math. Appl.*, 22(4-5):147–178, 1991.

[96] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266, 1984.

[97] A. Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51:1893–1903, 1986.

[98] M. E. Taylor. Reflection of singularities of solutions to systems of differential equations. *Communications on Pure and Applied Mathematics*, 28:457–478, 1975.

[99] A. P. E. ten Kroode, D.-J. Smit, and A. R. Verdel. A microlocal analysis of migration. *Wave Motion*, 28:149–172, 1998.

[100] L. Thomsen. Weak elastic anisotropy. *Geophysics*, 51(10):1954–1966, 1986.

[101] F. Treves. *Introduction to pseudodifferential and Fourier integral operators*, volume 2. Plenum Press, New York, 1980.

[102] F. Treves. *Introduction to pseudodifferential and Fourier integral operators*, volume 2. Plenum Press, New York, 1980.

[103] I. Tsvankin. *Seismic signatures and analysis of reflection data in anisotropic media*. Elsevier Science, Amsterdam, 2001.

[104] G. Uhlmann. Light intensity distribution in conical refraction. *Comm. Pure Appl. Math.*, XXXV:69–80, 1982.

[105] L. P. Vinnik. Detection of waves converted from P to SV in the mantle. *Phys. Earth Planet. Inter.*, 15:39–45, 1977.

[106] K. Walton. The oblique compression of two elastic spheres. *J. Mech. Phys. Solids*, 26:139–150, 1978.

[107] K. Walton. The effective elastic moduli of a random packing of spheres. *J. Mech. Phys. Solids*, 35:213–226, 1987.

[108] M. Weber. Die bestimmung einer beliebig gekruemmten schichtgrenze aus seismischen reflexionsmessungen. *Geofisica pura e applicata*, 32:7–11, 1955.

[109] D. N. Whitcombe and R. J. Carroll. The application of map migration to 2-d migrated data. *Geophysics*, 59:1121–1132, 1994.

[110] B. S. White, B. Nair, and A. Bayliss. Random rays and seismic amplitude anomalies. *Geophysics*, 53:903–907, 1988.

[111] J. H. Woodhouse. Surface waves in laterally varying layered media. *Geophys. J. R. Astr. Soc.*, 37:461–490, 1974.

[112] H. Wözniakowski. Average case complexity and multivariate integration. *Bull. Am. Math. Soc.*, 24:185–194, 1991.

[113] K. K. Zhang and F. H. Busse. Convection driven magnetohydrodynamic dynamos in rotating spherical-shells. *Geophys. Astro Fluid*, 49(1-4):97–116, 1989.

[114] K. K. Zhang and D. Gubbins. Convection in a rotating spherical fluid shell with an inhomogeneous temperature boundary-condition at infinite prandtl number. *J. Fluid Mech.*, 250:209–232, 1993.

Maarten V. de Hoop
Center for Wave Phenomena
Colorado School of Mines
Golden, CO 80401-1887 United States
mdehoop@Mines.EDU